Environmental determinants of chromosomal translocations and DNA methylation: towards identifying and understanding modifiable risk factors of childhood leukaemia

J R Saville

PhD

2022

Environmental determinants of chromosomal translocations and DNA methylation: towards identifying and understanding modifiable risk factors of childhood leukaemia

Jessica Rose Saville

A thesis submitted in partial fulfilment

of the requirements of the

University of Northumbria at Newcastle

for the degree of

Doctor of Philosophy

Research undertaken in the Faculty of Health and Life Science

December 2022

Abstract

Leukaemia is the most common type of cancer found in children, with incidence rates increasing. Chromosomal translocations are considered one of the leukaemia initiating events. Causes of many chromosomal translocations remain unknown. Further molecular events, such as epigenetic changes, are considered necessary for disease progression. Epidemiological studies have identified many environmental exposures associated with increased risk of childhood leukaemia. There is little understanding of the molecular role these exposures play in the induction of chromosomal translocations or epigenetic changes, such as DNA methylation.

The aim of this project was to use *in vitro* techniques to screen for the induction of chromosomal translocations in response to environmental exposures i.e. caffeine, benzene (smoking/air pollution), cotinine (smoking) and folate. To explore the potential role of DNA methylation in childhood leukaemia development, a bioinformatic analysis investigated the correlation between altered methylation in childhood leukaemia subtypes and environmental risk exposures.

The NALM6 cell line was used as an exposure model to detect *TCF3-PBX1* and *RUNX1-RUNX1T1* translocations via RT-PCR assays. Preliminary experiments investigating physiologically relevant concentrations of caffeine, benzene, cotinine, and folate, observed translocation events with each risk exposure. Further cell line models were optimised to investigate the impact of daily exposures and first exposure during exponential cell growth, which also observed translocation events for each risk exposure. However, in these further studies, translocation events were also observed in DMSO treated control cells suggesting DMSO could be a contributing factor to the translocation events observed.

Using a meet-in-the-middle approach, methylation patterns were found to be significantly overlapping between specific childhood leukaemia subtypes and radiation, alcohol, smoking, nursery attendance, reported colds and maternal plasma folate.

Understanding the environmental risk contributing to chromosomal translocations and epigenetic events that aid disease progression may be useful to influence public health policy to assist and tailor prevention strategies for childhood leukaemia.

List of contents

List of Figures	***Page***

List of Tables	***Page***
Table 1.1. The most commonly occurring chromosomal translocations found in childhood B-ALL and their frequency	*6*
Table 1.2. The most commonly occurring chromosomal translocations found in childhood AML and their frequency	*7*
Table 2.1. Preparation of chemicals	*38*
Table 2.2. NALM6 cell line information	*38*

List of abbreviations

5-mTHF	5-Methyltetrahydrofolate
ALL	Acute Lymphoblastic Leukaemia
ALSPAC	Avon Longitudinal Study of Parents and Children
AML	Acute Myeloid Leukaemia
APL	Acute Promyelocytic Leukaemia
ARIES	Accessible Resource for Integrated Epigenomic Studies
B-ALL	B-Lineage Acute Lymphoblastic Leukaemia
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary DNA
CML	Chronic Myeloid Leukaemia
CYP1A2	Cytochrome P450 1A2
DAPI	4',6-Diamidino-2-Phenylindole
DDR	DNA Damage Response
DHF	Dihydrofolate
DMC	Differentially Methylated CpG
Dmnts	DNA Methyltransferases
DSB	Double Strand Breaks
dTMP	Deoxythymidine Monophosphate
dUMP	Deoxyuridine Monophosphate
EWAS	Epigenome Wide Association Studies
FBS	Fetal Bovine Serum
FISH	Fluorescent In Situ Hybridisation
GFP	Green Fluorescent Protein
GIPFEL	Genomic Inverse PCR For Exploration of Ligated Breakpoints
GO	Gene Ontology
Heh	High Hyperdiploidy
HR	Homologous Recombination
IUGR	Intrauterine Growth Retardation
KEGG	Kyoto Encyclopedia Of Genes and Genomes
LMIC	Low- And Middle-Income Countries
MLL	Mixed Lineage Leukaemia
MTHFR	5,10-Methylenetetrahydrofolate Reductase
MTRR	Methionine Synthase Reductase
MWM	Molecular Weight Marker

NHEJ	Non Homologous End Joining
NTD	Neural Tube Defect
PACE	Pregnancy And Childhood Epigenetics
PCR	Polymerase Chain Reaction
PML	Promyelocytic Leukemia
ppm	Part Per Million
qPCR	Quantitative PCR
RARA	Retinoic Acid Receptor Alpha
RFU	Relative Fluorescence Unit
ROS	Reactive Oxygen Species
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RUNX1T1	Runx1 Partner Transcriptional Co-Repressor 1
SAM	*S*-Adenosylmethionine
SAM domain	Sterile Alpha Motif
SHMT1	Serine Hydroxy-Methyltransferase
SSC	Saline-Sodium Citrate
T-ALL	T-Lineage Acute Lymphoblastic Leukaemia
THF	Tetrahydrofolate

Acknowledgements

The last four years have felt like both a whirlwind and the longest journey of my life so far. Undertaking a PhD in the middle of a worldwide pandemic has been challenging at the best of times but the people who have supported and helped me through it have made all the stress and difficulties worthwhile and encouraged me to the end goal.

To start, I would like to thank my primary supervisor Dr Jill McKay. Jill has been a tremendous role-model and a beacon of support, always there when I need her and has provided invaluable encouragement, ideas and discussions that have shaped me as a scientist. Throughout the difficulties of Covid-19 lockdowns Jill was fantastic in helping me pivot my PhD project to include bioinformatic analysis that could be performed whilst labs were closed and provided some of the analysis used in this thesis.

I would also like to thank the rest of my supervisory team for all of their help and support. Dr Lisa Russell has been a fantastic fountain of knowledge, helping me with designing experiments, completing FISH analysis and providing feedback. Dr Kay Padget was also such a welcoming presence at Northumbria University and a wonderful help.

I am also extremely grateful to the wonderful friends I have made in the lab. Through lockdown Teams check-ins to coffees and afternoon tea once the world started to open back up, my amazing group of lab mates have been there for me every step of the way, to chat, to laugh and to be a shoulder to cry on. A special thanks is also given to Gina Abdelaal, an amazing friend, who has been a fountain of advice in tissue culture and has provided invaluable help with bioinformatics and writing scripts in R.

I am extremely grateful to Northumbria University for providing the funding for this PhD position, and to Children with Cancer UK who provided £25,000 in funding, allowing us to carry out many more experiments than would have been possible otherwise and disseminate our research at international conferences.

Finally, I would like to take a moment to thank my family and friends for all of their love and support, which I wouldn't have been able to finish this PhD without. A special thanks to my Mam who has been my rock and my inspiration for my whole life.

And to my husband, Sean, this PhD has spanned our entire marriage so far. We survived lockdowns and working from home together, you were my biggest cheerleader keeping me motivated to keep going and keep writing, and you always made sure I had all the chocolate I need to get through the day. I hope you realise just how special you are.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. When necessary, permission was requested for the reuse of Figures in the printed and electronic copies of this thesis that are subject to copyright.

Any ethical clearance for the research presented in this commentary has been approved. Approval has been sought and granted through the Researcher's submission to Northumbria University's Ethics Online System on 2/11/2018.

I declare the word count of this thesis is 49,496 words.

Name: Jessica Rose Saville

Date: 22/12/2022

1 Introduction

1.1 Childhood leukaemia

Childhood leukaemia is the most common form of cancer in children, and it accounts for nearly a third of all childhood cancers [1]. A cancer of white blood cells, leukaemia is a heterogeneous disease, meaning that differences in histology, genetic and chromosomal abnormalities can not only be found between patients with the same phenotype, but can also have intratumoral differences within the same patient [2]. Childhood leukaemia can be segregated into 2 main subgroups, acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML), categorised by lymphoid or myeloid blood cells that divide uncontrollably but are unable to mature. Within these subgroups there is further diversity in patient age, with AML common in infants under 2 compared to 2-5 in ALL, molecular abnormalities, such as copy number variants and chromosomal aberrations, and morphology, such as the maturation of cells [3, 4]. The heterogeneity of childhood leukaemia plays an important role in risk stratification and disease management, with developments in treatment targeted to become more personalised based on cytogenetics, age and sex, which has improved survival rates [4, 5]. With major improvements in diagnosis and treatment over the past few decades, overall survival rates for childhood leukaemia are over 90% [6, 7].

ALL accounts for 80% of childhood leukaemias, making up 25% of all childhood cancers [1]. The peak age for children with ALL is 2 to 5 years of age, with over 450 cases per year diagnosed in the UK [1, 8]. Whilst 5-year event free survival rates for childhood ALL are reaching over 90%, overall relapse rates are still 25% [1]. Although survival rates are high, there are many long term side effects of leukaemia treatment including central nervous system relapse, cardiac and fertility problems plus an increased risk of therapy related cancers [1]. The majority of cells involved in childhood ALL are immature precursor lymphoid cells, consisting of 85% B-lineage and 10%-15% T-lineage cells [9].

AML makes up 15% of all childhood leukaemias, with 70 cases per year in the UK, accounting for 5% of total childhood cancers [1]. The peak age for children with AML is infants under 2 years old, with a higher ratio of boys diagnosed than girls [1]. Childhood AML survival has a lower rate than ALL at 66% [1].

A number of genetic abnormalities are involved in leukaemia including chromosomal translocations and inversions, chromosome number changes (aneuploidy), single nucleotide mutations and deletions [10]. As a consequence of genetic rearrangements misregulation of cellular processes can occur through novel fusion proteins or disruption of wild type proteins [11]. A gain of non-random chromosomes, exceeding 50 chromosomes, known as high hyperdiploidy, has been associated with secondary RAS mutations and chromatin modifier

genes which lead to oncogenic effects [12, 13]. The regulation of gene expression by altered methylation can lead to inhibition of tumour suppressor genes amongst other disregulation [14]. Many of these genetic abnormalities are key for patient treatment plans and contribute in a variety of ways to leukaemia onset. Although leukaemia originates as a clonal disease, progression of the disease leads to diverse mutant subclones, which gain a variety of these abnormalities, driving the leukaemia forward.

1.1.1 Chromosomal translocations

One of the major structural abnormalities found in childhood leukaemia are chromosomal translocations, with over 200 identified so far [3]. Chromosomal translocations occur when the chromosome breaks and is reattached to a different chromosome. Non-reciprocal translocations are caused by a small section of one chromosome attaching to a non-homologous chromosome [15]. On the other hand, a reciprocal translocation occurs when segments from two non-homologous chromosomes swap and reattach, generating two chromosome translocations. If no genetic material is lost in the transfer this is known as a balanced translocation, as shown in Figure 1.1 [16].

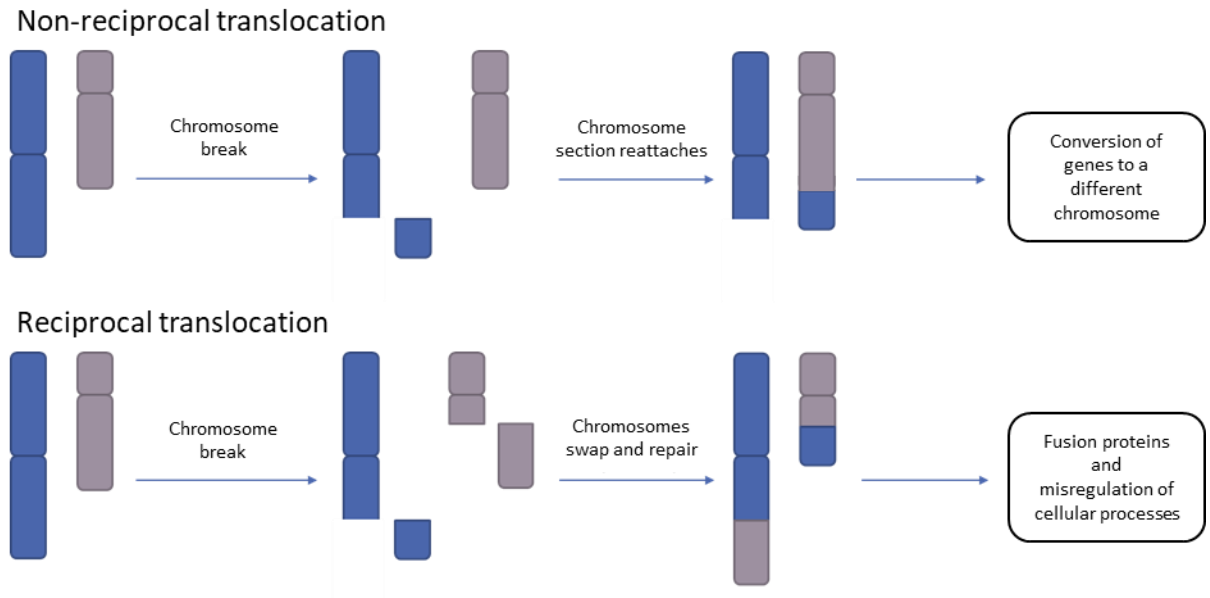


Figure 1.1. Overview of translocation formation. Translocations occur when chromosomes break, then swap genetic material with no loss of genetic material. Non-reciprocal translocations occur when a section of one chromosome is attached to a non-homologous chromosome, without any further exchange of genetic material, allowing for the conversion of genes from one chromosome to another. Reciprocal translocations occur following the repair of double strand breaks in 2 chromosomes creating 2 balanced chromosomal translocations that can lead to fusion proteins and misregulation of cellular processes.

The originating double strand breaks (DSB) can occur through various means. Spontaneous breaks can occur during replication stress, such as when the replication fork has a prolonged

arrest due to endogenous obstacles, that are then processed by either homologous recombination (HR) or non-homologous end joining (NHEJ) [17]. Interference of transcription and replication can cause blockages and lead to supercoiling, and some areas of DNA have been shown to be prone to fragility [17]. DSBs are also formed during lymphocyte development, where V(D)J gene segment recombination occurs to form antigen binding domains. Errors in target recognition and joining can occur during V(D)J rearrangement leading to translocations and insertions [18]. Exogenous stress from sources such as ionising radiation, which has the energy to break DNA molecules and chemotherapy agents, such as etoposide and cisplatin, that generate DNA adducts, have been shown to lead to DNA DSB formation [17].

DNA DSBs can be dangerous, so when they occur cells mount DNA damage response (DDR) measures. During S-phase, when the cell is replicating, DSBs are usually repaired by HR. HR depends on a homologous template to repair the DSB, using a homologous chromosome or sister chromatid to form a D loop, allowing for DNA synthesis and repair [19]. Non-homologous end joining differs as it is used throughout the whole cell cycle, and allows for repair of DSBs without the need for a homologous template [20]. NHEJ requires a multitude of proteins including polymerases, nucleases and ligases, which process the two ends of the DSB allowing for re-ligation [20]. Alternative-NHEJ (a-NHEJ) occurs through pairing of short homologous sequences close to the breakpoint ends [21]. Due to their nature, DNA repair pathways can both prevent translocations, by repairing DNA, through mechanisms such as HR or promote translocations, by ligating non-homologous DSBs, via NHEJ and a-NHEJ. Recognition of chromosome breaks is important in translocation prevention, through activation of cell cycle checkpoints, such as the activation of Ataxia Telangiectasia Mutated (ATM) by Mre11, Rad50 and Nbs1 [22]. Members of the PARP family are also activated by DNA strand breaks, which in turn recruit DNA repair factors for repair via NHEJ and a-NHEJ, which may promote translocations [22]. With an estimated 10 DSBs per cell per day, DNA repair mechanisms are incredibly efficient, however, when repair takes too long or encounters a problem, then these DSBs can lead to chromosomal rearrangements [20].

The exact causes of many translocations are still unknown, however many factors within the cell are thought to contribute to their formation. Although translocations are extremely rare, many translocations have common gene partners, as observed in leukaemia subtypes, suggesting that genome organisation may play a role in driving translocations. Two main theories of translocation induction include the “contact-first” and “breakage-first” mechanisms [23]. The contact first mechanism suggests that translocations occur between chromosomes in close proximity [24]. On the other hand, the breakage first mechanism suggests that once

double strand breaks are induced then the chromosome ends are able to travel through the nuclear space and come into contact [24]. The mobility of DSBs have been shown to be influenced by cell cycle stage and chromosome association with the nucleolus and nuclear periphery [16]. Proximal chromosomes have been shown to have a higher translocation frequency, with 80% of translocation breaks within a distance of 2.5μm [25].As DSBs have limited movement, this restricts potential translocation partners that they may come in to contact with. DSBs have been shown to undergo pairing and dissociation with other DSBs, but persistent pairs are more susceptible to forming translocations [16]. Common repair centres and transcription factories, areas of the nucleus where multiple genes are brought together with repair and transcription machinery, may provide a centre for pairing of persistent DSBs [16]. The stability of the chromosome following a DSB may influence the movement of the broken ends and whether translocations occur between proximal or distant chromosomes [24].

Adults and children who have received chemotherapy treatment for a primary cancer can go on to develop therapy-related leukaemias, as a secondary cancer. Like childhood leukaemias, many therapy-related leukaemias are characterised by a variety of chromosomal translocations including the 11q23 *MLL* translocation, t(8;21)(q22,q22) *RUNX1-RUNX1T1* fusion and t(15;17) *PML-RARA* fusion.

A common chemotherapy drug associated with therapy related leukaemia is etoposide, a Topoisomerase II poison [26]. Topoisomerase II is an important enzyme catalysing the topological conversion of DNA during various biological processes. Due to the length of the DNA double helix, it is coiled tightly in the cell, and as such it can be difficult to unwind and access required parts of the genome. During replication, enzymes such as polymerases can increase the coiling of DNA as they track along the double helix leading to twists and tangles in the DNA that can inhibit biological processes [27]. DNA segments can also become intertwined and knotted during recombination and replication which can be detrimental to the cell [27].

Topoisomerase proteins are cleavage enzymes used to regulate the topology of DNA. Type I topoisomerases induce single strand breaks into DNA, allowing for strand passage or helix rotation to regulate under and overwinding of DNA [27]. Type II topoisomerases use DSBs to allow transport of double stranded DNA, removing DNA tangles and supercoiling [27]. There are 2 isoforms of topoisomerase II enzymes in humans, topoisomerase IIα and topoisomerase IIβ. Topoisomerase IIα is found at high levels during replication and is thought to be important for proliferating cells [27]. Expression of topoisomerase IIβ is independent of cell cycle but is important in neural development [27].

Topoisomerase II introduces a DSB to allow the unwinding of tightly coiled DNA, Figure 1.2. A cleavage complex is formed with topoisomerase II and the DSB, allowing for the second DNA strand to pass through the middle of the complex. Once the uncoiling has occurred then the complex is processed and the broken DNA strand is religated. Topoisomerase II poisons, such as etoposide, prevent topoisomerase II-DNA complexes from proceeding to the next step, which is re-ligation of the DSB, Figure 1.2. Once these complexes have been processed, DNA DSBs are produced, initiating repair mechanisms, which can lead to chromosomal translocations [28].

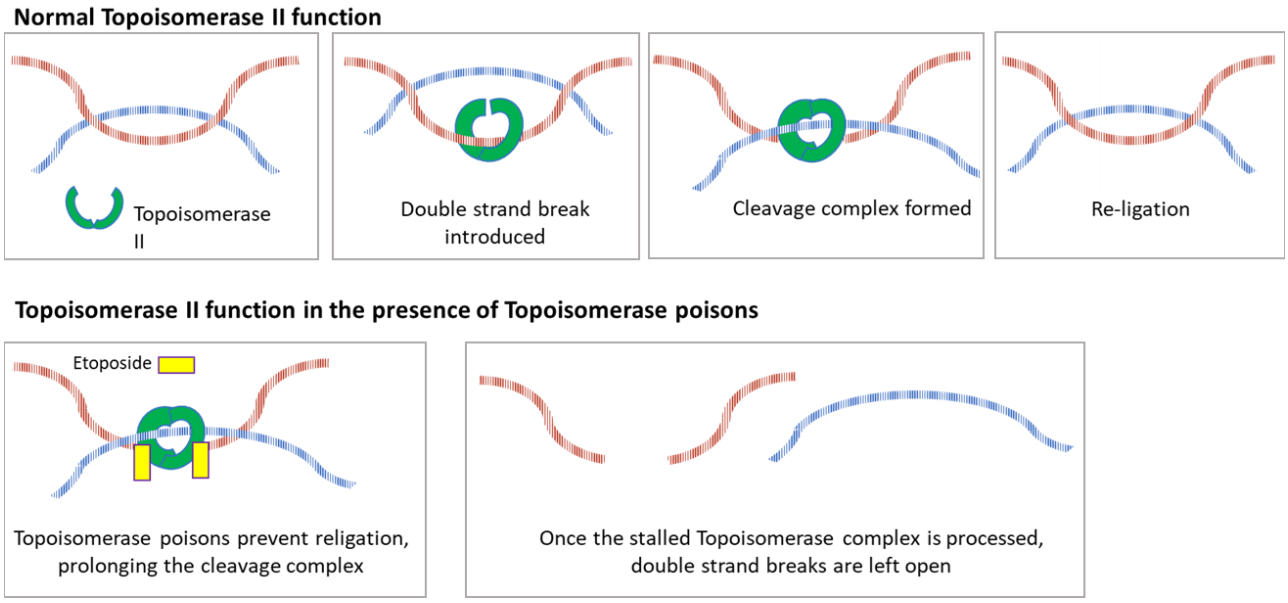


Figure 1.2. Action of Topoisomerase II in the presence and absence of Topoisomerase poisons.

1.1.2 Chromosomal translocations associated with Acute Lymphoblastic Leukaemia

Although some chromosomal translocation gene pairs are found at a high frequency in childhood leukaemia, there have been over 200 genes associated with childhood leukaemia translocation pairs [3]. The most common leukaemia translocations and their frequency in childhood leukaemia are described in Table 1.1. Other genetic abnormalities such as hyperdiploidy, hypodiploidy, T-cell abnormalities, plus many other translocation pairs, make up the rest of the main biological subtypes found in childhood leukaemia [3].

Table 1.1. The most commonly occurring chromosomal translocations found in childhood B-ALL and their frequency [3].

Chromosomal translocation	Fusion product (also known as)	Frequency (%)
t(12;21)(p13;q22)	*ETV6-RUNX1 (TEL-AML1)*	25
t(4;11)(q21;q23)	*KMT2A-AFF1 (MLL-AF4)*	5
t(1;19)(q23;p13)	*TCF3-PBX1 (E2A-PBX1)*	5
t(9;22)(q34;q11)	*BCR-ABL1* (*BCR-ABL*, Philadelphia Chromosome)	3

The translocation t(12;21)(p13;q22) leads to a *ETV6-RUNX1* fusion which is the most common genetic abnormality in B-ALL, accounting for 25% of cases [9]. Patients with *ETV6-RUNX1* have a high overall survival rate, reaching 93%, leading to a reduction of intensive therapy [9]. The *ETV6* gene on chromosome 12 encodes a transcription factor from the *ets* family that is involved in transcriptional repression through a Sterile alpha motif (SAM) domain (N-terminal pointed domain) and a central domain [29]. The *RUNX1* gene on chromosome 21 is a transcription factor involved in haematopoietic differentiation and cell cycle regulation [30]. Fusion of the SAM domain of *ETV6* to the *RUNX1* gene results in a fusion protein that can disrupt normal *ETV6* activity through dimerization with the wild-type protein and is known to interact with other transcription factors as well as repress *RUNX1* target genes [29, 30]. Immature B-cells with *ETV6-RUNX1* accumulate, infiltrating vital organs, suppressing normal blood cell formation and differentiation [31].

Rearrangements involving the mixed lineage leukaemia (*MLL*) locus, also known as *KMT2A*, are involved in over 5% of ALL cases, accounting for 70% of infant ALL [9, 32]. *KMT2A* translocations are found in both B-ALL and T-ALL leukaemia [9]. The *KMT2A* locus has been found to partner with over 40 different genes in childhood leukaemia [33]. The most common translocation partner of *KMT2A*, with a frequency nearing 40% is *AFF1* [34]. Patients with a *KMT2A-AFF1* fusion often have a poor prognosis and low survival rate [32]. *KMT2A* contains a domain involved in the methylation of histone H3, required in transcription [32]. Studies have shown that *KMT2A* is important for embryogenesis, normal haematopoiesis, and regulation of HOX genes, suggesting that a translocation at this locus could contribute to leukemogenesis [32, 33]. With a role in gene transcription regulation and chromatin remodelling, *AFF1* acts as scaffolding proteins for transcription elongation factors P-TEFb and ELL1/2 [35].

The translocation t(1;19) (q23;p13.3) creates a *TCF3-PBX1* fusion, which is one of the most common pre-B cell translocations involved in 5% of ALL cases [9]. Patients with *TCF3-PBX1*

often require an aggressive treatment course, but overall survival rates are currently at 85% [36, 37]. *TCF3*, a transcription factor 3 gene, is found on chromosome 19, and encodes E12 and E47 transcription factors [38]. These E proteins are ubiquitously expressed members of the helix-loop-helix family that are able to bind DNA as homodimers or heterodimers as regulators of gene transcription [38, 39]. Numerous studies have shown *TCF3* has an important role in B cell lineage and development, with mouse knockouts showing B-cell arrest before immunoglobulin rearrangement, as well as a functional role in T cell differentiation [39, 40]. The homeodomain transcription factor, *PBX1*, encoded on chromosome 1, is involved in the regulation of various embryonic processes [39, 41]. *PBX1* interacts with HOX proteins, regulating expression of developmental genes involved in haematopoiesis, suggested to be an early stage regulator of precursor commitment to B-cell lineage [41]. *TCF3-PBX1* fusion protein effects transcriptional activation, it can interact with HOX proteins inhibiting gene regulation, and reduces normal *TCF3* activity [39].

The translocation t(9;22)(q34;q11), creates a BCR-ABL1 fusion, known as the Philadelphia Chromosome, and is most commonly associated with adult Chronic Myeloid Leukaemia (CML) but is also found in childhood ALL and AML [42, 43]. *BCR-ABL1* is found in 3% of B-ALL cases [9]. Patients with Philadelphia positive ALL have a poorer prognosis than Philadelphia negative ALL and are often treated with a combination of tyrosine kinase inhibitors and chemotherapy [42, 44]. The *ABL1* gene on chromosome 9 encodes a tyrosine kinase that is joined to the *BCR* gene on chromosome 22, resulting in aberrant tyrosine kinase activity [42]. The *BCR* gene can break in different places (major and minor breakpoint cluster regions) resulting in different sized fusion genes, the most common in ALL being the 190 kDa *BCR-ABL1* fusion protein [42].

1.1.3 Chromosomal translocations associated with Acute Myeloid Leukaemia

Most AML cases harbour a variety of chromosomal alterations including high hyperdiploidy and recurrent translocations. The most common translocations found in AML are shown in Table 1.2.

Table 1.2. The most commonly occurring chromosomal translocations found in childhood AML and their frequency [3, 45, 46].

Chromosomal translocation	Fusion product (also known as)	Frequency (%)
t(8;21) (q22;q22)	*RUNX1-RUNX1T1 (AML1-ETO)*	*10-15*
11q23	*KMT2A* (*MLL*)	*8-14*
t(15;17)(q22;q21)	*PML-RARA*	*10-15*

In AML, *KMT2A* (*MLL*) translocations are found in up to 14% of cases [46]. In infants, *KMT2A* translocations account for 65% of AML cases [3, 46]. Unlike the same translocation in ALL, which has a poor prognosis, *KMT2A* translocations in AML are associated with intermediate risk [46]. In AML, *KMT2A* is found with fusion partners such as the *MLLT*3 and *MLLT10* genes [47]. These translocation partners may play an important role in leukemogenesis through Dot1, a histone methyltransferase, that has been shown to interact with *KMT2A* fusion partners and activate oncogenes [48, 49].

The *RUNX1-RUNX1T1* translocation occurs in 15% of all AML cases, involving the same translocation partner gene *RUNX1*, which is found in childhood ALL cases [3, 45]. This cytogenetic marker has been associated with good prognosis in childhood AML [50]. *RUNX1* Partner Transcriptional Co-Repressor 1 (*RUNX1T1*) is found mutated in many different cancers [51]. It acts as a transcriptional repressor by interacting with transcription factors [51]. The *RUNX1-RUNX1T1* fusion protein interferes with normal *RUNX1* activity and has been shown to mediate alternative splicing and chromatin remodelling [52].

The *PML-RARA* gene is another common translocation found in acute promyelocytic leukaemia (APL) and accounts for 10-15% of AML cases [50]. The *RARA* gene on chromosome 15 encodes retinoic acid receptor alpha, a nuclear transcription factor that is involved in myeloid cell differentiation [53]. The promyelocytic leukaemia gene, *PML*, located in chromosome 17, interacts with many proteins in different pathways to regulate tumour suppression and haematopoietic stem cell differentiation [53]. The fusion protein formed from the *PML-RARA* fusion gene acts as an oncoprotein with altered retinoic acid receptor function and transcription repression [53, 54].

1.1.4 Evidence for the “two-hit” hypothesis and *in utero* development of chromosomal translocations in leukaemia

Due to the early onset of childhood leukaemia and concordance in twins, evidence has grown to support the hypothesis that many of the initiating translocations that lead to childhood leukaemia occur *in utero* [55]. Using cord blood and Guthrie cards (blood from newborn heel pricks), multiple studies have shown that leukaemia associated translocations occur in the population at a range of frequencies, from 0.01% to ~8% [56]. One study using reverse transcription PCR, found an incidence of 1% for *ETV6-RUNX1* translocations in cord blood of newborns with normal development [10], and another study found the incidence of the same translocation to be 5% using a DNA based method called GIPFEL [56]. GIPFEL, genomic inverse PCR for exploration of ligated breakpoints, utilises restriction enzymes and ligases to cut genomic DNA and allow for ligation of known points in the presence of a translocation, which can then be detected through multiple PCR, qPCR and Sanger

sequencing steps [57]. A study using cord blood and spleen tissue samples from aborted foetuses, found samples positive for the *ETV6-RUNX1* translocation, at a rate of 2%, which provides further evidence for the initiating translocations occurring *in utero* [58]. With the incidence of *ETV6-RUNX1* positive leukaemia at 0.01-5%, the presence of the *ETV6-RUNX1* translocation in the population is estimated to be at least 100-500-fold higher than incidence of leukaemia.

The development of childhood leukaemia is thought to occur through a “two-hit” hypothesis which has been described and updated by Greaves *et al.* throughout the years [3, 10, 59-61], in which a first-hit mutation occurs, such as a chromosomal translocation, but requires further molecular events, such as changes to DNA methylation, before the development into leukaemia [3, 10, 55, 59, 62]. This is similar to Knudson’s “two-hit” hypothesis, which was first described in Retinoblastoma, and suggests two mutations are required to occur in a single tumour suppressor gene [63, 64]. Knudson’s “two-hit” hypothesis suggests that mutations are required in both alleles of a tumour suppressor gene either through the gain of two individual mutations in somatic cells for the case of sporadic cancers, or in inherited cancers, a mutation is passed on in germ line cells and only requires a single gain of mutation in somatic cells. This subsequently leads to loss of tumour suppressor activity and cancer progression. In leukaemia however, the “two-hit” hypothesis described by Greaves focusses on the timing of mutations and this 2-step development does not necessarily need to occur in the same tumour suppressor gene, although many mutations are expected to be complementary to initiate disease progression, such as in some AML cases, where cooperation between FLT3 mutations and *RUNX1-RUNX1T1* chromosomal translocations is expected to contribute to the disease phenotype, [62, 65, 66]. Evidence of the development of leukaemia via Greaves’ “two-hit” hypothesis comes in part from the difference between chromosomal translocation presence in the population and the incidence of leukaemia, suggesting that further mutations are required following the “initiating” chromosomal translocations to induce leukaemia, which is why disease incidence is less than translocation rates in the population. This is supported by the early onset of disease, in such that the initiating “hit” is caused *in utero* with further “hits” occurring *in utero* or early childhood. The biology of cancer is often summarised by six key hallmarks that underlie the progression of cancers: resisting cell death, sustaining proliferative signalling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality and inducing angiogenesis [67]. Childhood leukaemias mainly derive from naïve stem-like pre-B and pre-T cells that remain immature as they divide, developing blast-like leukaemic precursors [68]. These pre-cursor cells already encompass traits described by the hallmarks of cancer, such as the ability to travel in the blood stream and throughout the body, as well as growth and

extravasation. This suggests that these pre-cursor cells do not need to gain as many mutations as other forms of cancer, providing an explanation for the short latency period, with mutations occurring during development and childhood [68].

Further support that the first “hit” occurs *in utero* comes from studies of leukaemia in twins. Fusion genes have been found before and at birth in circulating blood and retrospective neonatal blood spots from twins with concordant leukaemia [69]. Monzygotic twins share a blood supply, so concordant leukaemia in twins may be initiated by a mutation in one twin that passed to the other twin through the placenta. The same clonal expansions and markers have been found within monozygotic twins, further supporting the idea that initiating events occur *in utero* [69]. In a study of triplets, a monozygotic pair sharing a placenta developed concordant ALL, whereas the 3rd dizygotic triplet with a separate placenta was leukaemia free [70]. The *ETV6-RUNX1* translocation was PCR amplified in neonatal blood spots from all 3 triplets, but blood taken after leukaemia diagnosis at 21 months, showed that the translocation was present in the monozygotic pair, along with an *ETV6* deletion, but not in the 3rd healthy triplet [70]. The author’s of the study addressed that the disappearance of the translocation in the 3rd triplet could, although unlikely, be due to laboratory cross-contamination during the neonatal blood spot testing, or could suggest that fetal cells are able to travel in the maternal circulation from one placenta to the other, and the pre-leukaemic cells in the 3rd triplet were eliminated after birth [70]. A further study of monozygotic triplets found that all three triplets developed ALL, with shared rearrangements and further subclonal differences, suggesting that monozygosity and sharing a placenta are important risk factors for multiple birth leukaemia [71, 72].

Concordant ALL in monozygotic twins is found at a rate of 5-10%, with a variable latency, from infant to late teens [69]. A study of a twin set with the same *ETV6-RUNX1* fusion translocation, showed twin one was diagnosed with ALL at 5 years old, whilst twin two was haematologically normal, but twin two was then diagnosed with ALL at 13 years old [55]. Historic bone marrow from both twins at the time of twin one diagnosis showed that both twins had the same translocation at that time. This variability in the onset of leukaemia may be due to the requirement for further postnatal events, which could occur at different times. A short latency period in infants may be propagated by the initiating mutation making the cell more susceptible to further mutations.

1.1.5 DNA methylation and leukemogenesis

DNA methylation plays a key role in the epigenomic landscape, providing essential gene regulation, allowing for gene expression in specific cell types. During DNA methylation a methyl group from the co-factor *S*-adenyl methionine is transferred to a cytosine nucleotide

that directly precedes a guanine nucleotide, known as a CpG site [73]. This reaction is catalysed by DNA methyltransferases (Dnmts), with *de novo* Dnmts catalysing initial methylation patterns on DNA, and maintenance Dmnts copying methylation patterns on new strands after replication [73, 74]. When stretches of DNA have a high CpG density, this is known as a CpG island, and they are frequently found in key locations of gene promoter regions, and as such are often not methylated [73]. This is key to the role DNA methylation plays in regulating gene expression during development, with methylation of CpG islands generally associated with silencing of gene expression [73, 75, 76]. DNA methylation interacts with a variety of binding proteins which can modify and maintain genomic integrity and alter transcriptional activity [77]. By influencing chromatin structure, DNA methylation can affect the accessibility of DNA to enzymes and potentially damage inducing molecules [77].

Altered methylation has key links to many diseases, including cancer, where aberrant hypermethylation has been associated with silencing of tumour suppressor genes [78]. Methylated cytosine can also spontaneously deaminate, resulting in a mismatch mutation of uracil or thymidine to the original guanine pair, which have been associated with various cancers, including AML [77]. Altered DNA methylation has also been observed in childhood ALL [79]. A study in paediatric B-cell ALL patients observed common DNA methylation signatures that were found across all subtypes, and in individual subtypes, with hypermethylation the predominant alteration [80]. This study found deregulation of genes involved in the prevention of apoptosis and increased proliferation which could play a vital role in disease progression [80].

1.2 Environmental factors associated with an increased risk of childhood leukaemia

Incidence of childhood leukaemia is increasing year by year [81], with a European population-based study showing that the incidence of childhood leukaemia has increased by nearly 1% each year from 1970 to 1999 [82]. In economically developed countries, leukaemia accounts for around 33% of total children's cancer, a higher percentage than in lower economically developed countries, where leukaemia accounts for less than 20% of children's cancers [83]. Some ethnicities, such as Latino children vs. non-Latino white children living in California, USA, show a higher incidence of childhood leukaemia [68]. Studies in California observed that Latino children may have a higher exposure to environmental hazards associated with an increased risk of childhood leukaemia. Higher levels of persistent organic pollutants including polychlorinated biphenyls, were found in dust at the homes of Latino children, and of fathers exposed to occupational pesticides, the majority were parents to Latino children [84, 85]. Once thought to be a disease of genetic

origin, there is growing evidence to suggest that various environmental factors may be associated with childhood leukaemia.

Epidemiological studies allow us to identify risks associated with diseases. These studies have allowed for the identification of many different environmental exposures as risk factors for the incidence of childhood leukaemia. Factors implicated in an increased risk of childhood leukaemia include maternal and paternal smoking, maternal alcohol consumption, maternal caffeine intake, maternal nutrition, such as folate deficiency during pregnancy, exposure to pesticides, paints and solvents, and a developing immune system [14]. There are limitations to the use of epidemiological studies in assessing the exposure-risk relationship due to the limited case numbers due to the rarity of childhood leukaemia. As many epidemiological studies are performed retrospectively, results may be influenced by reporting bias, with parents unable to accurately recall exposures during pregnancy which may have occurred many years previously [86]. Some parents may also be unwilling to report exposure due to the perceived stigma of blame associated with the exposure, such as smoking during pregnancy [87, 88].

To date, there is little understanding of the molecular role these exposures play in the induction of childhood leukaemia associated chromosomal translocations or epigenetic changes, such as DNA methylation.

1.3 Folate deficiency

1.3.1 Folate and one carbon metabolism

Folates are B-vitamins found naturally in many foods such as vegetables, fruit and eggs, and synthetically as folic acid in supplements and refined grain products including cereals and flour. Folic acid is a monoglutamate folate parent structure, whereas most dietary folates are polyglutamated [89]. Once ingested and entered into the small intestine, polyglutamated dietary folates must be converted to monoglutamyl folates through hydrolysis catalysed by glutamate carboxypeptidase II at the intestinal apical brush membrane border. Transport of monoglutamated folates into cells can then be facilitated by 3 different transporters: reduced folate carriers, with a high affinity for reduced folates; folate receptors, with a high affinity for folic acid; and proton coupled folate transporters, with high affinity for folic acid and reduced folates at low pH, as described in Figure 1.3 [90]. Proton-coupled folate transporters are widely expressed, with high expression seen in the duodenum and jejunum of the intestinal tract, where a low pH is optimum for folate transport [91]. Passage across the apical brush membrane into the vascular system is largely facilitated by proton-coupled folate transporters [90, 91]. Monoglutamated folates are then transported to the liver where they can either be polyglutamated for storage, secreted into bile or transported to systemic

tissues [89]. Reduced folate carriers are ubiquitously expressed, the primary route for folate transport to systemic tissues in neutral pH [92]. Folate receptors are expressed in epithelia, hematopoietic tissues and T-cells, and work through endocytosis-mediated transport, and are often found highly expressed in leukaemia and tumour cells [91, 92]. Once inside cells, folates are returned to a polyglutamated state by foly-poly-glutamate synthase, allowing folates to continue as a substrate for enzyme dependent reactions [90].

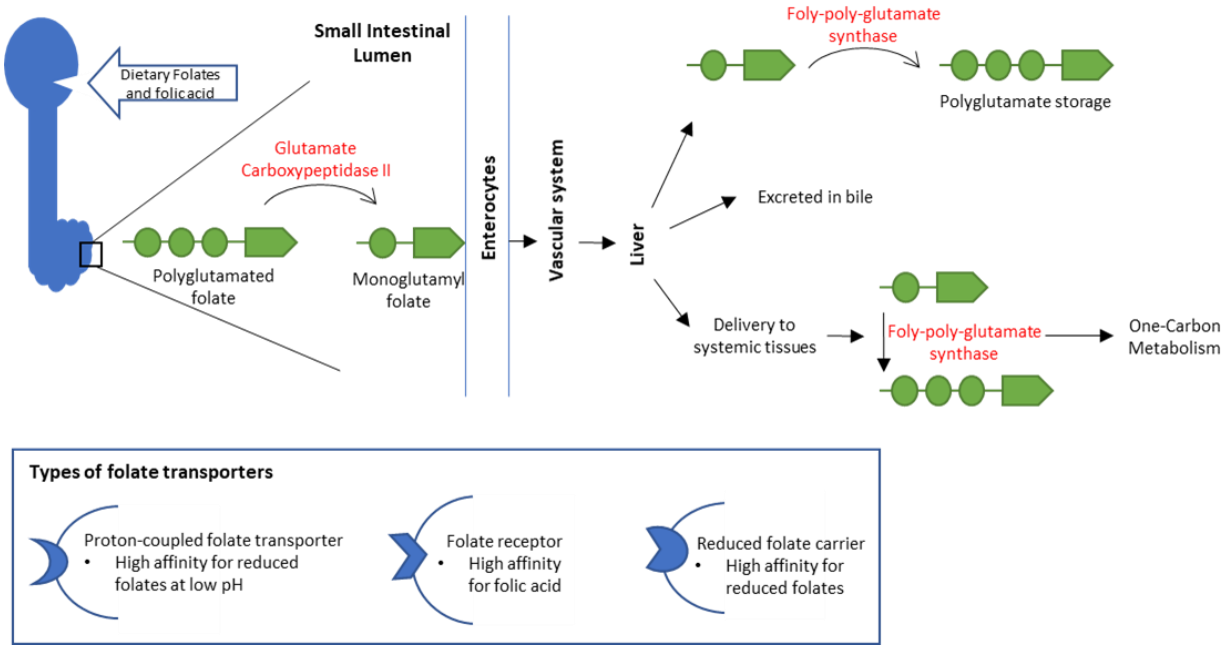


Figure 1.3. Overview of folate transport, with enzymes in red [89, 91, 93, 94]. Most dietary folates enter the body as polyglutamated 5-methylTHF and are subsequently hydrolysed by glutamate carboxypeptidase III at the intestinal apical brush membrane border into monoglutamyl form. Folic acid, a monglutamate, is the main folate source found in vitamins and cell culture medium. These folates are transported into cells via 3 transporters: folate receptors, which have a high affinity for folic acid, reduced folate carriers, which have a high affinity for reduced folates and proton coupled folate transporters which have a high affinity for folic acid and reduced folates at low pH. Monoglutamyl folates are primarily transported from the intestinal lumen across the apical brush membrane by proton coupled folate transporters on the enterocytes, and then enter the vascular system across the basolateral membrane. Folates are then directed to the liver and either stored as polyglutamed folates, secreted into bile or released into the bloodstream to be delivered to systemic tissues and polyglutamated for use in enzyme dependent reaction. Ubiquitously expressed reduced folate carriers are the main transporters of folates into systemic tissues, whereas folate receptors are found largely on epithelia, hematopoietic tissues and T-cells.

Metabolism of folate is important for many biological processes including DNA synthesis and repair. One carbon metabolism involves multiple enzymes and vitamin substrates involved in folate metabolism, allowing for one carbon moieties such as methyl, formyl and methenyl to be transferred, mediating methylation reactions and nucleotide synthesis. A simplified summary of the main pathways is shown in Figure 1.4 [95].

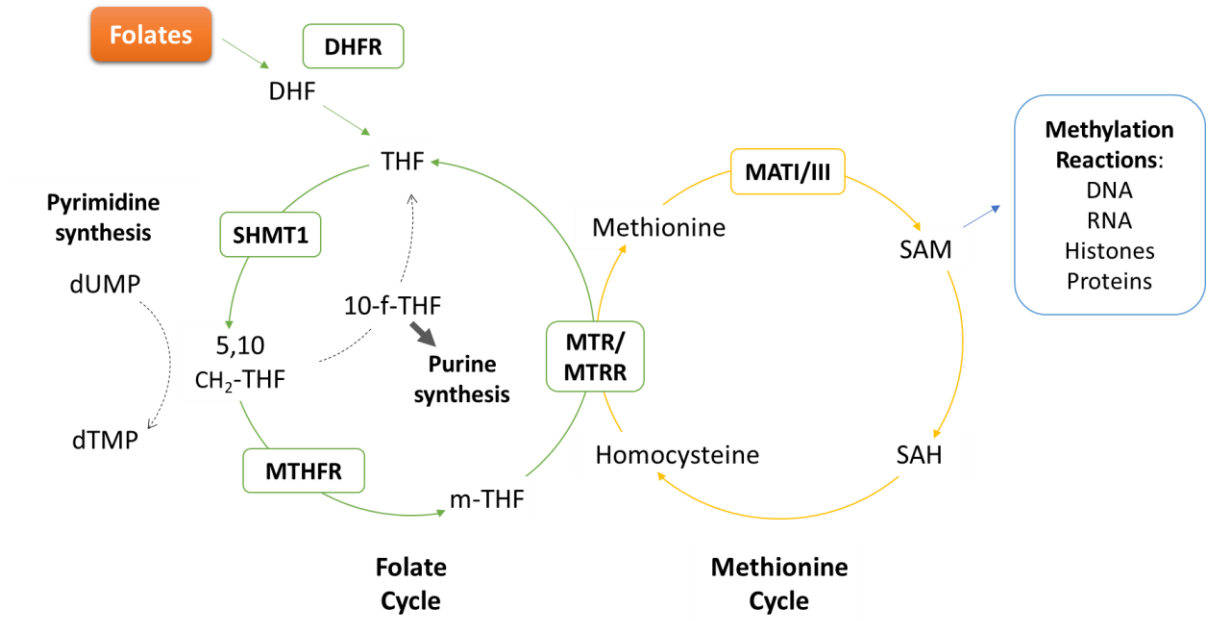


Figure 1.4. Simplified overview of one-carbon metabolism. Folate/methionine cycle enzymes in boxes: DHFR, dihydrofolate reductase; SHMT1, serine hydroxy-methyltransferase; MTHFR, 5,10-methylenetetrahydrofolate reductase; MTR, methionine synthase; MTRR, methionine synthase reductase; MATI/III, methionine adenosyltransferase. Substrates: DHF, dihydrofolate; THF, tetrahydrofolate; 5,10-CH_2-THF, 5,10-methylene-tetrahydrofolate; 5-mTHF, 5-methyltetrahydrofolate; 10-f-THF, 10-formyl-tetrahydrofolate; dUMP, deoxyuridine monophosphate; dTMP, deoxythymidine monophosphate; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine.

Once inside the cell, folic acid is reduced to dihydrofolate (DHF) and then tetrahydrofolate (THF) by DHF reductase [95]. The folate cycle continues with conversion of THF to 5,10-methylene-tetrahydrofolate (5,10-CH_2-THF) by serine hydroxy-methyltransferase (SHMT1) [95]. Finally, reduction of 5,10-CH_2-THF by 5,10-methylenetetrahydrofolate reductase (MTHFR) produces 5-methyltetrahydrofolate (5-mTHF), the substrate required for the methionine cycle [95].

Demethylation of 5-mTHF by methionine synthase reductase (MTRR) completes the folate cycle, returning 5-mTHF to THF. This allows for the transfer of a methyl group into the methionine cycle, converting homocysteine into methionine, which is then further adenylated to *S*-adenosylmethionine (SAM), the universal methyl donor [95]. Methylation is an important mechanism involved in many vital biological processes, catalysed by methyltransferases, which add a methyl group onto substrates such as DNA, histones, RNA and proteins. For example, DNA methylation is one epigenetic mechanism which provides regulation of genes through the addition of methyl groups onto CpG sites in the genome. Following methyl donation, SAM is converted into SAH, which is a competitive inhibitor of methyltransferases [90]. SAM is also an inhibitor of MTHFR, so high levels of SAM results in MTHFR inhibition and decreased production of 5-mTHF, therefore homocysteine is not converted to

methionine [90]. When there are low levels of SAM, homocysteine methylation is favoured, providing further regulation of the one-carbon cycle.

Nucleotide biosynthesis is also regulated by folate metabolism as 5,10-CH_2-THF is used as a substrate for the conversion deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP) during pyrimidine synthesis [95]. Further conversion of 5,10-CH_2-THF to 10-f-THF and then THF is also involved in the pathway for purine synthesis [95]. Nucleotide synthesis is required for effective DNA synthesis and for repair of DNA damage, and as the conversion of dUMP to dTMP is the only source of thymidine, it makes this reaction a rate limiting step in DNA synthesis. If dUMP is not converted and dTMP levels are low, then excess uracil can be incorporated into DNA instead [96]. Increased uracil misincorporation has been observed in lymphocytes from both animal and cell models grown in folate deficient conditions [97, 98]. In both studies, along with an increase in uracil misincorporation, an increase in DNA strand breaks was also seen [97, 98]. A study in *Trypanosoma brucei* showed that knockout of the enzymes dUTPase, an enzyme that hydrolyses dUTP to dUMP for use as a substrate in thymidine synthesis and is linked to increased uracil misincorportation, and Uracil-DNA glycosylase, an enzyme involved in the repair of uracilated DNA, leads to increased levels of uracil incorporation which showed a significant increase in DNA fragmentation and mutations thought to be due to attempted excision and repair of the mis-incorporated uracil [99].

1.3.2 Folate requirements, factors influencing folate status and associations between folate status and health

The recommended daily folate intake for adults is 400µg/day [100]. A lack of folate rich foods in the diet, such as leafy greens, fruit and beans, can impact on folate status. Due to folate deficiencies in the general population, many foods are now fortified with synthetic folic acid, such as white flour, rice, pasta and cereals to increase folate levels through foods that are regularly consumed across the population. In terms of folate bioavailability from folate consumption in the diet, bioavailability from natural food is 50%, compared to 85% bioavailability for folic acid fortified foods [101]. Diseases such as alcoholism can contribute to folate deficiency due to impairments in folate carrier transcription, folate storage and increase in excretion [101]. Several medications have also shown to contribute to folate deficiency such as the folate antagonist methotrexate, used to treat some cancers and autoimmune diseases, anti-convulsants and malaria medication [101].

Disease outcomes that have been associated with folate deficiency include vascular diseases, depression, neural tube defects (NTDs) and Down's syndrome [90]. Folate deficiency has been linked to an increased risk of several cancer types such as cervical,

lung and breast cancers, through polymorphisms, altered DNA methylation and uracil misincorporation, with some studies showing that high folate intake reduces the risk of developing colorectal cancer [96]. However, an excess consumption of folic acid has been highlighted in potentially progressing established tumours, by providing nucleotides for proliferation and altering DNA methylation status [96].

Polymorphisms within genes involved in folate metabolism and one-carbon metabolism pathways have been shown to have an influence on folate status and impact on biological processes leading to disease [90]. For example, individuals harbouring homozygous polymorphisms in the *MTHFR* gene have presented with lower folate levels in plasma compared to individuals who are homozygous wild type [90, 102].

1.3.3 Influence of folate deficiency pre-conception and during pregnancy for offspring health

Plasma folate concentration in pregnant women decreases throughout pregnancy, in part due to the increased folate demand of the foetus as it is rapidly growing, but also due to haemodilution, renal function and hormonal changes as a result of the pregnancy [103]. Folic acid supplementation is recommended during pre-conception and pregnancy at 400µg/day to protect against NTD in infants [104, 105]. NTD's are caused by a malformity in the central nervous system during early gestation, such as anencephaly and spina bifida. An increased risk of NTDs has been associated with maternal folate deficiency and with numerous polymorphisms in folate metabolism genes, including *MTHFR* and *MTR* [106].

Many other adverse pregnancy outcomes have been associated with low maternal folate including placental abruption (which has been linked to folate gene polymorphisms and plasma homocysteine concentration), recurrent pregnancy loss, low birth weight and foetal growth restriction [103]. *In vitro* studies have shown that folate deficiency can increase the risk of aneuploidy, specifically in chromosomes 17 and 21, which are abnormalities associated with Down's syndrome and Alzheimer disease [107]. Meta-analyses have shown that polymorphisms in folate metabolism genes (which are associated with insufficient levels of biologically active folate) are a potential maternal risk factor for having a child with Down's syndrome, along with changes in methylation status of the infant [108]. Studies have also shown that folic acid supplementation can reduce the risk of cardiovascular malformations [109]. Folate deficiency in men may also contribute to an increased NTD risk during pregnancy due to gamete exposure to dioxins that can cause mutations in spermatazoids from folate deficiency mechanisms [109]. Folate intake during pregnancy has also been suggested to influence childhood leukaemia risk.

1.3.4 Folate deficiency and childhood leukaemia

Epidemiological studies have suggested that folic acid supplementation before and/or during pregnancy lowers the risk of developing leukaemia in the offspring, although there has been some disparity between studies [96]. The ESCALE study used telephone interviews to collect data on maternal folic acid and vitamin supplements (that do or do not contain folic acid) from 764 acute leukaemia cases and 1681 controls, whilst also investigating some samples for polymorphisms [110]. This study found that maternal folic acid supplementation was inversely associated with acute leukaemia in offspring [110]. Polymorphisms in folate cycle genes *MTHFR* and *MTRR* (any *MTHFR* homozygous polymorphism plus heterozygous *MTRR* polymorphisms A66G and C524T), which lead to reduced enzyme activity and decreased availability of biologically active folate, were positively associated with acute leukaemia and ALL [110]. A further study, ESTELLE, also used telephone interviews to collect data on folic acid supplementation 3 months pre-conception, 1st trimester and 2nd/3rd trimester, from 747 cases and 1421 controls [111]. The ESTELLE study did not find any association with folic acid supplementation during pregnancy and childhood leukaemia but did find a weak association between maternal folate acid supplementation pre-conception and a reduced risk of childhood leukaemia in the offspring [111].

The Australia-ALL study used mailed questionnaires to collect data on maternal folate supplements taken pre-pregnancy, 1st trimester and 2nd/3rd trimester, from 416 ALL cases and 1361 controls [112]. This study found no association between folic acid supplementation at any stage of pre-conception or pregnancy, which was in disagreement with an earlier study in Western Australia, which had 83 cases and 166 controls, and found an association between folate supplementation and childhood leukaemia risk after asking mothers about their medication usage during pregnancy [112, 113]. Different outcomes from these studies may be due to: reporting or recall bias from the mother as data was collected retrospectively; variation in levels of consumed folate across geographical regions; many of the studies collected data on the use of vitamin supplements that may or may not have contained folate; and differences in experimental design which could all influence the overall conclusion of these studies. Differences in case numbers can also contribute to different outcomes, as generally larger studies are more robust, however with vague exposure measures, such as supplementation vs. no supplementation, the differences in dietary folate status and genetic variation, larger numbers may just increase the range of heterogeneity. Meta-analyses can help to increase power and streamline heterogeneity across populations allowing for more robust analysis of risk factors.

Due to the rarity of childhood leukaemia, the number of positive cases in these studies can be quite low, especially for the rarer AML. Metayer *et al.* 2014 conducted a pooled analysis of 12 case-controlled studies, from 10 countries, looking at self-reported folic acid and

vitamin supplement intake (that do or do not contain folic acid) before and during pregnancy [114]. The pooled analysis included information for 6,963 cases of ALL, 585 cases of AML and 11,635 control participants, overcoming the limitations of many smaller studies with very few case numbers, especially in the less prevalent AML subset [114]. Only 7 of the studies focussed specifically on folic acid supplementation, accounting for 47% of the total subjects [114]. Specific analysis was conducted for ALL and AML separately. Analysis for ALL showed that maternal consumption of vitamin supplements and folic acid supplements reduced the risk of ALL, with no significant difference for when the supplements were taken (i.e. preconception and during pregnancy). For AML, vitamin supplementation overall was not significant in reducing AML risk, however folic acid supplementation was specifically related to reduced risk, suggesting folate levels may be key for the rarer AML subtype.

One Californian based study observed that ALL risk reduction was greatest in children of Hispanic mothers who consumed B vitamin supplements a year before pregnancy [115]. Since Mexican-Americans have a higher prevalence for *MTHFR* polymorphisms than non-Hispanic Americans, which is likely to affect their available active folate levels, this may have been the reason for this observation [116]. Indeed, a study of genetic variants in the folate metabolism cycle found that a homozygous *MTR* 2756 polymorphism was significantly associated with an increased risk of ALL and AML and was most pronounced for leukaemias that had a *KMT2A* translocation [117]. The MTR enzyme is involved in the transfer of a methyl group from the folate substrate 5-mTHF onto homocysteine to produce methionine, required for methylation. Variations in MTR could reduce the amount of methyl transfer, and therefore affect the regulation of methylation, a scenario that is likely mimicked by folate deficiency [118]. Data from these studies highlights the complexities of the relationship between folate and risk and therefore the weakness of using supplementation alone as a measure of exposure.

1.3.5 Mechanisms through which folate deficiency may contribute to leukemogenesis

In vitro and animal studies may offer support to suggest how maternal folate deficiency may mechanistically influence leukaemia risk in the offspring via its involvement in nucleotide biosynthesis and methylation, which in turn may contribute to the chromosomal translocation events.

The role of folate in both nucleotide synthesis and methylation of biological molecules could play a part in leukemogenesis. Folate deficiency can impact on thymidine synthesis, limiting the conversion of dUMP to dTMP. If dUMP is not converted, then DNA synthesis is slowed down and excessive uracil can be incorporated into DNA. This can lead to DNA strand

breaks, point mutations and micronuclei formation, preventing effective DNA synthesis and repair [96].

A study by Blount *et al.* 1997 found that the uracil content of bone marrow DNA was 9-fold higher in folate deficient patient samples compared with individuals with normal folate plasma levels but could be reduced with folic acid supplementation [119]. The study also found the that folate deficient patients had higher levels of micronucleated reticulocytes and erythrocytes, associated with double strand DNA breaks, that could be reduced with folic acid supplementation [119]. Furthermore, mice fed a folate deficient diet have been observed to have higher frequencies of micronucleated reticulocytes and erythrocytes, they also had an increased mutation frequency in bone marrow [120]. Additionally, an *in vitro* study in prostate cancer cells found that mild folate depletion increased uracil misincorporation and single strand breaks, and induced of novel chromosomal rearrangements, with some translocations mapping to known fragile sites [121]. Aneuploidies influenced by folate deficiencies have also been seen in chromosome 8, a recurrent abnormality in AML [96, 122].

Structural aberrations, such as chromosomal translocations, can also be caused by altered DNA methylation, such as demethylation of centromeres leading to segregation abnormalities, altering gene expression, and chromatin remodelling resulting in breakages [96, 123]. A study by Potter *et al.* 2018 observed significant overlaps between altered methylation associated with childhood ALL and altered methylation associated with a mouse model of maternal folate depletion [124]. DNA methylation was measured for a subset of these overlaps in ALL patient samples to confirm aberrant methylation, and then measured in a cohort population where folate status was known, observing an inverse association was between mean percentage methylation in infants with red blood cell folate status (maternal 1st trimester) and mean percentage methylation with vitamin B_{12} status (cord blood) in ALL samples for 2 of 4 genes, observing hypermethylation with low folate status [124]. An epigenome wide association study, using a meet in the middle approach, also found significant overlaps between altered CpG methylation in ALL cases and maternal folate status [125]. Maternal folate deficiency has also been observed to impact telomeres lengths, with shorter telomeres in newborns [96]. Telomeres are important for maintaining chromosome stability, and are rich in thymidine, and as such may be prone to uracil incorporation in folate deficient conditions, leading to chromosome damage.

Due to its influence on DNA synthesis and repair, folate deficiency may make cells more susceptible to mutagenic exposures that can cause mutations in DNA. A mouse study investigating the effect of benzene on folate levels found that a low folate status was

exacerbated by benzene and exposure to benzene further increased the level of micronuclei in folate deficient mice, a measure of chromosome damage [126]. Moreover, *in vitro* studies using Chinese hamster ovary cells observed that cells grown in folate deficient medium had higher levels of DNA damage when exposed to low dose ionizing radiation compared to cells grown in high levels of folate [127].

Although evidence suggests that folate might influence various mechanisms involved in the leukaemogenic process, to the best of our knowledge no studies have linked folate with initiating events i.e. chromosomal translocations.

1.4 Caffeine

Caffeine is a natural alkaloid (organic substance containing basic nitrogen) consumed in coffee, tea, carbonated drinks and chocolate, and is also added as a stimulant to many pain medications, making it one of the most widely accessible psychoactive substances worldwide [128]. Caffeine is absorbed rapidly from the gastrointestinal tract into the circulatory system, where it can distribute throughout the whole body, crossing brain, placental and testicular barriers, reaching peak plasma concentrations an hour after consumption, shown in Figure 1.5 [129]. The main enzyme responsible for caffeine metabolism is cytochrome P450 1A2 (CYP1A2), which is found in the liver and is responsible for demethylation of caffeine into its metabolites. It plays a further role in the metabolism of caffeine metabolites before excretion into the urine, clearing caffeine from plasma in an average of 2.5-5 hours [130]. Some factors have been shown to affect the activity of CYP1A2, such as smoking which can induce CYP1A2 activity, increasing caffeine metabolism, and oral contraceptives which have been shown to double caffeine clearing time, slowing CYP1A2 activity [131-134]. Only 2% of caffeine is excreted in the urine unmetabolised, so CYP1A2 activity is the limiting factor for caffeine clearance from the body [129]. Metabolites of caffeine include: paraxanthine, which is biologically active with similar effects as caffeine; theobromine, which has effects on diuresis, cardiovascular and glandular processes; and theophylline, a potent caffeine homologue [129].

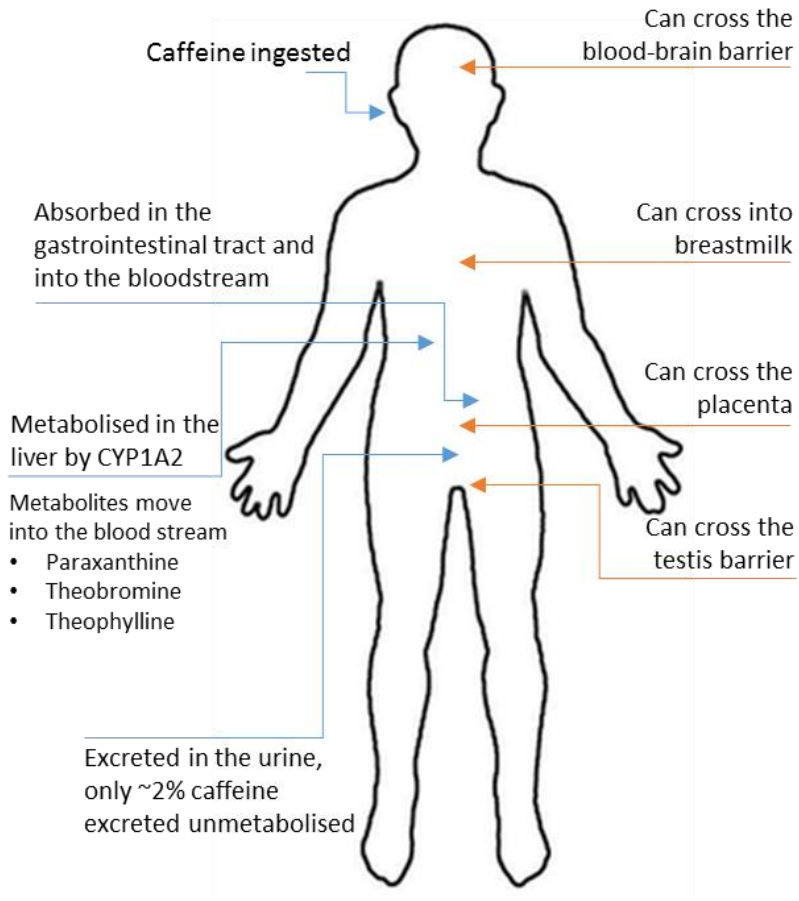


Figure 1.5. The pathway of caffeine metabolism in the body

1.4.1 Caffeine consumption during pregnancy

Caffeine consumption during pregnancy has been associated with a variety of adverse effects. During pregnancy, CYP1A2 is partially inhibited in the mother, decreasing CYP1A2 activity and increasing the half-life of caffeine up to 15 hours [135, 136]. As caffeine can freely cross the placenta, and neither the foetus or the placenta possess the CYP1A2 enzyme required for caffeine metabolism, the foetus is likely to be exposed to caffeine for long periods of time depending on maternal intake [137]. Caffeine is also able to enter milk produced by breastfeeding mothers [130]. Epidemiological studies have shown an association with maternal caffeine consumption and an increased risk of low birth weight and loss of pregnancy, supported by further evidence from animal and *in vitro* studies [136, 138, 139]. Recommendations are already in place in the UK to limit caffeine intake during pregnancy, with a recommend daily limit of 200mg caffeine, equivalent to 2 mugs of instant coffee [140].

Animal studies have gained mechanistic insight into the impact of caffeine on pregnancy [136]. A range of studies have shown that caffeine exposure can influence delayed conception, pregnancy outcomes and beyond. Mice exposed to caffeine prior to conception

had impaired embryo implantation leading to reduced fertility, which could be a result of delayed oviductal embryo transport or altered regulation of steroid hormone-regulated genes [136].Caffeine exposure during pregnancy has been shown to impact cardiovascular development and growth, which could affect heart health in later life, and can reduce blood flow to the uterus which may decrease the transfer of nutrients and lead to intrauterine growth retardation (IUGR) [136]. Furthermore, exposure to caffeine during pregnancy in humans has been associated with adverse outcomes on the long-term health of the foetus such as childhood leukaemia, childhood obesity, impaired cognitive development and cardiovascular and metabolic diseases in adulthood [136].

1.4.2 Maternal caffeine intake and childhood leukaemia

A pooled analysis of 5 epidemiological studies have shown an association between high coffee consumption in pregnant mothers with an increased risk of childhood leukaemia [141]. Investigating these 5 individual studies in more detail, all data was collected retrospectively following diagnosis of patients under the age of 15 years old with childhood leukaemia, with 4 studies in France and 1 study in Australia. Menegaux *et al.* 2005 [142] conducted face-to-face interviews to collect data on consumption of tea, coffee and cola during pregnancy and found an association for childhood leukaemia with increasing coffee consumption, but not with tea or cola. This study contained the fewest cases and controls, at 280 and 288 respectively, with data collected from 4 regions in France, and analysis conducted at 80% power and 5% significance [142]. The controls for this study were chosen from children with matched age admitted to the same case hospitals for reasons other than cancer and birth defects [142]. A larger study by Menegaux *et al.* in 2007 covered over 14 regions in France with 472 cases and 567 controls, with controls randomly selected to match for age, gender and region [143]. This study used self-reported written maternal questionnaires recording cups of coffee consumed per day during pregnancy (caffeinated or decaffeinated not differentiated), rather than tea and caffeinated sodas, and found a slight association between more than 3 cups of coffee consumed per day and risk of childhood leukaemia using odds ratios and 95% confidence intervals [143].

The 2 largest studies, ESCALE and ESTELLE, both conducted in France, employed maternal telephone interviews, with close to double the number of controls to case participants [144, 145]. The ESCALE study had 764 participants, with 1681 controls selected randomly to match age and gender, with analysis conducted at 80% power and 5% significance [144]. Results from the ESCALE study, where data was collected for consumption of coffee, tea and cola during pregnancy and breastfeeding, showed an association between childhood leukaemia and increasing quantities of coffee consumed

during pregnancy, no association for tea consumption, but a slight association between ALL and maternal cola consumption [144]. The ESTELLE study had 747 participants and 1421 controls, with controls selected through quota sampling to match age and gender of childhood cancers, with an over-representation of infants less than 1 year old to gain power in this subset [145]. This study assessed weekly and daily coffee or tea consumption during the first trimester specifically and throughout the whole pregnancy, and found a significant association between mother's who consumed more than 2 cups of coffee per day and an increased risk of ALL [145].

Furthermore, an Australian study used written maternal questionnaires to assess tea and coffee consumption during the last 6 months of pregnancy, using odds ratio and 95% confidence intervals [146]. Smaller than the ESCALE and ESTELLE studies, this study had 337 case participants but again had double the number of controls to cases with 697 controls matched to age, gender and location [146]. Results of the study found little association between tea/coffee consumption and ALL, with an OR (95% CI) of 0.87 (0.61, 1.24) for consumption of coffee and/or tea in all mothers [146]. However, when limited to ALL cases containing a balanced translocation, including *ETV6-RUNX1* and *KMT2A* translocations, accounting for 105 cases out of 337, there was some association seen between tea and/or coffee consumption in all mothers with an OR (95% CI) of 1.27 (0.67, 2.38), and this was observed more prominently in non-smoking mother's, especially when 2+ cups of coffee are consumed per day, with an OR (95% CI) of 2.55 (1.04, 6.22) [146].

1.4.3 Mechanisms through which caffeine may contribute to leukemogenesis

The potential mutagenic properties of caffeine have been discussed for many years, and early studies *in vitro,* exposing HeLa cells to caffeine concentrations between 0-10,000µg/ml of caffeine for 1 hour, found increasing chromosomal breaks and translocations with increasing caffeine concentrations [147]. A review of caffeine carcinogenicity by Porta *et al.* 2003, described numerous studies that have shown that caffeine can impact DNA repair, including altering checkpoints and increasing the carcinogenicity of other DNA damage causing compounds including ionising radiation and alkylating agents [148]. A specific example of this is from a study by Muller *et al.* 1996, where mouse embroys exposed to caffeine had persistent DNA damage over time as measured by the COMET assay following exposure to ionising radiation, compared to embryos which were not exposed to caffeine, suggesting that caffeine delays repair of DNA damage [149]. Caffeine is an inhibitor of phosphodiesterases and protein kinases such as ATM and ATR, which are components of cell cycle checkpoint pathways. DSBs are recognised by a complex of proteins including Mre11, Rad50 and Nbs1, which lead to activation of ATM, triggering DNA damage

responses [22, 150]. ATM activation leads to the phosphorylation of substrates including p53, Brca1 and Chk2, which allows for activation of cell cycle arrest and apoptosis [151]. ATR is important in replicating cells and responds to a range of DNA damage through recruitment to RPA-coated ssDNA via ATRIP, leading to the phosphorylation of Chk2 to slow cell-cycle progresion [151], Inhibition of ATM and ATR can be oncogenic by preventing delays in cell cycle checkpoint pathways, inhibiting repair time and allowing damage to continue [148]. Inactivation of ATM has been shown to affect genomic stability and cell viability, which could lead to chromosomal abnormalities [152]. The G2 checkpoint is important for detecting damaged cells and allowing in arresting cells for repair or apoptosis, however caffeine exposure has been shown to allow cells to evade this checkpoint, reducing the amount of time cells have to repair, and allowing entry to further phases of the cell cycle Furthermore, caffeine has been shown to disrupt HR, and a study by Oda *et al.* 2017 observed a potentiation effect between caffeine and chemotherapy treatments such as the alkylating agent cisplatin, increasing apoptosis and reducing the number of cells in S phase *in vitro,* increasing the anti-cancer effects of cisplatin [153]. The study exposed hepatocellular carcinoma cells, HepG2, to 3.8µg/ml of cisplatin in the presence or absence of up to 200µg/ml of caffeine and measured viability by CCK-8 assay, apoptosis via flow cytometry and expression of FANCD2-Ub and RAD51 by western blotting [153]. The results observed that in the presence of caffeine, apoptosis increased compared to cisplatin alone, cell viability decreased in cells exposed to cisplatin plus more than 100 µg/ml of caffeine, and expression of FANCD2-Ub and RAD51 was decreased in cells exposed to cisplatin plus caffeine compared to cisplatin alone [153]. FANCD2-Ub and RAD51 are proteins involved in the Fanconi anemia pathway, which is important for the activation of HR, suggesting that caffeine may play a role in HR regulation through inhibition of these proteins [153].

Through abrogation of G2 checkpoints, caffeine is able to override etoposide-induced G2 arrest allowing cells to progress to mitosis without repairing DNA [154, 155]. Caffeine has also been shown to impair the resection of DSBs through inhibition of single strand annealing, preventing formation of the replication proteins RPA and Rad51 foci, which are involved in homologous sequence searching and strand invasion during homologous repair of DSBs [156]. *In vitro* studies have also suggested that caffeine acts as a topoisomerase II antagonist [157]. Topoisomerases form a complex with DNA, inducing DSBs to aid the unwinding of DNA during replication, before religation. Topoisomerase antagonists prevent topoisomerase II covalently binding to DNA, which could lead to chromosomal abnormalities due to DNA being unable to rewind. As a topoisomerase inhibitor, caffeine containing dietary factors have been associated specifically with AML leukaemia, where 80% of cases contain 11q23 abnormalities [158]. Other topoisomerase II inhibitors such as etoposide, which acts

as a topoisomerase II poison, are associated with translocations including 11q23, which lead to therapy-related leukaemias [159].

1.5 Smoking

It is estimated that over 1.3 billion people worldwide use tobacco, most commonly smoked as cigarettes, with over 80% of users from low- and middle-income countries (LMIC) [160]. Although smoking levels are decreasing in high-income countries, smoking incidence in LMIC's are increasing [161]. Over 8 million deaths per year are associated with tobacco use, from both direct use and from second-hand smoke exposure [160]. The main causes of death from tobacco smoke include cardiovascular disease, respiratory disease and cancer [162]. Smoking is associated with at least 15 different types of cancers, accounting for 70% of lung cancers in the UK, and is associated with an increased risk of adult AML, specifically with cases of t(8;21) translocations [163-165]. Tobacco smoke contains over 60 known carcinogens including nitrosamines, benzopyrene and benzene, as well as toxic gases such as carbon monoxide [162, 166]. Smoking cigarettes allows the smoker to ingest the addictive drug, nicotine, which makes smoking cessation difficult for many people [162].

Although smoking among women is less prevalent than smoking among men worldwide, it is recommended that pregnant women stop smoking completely during pregnancy due to the detrimental effects on the foetus [167]. Maternal cigarette smoking is associated with infertility and placental complications during pregnancy, including placental abruptions [168]. Maternal smoking also increases the risk of foetal growth restrictions, early and still births and sudden unexplained death in infancy [168]. Furthermore, smoking during pregnancy has been associated with lower folate levels and NTDs, along with an increased risk of other adverse health outcomes for the child including respiratory problems, cardiovascular disease, birth defects and cancer [169].

1.5.1 Smoking and the risk of childhood leukaemia

Many epidemiological studies have been conducted to investigate the association between tobacco smoke exposure and childhood leukaemia, with varied and often contradicting results for maternal and paternal risks, as shown in Figure 1.6. A meta-analysis of 19 case-controlled studies examined the effect of parental smoking on the risk of both childhood ALL and AML, concluding that no association is seen for maternal smoking and increased risk of ALL or AML [170]. Conversely, the study did find an association between paternal smoking and increased risk of ALL, particularly in the paternal preconception smoking subgroup [170]. However, results from individual studies vary in their associations.

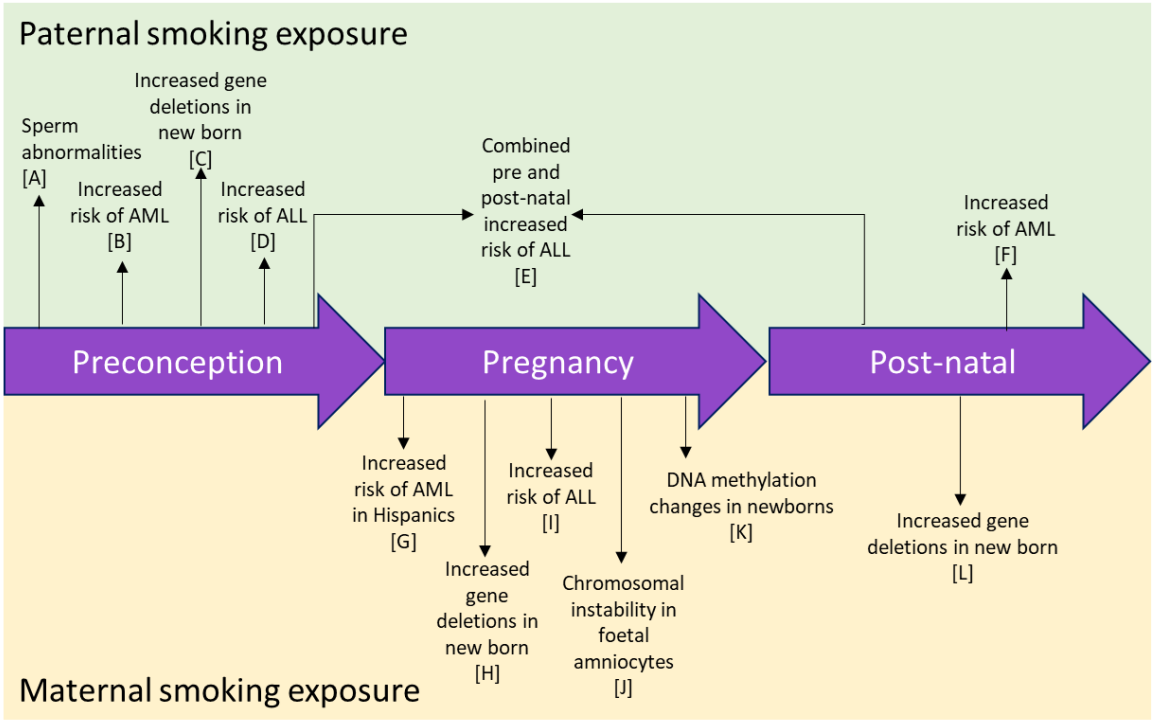


Figure 1.6. Differences in risk factor associations for ALL and AML seen between paternal and maternal smoking exposure at different stages throughout offspring development. Arrows denote the associated risk for each stage, with E denoting an increased risk of ALL when paternal smoking preconception and post-natally are combined. A: [171], B: [145, 172], C: [173], D: [170, 174, 175], E: [172, 176], F: [177], G: [177], H: [173], I: [174], J: [178], K: [178], L: [173].

One of the earliest studies, conducted in Colorado, John *et al.* 1991, used home visit interviews to collect maternal and paternal preconception and pregnancy smoking data from 73 leukaemia cases with random age matched controls [174]. It was one of the first studies to suggest an association between maternal smoking throughout pregnancy and increased risk of childhood ALL, along with an increased risk with paternal smoking preconception [174]. However, more recent studies, which include larger numbers of leukaemia cases, have found no association between maternal smoking and increased risk of childhood leukaemia. French study ESTELLE used telephone interviews to assess parental cigarette consumption preconception and during pregnancy in 747 childhood leukaemia cases and 1421 control cases [145]. The ESTELLE study concluded that there was no association with maternal smoking and increased risk of childhood leukaemia, but an association was seen for paternal smoking and increased risk of childhood leukaemia, with a higher association for AML over ALL, and for preconception over pregnancy exposure [145].

This was further supported by an Australian ALL study which used questionnaires to evaluate parental smoking during pregnancy and prior to conception for 388 cases and 868 controls, and again found no association with maternal smoking, but did see an association between paternal smoking preconception and an increased risk of ALL [175]. A similar sized study in Northern California used questionnaires to determine maternal and paternal

smoking exposures in 327 cases and 416 controls, finding an association between paternal smoking preconception with a significant increased risk of AML and a non-significant increased risk of ALL [172]. This study also found that the risk of ALL was greater when paternal smoking preconception was combined with children exposed to postnatal passive smoking [172]. A further study in Northern California in 2013 again showed an increased risk of ALL when paternal smoking preconception was combined with postnatal passive smoking, but also showed that this combination was specifically associated with the cytogenetic subtype t(12;21)/*ETV6-RUNX1* [176]. Although not significant, the study also found an increased risk of AML with recurrent chromosome aberrations associated with prenatal parental smoking and postnatal passive smoking [176]. Second-hand smoke exposure has been shown to worsen the overall survival of children diagnosed with ALL [179]. Maternal smoking during pregnancy and breastfeeding and paternal smoking preconception has been associated with gene deletions in children with ALL, with significantly high levels of deletions in cases with t(12;21)/*ETV6-RUNX1* translocations [173].

The variation seen between parental smoking risk factors and childhood leukaemia could be due to individual studies being conducted in different geographical locations, where other potential compounding factors, such as genetics and environment, could influences results. For example, a meta-analysis focussing specifically on AML found an association between maternal smoking of Hispanic mothers during pregnancy and increased risk of AML, along with an increased risk with any paternal smoking [177]. Due to the stigma and inference of guilt surrounding smoking during pregnancy, reporting bias may also impact the outcome of these epidemiological studies, and provide a reason for why an increased risk of childhood leukaemia is seen repeatedly with paternal and second-hand smoke exposure, but not with maternal smoking [86-88]. Other limitations of these studies include potential recall bias as data is collected retrospectively following a diagnosis of leukaemia and small sample sizes due to the rarity of the disease.

1.5.1.1 Mechanisms through which parental smoking exposure contributes to leukemogenesis

Cigarette smoke has been shown to cause DNA damage and chromosomal structural abnormalities [171, 180-182]. Using single cell gel electrophoresis, it has been shown that smokers have significantly more DNA damage than non-smokers [182]. Although there is little epidemiological evidence of an increased risk of childhood leukaemia with maternal smoking exposure, chromosomal instability has been shown to be present at an increased frequency in foetal amniocytes from mothers who smoked during pregnancy [178]. In the study, amniocytes were taken from 25 smoking and 25 non-smoking pregnant women and

underwent cytogenic evaluation, and significant increases in chromosomal instability, chromosomal legions and structural chromosomal abnormalities were observed in smoking mothers compared to non-smoking controls [178]. The most frequent structural chromosomal abnormalities observed included deletions and translocations [178]. This study also found that some chromosomal regions were particularly sensitive to smoke exposure, including 5q31, 17q21 and most significantly, as it was only seen in the smoke exposed sample, 11q23, a common translocation in childhood leukaemia [178]. Maternal smoking during pregnancy has also been linked to changes in DNA methylation in newborns, implicated in the development of childhood leukaemia [14].

Many epidemiological studies have reported an association between paternal smoking preconception and an increased risk of childhood leukaemia, suggesting that leukaemic abnormalities may occur in male gametes. An animal study exposing male mice to cigarette smoke condensate found changes in gene expression of sperm, activation of DNA damage pathways and apoptosis in the testis [180]. The study also found phenotypic changes in the offspring of cigarette smoke condensate exposed fathers, including decreased weight [180]. Chromosomal abnormalities have also been found in sperm of fertile male smokers at a higher frequency compared to non-smokers [171].

1.5.2 Benzene

One of the major carcinogens found in cigarette smoke is benzene. This chemical is released during the incomplete combustion of carbon materials, such as wood and tobacco, and since the early 20th century has been produced in vast quantities through isolation from petroleum [183]. Benzene can be found in gasoline, air pollution and solvents used for paints and ink [183]. Since the mid-20th century, an increase in adult leukaemias has been associated with workers exposed to benzene leading to guidelines on benzene workplace exposure limits of 1 part per million (ppm) over an 8-hour period [183]. Benzene exposure levels vary depending on different microenvironments, with contributions from cigarette smoke and gasoline engine emissions. People who work in petrochemical, manufacturing and transport industries are exposed to much higher levels of benzene than the general population, and general indoor/urban outdoor exposure in the USA and Europe is lower than general indoor/urban outdoor exposure in Asia [184]. Restrictions on car emissions and bans on smoking in public places have shown a reduction in general benzene exposure in environments around the world [184].

1.5.2.1 Benzene in air pollution

Benzene, along with formaldehyde, makes up 60% of the cancer risk for toxic air contaminants [185]. Ambient air pollution, fuelled by industry, power generation, motor

vehicle emissions and cigarette smoke contribute to 4.2 million deaths per year worldwide, with South East Asia and the Western Pacific most encumbered by ambient air pollution related deaths [186]. The main pollutants studied as constituents of air pollution include sulphur dioxide, carbon monoxide, nitrogen dioxide and particulate matter of different diameters PM_{10} and PM_{25}, with adverse health impacts including cardiovascular disease, lung cancer and respiratory diseases [186, 187].

1.5.2.2 Benzene exposure during pregnancy

Foetal exposure to benzene *in utero* has been detected through its presence in cord blood, showing that benzene can travel across and accumulates in the placenta at levels equalling or higher than maternal concentrations [188, 189]. Infants and children may be more at risk from benzene due to underdeveloped excretion pathways [190]. Occupational exposure to benzene during pregnancy has been associated with adverse pregnancy outcomes such as spontaneous abortions, low birth weight, decreased intrauterine growth and leukaemia [191-194]. An animal study on prenatal benzene exposure in rats found that offspring had changes in their motor activity and cognitive function [195].

Many recent studies have investigated the impact of air pollutants during pregnancy on adverse health outcomes for the foetus, including foetal loss, pre-term birth, low birth weight, developmental delays and congenital defects [187, 196-199]. Associations differ with different pollutants across studies, along with the use of different biomarkers and proxies for each pollutant. Two studies focussing on benzene as a constituent of air pollution exposure during pregnancy used passive samplers across the study area to measure ambient benzene levels and predicted women's exposure levels through land-use regression models [197, 200]. These studies found an association between prenatal ambient benzene exposure and an increased risk of preterm births [197, 200].

1.5.2.3 Benzene exposure and risk of childhood leukaemia

With its involvement in adult occupational AML, many epidemiological studies have investigated whether there is also an association between benzene exposure *in utero* and childhood leukaemia. A review by Pyatt and Hays in 2010 addressed all literature up until that point and found that whilst there were some positive associations, most of the literature pointed to no association for benzene exposure and childhood leukaemia [201]. However, much of the data was lacking in quantitative benzene data, small sample sizes and most did not analyse childhood leukaemia by subtype, which may greatly impact the true association between benzene and childhood leukaemia, and would benefit from analysis of a larger sample size. A meta-analysis in 2015 found many associations with benzene exposure including a greater risk for occupational and household use of benzene, with a higher risk for

maternal exposure and exposure during pregnancy and a stronger association for AML than ALL [202]. The meta-analysis provided a larger sample size for analysis which leads to greater statistical power for determining associations, and due to the larger sample sizes the analysis could more significantly determine a difference in association between ALL and AML.

More recent studies have provided additional evidence supporting the association between benzene exposure and childhood leukaemia. A Swiss based cohort study identified parental occupations from census records to estimate benzene exposure levels [203]. The study included 262 childhood leukaemia cases for maternal exposure and 380 cases for paternal exposure [203]. The study found no association for paternal exposure, but an association was seen for maternal exposure and childhood leukaemia, particularly for ALL [203]. More evidence supporting an increased risk for maternal exposure came from a cohort study in Norway which used questionnaires at 17 weeks gestation to identify maternal exposures to sources of benzene in the 6 months prior, and paternal exposure 6 months pre-conception [204]. From 1999 to 2009, 70 leukaemia cases were identified from this cohort, and associations were found between maternal exposure to “gasoline or exhaust”, as a benzene proxy, and increased risk of childhood leukaemia and ALL, with a weaker association for paternal exposure to “gasoline or exhaust” as a benzene proxy [204]. Furthermore, a Danish study investigated parental occupational exposure to benzene using data from pension records, with 217 childhood ALL cases for paternal preconception exposure and 169 childhood ALL cases for maternal pregnancy exposure, and observed an increased risk of childhood ALL with maternal exposure to benzene during pregnancy [205].

1.5.2.4 Exposure to air pollution and risk of childhood leukaemia

In recent years the impact of air pollution on childhood leukeamia has been the subject of many epidemiological studies. A recent meta-analysis found evidence to support an association between air pollution and an increased risk of childhood leukaemia, with a higher association for AML [206]. The studies in this meta-analysis use different methods for assessing exposure to air pollution, including measures of benzene, nitrogen dioxide, particulate matter, 1,3-butadiene, traffic count and road density [206]. There has been some contradictory evidence to support this association, with an earlier meta-analysis study finding no significant association between air pollution and childhood leukaemia, however many individual and newer studies and reviews support the relationship, including those with a focus on benzene [206, 207]. For instance, an Italian study found an association between benzene exposure and leukaemia risk, especially for children under 5 years old, with a stronger association for AML than ALL [208]. This study used ambient air modelling of

benzene and PM_{10} at subjects' residential address at birth, with estimates of traffic flow and emissions from vehicles for 83 childhood leukaemia cases and 332 controls [208]. This was further supported by a study in Oklahoma, USA, which used proximity to primary, secondary and tertiary roads at time of birth and a land-use regression model to measure nitrogen dioxide as a proxy for traffic related air pollution [209]. With 307 acute leukaemia cases and 1013 controls, this study found an association between urban children and increased risk of AML only [209]. This study was repeated using benzene measurements from the 2005 National-Scale Air Toxics Assessment, and although no significant associations were seen, a stronger association for AML was observed compared to ALL [210]. The association between air pollution and AML was shown again in a Danish study of childhood cancers, with 727 ALL and 125 AML cases, which used benzene concentrations at all addresses during pregnancy and until diagnosis [211]. Controls were case matched to individuals, selected from the general Danish population, who were free from cancer at the time of case diagnosis [211]. This study found a higher relative risk of AML with benzene exposure during pregnancy than ALL [211].

1.5.2.5 Mechanisms through which benzene may contribute to leukemogenesis

Animal and *in vitro* studies play an important role in determining the effect of benzene exposure *in utero* and the mechanisms that may lead to childhood leukaemia. A mouse study found that *in utero* exposure to benzene increased production of reactive oxygen species (ROS) without increasing oxidative stress markers but did decrease levels of IκB-α leading to activation of NFκB and altered haematopoietic progenitor cell growth in the foetal samples [189]. An *in vitro* study by Peng *et al.* 2012 supported the findings of increased ROS with benzene exposure and found that exposure of peripheral blood lymphocytes to benzene lead to an increase in chromosome aberrations (DNA breaks and fragmentation) as measured by the COMET and micronuclei assay, and protein damage, evaluated by ATR microspectroscope [212]. Furthermore, an increase in micronuclei was found in foetal mouse samples after benzene exposure *in utero*, along with an increased chromosomal recombination frequency in offspring postnatally [213]. A study using Chinese Hamster Ovary cells found an increase in DNA damage, micronuclei and chromosomal aberrations (breaks, fragments, and chromosome gaps) when cells were exposed to benzene and its metabolites *in vitro,* with a more potent effect from benzene metabolites [214]. Further *in-silico* molecular modelling was used to assess the molecular docking of benzene metabolites with Topoisomerase II α, suggesting that benzene exposure may have a genotoxic effect through topoisomerase inhibition [214].

Benzene has been linked with adult occupational AML for many years, however the exact mechanism of action required for the development of leukaemia following exposure to benzene remains vague [183]. Proposed mechanisms of action for benzene exposure in adult AML involve many pathways and include disruption of cell signalling pathways, such as Wnt and NF-κB, disruption of the hematopoietic stem cell niche leading to increased apoptosis and cell proliferation, gene mutations, oxidative stress, supressed immunosurveillance and epigenetic alterations [215, 216]. Some studies have also shown that workers exposed to benzene have altered DNA methylation patterns, which could play a role in benzene exposed leukemogenesis [216].

Benzene metabolism may also play a large part in hematotoxicity through the generation of toxic metabolites such as hydroquinone, and 1,4-benzoquinone. Exposure to benzene metabolites *in vitro* has been shown to cause DNA damage, strand breaks and topoisomerase inhibition [217]. Moreover, benzene metabolites have been observed to reach the bone marrow, which may be key in leukaemia development as the bone marrow microenvironment has been shown to influence haematopoietic stem cells and may be a site of leukaemia progression [217, 218]. Benzene metabolites can react with peptides and proteins, affecting cell function, and they can also generate oxygen radicals which could cause DNA strand breaks and point mutations [216]. Repair of DNA legions caused by metabolites could also lead to genomic instability, through DDRs such as base excision repair and nucleotide excision repair, that remove the legions and leave nicks in the DNA that can lead to strand breaks [216]. Secondary metabolism in the bone marrow to produce benzoquinones can lead to haematopoetic genotoxicity as benzoquinones have been shown to stall replication forks and interfere with topoisomerase I [216, 219]. *In vitro,* benzene metabolites have been associated with topoisomerase II inhibition, which can increase DNA cleavage and form translocations [216].

A study in Shanghai used fluorescent *in situ* hybridisation (FISH) to assess the cytogenetic abnormalities in 722 adult AML cases, 78 of which had chronic benzene exposure [220]. This study found that cytogenetic abnormalities were more prevalent in benzene exposed AML, with increased numbers of balanced translocations t(8;21) *RUNX1-RUNX1T1* and t(15;17) *PML-RARA*, mirroring *de novo* AML rather than therapy related AML [220, 221]. These 2 balanced translocations are also some of the most common translocations found in childhood AML [222]. The mechanism of action in benzene related adult AML may provide a foundation for the role of benzene exposure *in utero* on development of childhood leukaemia.

1.5.3 Nicotine and cotinine

The major chemical component of cigarette smoking is nicotine. It is not only found in tobacco but also e-cigarettes and nicotine replacement therapies. Although it is not classed as a carcinogen, nicotine is highly addictive and fuels the continued use of tobacco products known to cause cancer [223]. Once inhaled as cigarette smoke, nicotine can be absorbed into the bloodstream from the lungs, reaching peak concentrations quickly before distribution throughout body tissues [223]. Nicotine can also be absorbed through the oral cavity and the skin, with different nicotine products varying in bioavailability [223]. Nicotine is known to accumulate in gastric juices, saliva and breast milk, whilst also being able to cross the placenta [223]. Once consumed, nicotine stimulates nicotinic acetylcholine receptors in the central nervous system, causing the release of neurotransmitters such as dopamine [224]. The sustained intake of nicotine and stimulation of neurotransmitter release influences synaptic plasticity and reinforcement of rewarding behaviour in the brain, contributing to the addictive nature of nicotine [224].

1.5.3.1 Nicotine and cotinine metabolism

Nicotine metabolism occurs in the liver, with 6 known primary metabolites formed, shown in Figure 1.7 [223]. Between 70-80% of nicotine is converted to cotinine, making it the major primary metabolite in nicotine metabolism [223]. Metabolism of nicotine is much faster than cotinine metabolism with an average clearance of 1200ml/min for nicotine and 45ml/min for cotinine [223]. This leads to cotinine having a much longer half-life in the body than nicotine at around 12-20 hours compared to 1-3 hours, leading to much higher plasma concentrations of cotinine than nicotine [223]. Renal excretion of nicotine is influenced by urine pH, with much faster excretion when nicotine is ionised in acidic conditions, whereas cotinine is less affected by changes in urine pH [223]. Around 10% of nicotine is excreted unmetabolized in the urine, with a further ~15% excreted as nicotine metabolites (not derived from cotinine), but the majority, around 75%, is excreted as cotinine and cotinine metabolites [223]. A small amount of nicotine has also been found to be excreted in sweat and faeces [223]. Due to its longer half-life, sensitivity and specificity, cotinine is used as a biomarker to detect cigarette smoking [225].

Not only is cotinine the major metabolite of nicotine, but it also has its own metabolic pathways, with 6 reported primary metabolites, shown in Figure 1.7 [223]. Conversion of nicotine to cotinine starts with the production of nicotine-$\Delta^{1'(5')}$-iminium intermediate facilitated by cytochrome P450 enzymes and followed by aldehyde oxidase catalysis to form cotinine [223]. *Trans*-3'-hydroxycotinine is the most abundant metabolite found in the urine of smokers, often conjugated to glucuronide [223]. Oxidation of nicotine to cotinine and then to *trans*-3'-hydroxycotinine is facilitated by the cytochrome P450 enzyme CYP2A6, however

other cytochrome P450 enzymes are proposed to be involved throughout both nicotine and cotinine metabolism [223]. Other cotinine metabolites include norcotinine, cotinine *N*-oxide, 5'-hydroxycotinine, cotinine methonium ion and cotinine glucuronide [223].

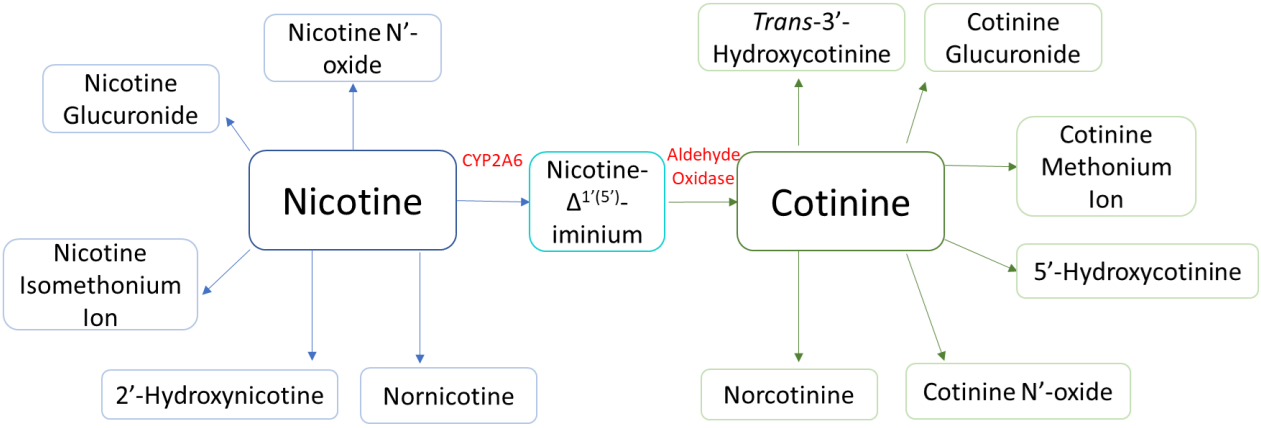


Figure 1.7. Overview of Nicotine and Cotinine metabolism and their associated metabolites. The main enzymes involved in nicotine metabolism are shown in red: Cytochrome P450 Family 2 Subfamily A Member 6 (CYP2A6) and Aldehyde Oxidase.

Polymorphisms in nicotine/cotinine metabolism enzymes is one mechanism that can affect nicotine and cotinine metabolism [223]. Other factors can also influence nicotine and cotinine metabolism leading to variability between individuals. The diet can affect nicotine metabolism, such as menthol, which has been shown to increase cotinine half-life in women who smoke menthol cigarettes [223]. Age can also affect nicotine metabolism with a longer nicotine half-life in neonates [223]. Further factors such as gender, pregnancy, diseases and medications can all impact nicotine/cotinine metabolism and clearance from the body [223].

1.5.3.2 Nicotine and cotinine exposure during pregnancy

Pregnancy has been shown to lead to quicker clearance of nicotine and cotinine in the mother compared to postpartum levels, which may be due to increased activity of CYP2A6 [223]. However, cotinine concentrations in newborns exposed to nicotine *in utero* have levels comparable to adults. A study in the USA used dried blood spots to detect cotinine levels in newborns as an indicator of exposure to cigarette smoke *in utero* and found that cotinine can be detected in newborns, with 12% of the 1414 newborns having cotinine levels comparative to active smokers [87]. Cotinine was detected in 83% of newborns whose mothers reported smoking during pregnancy and was also detected in 29% of newborns whose mothers reported no smoking during pregnancy, suggesting that maternal smoking reporting does not always accurately reflect the level of *in utero* exposure to offspring [87]. A Greek study used meconium from 45 newborns to measure cotinine and nicotine levels and found significant concentrations in newborns whose mothers were active smokers or were exposed to

second-hand smoke from the father, with levels correlating with a decreased birth weight and gestational age [226].

Nicotine and its metabolites are thought to be one of the main contributors to adverse foetal outcomes from *in utero* exposure to cigarette smoking, e-cigarettes and nicotine replacement therapies [227, 228]. Nicotine exposure *in utero* has been linked to a decrease in foetal weight, congenital abnormalities, respiratory conditions, neurological conditions and obesity [227]. Animal models have provided further understanding of the effects of nicotine exposure *in utero* with offspring phenotypes showing increased risk of cardiovascular and respiratory disease, alternations in neurotransmitter signalling and impacts on offspring metabolic and fertility outcomes [229]. There is also evidence to suggest that nicotine exposure *in utero* can have multigenerational effects, as female offspring exposed to nicotine *in utero* undergo gametogenesis during development and their subsequent offspring have been found to show adverse effects, such as respiratory problems, even when they were not exposed *in utero* [229].

1.5.3.3 Mechanisms through which nicotine and cotinine may contribute to leukemogenesis

Although not currently considered to be carcinogenic, there is growing evidence that nicotine exposure leads to DNA damage [230]. Binding of nicotine to nicotinic acetylcholine cell surface receptors leads to the activation of the kinase ATK which reduces apoptosis in some tissues and stimulates cell proliferation, angiogenesis and tumour growth *in vitro* and in animal models [166, 231, 232]. An *in vitro* study by Demirhan *et al.* 2011 found that amniocyte exposure to nicotine containing medium resulted in significantly increased chromosomal abnormalities, such as numerical and structural aberrations, compared to cells grown in control medium [233]. Furthermore, a mouse study found that nicotine exposure from e-cigarettes *in utero* altered DNA methylation patterns of the offspring and affected the lung inflammatory environment [234].

Over 70% of nicotine is converted into cotinine after consumption, which has a much longer half-life than nicotine, so the body is exposed to cotinine for much longer lengths of time and therefore could play a role in the DNA damage effects of nicotine and smoking. Cotinine is a partial agonist to nicotinic acetylcholine receptors and can affect receptor assembly and upregulation, much like nicotine [235]. However, the evidence for cotinine carcinogenicity is still inconclusive, with many epidemiological, *in vitro* and animal studies reporting conflicting conclusions [236]. A lack of evidence for nicotine/cotinine carcinogenicity was observed in an epidemiological study of nicotine replacement products, where nicotine replacement products were not a predictor for lung cancer, gastrointestinal cancer, or all cancers in a group of 3320 participants enrolled on an intervention study over 5 years [237]. However,

this study had a very limited time length for assessing cancer incidence and is compounded by a variety of factors such as past and current cigarette use. Animal studies of cotinine exposure have varied evidence, with Truhaut *et al.* 1964, observing toxicity and malignant tumours in rats exposed to cotinine through drinking water, whereas LaVoie *et al.* 1985 did not observe any carcinogenicity in rats exposed to cotinine through drinking water [236, 238]. Nevertheless, many studies have reported the DNA damaging effects of cotinine, which could support the role of cotinine in chromosomal translocation development. Various studies have shown that cotinine increases adenomas and cancer incidence, as well as stimulates tumour growth [236]. A recent *in vitro* study by Dalberto *et al.* 2020 using human neuroblastoma cells found that exposure to cotinine caused both cytotoxic effects on cell viability and genotoxic effects with induction of DNA damage, similar to the effects from nicotine exposure [239]. Cotinine exposure to endothelial cells *ex vivo* also demonstrated an increase in VEGF expression, a growth factor overexpressed in many cancers contributing to tumour growth and angiogenesis [240].

Epidemiological data shows that paternal smoking preconception increases the risk of childhood leukaemia [170]. FISH analysis on sperm found that male smokers had more chromosomal anomalies than non-smoking males, suggesting genotoxic damage could occur pre-conception in male sperm [171]. An *in vitro* study found that when sperm from non-smoking males was exposed to nicotine it caused instability of the sperm membrane, decrease in sperm viability and an increase in DNA strand breaks, mainly of double strand origin [241]. Cotinine itself can also cross the hematotesticular barrier in males, and exposure to high concentrations of cotinine have been shown to decrease sperm mobility, decrease membrane function and affect sperm maturation [242, 243]. As well as alterations to sperm genomes, nicotine has also been shown to alter sperm DNA methylation in rats exposed to nicotine concentrations equivalent to average human smokers [244]. More data is required to confirm the potential genotoxicity of nicotine and its major metabolite, cotinine to further understand its role in the development of childhood leukaemia.

1.6 Aims and objectives

Understanding the aetiology of childhood leukaemia is important to provide guidance on modifiable risk factors for public health with the aim of preventing childhood leukaemia. Epidemiological studies have identified a number of risk factors associated with risk of childhood leukaemia, including parental smoking, maternal caffeine intake and maternal plasma folate levels. However, there is little evidence of how these exposures influence the initiating chromosomal translocations involved in childhood leukaemia and how they

influence the secondary modulations, such as epigenetics, in leukaemia progression. This gap in the knowledge informs the main aims and objectives of this thesis:

1) Investigate if modifiable environmental exposures can trigger the induction of chromosomal translocations associated with childhood leukaemia

a. Use an *in vitro* cell model to investigate the effect of physiologically relevant levels of risk exposures: folate, caffeine, benzene, and cotinine on cell growth, viability, and chromosomal translocation induction

2) Explore the potential role of environmentally-associated DNA methylation to contribute to risk of childhood ALL in specific subtypes

a. Use the “meet in the middle” approach to determine significant overlapping methylation between environmental exposures and ALL and its subtypes
b. Use GO enrichment and KEGG pathway analysis to investigate potential underlying mechanisms of environmental exposure in ALL subtypes

2 General Procedures

2.1 Chemicals and reagents

All chemicals and reagents were purchased from Merck Sigma-Aldrich (Dorset, UK) or Fisher Scientific (Loughborough, UK) unless otherwise stated.

2.1.1 Preparation of exposure variables

Table 2.1. Preparation of chemicals. All exposure variables were prepared as stock solutions by dissolving in dimethylsulphoxide (DMSO) as a vehicle. Further dilutions were made in DMSO to give a 1000X concentration for working standards

Chemical	Amount of chemical	DMSO volume	Final concentration
Etoposide	25mg	4.25ml	10mM
Cotinine	250mg	2.84ml	500mM
Caffeine	0.971g caffeine dissolved in 20ml boiling water	N/A	250mM
Benzene	8.6µl	2ml	48mM
Folic acid	88.2mg	5ml	20mM

2.2 Cell culture

2.2.1 Cell lines

The NALM6 cell line was donated by Dr Gordon Strathdee of Newcastle University, Table 2.2, and used as the main cell line model. Cell line authentication was performed by Eurofins Genomics (Ebersberg, Germany), using STD profiling. Mycoplamsa testing on working cultures was performed periodically using Venor®GeM OneStep (Minerva Biolabs, Berlin, Germany) following manufacturer's instructions. NALM6 is a Pre-B acute lymphoblastic leukemia cell line that was derived from a 19 year old male in ALL relapse in 1976 [245, 246].

Table 2.2. NALM6 cell line information.

Cell line	Cell type	Main aberrations	Cell density maintenance	Ref.
NALM6	Pre-B acute lymphoblastic leukemia	46(43-47)<2n>XY, t(5;12)(q33.2;p13.2)	1.0-2.0 x 10^6 cells/ml	[245, 247]

Cell lines in Table 2.3 were donated by Dr Gordon Strathdee (NALM6, K562, HL60), Dr Lisa Russell (Pre-B 697, REH, Kasumi-1) and Dr Jim Allen (NB-4) (Newcastle University, UK).

These cell lines were used to extract RNA and produce cDNA to be used as positive controls for each type of translocation.

Table 2.3. Cell line information for cells used as translocation positive templates in RT-PCR assays.

Cell line	Cell type	Main aberrations	Cell density maintenance	Ref.
Pre-B 697	B cell precursor leukemia	t(1;19)(q23;p13)	0.5-1.5 x 10^6 cells/ml	[247, 248]
Kasumi-1	Acute myeloid leukemia	t(8;21)(q22;q22)	0.5 x 10^6 cells/ml	[247, 249]
REH	B cell precursor leukemia	t(12;21)(p13;q22)	0.5 x 10^6 cells/ml	[247, 250]
NB4	Acute promyelocytic leukemia	t(15;17)(q24;q21)	N/A	[247, 251]
MV4-11	Acute monocytic leukemia	t(4;11)(q21;q23)	N/A	[252]
HL60	Acute promyelocytic leukemia	45;–X,–6,–8,–16, + M1, + M2, + M3	N/A	[247, 253]

2.2.2 Cell culture

NALM6 was maintained in RPMI 1640 medium supplemented with 10%(v/v) Fetal Bovine Serum (FBS) and 1%(v/v) penicillin and streptomycin. Kasumi-1, REH and Pre-B 697 were maintained in RPMI 1640 medium supplemented with 20%(v/v) Fetal Bovine Serum (FBS) and 1%(v/v) penicillin and streptomycin. Cells were cultured at 37°C in a humidified atmosphere containing 5% CO_2.

For routine subculture, cells were diluted with fresh warmed medium every 2-3 days. Once cells reached a passage number of 25, they were discarded and a fresh cell stock with a low passage was thawed.

2.2.3 Cell harvesting

Suspension cells were transferred from the culture flask and centrifuged for 3 minutes at 1200rpm, followed by removal of the supernatant. The pellet was then washed with PBS, centrifuged for 3 minutes at 1200rpm, and the supernatant removed again before re-suspension in fresh medium or used for further experiments.

2.2.4 Cryopreservation of cell lines

Cells were harvested as described in 2.2.3 and re-suspended in up to 1ml of prepared RMPI 1640 media containing 20% FBS. For NALM6 cells, 1x10^6 cells/ml were transferred to a cryovial and 5% DMSO added one drop at a time. For Kasumi-1, REH and Pre-B 697, 1x10^7 cells/ml were transferred to a cryovial and 10% DMSO added one drop at a time. The

cryovials were stored in a Nalgene Mr Frosty freezing container for a minimum of 24 hours at -80°C and then transferred to -150°C freezer for long term storage.

2.2.5 Thawing of cell lines

Stored cells were thawed quickly at 37°C and added to 5ml of warmed medium. The cell suspension was centrifuged at 1200rpm for 3 minutes and the supernatant was removed. The pellet was re-suspended in 10ml warmed medium and split between 2 x T25 flasks for NALM6 cells, or in 24 well plates for Kasumi-1, REH and Pre-B 697.

2.2.6 Trypan blue assay for cell counting

Trypan blue enters dead cells, where the cell membrane has been compromised, and binds to intracellular proteins turning dead cells blue. Cell counts were performed using a Neubauer haemocytometer. A mix of Trypan Blue Solution 0.4% and cell suspension at an appropriate dilution factor was made and 10µl added to each chamber of the haemocytometer.

Cells were counted in each of the 4 large corner squares and the middle square (1mm^2 each), representing 10^4 cells/ml, Figure 2.1. To maintain consistency across counts, cells touching the left and bottom edges of each square were counted, and cells that touched the top and right edges of each square were not counted (Figure 2.1).

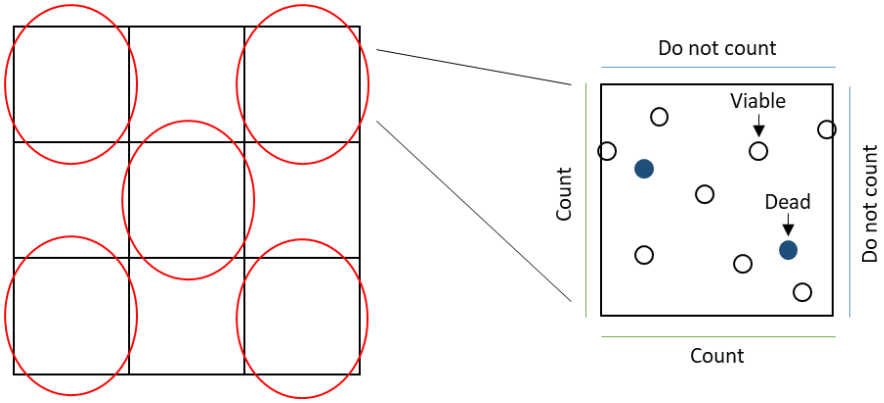


Figure 2.1. Trypan blue counting using a haemocytometer. All cells were counted in each large square circled red. Cells touching the left and bottom edge of each square were counted. Cells touching the top and right edge of each square were not counted. Halo shaped cells (not blue) were deemed as viable, any cells that were blue in colour were deemed not viable (dead).

All cells that appeared blue during cell counting with trypan blue were deemed non-viable. Halo shaped cells that were not blue were deemed as viable. The following calculation was used to determine the number of cells in the original suspension:

Mean cell count per square x dilution factor x 10^4 = cells/ml

2.2.7 Resazurin cell viability assay

The resazurin cell viability assay works by measuring the reduction of the redox dye resazurin by metabolically active, viable cells. Exhibiting both colorimetric and fluorometric changes, the intracellular conversion of oxidized (blue and non-fluorescent) to reduced (red and fluorescent) resazurin can be monitored at different wavelengths, 530-560nm excitation and 590nm emission for fluorescence and 540nm and 630nm for absorbance [254]. The rate of resazurin reduction is proportional to the number of viable cells in the well, allowing cell viability to be calculated.

A 2% resazurin solution was prepared with PBS and filter sterilised using a 0.2µm syringe filter. On the day of measurement, a 1 in 10 dilution was prepared with fresh medium, and further diluted 1 in 10 when added to each well of the 96 well plate, making a final concentration of 0.02% resazurin. A calibration curve was used to determine optimum cell density, section 2.2.7.1.

To set up the cell viability assay, treated and untreated cells were seeded in a 96 well plate with a total volume of 100µl per well, with 3 technical replicates. Cell free medium was used as a negative control, with 100µl per well. The plate was centrifuged at 1200rpm for 2 minutes and left to incubate at 37°C for the required time stated in each experiment design. Four hours prior to the measurement time, 10µl was removed from each well, and 10µl of resazurin added.

The plate was left to incubate for 4 hours, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm, or an absorbance wavelength of 570nm, on a Tecan Spark 10M Plate Reader with SparkControl Magellan software.

The relative fluorescence for each sample was calculated using the equation below, and cell viability was expressed as a percentage of the control.

Relative fluorescence unit (RFU) = A – B

A = absorbance of sample wells
B = average absorbance of negative control wells

Cell viability as a % of control = (C ÷ D) × 100

C = RFU of sample
D = average RFU of positive control

2.2.7.1 Calibration curve of cell density

To determine the optimum cell density required for the resazurin cell viability assay, a standard curve was set up for each cell type. A 96-well plate was set up with starting density

of 2x10^6 cells/ml and diluted 1 in 2 with normal medium. Cells were incubated for 44 hours, resazurin added, and the colour change was measured 4 hours later. Optimum cell density was chosen from the highest exponential point on the curve Figure 2.2.

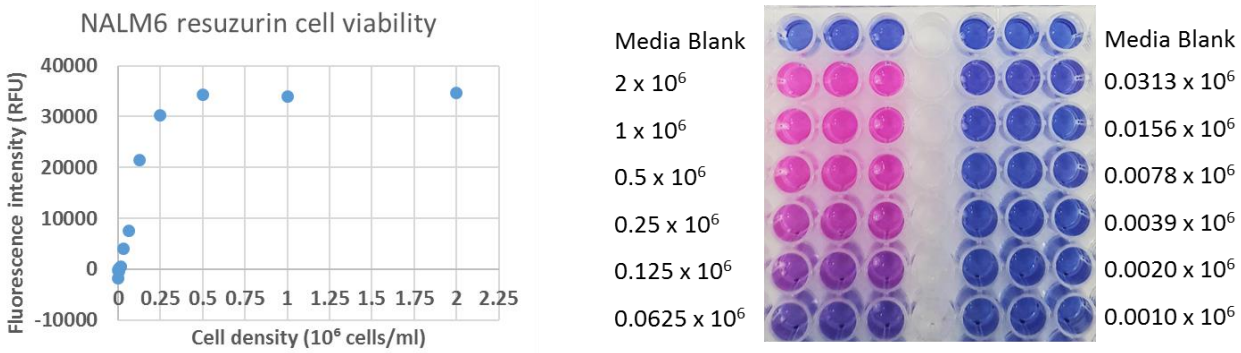


Figure 2.2. Resazurin standard curve for NALM6 cells. A) Relative fluorescence plotted against cell density. B) Visual colour change during resazurin assay.

2.2.8 Molecular Probes™ CyQUANT™ Cell Proliferation Assay

Molecular Probes™ CyQUANT™ Cell Proliferation Assay (Fisher Scientific, UK) relies on a strong fluorescent signal measured at 480/520 nm when CyQUANT™ green fluorescent dye binds to cellular nucleic acids. This assay allows for the detection of cell density in culture without relying on metabolic activity. Cell proliferation can be followed by measuring increases in nucleic acid content as cells proliferate.

The CyQUANT™ Cell Proliferation Assay can detect between 50 to 50,000 cells as standard. Samples for analysis were prepared by taking an aliquot from a cell culture solution, to give an expected cell density of under 50,000 cells (using previous cell growth curves) and transferred to a 96 well plate. The plates were spun at 3000rpm for 5 minutes to pellet cells and the supernatant was removed by blotting plates on tissue. To each well 100µl PBS was added to wash cells of any remaining medium that may interfere with fluorescence. Plates were spun again at 3000rpm for 5 minutes and supernatant removed by blotting on tissue. Sample plates were then stored at -80°C to aid cell lysis for up to 4 weeks.

On the day of analysis, the CyQUANT™ reagent was prepared by diluting Component B 20-fold in nuclease free water, followed by diluting Component A 400-fold into the diluted Component B, according to the manufacturer's protocol. To the 96-well sample plates thawed at room temperature, 200µl of prepared CyQUANT™ reagent was added to each sample well and allowed to incubate for a minimum of 5 minutes in the dark.

Fluorescent samples were measured on a Tecan Spark 10M Plate Reader with SparkControl Magellan software, using a 480nm excitation and 520nm emission.

To determine cell number, a standard curve for NALM6 cells was prepared in a 96 well plate with a starting cell number of 50,000 cells, determined by trypan blue cell count, and diluted 1 in 2 until a final concentration of 98 cells in a well, Figure 2.3.

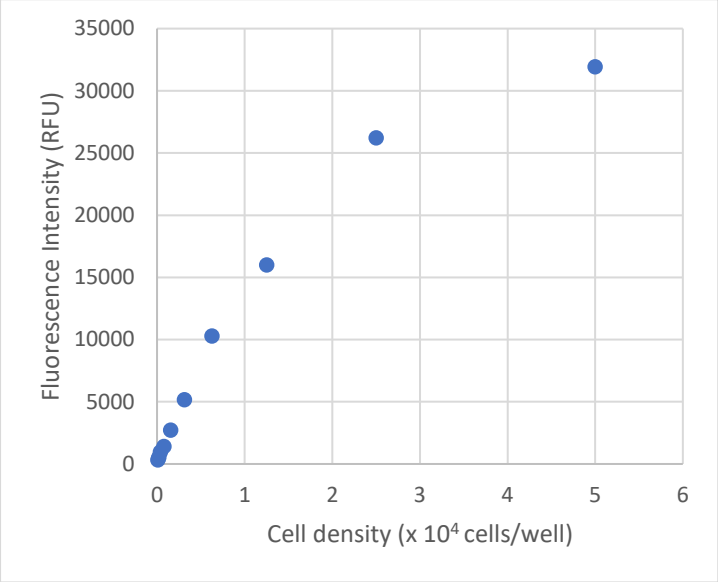


Figure 2.3. Cyquant standard curve for NALM6 cells. A starting cell number of 50,000 cells was diluted 1 in 2 to give a final cell number of 98 cells.

The relative fluorescence for each sample was expressed as a percentage of the control.

Cell viability as a % of control = (RFU of sample ÷ average RFU of positive control) × 100

2.3 Reverse transcription PCR assays for detection of fusion genes

Reverse transcription PCR assays allow for the detection of selected gene transcripts or fusion genes. By isolating RNA and using reverse transcription to convert mRNA to cDNA, fusion genes can be amplified across breakpoints using only coding exons, which shortens the target sequence for PCR allowing for amplification of fusion genes that would be otherwise too long to amplify by standard or long range PCR.

2.3.1 RNA extraction with EZNA Total RNA Kit

RNA extraction was performed using Omega Biotek E.Z.N.A.® Total RNA Kit I (VWR, UK), with all buffers prepared following manufacturer's instructions. Following harvesting of cells, 350µl of TRK lysis buffer was added to a pellet with a cell number <5 x 10^6. After vortexing, the samples were either stored at -80°C or proceeded to the next step. Cells were homogenized by pipetting up and down with a narrow tip and thorough vortexing. To each sample, 350µl of 70% ethanol was added and tubes vortexed before transferring the full sample to a HiBind® RNA Mini Column in a collection tube. The column was centrifuged at

10,000 x g for 30 seconds and repeated until all of the sample has passed through the column. With the filtrate discarded, 500µl of RNA Wash Buffer I was added and centrifuged for 30 seconds at 10,000 x g. The filtrate was discarded and the columns were washed twice with 500µl RNA Wash Buffer II, and centrifuged for 1 minute at 10,000 x g. The empty column was centrifuged at 10,000 x g for 2 minutes to remove any traces of ethanol. Placing the HiBind® RNA Mini Column in a new collection tube, 50µl of DEPC treated water was added, allowed to stand for 2 minutes and centrifuged at 10,000 x g for 2 minutes to elute the RNA. The RNA was either stored at -80°C or proceeded to DNase treatment.

2.3.2 DNase treatment

To remove DNA contamination from RNA samples, a Precision DNase kit (Primer Design, Eastleigh, UK) was used following RNA extraction. Five microlitres of 10X Precision DNase reaction buffer was added to the RNA sample, followed by 1µl of Precision DNase enzyme. The reaction was incubated at 30°C for 10 minutes and then deactivated for 5 minutes at 55°C.

The concentration of the RNA was measured using a NanoDrop spectrophotometer. A NanoDrop spectrophotometer measures the concentration and purity of nucleic acids including RNA and DNA by reporting the wavelength ratio of A260/A280. The buffer used to elute RNA was used as a blank measurement, then 1µl of each sample was used for measurement.

The treated RNA was then either stored at -80°C or proceeded to cDNA synthesis.

2.3.3 cDNA synthesis

Samples are incubated with primers, dNTPs and a reverse transcriptase enzyme allowing for the transcription of mRNA into a complementary DNA (cDNA) strand. Primers are used to initiate transcription points and can be either oligo dT primers which bind to the polyA tail of mRNA, random primers, which bind to random sequences along the RNA strand, and are beneficial to samples which may be degraded or without a polyA tail, or a combination of both. Synthesis of cDNA was performed with either the Primer Design Precision nanoScript2 Reverse Transcription Kit or Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit. cDNA was synthesised using Primer Design Precision nanoScript2 Reverse Transcription Kit, used a combination of oligo dT and random primers as this allowed conversion of RNA with or without a polyA tail. Due to the need for high throughput processing of large sample numbers, RNA from model optimisation experiments were cDNA synthesised with Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit,

using random primers. No differences were seen during PCR amplification when comparing positive control cDNA synthesised with each kit.

2.3.3.1 Precision nanoScript2 Reverse Transcription Kit

A 10µl reaction mix was prepared in PCR tubes containing 1µg of RNA template, 0.5µl each of random and oligo dT primers and RNAse/DNAse free water. The mix was heated to 65°C for 5 minutes and then transferred to ice. A further 10µl reaction mix was prepared and added to the sample containing 5µl of nanoScript2 4X Buffer, 1µl 10mM dNTP mix, 1µl nanoScript2 enzyme and 3µl RNAse/DNAse free water. Tubes were vortexed followed by a pulse spin and incubated at room temperature for 5 minutes then 42°C for 20 minutes. Samples were then heat inactivated at 75°C for 10 minutes. The cDNA was stored long term at -20°C.

2.3.3.2 Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit

Each RNA sample was diluted in DNase/RNase free water to contain 1µg RNA in 10µl. A 10µl reaction mix was prepared of 2µl 10X RT Buffer, 0.8µl 25X dNTP Mix (100mM), 2µl 10X RT Random Primers, 1µl MultiScribe™ Reverse Transcriptase and 4.2µl Nuclease free H_20. This was then added to the prepared 10µl of RNA on ice, in either PCR tubes or 96 well PCR plates and briefly spun town to remove air bubbles. The reverse transcription reaction was performed using a thermal cycler, either Applied Biosystems™ ProFlex™ 3 x 32-well PCR System for reactions in PCR tubes or Eppendorf 6331 Nexus Gradient MasterCycler Thermal Cycler for reactions in PCR plates. Reaction temperatures and times are detailed in Table 2.4. Following cDNA synthesis samples were stored at -20°C

Table 2.4. Thermal cycler temperatures and times for the High-Capacity cDNA Reverse Transcription Kit.

Temperature (°C)	**Time (minutes)**
25	10
37	120
85	5
4	∞

2.3.4 Primer Sequences and positive control samples for PCR

For PCR reactions, a positive control was provided by extracting RNA from frozen cell line pellets that were known to be positive for the translocation or reference gene of interest. The RNA was then reverse transcribed into cDNA for use in PCR reactions, Table 6. All positive

cell lines as shown in Table 4 were kindly donated by Dr Lisa Russell, Dr Gordon Strathdee and Dr James Allan (Newcastle University, UK).

Oligonucleotide sequences were taken from the literature or custom designed based on literature and supplied by Eurofins (Germany), Table 2.5. Oligonucleotides were reconstituted with autoclaved DI water to a stock concentration of 100pmol. Working standards were further diluted using RNAse/DNase free water to 10pmol unless otherwise stated.

Table 2.5. Oligonucleotide sequences and positive control cell lines used for PCR assays.

Translocation/ Gene	Type of leukaemia observed in	Positive cell line used	Primer name	Primer sequence	Source
t(1;19)(q23;p13) ***TCF3-PBX1***	Childhood ALL	Pre-B 697	*TCF3*_F PBX1_R Nested primers: *TCF3*_FNE PBX1_RNE	GGCCTGCAGAGTAAGATAG CACGCCTTCCGCTAACAG CCAGCCTCATGCACAACCA ATGTTGTCCAGCCGCATCAG	[255] Custom design
t(8;21)(q22;q22) ***RUNX1-RUNX1T1***	Childhood AML / Therapy-related	Kasumi-1	*RUNX1*_F *RUNX1*T1_R Nested primers: *RUNX1*_FNE *RUNX1*T1_RNE	GAGGGAAAAGCTTCACTCTG GCGAACTCTTTCTCCTATC CTGTCTTCACAAACCCACCG ACTGCTGCAGGGTAGTAAGG	[256] Custom design
t(12;21)(p13;q22) ***ETV6-RUNX1***	Childhood ALL	REH	*ETV6*_F *RUNX1*_R	CACCATAACCCTCCCACCA CATTGCCAGCCATCACAGTGAC	[257]
t(15;17)(q24;q21) ***PML-RARA***	Therapy Related	NB4	PML_F RARA_R PML_F2 RARA_R2	AGCTGCTGGAGGCTGTGGACGCGCGGTACC CAGAACTGCTGCTCTGGGTCTCAAT AGAGGATGAAGTGCTACGCCT TTCCGGGTCACCTTGTTGAT	[258] [259]
t(9;22)(q34;q11) ***BCR-ABL1***	Childhood ALL	-	BCR_F ABL1_R BCR_F2 BCR_F3 ABL1_R2	CTCCAGCGAGGAGGACTTCTCCT CCATTGTGATTATAGCCTAAGACCCGGAG GACTGCAGCTCCAATGAGAAC GAAGTGTTTCAGAAGCTTCTCC TGACTGGCGTGATGTAGTTGCTT	[260] [261]
t(4;11)(q21;q23) ***KMT2A-AFF1***	Childhood ALL / Therapy-related	MV411	*KMT2A*_F AFF1_R	AGAGCAGAGCAAACAGAA GCTGAGAATTTGAGTGAG	[262]
GAPDH	Reference gene	HL60	GAPDH_F GAPDH_R	TGAAGGTCGGAGTCAACGGATTTG CATGTAAACCATGTAGTTGAGGTC	[263]
Beta Actin	Reference gene	HL60	ACTB_F ACTB_R	TCCCTGGAGAAGAGCTACGA AGGAAGGAAGGCTGGAAGAG	[264]

2.3.5 RT-PCR for the detection of fusion genes and reference genes

RT-PCR was performed using an Eppendorf 6331 Nexus Gradient MasterCycler Thermal Cycler or Applied Biosystems™ ProFlex™ 3 x 32-well PCR System. The main enzyme and buffer system used for PCR, unless stated otherwise, was PCRBIO HS Taq Mix (PCR Biosystems, UK), a 2X hot start taq mix containing 6mM $MgCl_2$ and 2mM dNTPs. Some PCR reactions, where stated, were performed using AmpliTaq Gold™ 360 PCR Master Mix, following manufacturer's protocol, with PCR conditions stated where applicable. All primers and positive controls used are shown in Table 2.4. For negative controls, the cDNA template was replaced with the same volume of H_2O. The final master mix reactions and cycle conditions are reported in Tables 2.6-8.

Following PCR purification, positive samples were sent for Sanger Sequencing to confirm the correct fusion gene sequence was amplified. For some *RUNX1-RUNX1T1* and *TCF3-PBX1* PCR samples, due to low copy number, further amplification was required using nested primers, Table 2.5, to produce a high enough DNA concentration for sequencing. The same reaction mix was used for nested primers, however, the template used was the PCR amplified sample diluted 1 in 10 with water.

Table 2.6. Master mix volumes for RT-PCR reactions

Reagent	**20µl reaction**	**Concentration**
2X HS Taq Mix	10µl	1X
Forward primer	1µl	Up to 0.5µM
Reverse primer	1µl	Up to 0.5µM
H_2O	6µl	-
Template	2µl	Max. 100ng cDNA

Table 2.7. Cycle conditions for KMT2A-AFF1 PCR reactions.

Step	**Temp**	**Time**	
Initial denature	95°C	2 minutes	
Denaturation	95°C	15 seconds	40 cycles
Anneal	50°C	15 seconds	
Extension	72°C	30 seconds	
Final extension	72°C	2 minutes	
Hold	4°C	Hold	

Table 2.8. Cycle conditions for TCF3-PBX1, RUNX1-RUNX1T1, ETV6-RUNX1, GAPDH and Beta Actin.

Step	Temp	Time	
Initial denature	95°C	2 minutes	
Denaturation	95°C	15 seconds	40 cycles
Anneal	60°C	15 seconds	
Extension	72°C	30 seconds	
Final extension	72°C	2 minutes	
Hold	4°C	Hold	

2.3.6 Gel electrophoresis

PCR products were visualised on a 1% agarose gel in TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA) with the addition of Invitrogen™ SYBR™ Safe DNA Gel Stain (Fisher Scientific, UK) diluted 1 in 10,000 in agarose gel. Before loading, Thermo Scientific™ DNA Gel Loading Dye (6X) was diluted 1 in 6 in the PCR product, and 10µl of the final reaction mix loaded in the gel wells. A 100bp DNA marker ladder was added to each gel, New England Biolabs 100 bp DNA Ladder or PCRBio Ladder IV. Small electrophoresis tanks were ran at 100V for 40 minutes and large tanks were ran at 100V for 50 minutes. Gels were imaged under UV light using a Syngene G:BOX F3 gel doc system.

2.3.7 RT-qPCR for fusion genes

RT-qPCR was performed using Primer Design Precision®PLUS Master Mix with SYBR®Green (Primer Design, UK) on a C1000 Thermal Cycler BIO-RAD CFX96 Real time system. All reagent handling was performed in a sterile PCR hood, with UV treated equipment. All primers and positive controls used are shown in Table 2.5. For negative no template controls, the cDNA template was replaced with the same volume of H_2O. For no reverse transcription (no-RT) controls, DNase treated RNA from positive samples followed the cDNA synthesis protocol, section 2.3.3.1, substituting the nanoScript2 enzyme for an equal volume of H_2O to identify any potential contamination of genomic DNA at this stage. The qPCR master mix for each reaction was prepared as described in Table 2.9. All samples were repeated in triplicate in a 96-well plate, with 2 repeats for no-RT and negative water controls during primer validation. QPCR cycle conditions can be found in Table 2.10 with optimum data collection temperature for each primer pair corresponding with the annealing temperature conditions reported for standard PCR in Table 2.9.

Table 2.9. Master mix volumes for qPCR reactions

Reagent	**20µl reaction**	**Final conc.**
Precision®PLUS Master Mix	10µl	1X
Forward primer	1µl	0.5µM
Reverse primer	1µl	0.5µM
DNase/RNase free H_2O	6µl	-
Template	2µl	Max. 100ng DNA

Table 2.10. Cycle conditions for qPCR reactions

Step	**Temp**	**Time**	
Enzyme activation	95°C	2 minutes	
Denaturation	95°C	10 seconds	40 cycles
Data collection	60°C	60 seconds	
Melt curve			

To evaluate primer pair efficiency, standard curves were set up using positive control cDNA. The positive cDNA was diluted either 1 in 10, 1 in 5 or 1 in 2 to give 5 points on a standard curve. The Ct values measured for each standard curve sample were then plotted on a logarithmic scale to determine the linear regression curve. Using the value for the slope of the curve efficiency can be calculated using the equation below [265]:

$E = 10^{(-1/slope)} - 1$

An efficiency of 1 (or 100%) indicates that for each cycle the amount of product doubles [265].

2.3.8 PCR purification

2.3.8.1 PCR clean up

For single PCR bands between 100bp and 10kb in size, the Thermo Scientific GeneJET™ PCR Purification Kit (Fisher Scientific, UK) was used to purify DNA, removing primers, dNTPs and other impurities from the PCR product. An equal volume of binding buffer to PCR product was added to the completed PCR mixture and mixed thoroughly. If the DNA

fragment was less than 500bp, an equal volume of 100% isopropanol was also added to the PCR product. The solution was transferred to a GeneJET purification column and centrifuged for 1 minute at 12,000 x g. After the filtrate was discarded, 700µl of Wash Buffer was added to the column and centrifuged at 12,000 x g for 1 minute. The empty column was then centrifuged for 1 minute at 12,000 x g to remove any traces of wash buffer. Once transferred to a clean 1.5ml microcentrifuge tube, 20µl of RNase/DNase free water was added to the purification column, incubated for 2-5 minutes and centrifuged for 1 minute at 12,000 x g. Purified DNA was stored at -20°C.

2.3.8.2 PCR gel extraction

For PCR products with multiple bands or high primer dimer concentration, Thermo Scientific GeneJET™ Gel Extraction Kit (Fisher Scientific, UK) was used to purify DNA from each band. PCR products were ran on a 1% agarose gel with SybrSafe DNA stain. Bands were excised under UV light using a scalpel and transferred to a 1.5ml microcentrifuge tube. An equal volume of binding buffer to weight of the gel slice was added to the tube and incubated for 10 minutes at 50-60°C. If the DNA fragment was less than 500bp, an equal volume of 100% isopropanol to gel slice weight was added. The solubilised gel solution was added to a GeneJET purification column and centrifuged for 1 minute at 12,000 x g. For sequencing samples, a further 100µl of binding buffer was added to the empty column and centrifuged for 1 minute at 12,000 x g. After the filtrate was discarded, 700µl of Wash Buffer was added to the column and centrifuged at 12,000 x g for 1 minute. The empty column was then centrifuged for 1 minute at 12,000 x g to remove any traces of wash buffer. Once transferred to a clean 1.5ml microcentrifuge tube, 20µl of RNase/DNase free water was added to the purification column, incubated for 2-5 minutes and centrifuged for 1 minute at 12,000 x g. Purified DNA was then stored at -20°C.

2.3.9 Sanger Sequencing

Purified PCR products were sent to DBS Genomics (Durham University, UK) for Sanger Sequencing using Applied Biosciences 3730 capillary instrument. The forward and reverse sequencing reactions were performed using 3.2pmol/µl of the corresponding PCR primers. Sequencing data was submitted to the Basic Local Alignment Search Tool (BLAST, National Center for Biotechnology Information) to compare with known nucleotide sequences to identify amplified products.

2.4 Fluorescence *in situ* hybridization (FISH)

FISH is used to identify chromosome abnormalities by using fluorescent probes that localise to specific gene loci and can be visualised using a fluorescent microscope [266]. Two

commercial FISH probes were obtained from Cytocell (Cambridge, UK), an E2A (*TCF3*) breakapart probe and an AML1 (*RUNX1*) breakapart probe.

2.4.1 Cell fixation for FISH

When cells were ready for fixing post exposure, they were harvested and the pellet washed in PBS, centrifuged at 1200 rpm for 3 minutes. The supernatant was removed, leaving a small volume behind to allow resuspension of the pellet. Carnoy's fixing solution (3:1 methanol to acetic acid) was added to the cell solution and vortexed to mix before storage at -20°C [267].

2.4.2 Slide preparation

Cell suspensions were spun down for 3 minutes at full speed and the pellet resuspended in fresh Carnoy's fixative until the solution appeared clear. Using a 10µl pipette tip, 3µl of resuspended cells were dropped onto a SuperFrost microscope slide sitting on wet tissue and allowed to air dry.

Temperature and relative humidity are crucial to the spread of chromosomes in the metaphase stage on the slide. As the Carnoy's fixative evaporates, first methanol followed by the acetic acid, the cells flatten, and the chromosomes elongate [268-270]. Once the slides are dried, they are checked under a phase contrast microscope to look for the appearance of singular dark grey nuclei that can be selected for analysis.

2.4.3 Hybridisation

On a 10mm coverslip, 1µl of commercial FISH probe (Cytocell) was mixed with 1µl of hybridization solution containing: formamide; dextran sulphate; saline-sodium citrate (SSC). Formamide reduces the melting temperature of nucleic acids allowing hybridisation to occur at lower temperatures [271]. Dextran sulphate is a polymer used to increase hybridisation rates by accelerating DNA renaturation [272]. SSC is a salt solution used to improve stringency between the probe and target chromosomes by regulating ionic strength.

The prepared sample slide is inverted over the coverslip allowing adherence to the sample cell area by surface tension. Rubber cement was used to seal the edges of the coverslip preventing air drying out the sample. Slides were placed in a humid temperature-controlled hotplate, (Thermobrite: Leica Biosystems, Hybrite: Abbott Diagnostics) at 75°C for 5 minutes, followed by overnight incubation at 37°C.

2.4.4 Post-hybridisation washes

Following hybridisation, the rubber cement was removed from the slides and the slides were soaked in 2 x SCC to aid removal of the coverslip. Slides were then soaked for 2 minutes in Wash Buffer 1 (980ml H_2O, 20ml 2X SCC, 3ml Igepal) that was pre-heated to 72°C. Washing of the slides in a low salt buffer helps to remove background fluorescence from bound probes to mis-matched DNA by reducing the stability of the DNA complex. A further 2-minute wash in Wash Buffer 2 (900ml H2O, 100ml 20X SSC, 1ml Igepal) removes any remaining loose or unbound probe.

2.4.5 Counterstaining

The DNA-specific fluorescent probe 4',6-diamidino-2-phenylindole (DAPI) is used as a counterstain [273]. DAPI attaches to A-T rich regions of the DNA to produce a reversed banding pattern seen by G-banding. Slides were mounted in 10µl of DAPI and stored in the dark at 4°C until required for analysis.

2.4.6 Analysis

Slides were scored using a Zeiss Axioskop Fluorescence Microscope and an Olympus BX61 Fluorescence microscope, with a dual red/green filter and captured with Cytovision software. From each slide, 100 cells were viewed for the presence or absent of breakapart probe signals. As a negative control, bone marrow samples from four healthy patients were prepared as above, with 100 cells counted to identify the percentage of abnormal cells to be used as a comparison for positive translocation samples. When cells contain a wild type gene the probe signals are visualised together, and when a translocation is present the 2 probe signals are separated, at least 2 signal widths apart.

2.5 Statistical analysis

Analysis by one-way ANOVA with Bonferroni post-hoc tests were performed using Graphpad Prism and SPSS. Two-tailed Fisher's exact tests were performed using Graphpad Prism and https://www.socscistatistics.com/tests/fisher/. Hypergeometric testing was performed within *R* software.

2.6 Experimental design

2.6.1 Preliminary tests to examine the effects of environmental exposures on translocation events

To investigate the influence of caffeine, benzene, cotinine, and folic acid on translocation induction, NALM6 cells were exposed to a range of physiologically relevant concentrations (further described in each chapter). Exposure experiments were repeated in tandem, with

DMSO used as a common solvent to reduce variables across exposures and controls. An overview of the preliminary experiment design is outlined in figure 2.4.

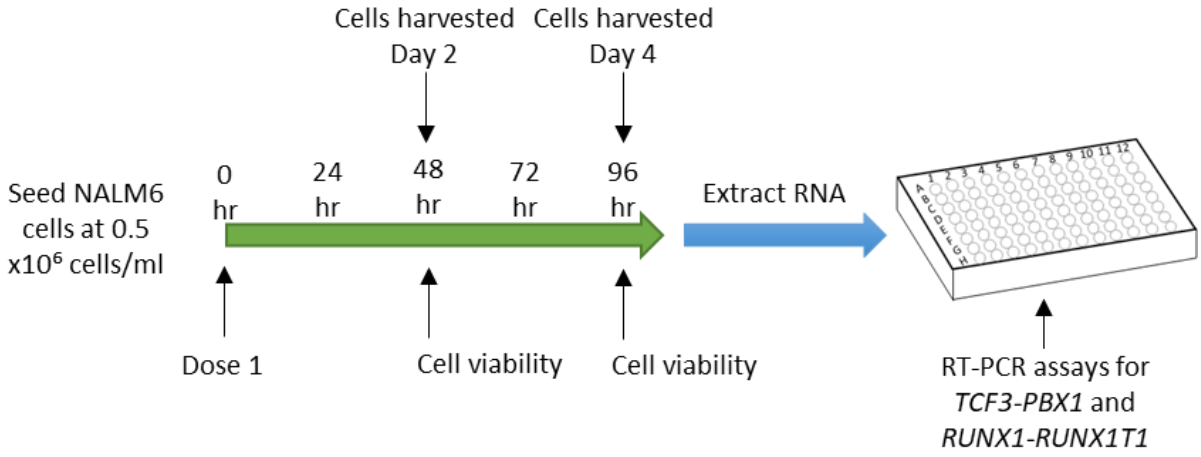


Figure 2.4. Experimental design of preliminary tests of environmental exposures on translocation induction. NALM6 cells were seeded into 12-well plates at a density of 0.5 x10^6 cells/ml. A single dose of exposure or DMSO was added into the wells incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2) before cells were harvested for RNA extraction. The resultant RNA was reverse transcribed for use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations. Trypan blue counts occurred every 24 hours and cell viability was measured via resazurin assay on day 2 and day 4.

NALM6 cells were seeded in 12-well plates at a density of 0.5 x 10^6 cells/ml in standard RPMI 1640 media for caffeine, benzene and cotinine, or folic acid free RPMI 1640 for folic acid exposures. A single dose of exposure (caffeine, benzene, cotinine or folic acid diluted in DMSO) or DMSO alone was added to one well for each concentration (one technical replicate), signifying time point 0. The experiment was repeated for three biological replicates. For caffeine, benzene and cotinine, two 12-well plates were set up with one plate left to incubate at 37°C for 48 hours, and the second plate left to incubate at 37°C for 96 hours, with cells harvested at each time point for RNA extraction. For folic acid, one plate was left to incubate at 37°C for 96 hours, before harvesting of cells for RNA extraction. RNA was extracted for each of the harvested cell samples, followed by DNase treatment and cDNA synthesis with Precision nanoScript2 Reverse Transcription Kit. The cDNA was used as a template for *TCF3-PBX1* and *RUNX1-RUNX1T1* RT-PCR assays to identify any positive translocation events, along with a H_2O negative control and cell line positive controls (Pre-B 697 and Kasumi-1 respectively), using HS Taq polymerase with 10pmol primer concentrations for 40 cycles. Positive RT-PCR results were purified and underwent Sanger sequencing by DBS Genomics, with resultant sequences analysed by Blast (Basic Local Alignment Search Tool).

To ensure that the exposure concentrations chosen did not significantly reduce growth and viability of NALM6 cells compared to the DMSO control, cell counts were performed every 24

hours using trypan blue. For caffeine, cotinine, and folic acid, a resazurin cell viability assay was also performed on day 2 and day 4. Results were presented as percentage viability of the DMSO control.

2.6.2 Experiments to investigate translocation events in response to environmental exposures in NALM6 cells for different models

To optimise the preliminary NALM6 cell model, two variables were considered: i) the impact of treating cells either immediately or 24 hours after initial seeding, and ii) single vs daily dosage of exposure. In the preliminary model, cells were exposed to a single dose of environmental exposure during the lag phase immediately after seeding, when cells have just been seeded at a low density and are adjusting to a new environment. NALM6 cells enter the exponential phase 24 hours after seeding, at which point cells are actively replicating which allows for a greater chance of mutations occurring [17]. Therefore, to investigate if translocations were more likely to be induced by environmental exposures when cells are actively replicating, cells were either seeded and rested for 24 hours to increase the number of actively replicating cells before exposure or treated straight after seeding. To reflect human exposure to caffeine, benzene, cotinine and folic acid, which likely occurs on a daily basis, cells were treated with daily doses of each exposure. An overview of the experimental design is shown in Figure 2.5.

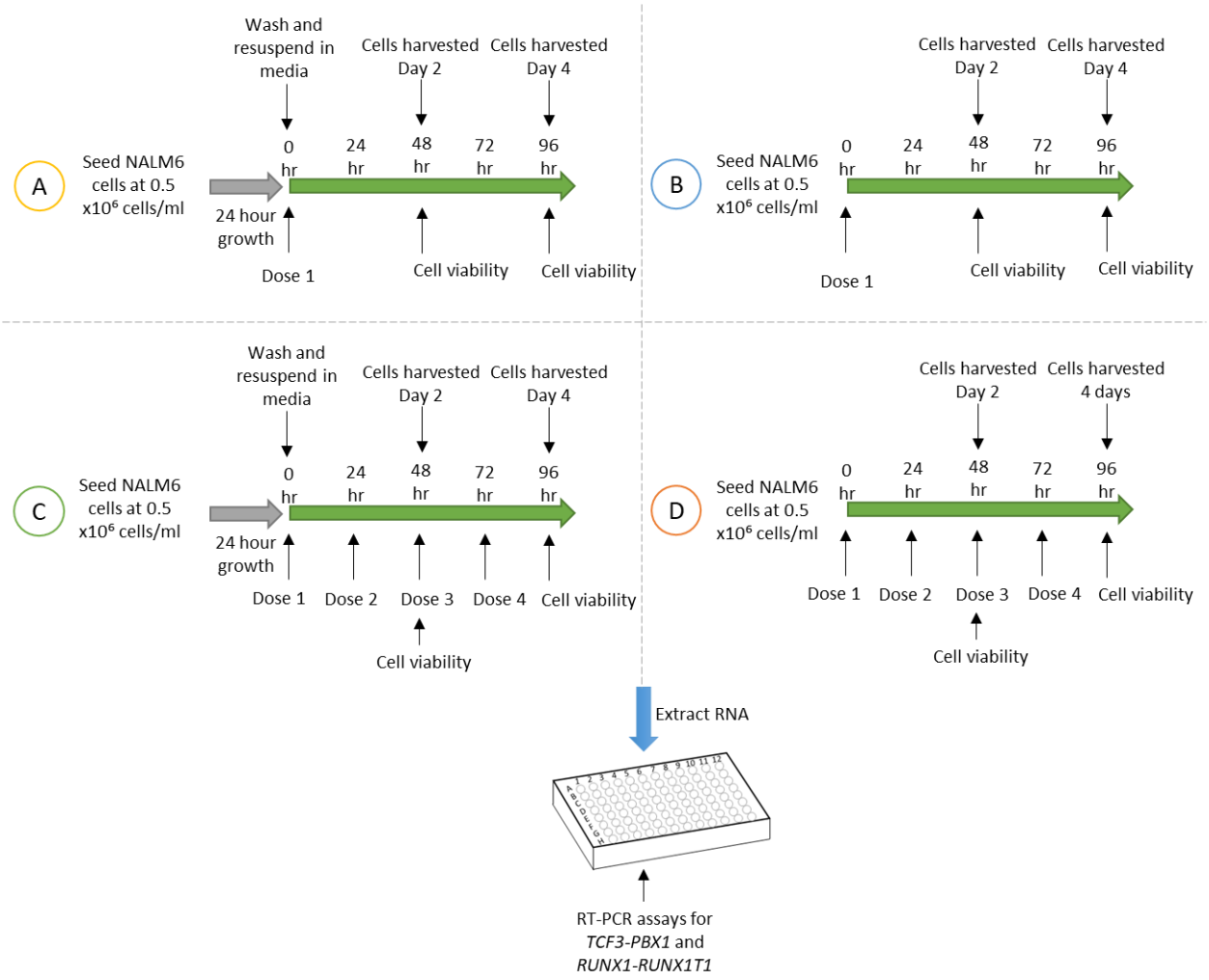


Figure 2.5. Model optimisation design to investigate the impact of daily dosing and exponential growth phase on NALM6 cells exposed to environmental exposures. NALM6 cells were seeded at a density of 0.5 x106 cells/ml in normal RPMI- 1640 media and incubated at 37°C. Model A: cells were allowed to grow for 24 hours, resuspended in fresh media and a single dose of exposure or DMSO was added. Model B: cells were administered a single dose of exposure or DMSO at the same time cells were seeded. Model C: cells were allowed to grow for 24 hours, resuspended in fresh media, then a dose of exposure or DMSO was administered every 24 hours for 48 hours or 96 hours. Model D: a daily dose of exposure or DMSO was administered every 24 hours for 48 hours or 96 hours, starting at the same time as seeding. Cells were harvested for RNA extraction after 2 or 4 days exposure. The resultant RNA was reverse transcribed for use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations. Trypan blue counts occurred every 24 hours following the first dose. On day 2 and day 4, cell viability was recorded via resazurin assay and samples frozen for Cyquant analysis.

For all models, NALM6 cells were seeded at a density of 0.5 x 10^6 cells/ml. In Model's B and D, cells are seeded in standard RPMI 1640 media for caffeine, benzene and cotinine, or folic acid free RPMI 1640 for folic acid exposures. In Model's A and C, all cells are seeded in standard RPMI 1640, then washed and resuspended in either standard RPMI 1640 media for caffeine, benzene and cotinine, or folic acid free RMPI 1640 for folic acid before first exposure. Each condition was repeated in three wells as technical replicates with each whole experiment repeated at two different times as biological replicates in which to assess translocation events, giving six replicates in total. Exposure experiments were repeated in tandem, with DMSO used as a common solvent to reduce variables across exposures and

controls. The first dose of exposure or DMSO was assigned time point 0. For caffeine, benzene and cotinine, two 12-well plates were set up, and incubated at 37°C for either 48 hours following first exposure (day 2 exposure), or 96 hours following first exposure (day 4 exposure) with cells harvested at each time point. For folic acid, cells were left to incubate at 37°C for 96 hours following first exposure (day 4 exposure) before harvesting. Once cells were harvested, half were used for RNA extraction and half were resuspended in Carnoy's solution for storage at -20°C. Following RNA extraction, samples underwent DNase treatment and cDNA synthesis with High-Capacity cDNA Reverse Transcription Kit. The cDNA was used as a template for *TCF3-PBX1* and *RUNX1-RUNX1T1* RT-PCR assays to identify translocation events, using HS Taq polymerase with 10pmol primer concentrations for 40 cycles. For each positive sample, the PCR product was used as a template for a nested PCR to amplify the breakpoint to concentrations sufficient for PCR purification and Sanger sequencing analysis by DBS genomics. PCR sequences were analysed using Blast (Basic Local Alignment Search Tool).

To ensure that the conditions in each model did not significantly reduce the growth and viability of NALM6 cells compared to a DMSO control, viability was measured for a single biological replicate using resazurin and Cyquant assays. The resazurin assay was performed on day 2 and day 4 following first exposure. For the Cyquant assay, small aliquots of cells were harvested on day 2 and day 4 and stored at -80°C, with the assay performed at a later date. Results are presented as percentage viability of the DMSO control. Cells were counted using trypan blue every 24 hours following the first exposure.

3 Optimisation of a model for the measurement of translocation events in NALM6 cells

3.1 Introduction

There have been few *in vitro* studies conducted that investigate the effect of environmental exposures on the induction of translocations involved in childhood leukaemia, and those that have used different methods for inducing and analysing translocation events. Strick *et al*. 2000 investigated *MLL* cleavage on exposure to bioflavonoids using haematopoietic progenitor cell lines BV173 and K562 alongside primary haematopoietic progenitor cells, cultured for 6 or 16 hours in different bioflavonoids or chemotherapy drugs, previously reported to induce *MLL* cleavage [274]. Southern blot analysis was conducted on genomic DNA from each sample to identify cleavage along the *MLL* gene, and a mechanism of targeted cellular topoisomerase II inhibition was confirmed through further *in vitro* assays. Further studies have also focused on bioflavonoids and *MLL* translocations, including van Waalwijk van Doorn-Khosrovani *et al*. 2007 who used umbilical cord blood treated overnight with bioflavonoids or etoposide, first assessed for DNA damage by Comet Assay, then left to recover for 48 hours for analysis of wild type *MLL* by inverse PCR and FISH [275]. Bariar *et al*. 2019 used mouse embryonic stem cells with GFP constructs to produce a reporter system that was able to identify and quantify translocation events between *MLL* and *AF9* [276]. Cells were incubated with flavonoids or etoposide for 1 hour and then monitored for GFP+ fluorescence every 24 hours. Translocations were confirmed by DNA extraction and PCR amplification of the breakpoint using GFP primers.

For use as a model to identify environmental risk factors that induce translocations associated with childhood leukaemia, these previous studies are limited to having only investigated bioflavonoids and only target translocations involving *MLL*. These studies are limited to only detecting 1 translocation event, whereas there are many translocation events that have been associated with childhood leukaemia. These studies work on the basis that bioflavonoids are topoisomerase II inhibitors, which prevent religation of DSBs induced by topoisomerase II, which could then go on to form chromosomal translocations. However, some environmental exposures could induce translocations via different mechanisms, which means that the study design would be limited to only exposures that work as topoisomerase inhibitors. The concentrations of exposures used in this study were often much higher than a person would be exposed to daily, and exposure length varied between each study.

The key aim of this chapter was to optimise an *in vitro* translocation model that could detect a variety of chromosomal translocation events associated with childhood leukaemia, using a cell line that could be exposed to a wide number of environmental risk factors.

Reverse transcription PCR (RT-PCR) is often used for clinical diagnosis of leukaemia cytogenetic subtypes and detection of minimal residual disease [261]. As chromosomal translocation breakpoints can vary, using genomic DNA to detect translocations via PCR can be difficult due to potentially very large target sequences that require amplification. RT-PCR can analyse the fusion genes that are produced by chromosomal translocations providing shorter target sequences for PCR amplification across the breakpoint junction.

To develop PCR assays for common childhood leukaemia translocations, primer pairs were chosen from the literature where they had previously been used to detect translocations using RT-PCR (Table 2.4). The primer sequences were entered into the Basic Local Alignment Search Tool (BLAST) to ensure the sequence covered the expected region within the respective translocated genes. Cell lines known to contain the selected translocation were used as positive controls in PCR reactions, with RNA extracted from cell pellets and reverse transcribed to cDNA (Table 2.4). Each RT-PCR assay was optimised using the respective positive templates and the sensitivity of each assay to detect translocations was analysed.

A defective DDR is more likely to contribute to chromosomal translocations [16]. NALM6 is a pre-B acute lymphoblastic leukemia cell line which is wild type for p53 but harbours a homozygous mutation in *MSH2*, meaning cells lack the protein, which is a key component in the mismatch repair mechanism. The mismatch repair response is involved in repairing mismatched bases and small insertions–deletion loops by inducing a single strand nick to promote removal of the mispair, followed by religation [277]. Deficiencies in mismatch repair responses are associated with cancers including colorectal cancer and leukaemia and have been shown to increase mutation rate [278, 279]. When cells are mismatch repair deficient, double strand breaks are often repaired by microhomology-mediated end joining, which leads to microsatellite instability, which in turn supresses chromosomal instability and may reduce the likelihood of translocations forming, due to double strand breaks being repaired. The mismatch repair response is also involved in supressing HR, preventing potential incorrect recombination, that could lead to translocations, and as such, deficiencies in mismatch repair may allow for incorrect recombination by HR [279, 280]. Some studies in mismatch repair deficient cells found a lack of chromosomal abnormalities, using methods such as conventional karyotyping and flow cytometrical ploidy analysis, suggesting that deficient mismatch repair may supress the formation of abnormalities such as chromosomal translocations [281, 282]. A study by Abdel-Rahman *et al.* 2001 however, identified multiple chromosomal translocations in colon cancer cell lines with defective mismatch repair, using spectral karyotyping [283]. However, a later study, Abdel-Rahman *et al.* 2005, found that restoring the mismatch repair response in these cells did not prevent translocations from

forming and actually had increased chromosomal instability and sensitivity to radiation, suggesting other mechanisms may be involved [284]. NALM6 cells have been shown to have a greater resistance to irradiation and compromised DSB repair [285]. Previously, NALM6 has been used to detect *MLL* translocations with exposure to the chemotherapy drug, etoposide [159]. Furthermore, although NALM6 has been found to have a deletion of the long arm of chromosome 5 and FISH analysis has shown the presence of translocation, t(5;12)(q33.2;p13.2), the breakpoint of the translocation does not affect *ETV6*, one of the genes of interest [286]. As such, NALM6, was chosen to investigate the induction of chromosomal translocations due being a pre-B cell type, the most common cell type in childhood leukaemia, with DDR deficiencies and without the presence of the translocations under investigation [9, 286].

Etoposide, a known translocation inducer [159], was used in the current study to assess the ability of the RT-PCR assays developed to detect childhood leukaemia translocations in NALM6 cells. As etoposide has already been associated with known leukaemia translocations *MLL* and *RUNX1*, it would act as a positive exposure control to understand if the NALM6 translocation model would be able to detect translocations before moving onto exposures associated with childhood leukaemia risk that have previously not been studied in the context of inducing translocations [159, 287].

3.2 Development of RT-PCR assays

3.2.1 Optimisation of *GAPDH* and *Beta Actin* RT-PCR assays

RT-PCR assays for *GAPDH* and *Beta Actin* were developed to determine the optimal annealing temperature for translation to quantitative PCR assays (qPCR) so that these primer pairs could be used as reference genes. An annealing temperature gradient was performed using AmpliTaq Gold™ 360 polymerase for 30 amplification cycles. With a primer concentration of 2µM, cDNA from the HL60 cell line was used as a positive control at a concentration of 5ng/µl. An annealing temperature of 60°C produced successful amplification of both *GAPDH* and *Beta Actin* at 150bp, Figure 3.1A. RT-PCR assays were repeated using HL60 and Kasumi-1 cDNA to ensure successful amplification of reference genes across different cell types, Figure 3.1B.

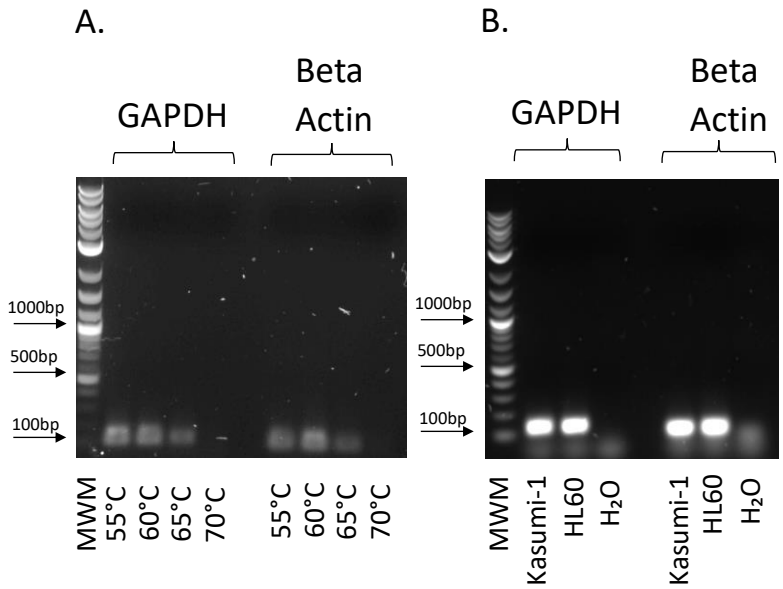


Figure 3.1. Gel electrophoresis visualisation of GAPDH and Beta Actin RT-PCR assay using AmpliTaq Gold™ 360 polymerase. A) temperature gradient, B) 60°C annealing temperature with Kasumi-1 and HL60 cDNA. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

The HL60 positive controls for *GAPDH* RT-PCR was sent for Sanger sequencing, Table 3.1, and the correct target sequence was identified.

Table 3.1. Sanger sequencing for GAPDH PCR.

Sample	Sequence of gene	Genes identified in BLAST
GAPDH	CTGCTTTTACTCTGGTAAAGTGGATATTGTTGCCATCAAT	*GAPDH*

3.2.2 Optimisation of *TCF3-PBX1* and *RUNX1-RUNX1T1* RT-PCR assays

To determine the optimal temperature to amplify *RUNX1-RUNX1T1* and *TCF3-PBX1* fusions, a gradient RT-PCR was performed using AmpliTaq Gold™ 360 polymerase and primers TCF3_F / PBX1_R and RUNX1_F / RUNX1T1_R. A primer concentration of 2µM was used with a cDNA template from translocation positive cell lines, 5ng/µl of Pre-B 697 for *TCF3-PBX1*, and 4ng/µl of Kasumi-1 for *RUNX1-RUNX1T1*. Annealing temperatures ranged between 50 to 60°C for *TCF3-PBX1* and 55 to 70°C for *RUNX1-RUNX1T1*, for 30 amplification cycles.

Under these conditions, bands were detected for *TCF3-PBX1* at 450bp across all annealing temperatures, Figure 3.2A. For *RUNX1-RUNX1T1* bands at 500bp were detected at 55°C and 60°C, Figure 3.2B. An annealing temperature of 60°C was chosen for all further RT-PCR reactions for *TCF3-PBX1* and *RUNX1-RUNX1T1*.

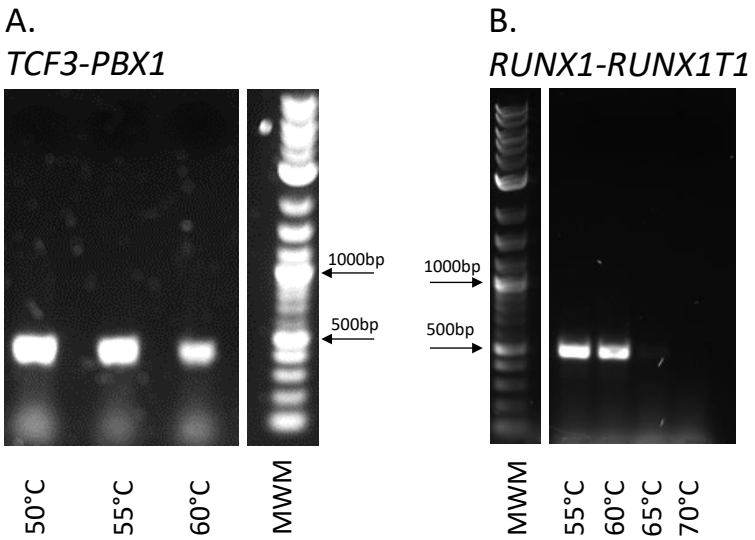


Figure 3.2. Gel electrophoresis visualisation of of RT-PCR assay for A) TCF3-PBX1 with Pre-B 697 positive cell line and B) RUNX1-RUNX1T1 with Kasumi-1 positive cell line . MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

To ensure the primer pairs only amplified the target fusion genes, the RT-PCR was repeated with AmpliTaq Gold™ 360 using 2.5ng/µl HL60 cDNA, which does not contain either translocation events. As high amounts of primer dimer were observed in previous RT-PCRs (Figure 3.2), the primer concentration was adjusted to 0.3µM with 2.5ng/µl cDNA template for *TCF3-PBX1* and 2ng/µl cDNA for *RUNX1-RUNX1T1*. RT-PCR product bands were only seen in the positive cell line controls, Pre-B 697 for *TCF3-PBX1* and Kasumi-1 for *RUNX1-RUNX1T1*, (Figure 3.3). The *ETV6-RUNX1* results will be discussed in the section below.

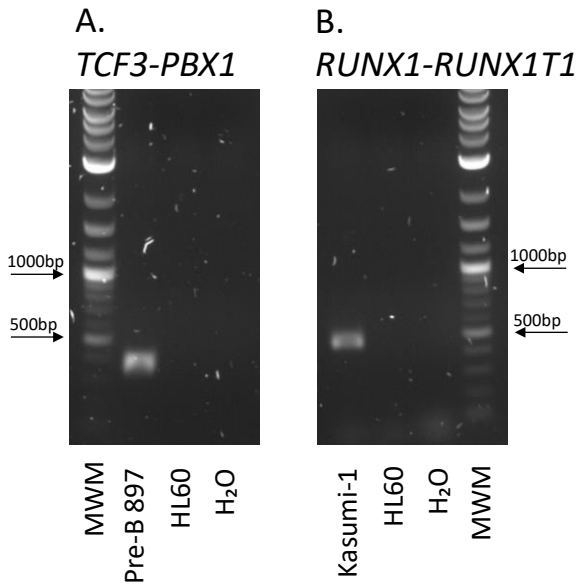


Figure 3.3. Gel electrophoresis visualisation of positive and negative cell line controls for A) TCF3-PBX1 and B) RUNX1-RUNX1T1 RT-PCR assays. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

To further optimise the RT-PCR reactions, a range of primer concentrations, cDNA concentrations and amplification cycles were explored using AmpliTaq Gold™ 360 to increase yield and reduce primer dimers. For *RUNX1-RUNX1T1* the primer concentrations used included 0.3µM, 0.5µM and 4µM. For each primer concentration a range of positive cDNA template concentrations were used from 4ng/µl to 0.1ng/µl of Kasumi-1 with 40 amplification cycles. Positive bands were detected across all primer concentrations, with intense primer dimer bands seen at 4µM primer concentration, Figure 3.4A. For *TCF3-PBX1* a primer concentration of 0.5µM was used with amplification cycles increased to 40 cycles. A range of positive cDNA template concentrations were used from 5ng/µl to 0.125ng/µl of Pre-B 697. Bands were detected for all positive samples as seen in Figure 3.4B. Bands were detected at the lowest cDNA concentration for both translocation events, demonstrating that the fusion genes can still be detected at lower copies of fusion template when using a 100% cell line population.

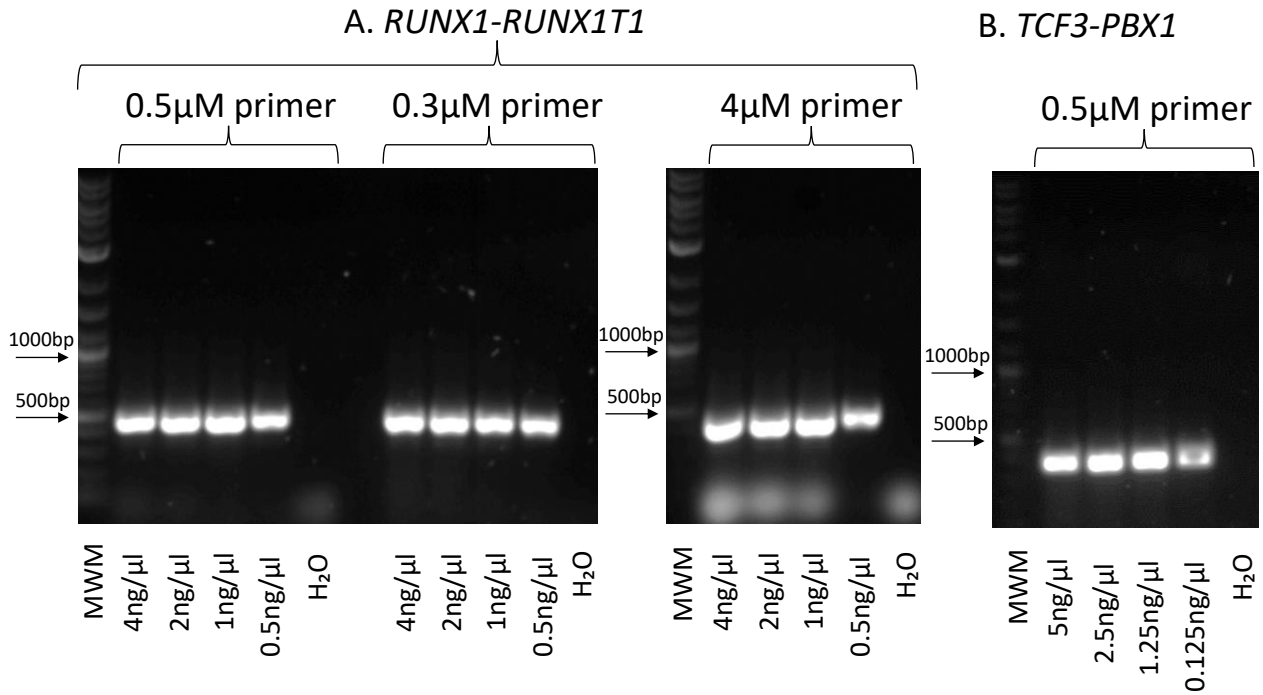


Figure 3.4. Gel electrophoresis visualisation of TCF3-PBX1 RUNX1-RUNX1T1 RT-PCR assay using AmpliTaq Gold™ 360 polymerase with a range of primer and cDNA concentrations for 40 amplification cycles. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

When switching to PCRBio HS Taq Mix for cost effectiveness, an RT-PCR was performed using 2.5ng/µl of positive control cDNA and 0.5µM of each primer to ensure the optimised temperature of 60°C would work for *TCF3-PBX1* and *RUNX1-RUNX1T1*, Figure 3.5. Positive bands were observed for both primer pairs. The *ETV6-RUNX1* and *KMT2A-AFF1* results will be discussed in the section below.

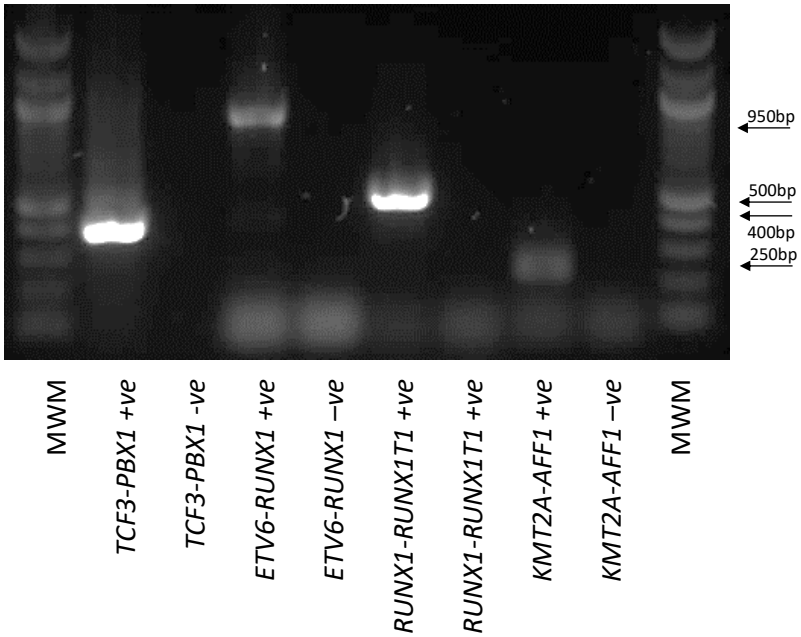


Figure 3.5. Gel electrophoresis results for PCR annealing temperature check using PCRBio HS Taq Mix. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder). ETV6-RUNX1 and KMT2A-AFF1 results will be discussed in the section 3.3.2.

To confirm positive samples were the correct target sequence, samples were PCR purified and sent for Sanger sequencing. Table 3.2 shows the sequence of the translocation junction for Kasumi-1 identifying *RUNX1* and *RUNX1T1* sequences, and Pre-B 697 identifying *TCF3* and *PBX1* sequences.

Table 3.2. Sanger sequencing of RUNX1-RUNX1T1 and TCF3-PBX1 positive controls. The sequence relating to each translocation partner is highlighted in red or blue.

	Sequence of translocation junction	**Genes identified in BLAST**
Kasumi-1	GATGGGCCCCGAGAACCTCGAAATCGTTACTGAGAAG	*RUNX1* - *RUNX1T1*
Pre-B 697	GGCCTCCCGACTCCTACAGTGTTTTGAGTATCCGAGGAGCC	*TCF3* - *PBX1*

To understand the sensitivity of the RT-PCR assay, positive cell lines were diluted with negative cell lines to give different percentages of positive fusions. For *TCF3-PBX1*, the positive Pre-B 697 cell line was diluted with negative Reh cell line. For *RUNX1-RUNX1T1*, the positive Kasumi-1 cell line was diluted with negative Pre-B 697 cell line. The RNA was then extracted from each cell mixture, allowing the amount of total RNA to be consistent

across different percentages of target fusion product. For *TCF3-PBX1*, the RT-PCR assay can visualise translocations in as low as 1 in 100 cells, Figure 3.6.

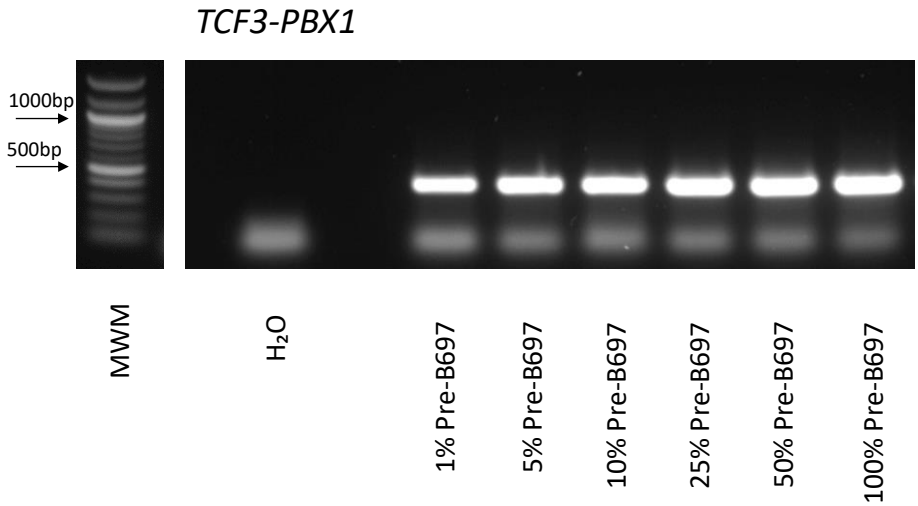


Figure 3.6. Gel electrophoresis image for RT-PCR sensitivity of TCF3-PBX1. PCRBio HS Taq polymerase, for 40 amplification cycles with a primer concentration of 0.5µM. Final cDNA concentration used in RT-PCR was 5ng/µl. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

For *RUNX1-RUNX1T1*, the RT-PCR assay can also visualise translocations as low as 1 in 100 cells, Figure 3.7, however the 1% band is smeary, so although it can be observed, the product is not as strong as higher percentages of cells so a 2nd confirmation would be beneficial in these cases. Although the bands are the expected molecular weight for the positive translocation, confirmation by Sanger sequencing would also be beneficial.

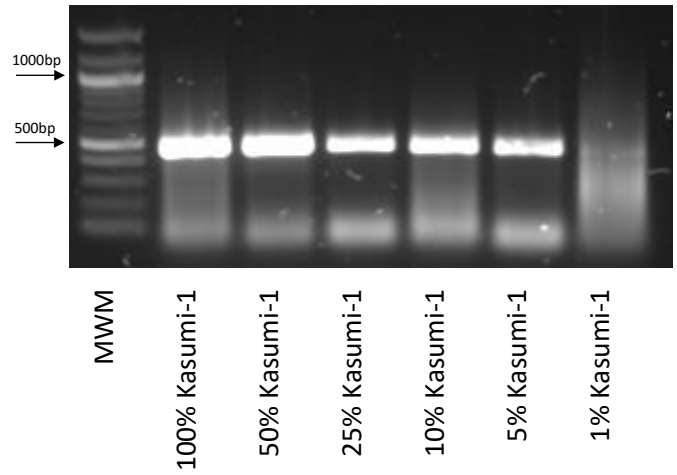


Figure 3.7. Gel electrophoresis image for RT-PCR sensitivity of RUNX1-RUNX1T1 using PCRBio HS Taq polymerase, for 40 amplification cycles with a primer concentration of 0.5µM. Final cDNA concentration used in RT-PCR was 5ng/µl. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

Due to the presence of primer dimer, the RT-PCRs were repeated using a gradient of forward and reverse primer concentrations for *TCF3-PBX1* and *RUNX1-RUNX1T1*, Figure 3.8. All primer concentration combinations gave amplification of the expected fusion product with visually low production of primer dimer (below 100bp), when using 100% positive cell line cDNA as a template.

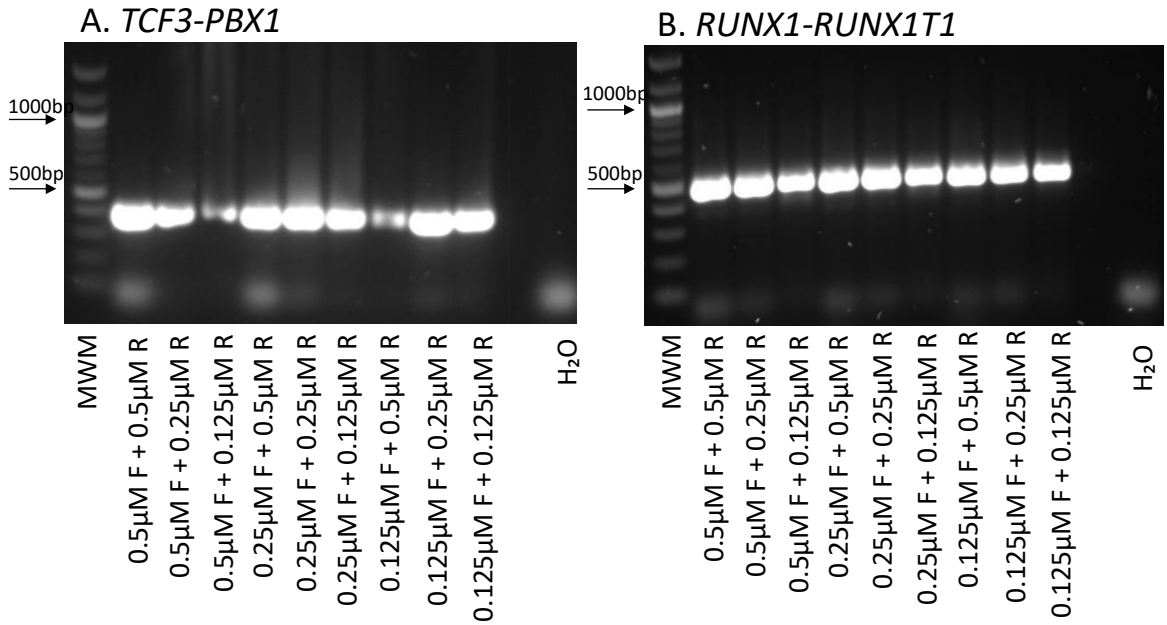


Figure 3.8. Gel electrophoresis image for RT-PCR amplification of A) TCF3-PBX1 and B) RUNX1-RUNX1T1 using a gradient of primer concentrations. PCRBio HS Taq polymerase, for 40 amplification cycles. Primer concentrations were varied using 2.5ng/µl of positive control templates, Pre-B 697 for TCF3-PBX1 and Kasumi-1 for RUNX1-RUNX1T1. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

However, when using samples which are expected to have a low copy number of the expected fusion compared to the 100% cell line controls, such as from NALM6 cells exposed to environmental exposures (benzene, cotinine, caffeine, folic acid), or when using 1% cell line cDNA, high amounts of primer dimer are still observed when using varying primer combinations, Figure 3.9. In some cases, positive samples were not amplified sufficiently for visualisation during gel electrophoresis, Figure 3.9. Due to this, unless stated otherwise, a higher primer concentration of 0.5µM was used for both forward and reverse primers for *TCF3-PBX1* and *RUNX1-RUNX1T1* fusion RT-PCRs.

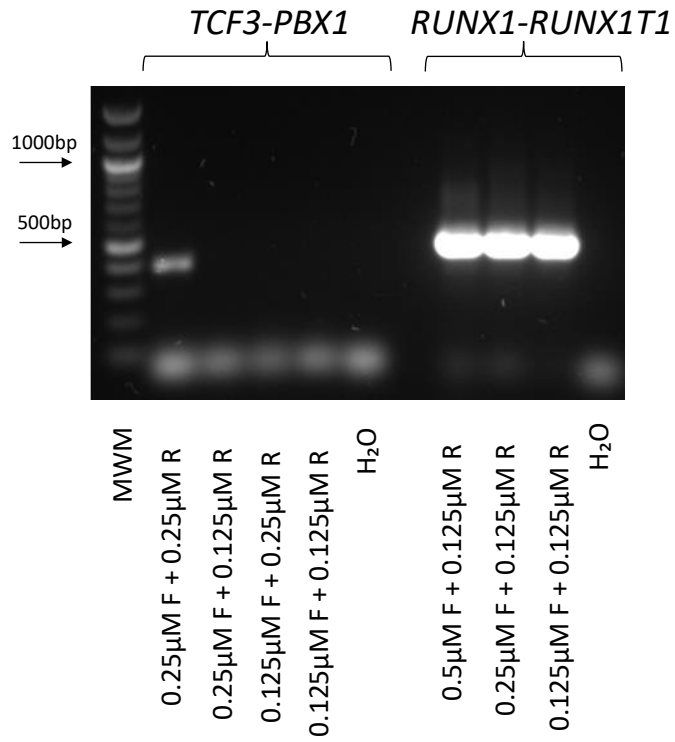


Figure 3.9. Gel electrophoresis image for RT-PCR amplification of TCF3-PBX1 and RUNX1-RUNX1T1 with varying primer concentrations. PCRBio HS Taq polymerase, for 40 amplification cycles. Primer concentrations were varied using 5ng/µl translocation positive low copy number template for TCF3-PBX1 and 1ng/µl Kasumi-1 for RUNX1-RUNX1T1. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

Due to low copy number and subsequently low DNA concentration obtained from gel extraction of PCR products, samples were further amplified using nested primers to increase amount of PCR product required for sequencing to confirm the breakpoint. Primers were designed to amplify within the original target sequence for *TCF3-PBX1* and *RUNX1-RUNX1T1* whilst remaining outside of the breakpoint region. For *TCF3-PBX1*, the primers used were TFC3_FNE forward primer and PBX1_RNE reverse primer, shown in Table 2.4. For *RUNX1-RUNX1T1*, RUNX1_FNE was used as a forward primer and RUNX1T1_RNE was used as a reverse primer. Using PCRBio HS Taq polymerase, an annealing temperature of 60°C, 40 amplification cycles and a primer concentration of 0.5µM, PCR products were diluted 1 in 10 with water and used as templates for the nested PCR reactions. Strong bands were seen using positive PCR products as the template for both *TCF3-PBX1* and *RUNX1-RUNX1T1*, Figure 3.10.

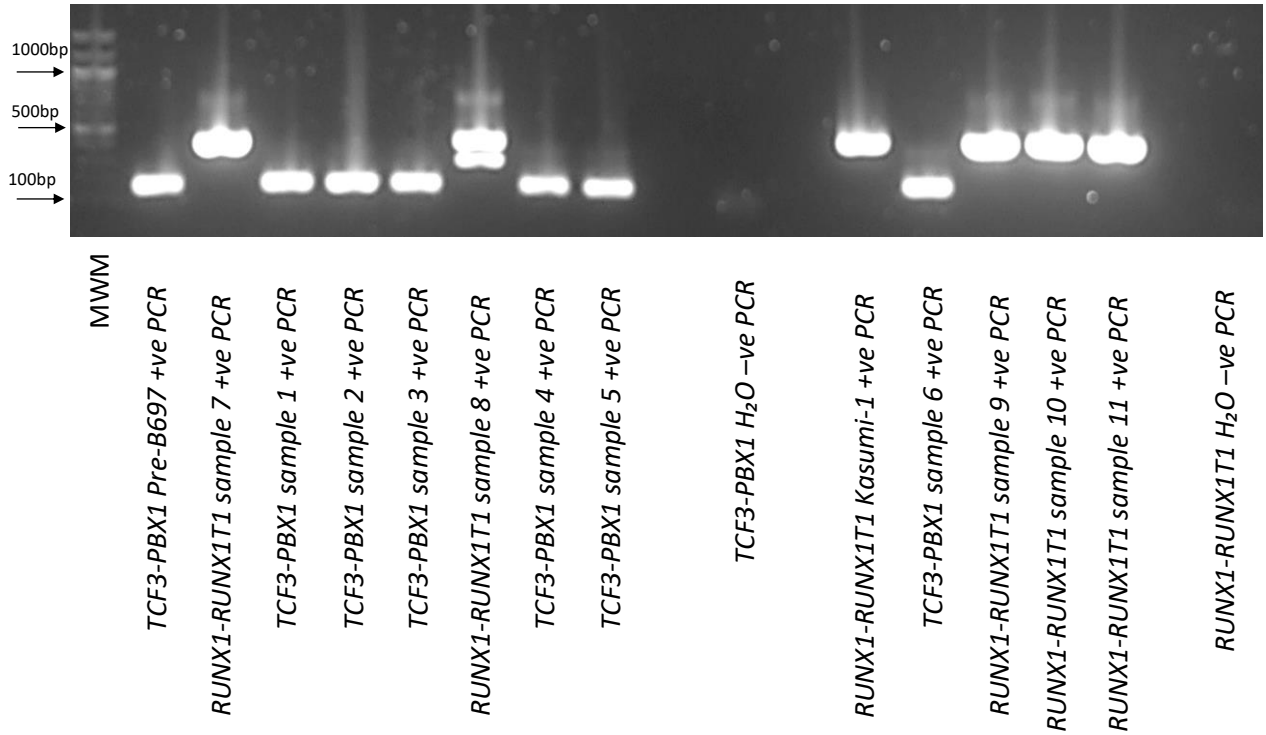


Figure 3.10. Gel electrophoresis image for PCR amplification of TCF3-PBX1 and RUNX1-RUNX1T1 using nested primers. The templates used as examples in this diagram are from positive PCR samples observed in chapters 4-7. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

Products were gel extracted for purification and sequencing. If multiple bands were present, then the visible bands were extracted and amplified. Sanger sequencing for the positive cell line templates show the target sequence at 400bp for Kasumi-1 identifying *RUNX1* and *RUNX1T1* sequences, and at 150bp for Pre-B 697 identifying *TCF3* and *PBX1* sequences, Table 3.3. The breakpoint junction for the cell line controls in the nested PCR is the same as the breakpoint junction using the original RT-PCR primers, as seen in Table 3.2, therefore it is appropriate to use the nested primers to amplify when needed.

Table 3.3. Sanger sequencing for RUNX1-RUNX1T1 and TCF3-PBX1 nested PCRs. The sequence relating to each translocation partner is highlighted in red or blue.

	Sequence of translocation junction	**Genes identified in BLAST**
Kasumi-1 PCR Nested	TGGGCCCCGAGAACCTCGAAATCGTACTG AGAAGCA	*RUNX1* - *RUNX1T1*
Pre-B 697 PCR Nested	CGGCCTCCCGACTCCTACAGTGTTTTGAG TATCCGAGGA	*TCF3* - *PBX1*

With reliable RT-PCR amplification, *TCF3-PBX1* and *RUNX1-RUNX1T1* were chosen as the translocation targets to be investigated in the NALM6 model.

3.2.3 Optimisation of *ETV6-RUNX1* and *KMT2A-AFF1* PCR assays

RT-PCR assays for *ETV6-RUNX1* and *KMT2A-AFF1* were also attempted and went through many optimisation steps with varying results. Annealing temperature optimisation was performed using AmpliTaq Gold™ 360, with 30 amplification cycles, 2µM primer concentration and 5ng/µl of positive cell line control cDNA template, Reh for *ETV6-RUNX1* and MV4-11 *KMT2A-AFF1*. At an annealing temperature of 55°C and 60°C, faint bands can be seen for *ETV6-RUNX1* close to 1000bp, Figure 3.11A. No bands were observed for *KMT2A-AFF1* between 55°C-65°C, Figure 3.11A. When the same conditions were repeated at an annealing temperature of 50°C for *KMT2A-AFF1*, a band can be seen at 250bp, Figure 3.11B. For *ETV6-RUNX1*, when the same conditions were repeated at 60°C, a clear band can be seen near 1000bp, Figure 3.11B.

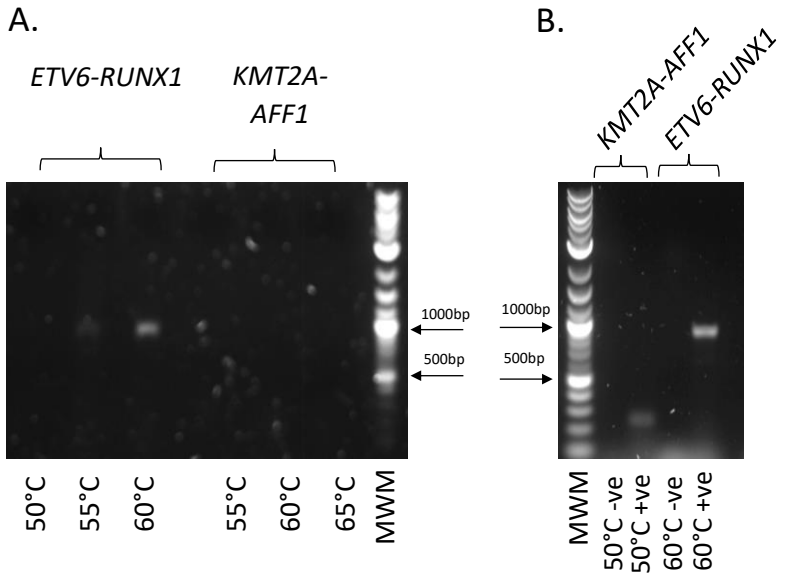


Figure 3.11. Gel electrophoresis visualisation of ETV6-RUNX1 and KMT2A-AFF1 PCR assay using AmpliTaq Gold™ 360 polymerase. A) Temperature gradient and B) Final annealing temperatures. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

To enhance the visibility of the *ETV6-RUNX1* PCR product, AmpliTaq Gold™ 360 was used at annealing temperatures of 60°C and 65°C, with and without additional 1mM Magnesium Chloride and GC enhancer. The AmpliTaq Gold™ 360 alone gave the brightest PCR product bands at both 60°C and 65°C, although these bands are still faint in comparison to PCR bands observed for other primer pairs, Figure 3.12. The annealing temperature of 60°C was used for further RT-PCR reactions.

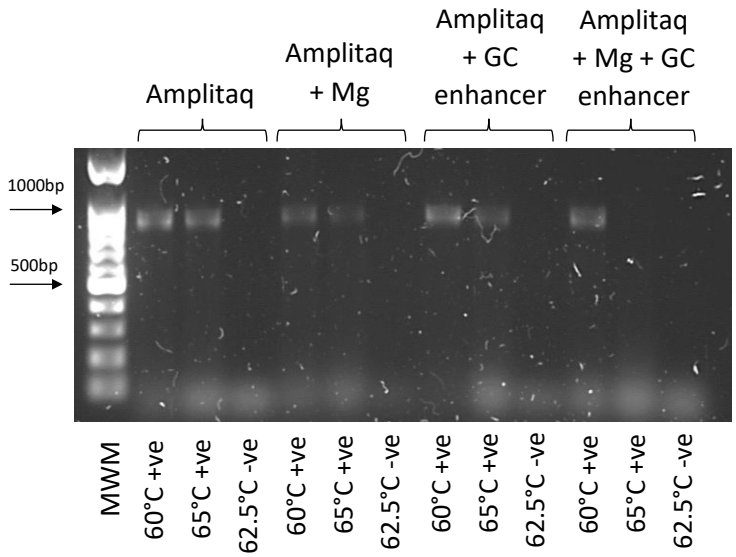


Figure 3.12. Gel electrophoresis visualisation of ETV6-RUNX1 PCR assay using AmpliTaq Gold™ 360 polymerase. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

Due to the presence of primer dimer in the *ETV6-RUNX1* and *KMT2A-AFF1* RT-PCR assays, the assays were further optimised by varying the concentration of primer and cDNA template used. To increase the visibility of PCR product bands, the number of amplification cycles was increased to 40 cycles. For *ETV6-RUNX1*, product bands were seen for cDNA template concentrations above 1.25ng/µl, across all primer concentrations, Figure 3.13. The brightest bands were seen using 0.5µM primer, although these bands are still faint in comparison to PCR bands observed for other primer pairs.

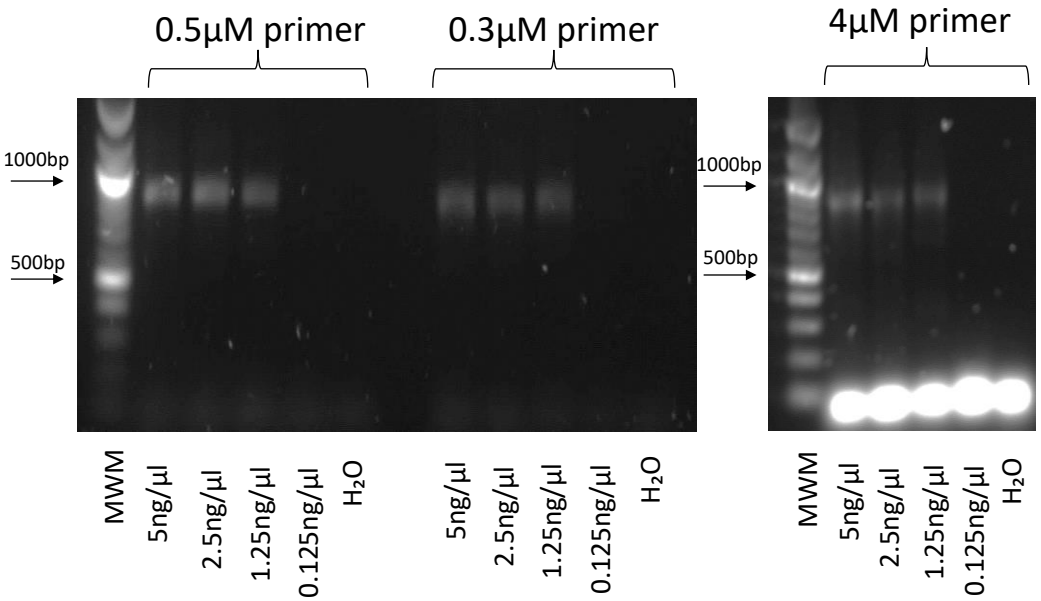


Figure 3.13. Gel electrophoresis visualisation of ETV6-RUNX1 PCR using AmpliTaq Gold™ 360 polymerase. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

For *KMT2A-AFF1*, PCR product bands were only seen at a cDNA template concentration of 5ng/µl and primer concentration of 0.5µM and 0.3µM, Figure 3.14. All bands were still very faint and appeared at less than 200bp.

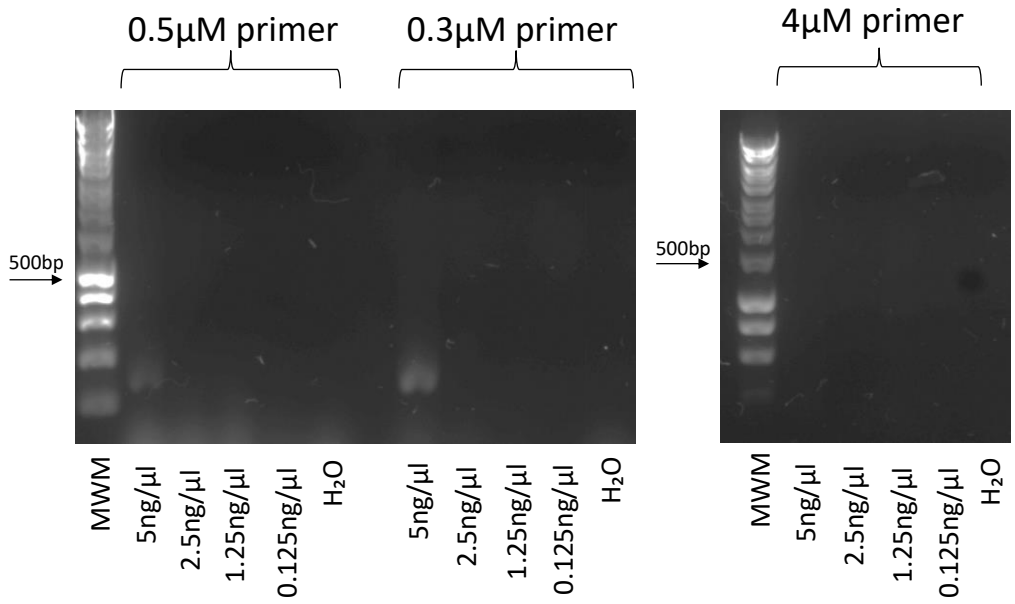


Figure 3.14. Gel electrophoresis visualisation of KMT2A-AFF1 PCR using AmpliTaq Gold™ 360 polymerase with a range of primer and cDNA concentrations. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

When switching to PCRBio HS Taq Mix, the RT-PCR was performed using 2.5ng/µl of positive control and 0.5µM of each primer following the conditions set out in section 2.3.5. The previous annealing temperatures of 60°C for *ETV6-RUNX1* and 50°C for *KMT2A-AFF1* were used. A band can be seen in Figure 3.5 at 250bp for *KMT2A-AFF1* using positive cell line MV4-11 cDNA as the template. Sanger sequencing for MV4-11 identified *KMT2A* and *AFF1* sequences confirming the product was the expected fusion, Table 3.4.

For *ETV6-RUNX1*, a bright band can be seen at 950bp using positive cell line Reh, Figure 3.5. However, the *ETV6-RUNX1* positive lane also contains some smearing and possible other faint bands. When the RT-PCR assay for *ETV6-RUNX1* was performed using cDNA templates from NALM6 cells exposed to DMSO that contained no or very low copies of the translocation, a bright band could be seen in some samples at 500bp, Figure 3.15A. Following gel extraction and Sanger sequencing of both bands, 500bp and 950bp, the band at 500bp was identified as Transportin 3 (*TNPO3*) located on chromosome 7, and therefore not related to the fusion gene, whilst the band at 950bp was the expected *ETV6-RUNX1* fusion, Table 3.4. Due to the presence of the unspecific binding when using samples with no or low translocation copies, the *ETV6-RUNX1* RT-PCR assay was not chosen as a target for the model.

When the RT-PCR assay for *KMT2A-AFF1* was performed using cDNA templates that contained no or very low copies of the translocation (such as cDNA from NALM6 cells exposed to DMSO, expected to contain no or very low numbers of translocations, used as an example in Figure 3.15) large amounts of smearing could be consistently seen in each lane, Figure 3.15B. Due to the smearing and very faint bands making true positives difficult to identify, the *KMT2A-AFF1* RT-PCR assay was not chosen as a target for the model.

Table 3.4. Sanger sequencing for ETV6-RUNX1 and KMT2A-AFF1 positive cell line controls. The sequence relating to each translocation partner is highlighted in red or blue. Underlined sequences in black refer to shared sequences between the translocation pair.

Sample	Sequence of translocation junction or gene	Genes identified in BLAST
MV4-11	CAAAAACCAAAAGAAAAGGAAATGACCCATTCATGGC	*KMT2A* - *AFF1*
Reh 950bp	CCCATTGGGAGAATAGCAGAATGCATACTTGGAATGAA	*ETV6* - *RUNX1*
Reh 500bp	GCCTCACCAANGCTGGAGTCTCTCCCTGCTCTGGCGGGA	*TNPO3*

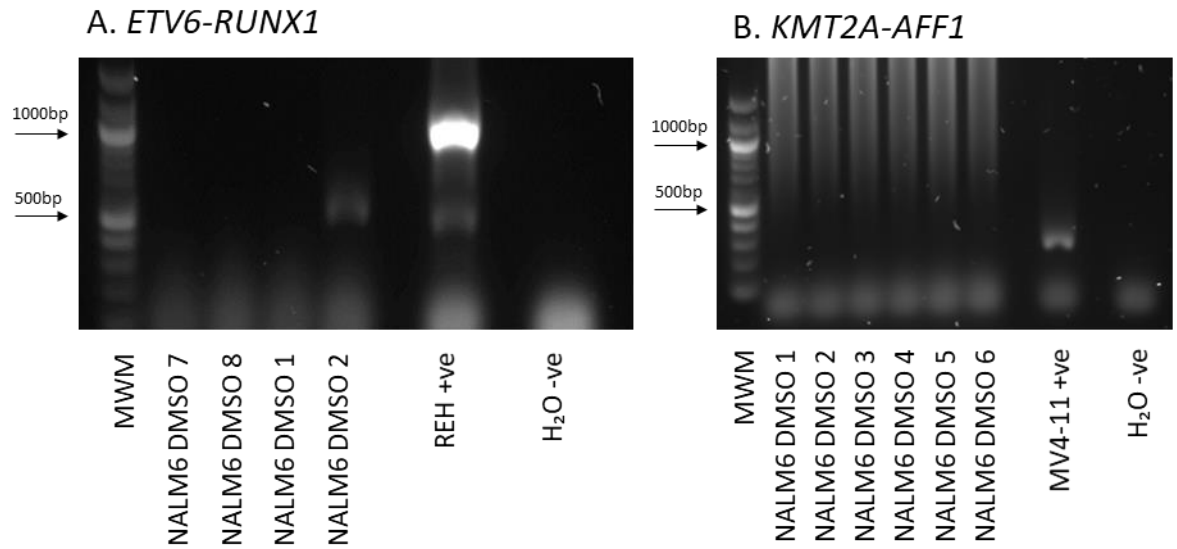


Figure 3.15. Gel electrophoresis visualisation of A) ETV6-RUNX1 and B) KMT2A-AFF1 RT-PCR using low or no translocation cDNA as template. Performed using PCRBio HS Taq polymerase, an annealing temperature of 60°C (ETV6-RUNX1) or 50°C (KMT2A-AFF1) for 40 amplification cycles and a primer concentration of 0.5µM primer. The reaction templates included 5ng/µl of cDNA from samples of NALM6 cells exposed to DMSO (no or low positive translocation), 2.5ng/µl of positive cDNA (Reh for ETV6-RUNX1 and MV4-11 for KMT2A-AFF1) or a negative control H_2O. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

3.2.4 Optimisation of *BCR-ABL1* and *PML-RARA* PCR assays

Optimisation of RT-PCR assays to detect *BCR-ABL1* and *PML-RARA* translocations started using AmpliTaq Gold™ 360 polymerase with BCR_F / ABL1_R and PML_F / RARA_R primers. An annealing temperature gradient was performed with 30 amplification cycles,

2µM primer concentration and 5ng/µl positive cell line cDNA template, K562 for *BCR-ABL1* and NB4 for *PML-RARA,* Table 2.4. No product bands were seen for either *BCR-ABL1* or *PML-RARA*, Figure 3.16.

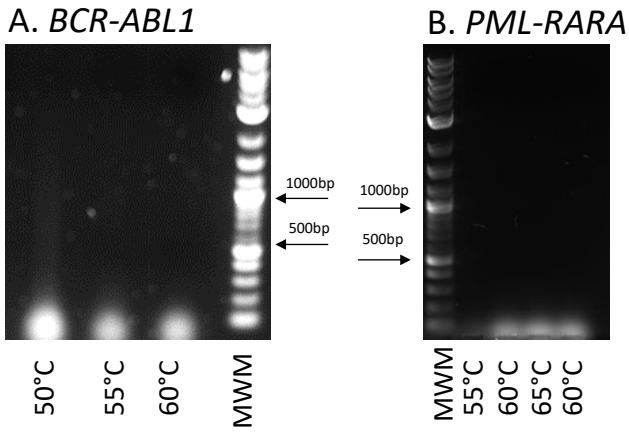


Figure 3.16. Gel electrophoresis visualisation of A) BCR-ABL1 and B) PML-RARA PCR assay using AmpliTaq Gold™ 360 polymerase. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

RT-PCR assays for *BCR-ABL1* and *PML-RARA* were repeated using AmpliTaq Gold™ 360 with and without the addition of 1mM Magnesium Chloride and GC enhancer. An annealing temperature gradient was performed with 30 amplification cycles, 2µM primer concentration and 5ng/µl positive cell line cDNA template or H_2O as a negative control. No product bands were seen for either *BCR-ABL1* or *PML-RARA*, Figure 3.17.

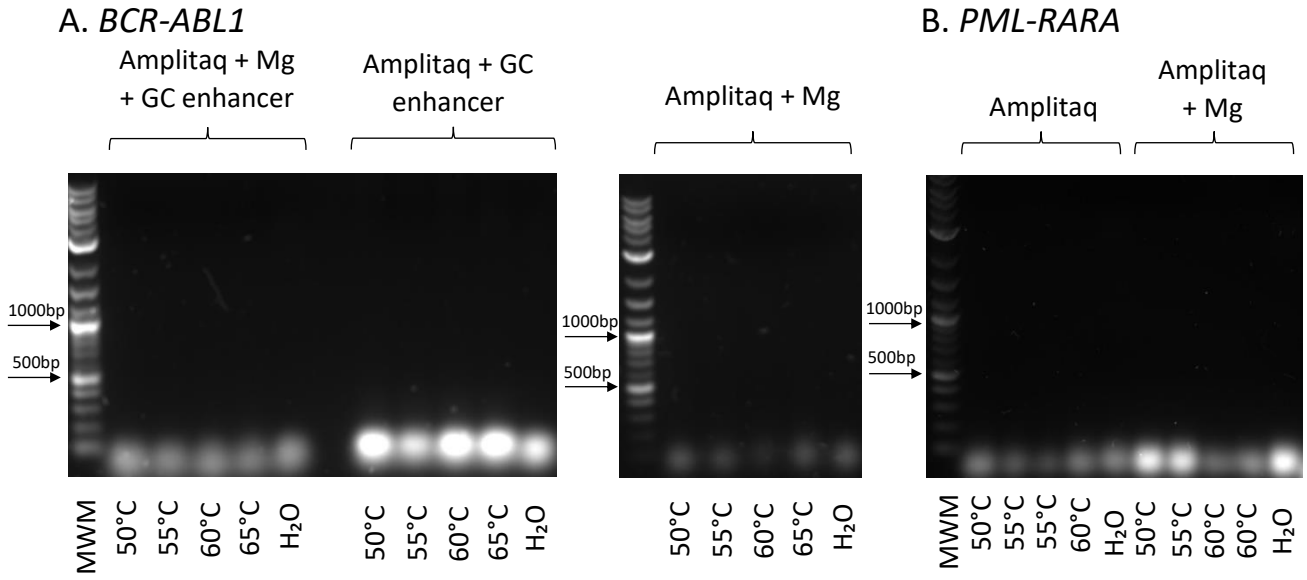


Figure 3.17. Gel electrophoresis visualisation of A) BCR-ABL1 and B) PML-RARA PCR assay using AmpliTaq Gold™ 360 polymerase. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

As no PCR amplification was detected for *BCR-ABL1* and *PML-RARA*, different primer pairs for each translocation were selected from previous literature, and the RT-PCR reaction

repeated with AmpliTaq Gold™ 360. Primer pairs BCR_F2 / ABL1_R2 and PML_F2 / RARA_R2 were used at a concentration of 2µM. An annealing temperature gradient was performed at 30 amplification cycles with 5ng/µl of positive cell line cDNA or H_2O. No PCR product bands were detected for *BCR-ABL1* or *PML-RARA*, Figure 3.18. A *GAPDH* RT-PCR assay produced product bands for both K562 and NB4, suggesting the cDNA was viable to use as a PCR template, although the product produced for NB4 was faint in comparison Figure 3.18.

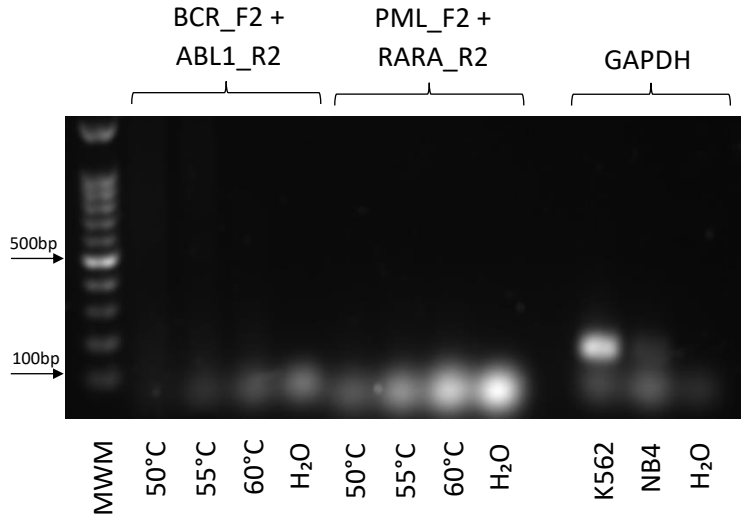


Figure 3.18. Gel visualisation of new PCR primer pairs for BCR-ABL1 and PML-RARA with GAPDH PCR on K562 and NB4 positive controls. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

Using primer pair BCR_F2 / ABL1_R2, various polymerases were used to amplify positive cell line control K562, including PCRBio Verify Taq Mix, PCR Bio HS Taq Mix Red and PCRBio HS Taq Mix. A primer concentration of 0.5µM was used across an annealing temperature gradient for 40 amplification cycles with 5ng/µl of cDNA template. Across all polymerases and temperatures, numerous bands could be seen in each lane, along with smearing, Figure 3.19A. Due to number of bands and smearing the *BCR-ABL1* assay was not chosen as a translocation target as it would be difficult to detect translocation breakpoints and to translate the assay into a qPCR assay. The *BCR-ABL1* translocation is found in both adult CML, ALL, and AML, and childhood ALL [288]. Many breakpoints have been identified across these disease types, resulting in fusion transcripts of different lengths due to different breakpoint junctions [261, 288]. Due to this, the cell line used may not have been optimal as a positive control for the amplification of the breakpoint junction expected for the primer pairs used.

The *PML-RARA* positive cell line NB4 was used as template for the PML_F2 / RARA _R2 primer pair using PCR Bio HS Taq Mix. An annealing temperature gradient was performed

for 40 amplification cycles with 0.5µM of primer and 5ng/µl of cDNA template. No PCR product bands were detected across any temperature, Figure 3.19B. As the *PML-RARA* RT-PCR assay was unsuccessful in amplifying the positive cell line NB4 following various optimisation steps, this assay could not be used to detect this translocation.

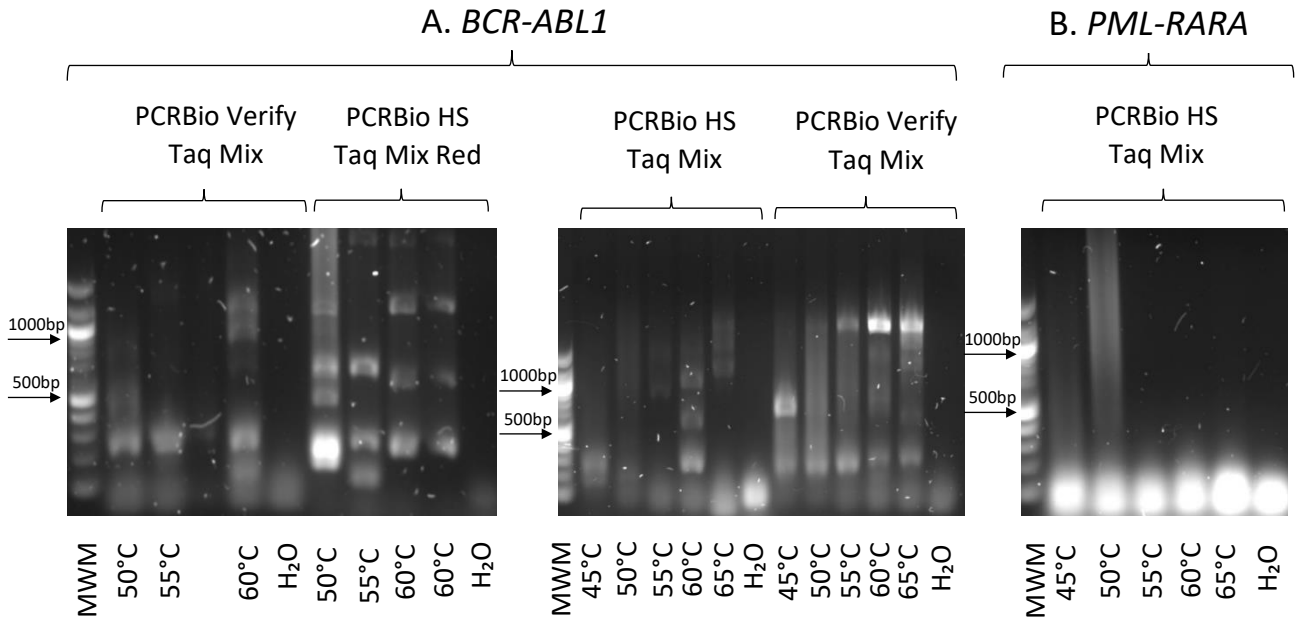


Figure 3.19. Gel electrophoresis visualisation of BCR-ABL1 and PML-RARA RT-PCR reactions using different PCRBio polymerases. A) BCR-ABL1 and B) PML-RARA. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

3.2.5 Development of qPCR assays to detect fusion genes

The use of quantitative PCR (qPCR) assays is useful in quantifying fusion gene expression in a given sample, as well as offering greater sensitivity than some other PCR assays. As extraction of RNA and the process of reverse transcription can vary from sample to sample, the use of reference genes allows for qPCR results to be normalised to report a more accurate result [289]. *GAPDH* and *Beta Actin* were chosen as reference genes for the qPCR assay. Using the annealing temperature of 60°C, a standard curve was set up for both *GAPDH* and *Beta Actin* using HL60 cDNA diluted 1 in 10 from 1.25ng/µl to 0.000125ng/µl. Once plotted on a logarithmic scale for each dilution point, Figure 3.20, *GAPDH* gave an efficiency of 96.2% and an R^2 value of 0.9965. For *Beta Actin*, the efficiency was 99.1% and the R^2 value was 0.9599. Working at an efficiency of 100%, the amount of PCR product should double with each PCR amplification cycle. An efficiency between 90-110% is desirable [290]. The R^2 value reports the linearity between each sample, with a desirable confidence >0.98[290]. The *GAPDH* results show an efficiency and R^2 value within the desirable limits. For *Beta Actin*, the efficiency is desirable but the R^2 value is less than the

desirable limit, which could be due to the preparation of the dilutions. As the confidence is above 0.95, the *Beta Actin* primer pair was deemed sufficient for the qPCR assay.

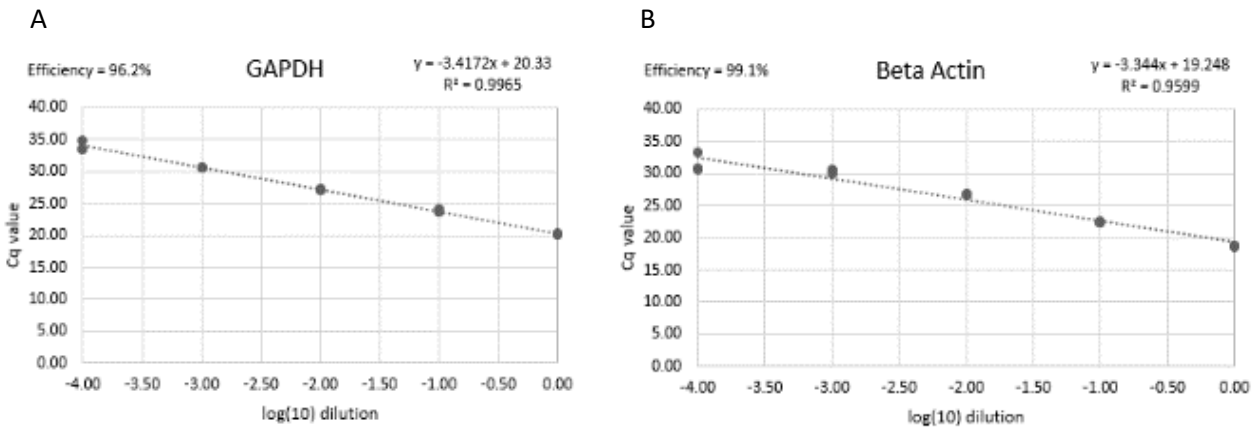
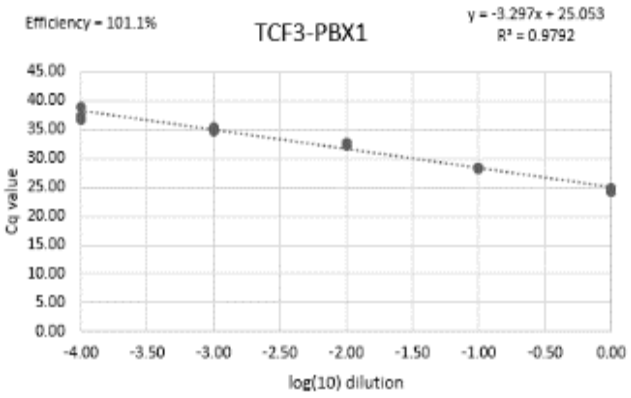


Figure 3.20. Standard curve for GAPDH and Beta Actin qPCR assay. Positive HL60 cDNA template was diluted 1 in 10 represented by each point of the graph (1.25ng/ul to 0.000125ng/ul). Graph A shows the standard curve for GAPDH, with an efficiency of 96.2% and an R^2 value of 0.9965. Graph B shows the standard curve for Beta Actin, with an efficiency of 99.1% and an R^2 value of 0.9599.

A qPCR standard curve was set up for both *TCF3-PBX1* and *RUNX1-RUNX1T1* using 60°C annealing temperature and positive cell line cDNA diluted 1 in 10 from 1.25ng/µl to 0.000125ng/µl for Pre-B 697 and 1 ng/µl to 0.0001ng/µl for Kasumi-1. Results for *TCF3-PBX1* showed an efficiency of 101.1%, and an R^2 value of 0.9792, Figure 3.21A, both close to the desirable range. However, after reporting low Cq values for the negative control, the qPCR products were ran on an agarose gel, showing contamination in the negative control therefore invalidating the curve. For *RUNX1-RUNX1T1*, the standard curve gave an efficiency of 80.5% and an R^2 value of 0.9633, Figure 3.21B. Although the R^2 value was close to the desirable confidence, the efficiency of the *RUNX1-RUNX1T1* qPCR was much lower, which could be due to ineffective primer pairs or contaminated cDNA template.

A



B

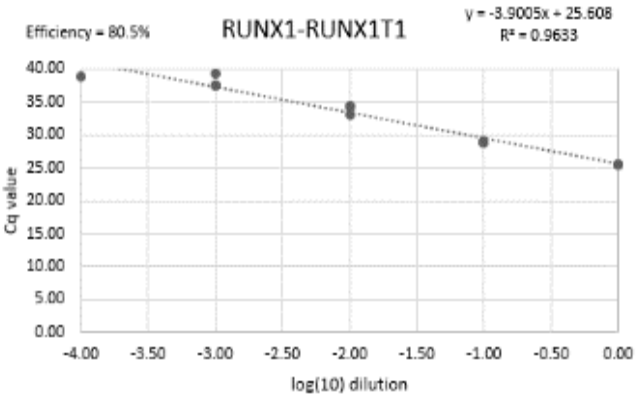


Figure 3.21. Standard curve for TCF3-PBX1 and RUNX1-RUNX1T1 qPCR assay. Positive Kasumi-1 and Pre-B 697 cDNA template was diluted 1 in 10 represented by each point of the graph (1.25ng/ul to 0.000125ng/ul for Pre-B 697 and 1ng/ul to 0.0001ng/ul for Kasumi-1). Graph A shows the standard curve for TCF3-PBX1, with an efficiency of 101.1% and an R^2 value of 0.9792. Graph B shows the standard curve for RUNX1-RUNX1T1, with an efficiency of 80.5% and an R^2 value of 0.9633.

To optimise the qPCR assays for *TCF3-PBX1* and *RUNX1-RUNX1T1*, the qPCR standard curves were repeated, altering conditions to include new cDNA templates, higher starting cDNA concentrations for dilution, different dilution factors, different primer concentrations and different sybr green taq mix (not reported). No conditions could produce sufficient efficiency or R^2 value.

As the *TCF3-PBX1* and *RUNX1-RUNX1T1* primer pairs could not produce a sufficient qPCR standard curve, these assays were not used subsequently. As consistent PCR results were produced with the standard PCR method, with a sensitivity of 1%, this method was chosen over the qPCR assay.

3.3 Tissue culture optimisation

3.3.1 NALM6 cell growth and cell viability

To evaluate the optimum density for NALM6 cell growth and viability, trypan blue exclusion was used to count viable and non-viable cells every 24 hours over a total of 96 hours, Figure 3.22. When seeded at densities of $0.5x10^6$ and $1x10^6$ cells/ml, at 0 to 24 hours the cells are in the lag phase, from 24 to 72 hours the cells are in the logarithmic phase, at 72 to 96 hours, the seeding density $0.5x10^6$ cells/ml reaches a plateau whereas the seeding density $1x10^6$ cells/ml enters the death phase with a reduction of cell growth and cell viability, Figure 3.22. As the seeding density $0.5x10^6$ cells/ml maintained a higher cell viability across 96 hours, it was chosen as the optimum seeding density for experiments involving NALM6.

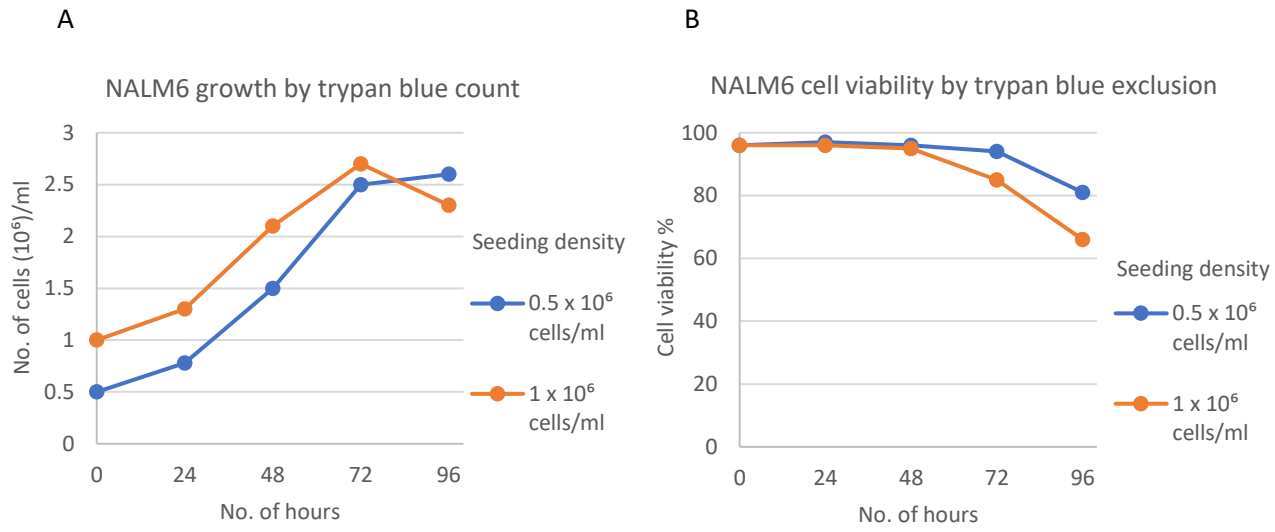


Figure 3.22. NALM6 cell growth and cell viability by trypan blue exclusion. NALM6 cells were seeded at 0.5 and 1 x 10^6 cells/ml and counted every 24 hours using trypan blue exclusion.

3.3.2 NALM6 cell line authentication and mycoplasma testing

Contamination of cell cultures can occur through cross contamination with other cell lines or through contamination with microorganisms such as mycoplasmas. To authenticate the use of the NALM6 cell line, a sample was taken from a live culture and sent to Eurofins for STR profiling. The results for the NALM6 cells used in the translocation model matched the identifiers reported in NALM6 STR profiles on ExPasy confirming authentication of the cells.

Contamination with mycoplasma can lead to altered RNA and DNA synthesis, chromosomal aberrations, altered cell metabolism, proliferation and morphology, which could all impact the reliability of assays performed on mycoplasma positive cells [291]. Using the Venor®GeM OneStep Mycoplasma Detection Kit, 12 different species of Mycoplasma can be detected through PCR amplification of a conserved 16S rRNA coding region. Routine mycoplasma testing was performed on growing cultures, to ensure that any cultures used for downstream assays were mycoplasma negative, Figure 3.23.

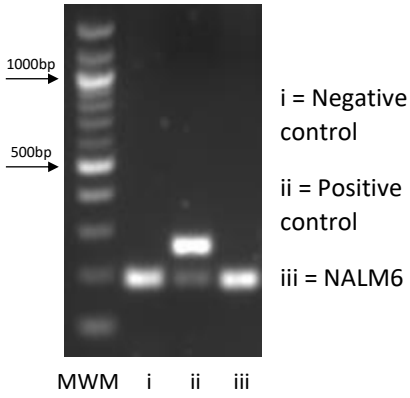


Figure 3.23. Gel electrophoresis visualisation of NALM6 cell mycoplasma PCR using Venor®GeM OneStep Mycoplasma Detection Kit. The mycoplasma negative control shows 1 band and the mycoplasma positive control shows 2 bands. MWM = molecular weight marker (New England Biosciences 100 bp DNA Ladder).

3.4 Etoposide as a model for translocation induction in NALM6 cells

To assess the ability of translocation induction in NALM6 cells, with detection by RT-PCR assays, etoposide was used as a translocation inducer, described in Figure 3.24.

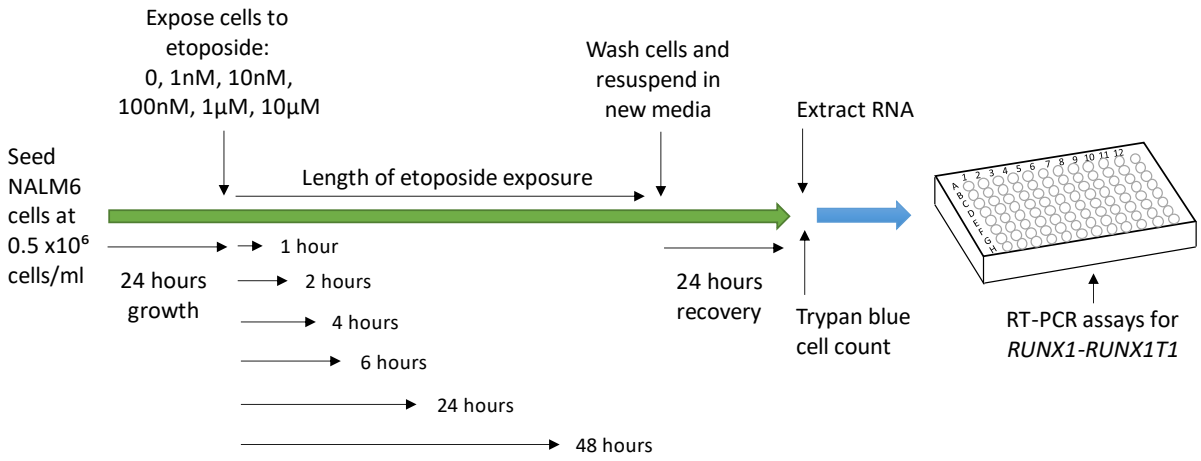


Figure 3.24. Study design for NALM6 etoposide exposure. NALM6 cells were seeded at 0.5x10^6 cells/ml in 6-well plates and allowed to grow for 24 hours. Once the cells were in the exponential growth phase, they were exposed to a range of etoposide concentrations for various lengths of time. After the exposure period, cells were washed, resuspended in fresh medium and allowed to recover for 24 hours, to give cells time to repair DNA damage and form translocations, before trypan blue exclusion count and RNA extraction. RNA was then reverse transcribed for use in the RUNX1-RUNX1T1 RT-PCR assay. One replicate was performed for each concentration and exposure length.

As expected, exposure to the strongest concentrations of etoposide, 1µM and 10µM, for longer than 2 hours resulted in a reduction in cell growth and reduced cell viability compared to the DMSO control, Figure 3.25. Exposure to 10µM of etoposide for any length of time resulted in less than 50% cell viability, with no increase in overall viable cell number above 0.5x10^6 cells/ml. For 1µM etoposide, as the length of exposure increased, the number of cells and percentage of cell viability decreased. For 100nM etoposide exposure, as the length of exposure time increased the viability of cells decreased, resulting in 70% viability

after 48 hours exposure. Exposure to 10nM and 1nM etoposide closely followed the growth and viability pattern of the DMSO control, with viability staying above 75% for all exposures.

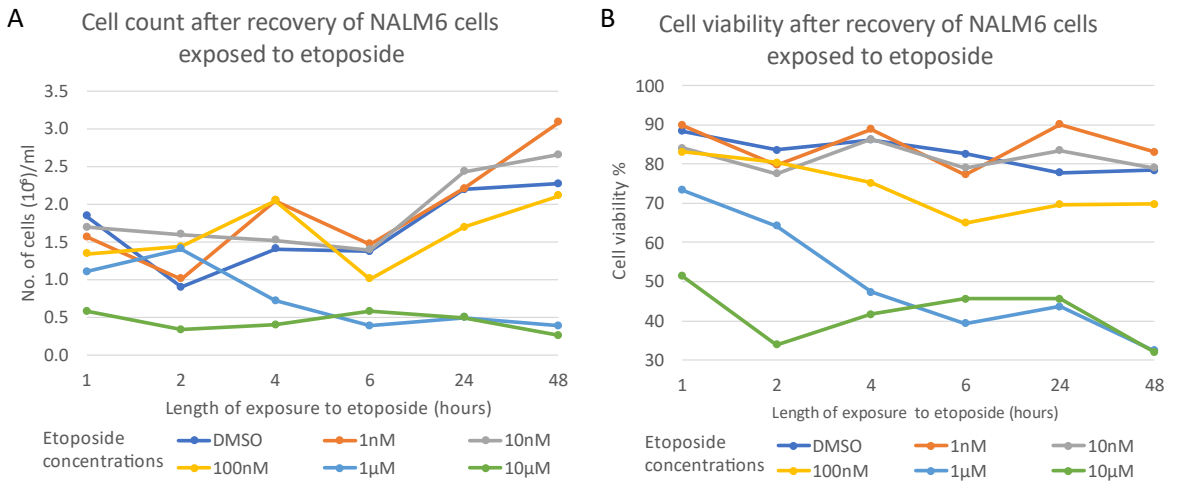


Figure 3.25. Cell count and viability via trypan blue for NALM6 cells exposed to etoposide. A) Number of NALM6 live cells following a 24-hour recovery after exposure to different concentrations of etoposide for different lengths of time. B) Cell viability of NALM6 following a 24-hour recovery after exposure to different concentrations of etoposide for different lengths of time. The graphs show a single replicate for each concentration.

Once assessed by the *RUNX1-RUNX1T1* RT-PCR assay, translocations were detected in NALM6 cells exposed to 1µM for 1 hour, and in NALM6 cells exposed to 10nM etoposide for 48 hours, Table 3.5. No translocations were detected in the DMSO controls, Table 3.5.

Table 3.5. Translocation event response to etoposide exposure. NALM6 cells were grown for 24 hours then exposed to various concentrations of etoposide for different lengths of time. Following exposure cells were washed and allowed to grow for 24 hours before RNA extraction and RT-PCR assays for RUNX1-RUNX1T1 using AmpliTaq Gold™ 360, 0.5µM primer concentration and 40 amplification cycles. Each X represents a translocation observed for a single replicate.

		Concentration of etoposide					
		DMSO	1nM	10nM	100nM	1µM	10µM
Length of exposure to etoposide	1 hour					X	
	2 hours						
	4 hours						
	6 hours						
	24 hours						
	48 hours			X			

A further 3 replicates were performed for the conditions where a translocation was detected. A *RUNX1-RUNX1T1* translocation was detected with exposure to 1µM etoposide for 1 hour in 1 out of 3 replicates, Table 3.6. No translocations were detected for the DMSO controls, or for 10nM etoposide for 48 hours, Table 3.6.

Table 3.6. Translocation event response to etoposide exposures previously observed to induce translocations. NALM6 cells were grown for 24 hours then exposed to 1μM etoposide for 1 hour and 10nM etoposide for 48 hours. DMSO was used as a negative control for each exposure. Following exposure, cells were washed and allowed to grow for 24 hours before RNA extraction and RT-PCR assays for RUNX1-RUNX1T1 using PCRBio HS taq, 0.5μM primer concentration and 40 amplification cycles. Each number represents the number of translocations observed for 3 replicates.

Etoposide exposure conditions	*RUNX1-RUNX1T1* translocations
DMSO for 1 hour	0
1μM for 1 hour	1
DMSO for 48 hours	0
10nM for 48 hours	0

The 3 samples where a *RUNX1-RUNX1T1* translocation was observed were gel extracted and sent for Sanger sequencing. Etoposide samples (original replicates) 10nM for 48 hours and 1μM for 1 hour, both identified *RUNX1* and *RUNX1T1* sequences, with the breakpoint junction shown in Table 3.7. Sequencing results for the positive repeat of 1μM for 1 hour, Table 3.7, identified a *RUNX1T1* sequence via BLAST, however forward and reverse sequencing was unable to identify *RUNX1* alone across the breakpoint junction, but hits were found for *AML1-ETO* RNA sequence (*RUNX1-RUNX1T1*).

Table 3.7. Sanger sequencing for RUNX1-RUNX1T1 PCR for positive etoposide exposure samples. The sequence relating to each translocation partner is highlighted in red or blue.

	Sequence of translocation junction	Genes identified in BLAST
Etoposide 1μM 1 hour	TGGGCCCCGAGAACCTCGA AATCGTACTGAGA	*RUNX1* - *RUNX1T1*
Etoposide 10nM 48 hour	TGGGCCCCGAGAACCTCGA AATCGTACTGAGA	*RUNX1* - *RUNX1T1*
Etoposide replicate 1μM 1 hour	CGTTAACTAATGGCACGAGC CATTCT	*RUNX1T1* (also identified *AML1-ETO* RNA sequence)

3.5 Discussion

As translocation breakpoints can vary, resulting in potentially very large DNA sequences that may be difficult to amplify via standard PCR, RT-PCR allows for the amplification of shorter sequences by using RNA. As mRNA consists of exons, genetic sequences are much shorter than the original DNA sequence. However, RNA is much less stable than DNA, making it difficult to work with, and requires conversion to cDNA by reverse transcription to be used as a template in RT-PCR reactions. The quality and purity of RNA is important for cDNA synthesis and DNA contamination is removed using a DNase treatment. This study used a 2-step RT-PCR method which allows for the conversion and storage of cDNA in 1 reaction before a 2nd reaction to undergo amplification by PCR. Although 2-step RT-PCR increases

the risk of contamination and experiment failure due to more independent steps than 1-step RT-PCR, where all reactions occur in the same tube, it allows for stored cDNA to be used in multiple RT-PCR reactions for different translocation primer pairs.

Although the RT-PCR assays were able to amplify across the fusion breakpoint junction, the exact breakpoint position in each gene is unknown as the breakpoints may have occurred within introns, which would have been spliced out on mRNA production. Due to the nature of translocations, not all translocations may be detected by RT-PCR as some breakpoints may interfere with gene transcription and as such a gene product may not be produced. However despite these limitations, RT-PCR is commonly used in detection of chromosomal translocations and monitoring of minimal residual disease, which is one of the reasons why it was chosen as the method of detection in this study [261]. The translocations selected were all common childhood leukaemia subtype identifiers, which are commonly detected by RT-PCR, so would be able to be detected in the translocation model [261].

Optimisation of RT-PCR assays for *TCF3-PBX1* and *RUNX1-RUNX1T1* translocation partners were successful, with the correct breakpoints sequenced for positive cell line controls. The assays were able to detect a minimum of 1 translocated cell per 100 negative cells. The positive cell line controls had high concentrations of the target breakpoint sequence, which allowed for adequate concentrations of purified PCR product necessary for Sanger sequencing. As translocations are rare events, positive translocation events in NALM6 samples were less common and less concentrated, and as such this often did not produce adequate concentrations required for Sanger sequencing. However, carrying out a nested PCR using positive PCR products as templates, provided suitable concentrations for successful sequencing, with the sequencing for positive cell line controls producing the same sequencing breakpoint junctions for the original and nested PCRs. This provided successful RT-PCR assays for 2 common childhood leukaemia associated chromosomal translocations for use in further experiments.

Optimisation for 4 other leukaemia associated translocation PCRs was unsuccessful. For *ETV6-RUNX1*, the primer pairs used produced unspecific binding to a gene identified as TNPO3, along with the target translocation. Further optimisation for an *ETV6-RUNX1* PCR assay would benefit from new primer design. The *KMT2A-AFF1* assay did produce the correct target sequence, although often with very faint bands. When NALM6 samples were used with low to no target sequence, the PCR assays resulted in large amounts of smearing that made identification of any potential bands difficult. Due to this, further optimisation would be needed to reduce contamination and non-specific amplification. No amplification was observed for any *PML-RARA* primer pairs with the NB4 positive cell line control. Further

optimisation, such as cell line authentication, would be required to ensure the positive cell line used was correct for the target sequence. For the *BCR-ABL1* assay, amplification was observed using PCRBio Verify and HS taq mixes, however multiple bands were produced with high levels of smearing, making it difficult to determine if the correct target sequence had been identified, or whether it was a result of non-specific amplification or contamination. As *BCR-ABL1* translocations can produce many different size fusion proteins with different disease types, further optimisation of this assay would require different positive cell lines that match the target sequence and further optimisation of multiple primer pairs [261, 288]. Although these assays require further optimisation, as cDNA from exposure samples can be stored, it would be possible to run the samples on further PCR assays in the future.

GAPDH and *Beta Actin* PCRs were successfully optimised for use as a reference gene in qPCR assays. However, optimisation of *TCF3-PBX1* and *RUNX1-RUNX1T1* qPCR assays was unsuccessful due to inefficient standard curves and contamination. Although qPCR is often more sensitive and able to provide quantitative information on gene expression, as the aim of this study was to identify the presence of translocations, standard RT-PCR was deemed sufficient, with its ability to detect translocations in 1 cell per 100, and as such were chosen for use in further experiments over further optimisation of the qPCR assays.

Selection of the NALM6 cell line in the translocation model was due to its defective DDR and lack of target translocations [285]. The NALM6 cell line STR profile was authenticated with known NALM6 STR profiles, and routine mycoplasma testing was performed on live cultures to ensure the cultures used in the translocation model were contamination free. An optimal seeding density of $0.5x10^6$ cells/ml was chosen due to only a slight loss of viability at 96 hours. The NALM6 cell line followed a typical growth curve, with a slow lag phase between 0 and 24 hours, an exponential growth phase between 24 and 72 hours, and plateau phase after 72 hours. However, it may be beneficial to explore the impact of a broader range of seeding densities on NALM6 growth to ensure an optimal balance between cell viability and length of time needed for exposure to different chemicals, as cells were no longer in the exponential growth phase after 72 hours.

The chemotherapy drug, etoposide, was used to assess the ability of the PCR assays to detect translocations induced in NALM6 cells due to its known association with therapy-related leukemias as a topoisomerase II inhibitor, with *MLL* translocations induced in cell lines exposed to etoposide [159, 287]. As etoposide is known to inhibit the topoisomerase II-DNA cleavage complex, NALM6 cells were exposed to etoposide during the exponential growth phase, when cells are actively replicating. The highest concentration of etoposide used, 10µM, is a peak serum concentration during etoposide treatment and as such, that

and all lower concentrations are physiologically relevant [292]. Length of exposure time is also important in etoposide treatment, so each concentration was left for time intervals between 1 hour and 48 hours. Following exposure, cells were washed and resuspended in new medium for 24 hours to remove any remaining etoposide from the growth medium before harvesting. Etoposide is a known topoisomerase II inhibitor, which prevents religation of DSBs induced by topoisomerase II [159]. As translocations can occur following repair of DNA DSBs, cells exposed to etoposide were grown for 24 hours in fresh medium to allow time for repair and translocation formation. To confirm that the exposure and repair time was sufficient to induce and repair double strand breaks, a COMET assay could have been performed at each time point to assess the DNA damage [293].

Overall, increased etoposide concentration and increased exposure time reduced cell growth and viability, for example, after exposure to 10µM etoposide for 48 hours, cell density was below 0.5 x 10^6 and cell viability was close to 30%. The *RUNX1-RUNX1T1* PCR assay specifically was used to assess translocation events induced by etoposide, as this translocation has been shown to be associated with therapy-related leukaemia [287]. Translocation events were observed for exposure to high concentrations of etoposide for short lengths of time, 1µM for 1 hour, and low concentrations of etoposide for longer periods of time, 10nM for 48 hours. These translocation events were also found in further replicates of 1µM for 1 hour. The presence of these translocation events in etoposide exposed cells, with no translocations observed in the DMSO control cells, provides confidence that the methods used in this study are sufficient to induce and detect translocation events in NALM6 cells. This chapter demonstrates the successful optimisation of *TCF3-PBX1* and *RUNX1-RUNX1T1* PCR assays, and detection of translocation events induced in NALM6 cells by RT-PCR assays and provides rationale for use of these methodologies for further experiments.

4 Caffeine exposure and translocation induction in NALM6 cells

4.1 Introduction

Caffeine is widely consumed across the world through intake of drinks, including tea, coffee and sugary carbonated drinks, and chocolate [128]. Absorbed into the blood stream through the gastrointestinal tract, caffeine is distributed throughout the body and can cross both the placenta and testicular barriers [129]. Caffeine is primarily metabolised in the liver by the enzyme Cytochrome P450 1A2 (CYP1A2), however during pregnancy, CYP1A2 activity is decreased leading to higher levels of serum caffeine [135, 136, 294]. Moreover, *in utero,* neither the foetus nor the placenta contain the CYP1A2 enzyme needed to metabolise caffeine, so caffeine mainly is cleared via urinary excretion but at a much slower rate than CYP1A2 metabolism in adults [137, 295].

In vitro studies have shown that caffeine exposure can impact DNA in a variety of ways, including inhibiting cell cycle checkpoint proteins ATM and ATR, which can allow damaged DNA to progress through the cell cycle, supressing HR, and perpetuating genomic instability [148, 156]. Caffeine has also been identified as a topoisomerase antagonist [157]. Other topoisomerase inhibitors, such as etoposide, are known to induce translocations found in childhood leukaemia [158, 159]. Whilst caffeine itself has not been directly found to induce translocations to date, other than in very early studies, this role suggests it may have the capacity to induce translocations associated with childhood leukaemia.

Most epidemiological studies investigating the association between caffeine and risk of childhood leukaemia use self-reported consumption of tea and coffee during pregnancy to assess caffeine exposure [141]. Associations have been observed for increased risk of childhood leukaemia with daily consumption of two or more cups of coffee during pregnancy [141-145]. Meanwhile, a study by Milne *et al.* 2011 observed little significant association between tea/coffee consumption and ALL overall, but when ALL cases were limited to those with a balanced translocation, some associations were observed with tea/coffee intake [146]. Consumption of caffeine during pregnancy has also been associated with various adverse health outcomes in offspring, including restricted growth and cardiovascular development [136, 138]. Current advice for pregnant women in the UK is to restrict caffeine intake to 200mg per day, equivalent to around two cups of instant coffee [140]. However, the exact levels of caffeine intake that may contribute to childhood leukaemia is unknown, so it is important to investigate a range of caffeine concentrations, to provide accurate advice on safe levels of caffeine consumption during pregnancy.

This chapter describes the use of an *in vitro* model to investigate the role of physiologically relevant levels of caffeine exposure on the induction of childhood leukaemia associated

translocations, *TCF3-PBX1* and *RUNX1-RUNX1T1.* Using NALM6, a pre-B acute leukaemia cell line, physiologically relevant caffeine exposures were assessed for their impact on cell growth and chromosomal translocation induction.

4.2 Methods

4.2.1 Experimental design of preliminary tests to examine the effects of caffeine exposure on translocation events

To investigate the influence of caffeine exposure on NALM6 cell viability and translocation induction, as described in Section 2.6.1, a range of physiologically relevant concentrations of caffeine were chosen using serum caffeine levels reported in literature, Figure 4.1. Beverages and food contain varying levels of caffeine which result in a range of exposure depending on the quantity consumed. Due to the decrease in CYP1A2 activity throughout pregnancy serum caffeine concentrations increase because of reduced clearance, which can be observed in a study by Yu *et al*. 2016 [294]. The study measured the caffeine serum concentrations of pregnant women in different trimesters following a low/medium caffeinated beverage equivalent to one cup of instant coffee [294]. To reflect the caffeine concentrations throughout pregnancy, 10µM was chosen to represent average caffeine serum concentrations in 1st trimester, 20µM caffeine was chosen to reflect the average serum concentration in 3rd trimester and 40µM was chosen to represent the maximum caffeine serum concentration observed in the 3rd trimester [294]. A study by Klebanoff *et al.* 1998 also measured caffeine serum content in pregnant non-smokers based on individual caffeine intake in mg/kg/day. In the Klebanoff *et al.* 1998 study, the caffeine serum concentration measured for a medium caffeine intake of two cups of filter coffee per day was equivalent to 10µM (the concentration previously chosen from the Yu *et al*. 2016 study) [296]. A concentration of 2µM, equivalent to one cup of tea per day, was also chosen based on the Klebanoff *et al.* 1998 study to represent a low caffeine intake [296]. As an extreme concentration, 80µM was also selected, expected to be equivalent to ~16 cups of filter coffee.

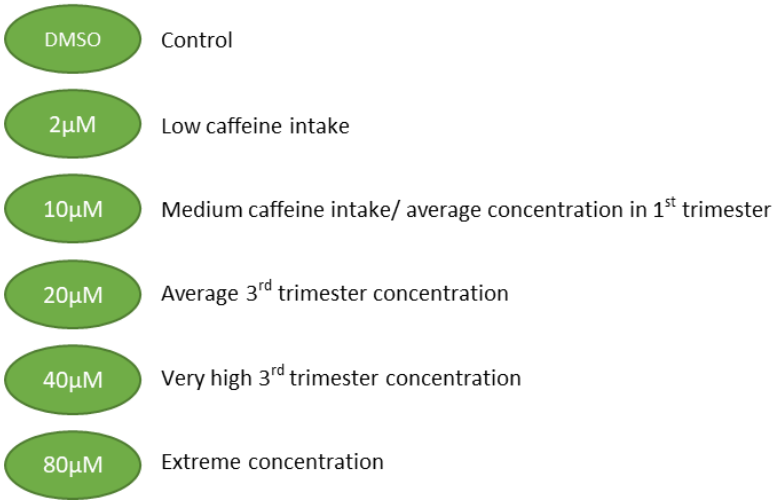


Figure 4.1. Caffeine concentrations used in preliminary experiments and their corresponding physiological range.

As caffeine experiments were performed in tandem with other exposures discussed in further chapters, DMSO was used as a common solvent for working stocks. The translocation events reported for DMSO control, Table 4.1, are the same results reported in subsequent chapters (Table 5.1 and 6.1).

4.2.2 Design of experiments to investigate the frequency of translocation events in response to caffeine exposure

To investigate the frequency that translocations may be expected to appear in NALM6 cells exposed to caffeine, the preliminary experiment was repeated for 20 technical replicates using 80µM caffeine exposed for 4 days. RNA was extracted from each of the 20 caffeine replicates and reverse transcribed into cDNA for use as a template in *TCF3-PBX1* translocation RT-PCR assays, using HS Taq polymerase with 2.5pmol forward and 5pmol reverse primers for 40 cycles. As a negative control for frequency experiments ran simultaneously across all exposure experiments (see chapters 5, 6 and 7 also), NALM6 cells were grown in 0.1% DMSO for 4 days, to allow for the maximum length of time for translocations to appear. The DMSO control replicates were used as a template in *TCF3-PBX1* translocation RT-PCR assays, using HS Taq polymerase with 10pmol primer concentrations for 40 cycles. The DMSO translocation events results reported in sections 4.3.2, 5.3.2, 6.3.2 and 7.3.2 are the same. All procedures are described in chapter 2.

4.2.3 Design of experiments to investigate translocation events in response to caffeine exposure in NALM6 cells for different models

To optimise the preliminary NALM6 cell model, two variables were considered: i) the impact of treating cells either immediately or 24 hours after initial seeding, and ii) single vs daily dosage of caffeine, as described in Section 2.6.2. Caffeine is suggested to be a

topoisomerase antagonist, disrupting replication processes, as well as altering cell cycle checkpoint pathways, which may impact the induction of translocation events when caffeine exposure occurs when cells are actively replicating [148, 158]. Caffeine exposure is also likely to occur on a daily basis through coffee, tea and caffeinated soft drinks intake. The caffeine concentrations used are outlined in Figure 4.2.

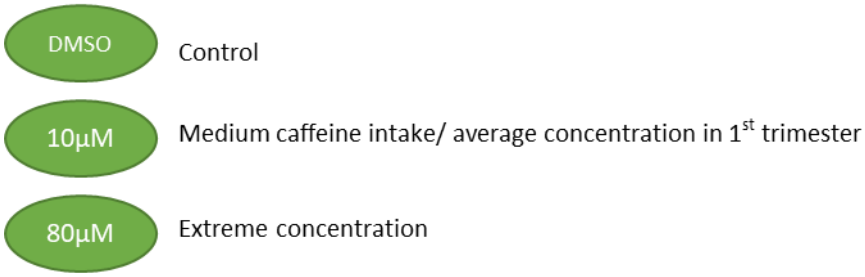


Figure 4.2. Caffeine concentrations used in Model's A-D and their corresponding physiological range.

A number of positive translocation events were selected for further analysis by fluorescence *in situ* hybridisation (FISH). Harvested cells fixed in Carnoy's solution were used to prepare slides for use with Cytocell breakapart probes. Cells with a *TCF3-PBX1* translocation were analysed with a *TCF3* breakapart probe and cells with a *RUNX1-RUNX1T1* translocation were analysed with a *RUNX1* breakapart probe. For each probe, four normal bone marrow samples were prepared and analysed to provide a threshold value over which the translocation events detected are considered to be genuine events associating with the exposure, rather than chance findings. The FISH procedure is described in chapter 2.

4.2.4 Investigating the influence of DMSO on translocation induction

Model C conditions were repeated with NALM6 cells grown in standard medium (no caffeine or DMSO) or treated with caffeine (with no DMSO present) or standard and low concentrations of DMSO. Stock caffeine solutions were originally prepared by dissolving caffeine in boiling water and then further diluting in DMSO for use in the preliminary, frequency and model experiments. Here, to investigate the effects of caffeine without DMSO, the caffeine stocks (dissolved in boiling water) were diluted in medium to concentrations of 80µM and 10µM and used to treat NALM6 cells. The standard and low concentrations of DMSO chosen to treat NALM6 cells include 0.1%, the standard concentration used in previous experiments as a control for caffeine exposed cells, and 0.00002%, a comparatively low concentration of DMSO. NALM6 cells grown in standard RPMI-1640 media were used as a negative control (0% DMSO). During daily dosing, NALM6 control cells were administered doses of medium. The exposures were repeated in three technical replicates and two biological replicates, giving six replicates total for each exposure. RNA was harvested for each sample, reverse transcribed and used as a template for *TCF3-PBX1*

and *RUNX1-RUNX1T1* RT-PCR assays, using HS Taq polymerase with 10pmol primer concentrations for 40 cycles.

4.2.5 Statistical analysis

For trypan blue, resazurin and Cyquant assays the data are presented as mean and standard error of the mean. The frequency of translocation events between caffeine exposed and DMSO control cells was analysed by two-tailed Fisher's exact tests using Graphpad Prism or https://www.socscistatistics.com/tests/fisher/. The frequency of FISH translocations in model experiments was analysed using ANOVA tests, followed by Bonferroni's multiple comparisons in SPSS.

4.3 Results

4.3.1 Preliminary experiment measuring cell growth and translocation events in NALM6 cells exposed to caffeine

To detect translocations by RT-PCR, NALM6 cells need to be able to proliferate when exposed to caffeine and maintain adequate cell viability for RNA harvesting. Trypan blue results show increasing growth for all concentrations from 24 to 72 hours, followed by a decrease in viable cell count at 96 hours for 10, 20, 40µM caffeine, and slight increases for 2µM and 80µM caffeine and DMSO control cells, Figure 4.3.

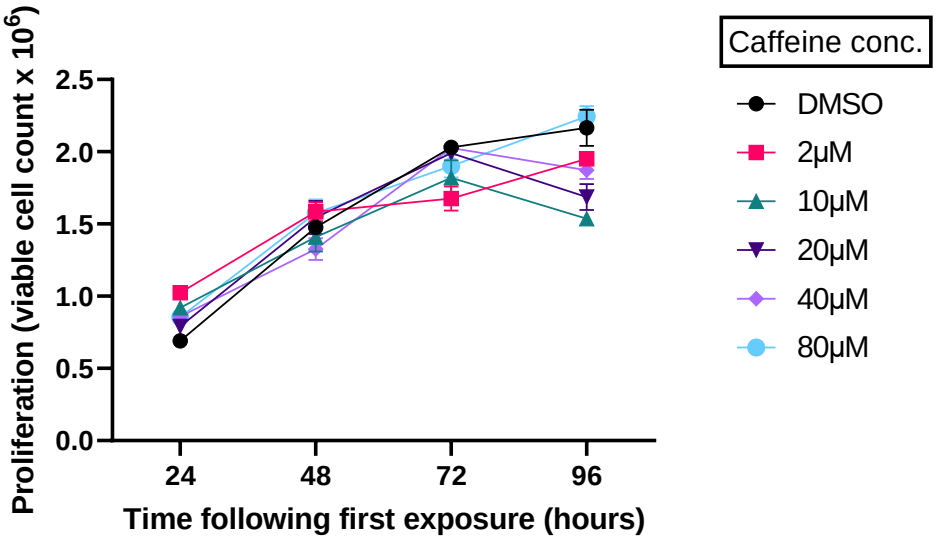


Figure 4.3. Trypan Blue assay of cell growth in NALM6 cells exposed to physiologically relevant concentrations of caffeine. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of caffeine or DMSO was added and then the cells were incubated at 37°C for up to 96 hours. Trypan blue counts occurred every 24 hours. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

The cell viability of caffeine exposed cells was measured via resazurin metabolic colorimetric assay, with no decrease in viability observed overall for all caffeine concentrations compared to the DMSO control, Figure 4.4.

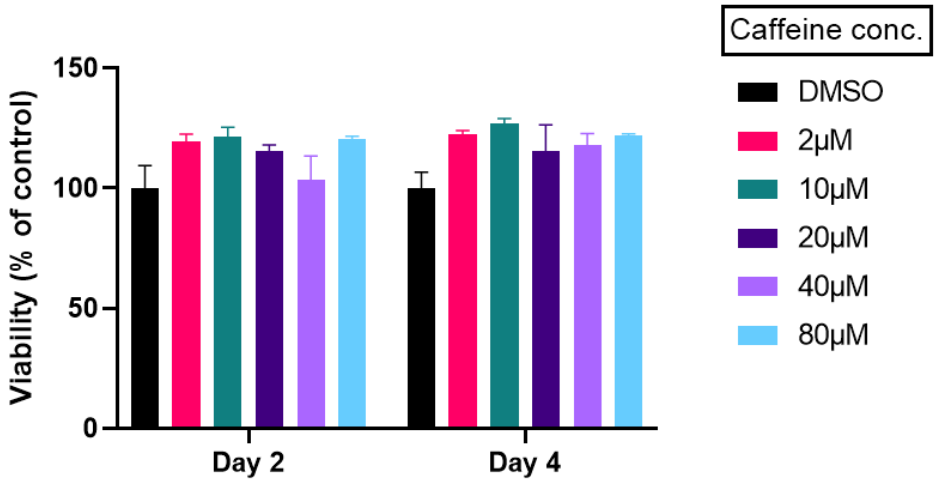


Figure 4.4. Resazurin assay measurement of cell viability of NALM6 cells grown in folic acid deficient conditions at day 2 and day 4. NALM6 cells were seeded into 2 well plates at a density of 0.5 x10^6 cells/ml in folic acid free RPMI-1640 media. A single dose of caffeine was added into the wells, and then they were incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2). On day 2 or day 4, resazurin was added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the 2μM control + standard error.

Translocation events were only observed for *TCF3-PBX1*, with 1 event in 1/3 replicates for each of the following conditions: 80μM exposure for 4 days and 10μM exposure for 2 days, Table 4.1.

Table 4.1. Translocation events observed for 3 technical replicates of NALM6 cells exposed to a range of physiologically relevant caffeine concentrations. NALM6 cells were seeded into two 12-well plates at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of caffeine or DMSO was added into the wells, and then they were incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2). Cells were harvested for RNA extraction and treated with DNase then reverse transcribed using Precision nanoScript2 Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq and 10pmol primers. The data presented shows 3 biological replicates. (DMSO control data is used in subsequent chapters 5 and 6).

Translocation	Day cells harvested	DMSO Control	2μM Caffeine	10μM Caffeine	20μM Caffeine	40μM Caffeine	80μM Caffeine
TCF3-PBX1 (3 replicates)	Day 2			1			
	Day 4						1
RUNX1-RUNX1T1 (3 replicates)	Day 2						
	Day 4						

Sanger sequencing of purified PCR product from caffeine 80μM Day 4 is shown in Table 2. The returned sequence from BLAST identified the correct expected translocated genes for each sample, allowing for the translocation junction to be identified. Due to low product concentration, sequencing was unable to be carried out for the translocation observed for 10μM Day 2 caffeine exposure.

Table 4.2. Sanger sequencing of purified PCR products from preliminary caffeine experiments. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified.

Sample	Sequence of translocation junction	Genes identified in BLAST
Caffeine 80µM Day 4 *(TCF3-PBX1)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*

4.3.2 Frequency of translocation events in response to caffeine exposure

As translocation events were only observed in two caffeine concentrations during preliminary experiments, it is difficult to establish a pattern of translocation induction. To investigate the frequency at which caffeine exposure may induce translocations in NALM6 cells, the preliminary experiment was repeated with 20 technical replicates of 80µM caffeine exposed for 4 days, alongside a DMSO control. This concentration was chosen as a caffeine exposure where a translocation event was observed and confirmed via sequencing. Thirteen *TCF3-PBX1* translocation events were observed across 13/20 replicates treated with 80µM caffeine for 4 days. The DMSO control observed two *TCF3-PBX1* events across 2/20 replicates. Analysis by two-tailed Fisher's exact test suggests there is a relationship between caffeine exposure and *TCF3-PBX1* translocations (p=<0.0001).

4.3.3 Cell growth, viability and translocation events measured in response to caffeine exposure in different models

To replicate the preliminary findings where translocations were observed in response to caffeine exposure and to address the issues of daily dosing and timing of those doses, further experiments were conducted using NALM6 cells treated with caffeine or DMSO across 4 different models.

Trypan blue assays were performed every 24 hours following the first exposure to caffeine to assess the impact of each condition on NALM6 growth, Figure 4.5. In Model A (single exposure, 24 hours after seeding), caffeine exposed cells were slow to proliferate from 24 to 48 hours compared to the DMSO control, but then viable cell count increased at 72 and 96 hours in line with the DMSO control, Figure 4.5. Cell growth in Model B (single dose at seeding) was slow in the first 24 hours then viable cell count increased every 24 hours in caffeine exposed cells similar to the DMSO control, Figure 4.5. In Models C (daily dose, 24 hours after seeding) and D (daily dose at seeding) cells were slow to proliferate for the first 72 hours followed by an increase in viable cell count for all concentrations at 96 hours, Figure 4.5.

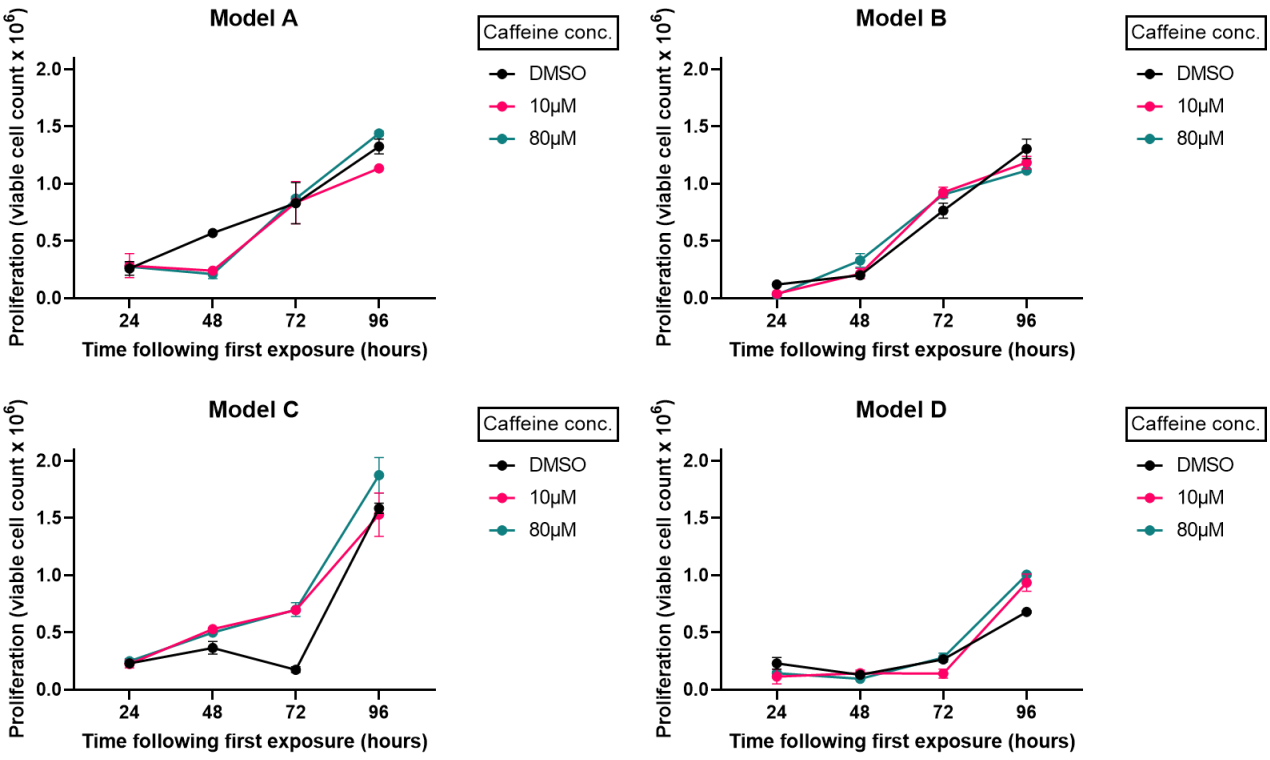


Figure 4.5. Trypan Blue assay for each of the 4 caffeine exposure models. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media and incubated at 37°C. Caffeine concentrations 10μM and 80μM were chosen to represent various physiologically relevant caffeine exposures, with the equivalent DMSO (0.1% v/v) used as a negative control. Trypan blue counts occurred every 24 hours following the first dose of benzene. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

Cell viability assays were performed at day 2 and day 4 following first caffeine exposure using resazurin and Cyquant. The resazurin assays show no decrease in viability for any caffeine concentration compared to the DMSO control, Figure 4.6.

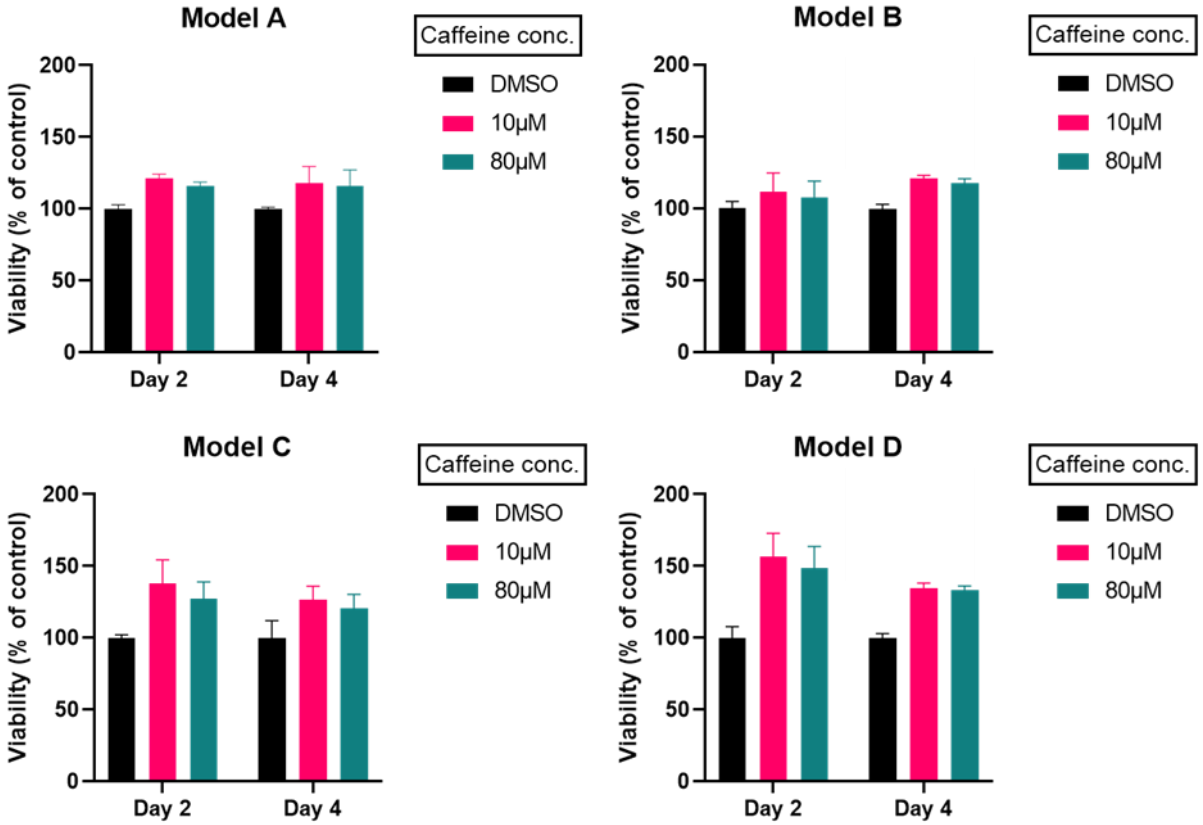


Figure 4.6. Resazurin assay measurements for each of the 4 caffeine exposure models. On day 2 or day 4 of each model, resazurin was added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

Cyquant viability assays also showed no reduction in cell viability across all of the caffeine concentrations compared to the control, Figure 4.7.

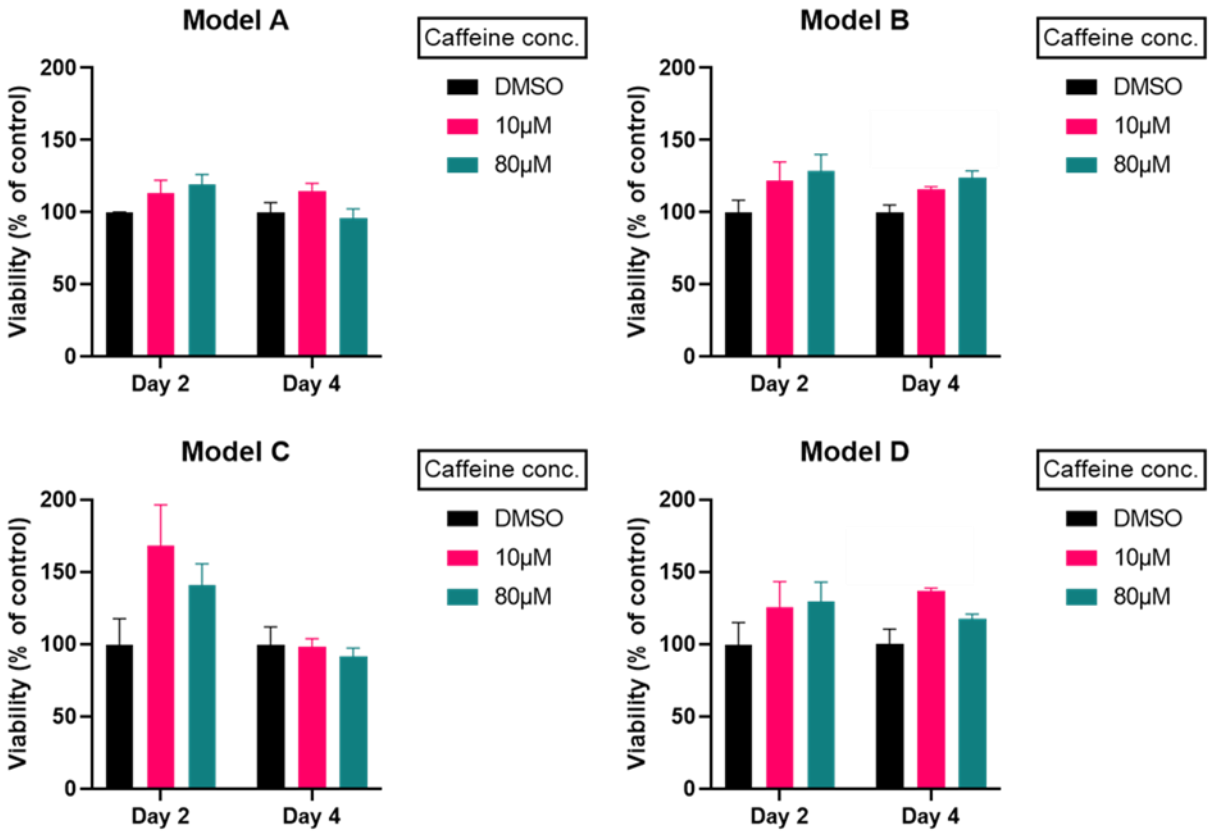


Figure 4.7. Cyquant assay measurement for each of the 4 caffeine exposure models. On day 2 or day 4 of each model, samples were taken from each well and frozen at -80°C. Cyquant was added to each sample, incubated overnight at 4°C and fluorescence monitored at 480nm excitation and 520nm emission. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

A total of 11 translocation events were observed across 96 replicates treated with caffeine overall, Table 4.3. The highest number of translocation events were observed in Model C with three *RUNX1-RUNX1T1* events across 6 replicates treated with 10μM caffeine for 4 days and one *RUNX1-RUNX1T1* event across 6 replicates for each condition of cells treated with 10μM caffeine for 2 days and 80μM for 4 days. A single *TCF3-PBX1* events was also observed in Model C across 6 replicates for each condition of cells treated with 10μM and 80μM caffeine for 2 days. A further three *TCF3-PBX1* events were observed in Model B, with two translocations across 6 replicates treated with 80μM caffeine for 4 days and a single translocation across 6 replicates treated with 80μM caffeine for 2 days. A single *RUNX1-RUNX1T1* event was observed in Model A across 6 replicates treated with 10μM caffeine for 2 days Table 4.3. Translocation events were also observed in DMSO controls, with 2 *TCF3-PBX1* across 6 replicates of cells treated with DMSO for 2 days in Model B, and 4 *RUNX1-RUNX1T1* events in Model C, with 3 out of 6 replicates in response to 4 days of DMSO exposure and 1 out of 6 replicates in response to 2 days DMSO exposure, Table 4.3.

Using Fisher's exact test, no significance was observed for translocation events in response to DMSO vs any caffeine treatment (p=1). No significance was observed between single dose vs daily dose (p= 0.3016) and for treatment following 24 hours growth vs at seeding (p= 0.1193) using Fisher's exact test. When considering any caffeine exposure, no significant difference was observed for the frequency of *TCF3-PBX1* and *RUNX1-RUNX1T1* events using Fisher's exact test (p=1).

Table 4.3. Model optimisation translocation events in response to caffeine. Each number represents the number of positive PCR bands seen during gel electrophoresis for 6 replicates of each condition. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media and incubated at 37°C. Model A: 24 hours growth, single dose of caffeine or DMSO. Model B: single dose of caffeine or DMSO at seeding. Model C: 24 hours growth, daily dose of caffeine or DMSO. Model D: daily dose of caffeine or DMSO starting at seeding. Caffeine concentrations 1.5nM, 3nM and 48nM were chosen to represent various physiologically relevant caffeine exposures, with the equivalent DMSO (0.1% v/v) used as a negative control. On day 2 and day 4 cells were also harvested for RNA extraction. The resultant RNA was treated with DNase and reverse transcribed using High-Capacity cDNA Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq with 10pmol primers. The data present shows 3 technical replicates and 2 biological replicates. (DMSO control data is also presented in chapter 5 (benzene) and chapter 6 (cotinine)).

Translocation	Model type	Day cells harvested	DMSO Control	10µM Caffeine	80µM Caffeine
TCF3-PBX1 **(6 replicates)**	Model A	Day 2			
		Day 4			
	Model B	Day 2			1
		Day 4	2		2
	Model C	Day 2		1	1
		Day 4			
	Model D	Day 2			
		Day 4			
RUNX1-RUNX1T1 **(6 replicates)**	Model A	Day 2		1	
		Day 4			
	Model B	Day 2			
		Day 4			
	Model C	Day 2	1	1	
		Day 4	3	3	1
	Model D	Day 2			
		Day 4			

As a confirmation of negative results, a selection of samples that did not contain visible bands during gel electrophoresis were used as a template for PCR using nested primers, and no amplification was observed. Sanger sequencing of the translocation breakpoints from caffeine exposed cells showed the expected fusion mRNA sequence, with the associated genes identified in BLAST, Table 4.4. Model C caffeine 80µM day 4 was not able to be sequenced due to low product concentration.

*Table 4.4. Sanger sequencing of translocation events observed in caffeine exposed cells for Model's A-D. Positive samples were used as a template for PCR using 10pmol of nested primers, and then purified for sequencing by DBS genomics. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence. (G->A)*

Sample	Sequence of translocation junction	Genes identified in BLAST
Caffeine 80µM Model B Day 2 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Caffeine 80µM Model B Day 4 #1 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Caffeine 80µM Model B Day 4 #2 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Caffeine 80µM Model C Day 2 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Caffeine 10µM Model C Day 2 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Caffeine 10µM Model A Day 2 *(RUNX1-RUNX1T1* Nested*)*	CCCCGTAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Caffeine 10µM Model C Day 2 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Caffeine 10µM Model C Day 4 #1 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Caffeine 10µM Model C Day 4 #2 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Caffeine 10µM Model C Day 4 #3 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*

The translocation events observed in the DMSO controls underwent Sanger sequencing to identify the breakpoint junction, all of which were confirmed to be the expected fusion gene for each translocation, Table 4.5.

*Table 4.5. Sanger sequencing of translocation events observed in DMSO control cells for Model's A-D. Positive samples were used as a template for PCR using 10pmol of nested primers, and then purified for sequencing by DBS genomics. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence. (G->A). The DMSO controls are referenced in subsequent chapters 5 and 6.*

Sample	Sequence of translocation junction	Genes identified in BLAST
DMSO control Model B Day 4 #1 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
DMSO control Model B Day 4 #2 *(TCF3-PBX1* Nested*)*	CCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
DMSO control Model C Day 2 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
DMSO control Model C Day 4 #1 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAAACCTCGAA*ATCGTACTGGAGAAG	*RUNX1 RUNX1T1*
DMSO control Model C Day 4 #2 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
DMSO control Model C Day 4 #3 *(RUNX1-RUNX1T1* Nested*)*	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*

As a confirmation of translocation events observed by RT-PCR, a number of samples were selected for fluorescent *in situ* hybridisation (FISH) analysis with breakapart probes, where a normal gene appears as a red/green fusion signal, and a translocation, where the chromosome breaks apart, appears as separated red and green signals. For *RUNX1* breakapart probes, nuclei with 2 fusion signals, as shown in Figure 4.8A, were scored as normal cells. Nuclei with 1 fusion signal and 1 red and green signal at least 2 signal widths apart, Figure 4.8B, were scored as *RUNX1* breaks. Normal bone marrow samples had a frequency of 0.5% *RUNX1* breaks, Table 4.6. The average number of *RUNX1* breaks in the DMSO control nuclei scored was 1.08%, based on individual data in Table 4.6. Comparatively, the average number of *RUNX1* breaks in the caffeine exposed nuclei scored was 3.4%, based on individual data in Table 4.6.

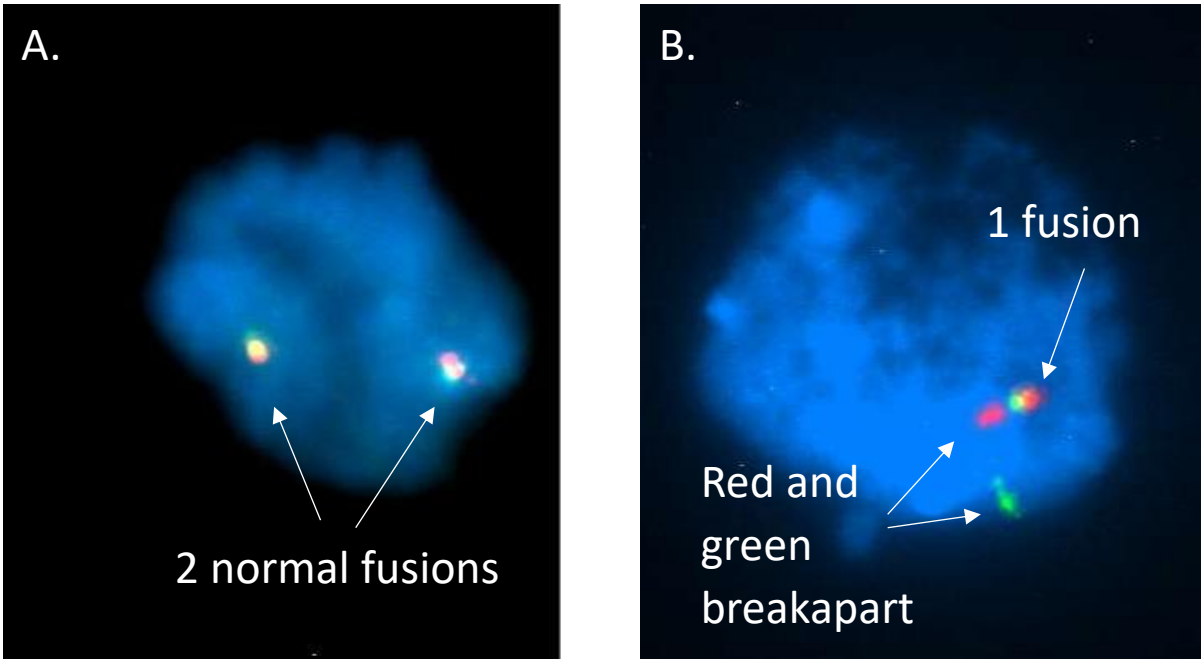


Figure 4.8. RUNX1 breakaway probe signals for normal and translocated cells. Images were taken as an example of normal and translocated cells using the sample Caffeine 10μM Model C Day 4 #2. A RUNX1 normal cell is expected to have 2 fusion signals, shown in A. With RUNX1 translocated cells, a single fusion is expected plus a red and green breakapart, indicating the translocation, where the red and green fusion signals are at least 2 signal widths apart, shown in B. Images were taken using an 100X objective on an Olympus BX61 microscope, with Cytovision software. Any signals that differed from these 2 expected signals were classed as abnormal signals.

Table 4.6. RUNX1 breakapart FISH probe results. A selection of samples underwent FISH analysis with RUNX1 breakapart probes. The table shows the sample and the expected translocation as detected by RT-PCR. Four normal bone marrow samples were prepared as a control. For each sample, the number of normal signals, breakapart signals and any other abnormal signals were counted in 100 cells, with the percentage of RUNX1 translocations, normal cells, and abnormal cells (not including translocation signals) presented. Details of the signal patterns in cells that were classed as abnormal, but not containing a standard breakapart translocation, can be found in the final right hand column.

Sample	Expected translocation	% normal cells	% *RUNX1* translocation	% abnormal (excluding translocations)	Notes on abnormal signals #
Normal Bone Marrow 1	-	97%	1%	2%	+red, 4 x fusion signals
Normal Bone Marrow 2	-	99%	1%	0%	(2 x breakapart)
Normal Bone Marrow 3	-	96%	0%	4%	2 x -red, -green, +green
Normal Bone Marrow 4	-	96%	0%	4%	2 x +red, 2 x +green
DMSO control Model C Day 2	-	97%	2%	1%	-red
DMSO control Model B Day 4 #2	*TCF3-PBX1*	98%	0%	2%	2 x 3 fusion signals
DMSO control Model C Day 4 #1*	*RUNX1-RUNX1T1*	92.9%	1.4%	5.7%	+red, -red, -fusion pair, 3 fusion signals
DMSO control Model C Day 4 #2	*RUNX1-RUNX1T1*	97%	2%	1%	-red
DMSO control Model C Day 4 #3	*RUNX1-RUNX1T1*	98%	0%	2%	2 x 3 fusion signals
Caffeine 10µM Model C Day 4 #1*	*RUNX1-RUNX1T1*	89.6%	1.3%	9.1%	2 x +green, +red, -red, -green, 2 x 3 fusion signals
Caffeine 10µM Model C Day 4 #2	*RUNX1-RUNX1T1*	91%	6%	3%	3 x +green
Caffeine 10µM Model C Day 2	*RUNX1-RUNX1T1*	91%	3%	6%	4 x 3 fusion signals, 2 x +red, +green

*** Less than 100 cells scored during analysis, 70 cells scored for DMSO control Model C Day 4 #1 and 77 cells scored for Caffeine 10µM Model C Day 4 #1.**
 # Abnormal signals with + or – green/red suggest gain or loss of *RUNX1* material, more than 2 fusion signals suggest *RUNX1* aneuploidy [297, 298]

Following ANOVA analysis with Bonferroni post-hoc tests, no significant differences was observed for the percentage of *RUNX1* translocations between DMSO control and bone marrow control (p=1) or caffeine exposed cells (p=0.123). A marginal significance was observed for the percentage of *RUNX1* translocations between bone marrow control and caffeine exposed cells (p=0.058). When comparing percentage of *RUNX1* translocations in DMSO control (with expected *RUNX1-RUNX1T1*) vs caffeine exposed (with expected *RUNX1-RUNX1T1*), no significant difference was observed (p=0.323). When considering the

percentage of overall abnormal cells (including translocations), a significant difference is seen for caffeine exposed cells between both DMSO control (p=0.002) and bone marrow control (p=0.002), with no significant difference between bone marrow control and DMSO control (p=1). A significant difference in percentage of overall abnormal cells is observed between caffeine exposed cells (with expected *RUNX1-RUNX1T1*) vs DMSO control (with expected *RUNX1-RUNX1T1*) (p=0.029).

A *TCF3* breakapart probe was used to assess *TCF3-PBX1* translocations in a selection of samples. Again, nuclei with 2 fusion signals, as shown in Figure 4.9A, were scored as normal cells. Nuclei with 1 fusion signal and 1 red and green signal at least 2 signal widths apart, Figure 4.9B, were scored as *TCF3* breaks. The frequency of *TCF3* breaks in normal bone marrow samples was observed to be 1%, Table 4.7. In the DMSO control samples scored, the average frequency of *TCF3* breaks was 3%, compared to a frequency of 7.5% for caffeine exposed samples, based on individual data in Table 4.7.

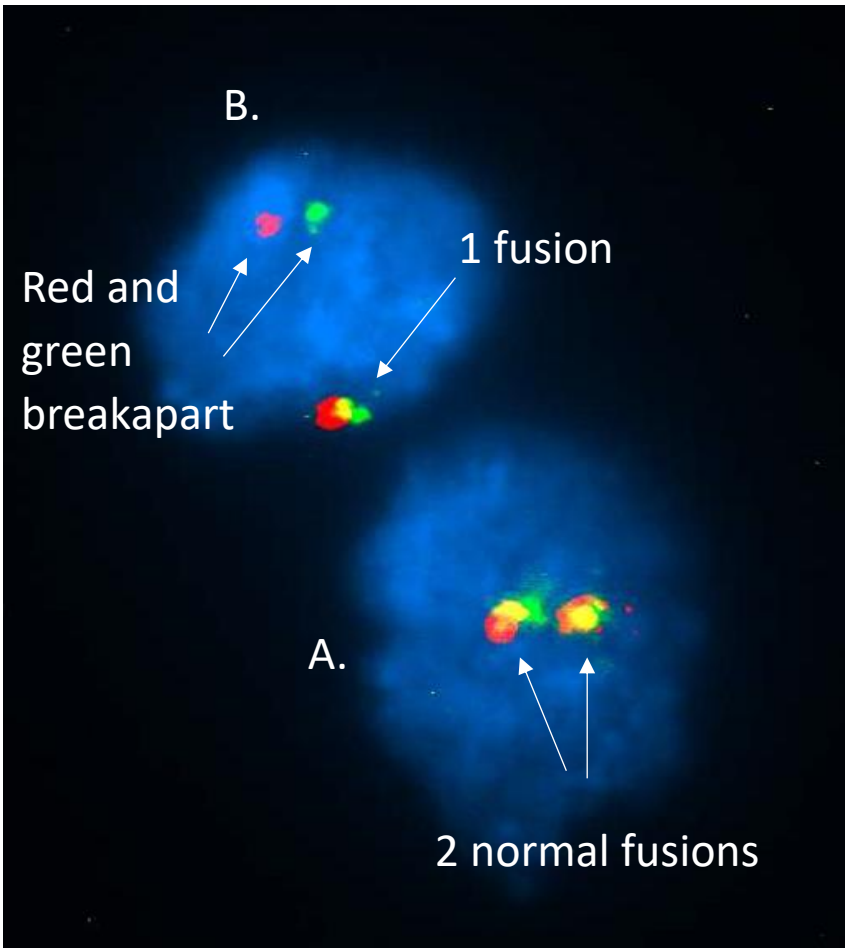


Figure 4.9. TCF3 probe signals for normal and translocated cells. Images were taken as an example of normal and translocated cells using the sample Caffeine 80µM Model B Day 4 #1. A TCF3 normal cell is expected to have 2 fusion signals, shown in A. With TCF3 translocated cells, a single fusion is expected plus a red and green breakapart, indicating the translocation, where the red and green fusion signals are at least 2 signal widths apart, shown in B. Images were taken using an 100X objective on a Zeiss Axioskop microscope, with Cytovision software. Any signals that differed from these 2 expected signals were classed as abnormal signals.

Table 4.7. TCF3 breakapart FISH probe results. A selection of samples underwent FISH analysis with TCF3 breakapart probes. The table shows the sample and the expected translocation as detected by RT-PCR. Four normal bone marrow samples were prepared as a control. For each sample, the number of normal signals, breakapart signals and any other abnormal signals were counted in 100 cells, with the frequency of TCF3 translocations, normal cells and abnormal cells (not including translocation signals) presented. Details of the signal patterns in cells that were classed as abnormal, but not containing a breakapart translocation, can be found in the final right hand column.

Sample	Expected translocation	% normal cells	% *TCF3* translocation	% abnormal (excluding translocations)	Notes on abnormal signals #
Normal Bone Marrow 1	-	99%	1%	0%	
Normal Bone Marrow 2	-	96%	2%	2%	+ green, - red
Normal Bone Marrow 3	-	95%	0%	5%	2 x + green, 2 x - red, + red
Normal Bone Marrow 4	-	93%	1%	6%	4 x + green, + red, - green
DMSO control Model C Day 2	-	93%	1%	6%	3 x + green, 2 x + red, - green
DMSO control Model B Day 4 #1	*TCF3-PBX1*	91%	5%	4%	3 x + red, (double breakapart)
DMSO control Model B Day 4 #2	*TCF3-PBX1*	93%	3%	4%	2 x + red, + green, - green
Caffeine 80µM Model B Day 4 #1	*TCF3-PBX1*	84%	10%	6%	2 x + green, 2 x + red, 2 x 3 fusion signals
Caffeine 80µM Model B Day 4 #2	*TCF3-PBX1*	93%	5%	2%	- green, + red, (breakapart + red)

 # Abnormal signals with + or – green/red suggest gain or loss of *TCF3* material, more than 2 fusion signals suggest *TCF3* aneuploidy [297, 298]

Following ANOVA analysis with Bonferroni post-hoc tests, no significant differences in the percentage of *TCF3* translocations were observed between DMSO control and bone marrow control (p=0.675) or caffeine exposed cells (p=0.131). However, a significant difference in the percentage of *TCF3* translocations was observed between bone marrow control and caffeine exposed cells (p=0.025). When comparing percentage of *TCF3* translocations in DMSO control (with expected *TCF3-PBX1*) vs caffeine exposed cells (with expected *TCF3-PBX1*), no significant difference was observed (p=0.323). When considering the percentage of overall abnormal cells, no significant difference is observed for caffeine exposed cells vs both DMSO control (p=0.717) and bone marrow control cells (p=0.121). For the percentage of overall abnormal cells, no significant difference is observed for caffeine exposed (with expected *TCF3-PBX1*) vs DMSO control (with expected *TCF3-PBX1*)(p=0.527).

When considering overall translocations across all of the samples (*TCF3* and *PBX1* combined), there was no significant difference observed between bone marrow control and

DMSO control (p=0.864). However, a significant difference was observed overall for caffeine exposed cells for both bone marrow control (p=0.003) and DMSO control (p=0.024). These findings were replicated when analysing the percentage of overall abnormal cells with no significant difference between bone marrow and DMSO (p=0.941), and a significant difference observed for caffeine exposed vs bone marrow (p=0.001) and DMSO (p=0.009).

4.3.4 Influence of DMSO on translocation induction in NALM6 cells

Model C conditions were chosen to investigate the influence of DMSO on translocation induction in NALM6 cells due to most translocation events being observed during model optimisation. Out of 6 replicates per exposure, a single *TCF3-PBX1* translocation event was observed for 10µM caffeine with no DMSO, after 4 days exposure, Table 4.8. No translocation events were observed for NALM6 cells with or without DMSO, Table 4.8.

Table 4.8. Translocation events observed in DMSO experiments.

Translocation	Model type	Day cells harvested	NALM6 untreated	10µM Caffeine No DMSO	80µM Caffeine No DMSO	0.00002% DMSO	0.1% DMSO
TCF3-PBX1 (6 replicates)	Model C	Day 2					
		Day 4		1			
RUNX1-RUNX1T1 (6 replicates)	Model C	Day 2					
		Day 4					

4.4 Discussion

Caffeine consumption varies from person to person depending on intake of caffeine containing substances, such as tea and coffee. However, during pregnancy, caffeine concentrations in the plasma can increase even if intake remains static, due to the decreased activity of CYP1A2. This may mean the impact of caffeine is heightened during pregnancy and *in utero* due to prolonged presence of caffeine, further intensified by the lack of the CYP1A2 enzyme in the foetus and placenta [137, 294]. It is therefore important to investigate a range of caffeine concentrations to understand the influence that caffeine may have on the potential induction of chromosomal translocations. However, as caffeine has been shown to significantly reduce cell viability in a number of different cell types, [299-301], it is important to ensure the range of physiologically relevant levels of caffeine investigated does not significantly impact cell death, so that the intended biological end points, i.e., translocation events, could be adequately measured. Therefore, cell growth and viability were measured by trypan blue, resazurin and Cyquant assays, to confirm cell viability was sufficient to assess these outcomes.

Similar growth patterns were observed in the preliminary experiments and Model B, a single dose of caffeine added at seeding, with trypan blue counts showing increasing growth up to 72 hours for caffeine treated and DMSO control cells, with a reduced rate of growth at 96 hours, Figure's 4.3 and 4.5. Although Model A had an extra 24 hours of growth before treatment with a single dose, caffeine exposed cells were slow to proliferate compared to Model B, and only reached a similar cell count to Model B after 72 hours, suggesting that caffeine may have a negative impact when there are more actively replicating cells, Figure 4.5. Caffeine has been shown to inhibit DNA synthesis, possibly through abrogation of cell cycle checkpoint kinase Chk1, and delay G1, reducing the progress into S-phase, which supports these findings [302, 303]. However, for Models C and D, where cells were treated daily with caffeine following a 24 hour growth period or upon seeding respectively, cell growth was slow or plateaued between 24 and 72 hours, Figure 4.5. This may be due to the daily treatment with caffeine providing a higher concentration of exposure, as increasing concentrations of caffeine have shown a greater reduction in DNA synthesis and cell survival [299-302]. Growth in Models C and D did not recover until 96 hours, observed more prominently for Model C, which may be due to the extra 24 hours growth in Model C, which would suggest that there were a larger density of cells to recover, Figure 4.5.

Both resazurin and Cyquant cell viability assays did not show any reduction in cell viability across the preliminary experiments and all models for caffeine exposed cells compared to controls, Figures 4.4, 4.6 and 4.7. Although decreased cell viability has been observed in response to caffeine exposure in multiple cell types, the concentrations used were much higher than those used in this study, i.e., millimolar rather than micromolar, which may be why we do not see any significant reduction in cell viability in this study [299-302]. As such, to detect smaller differences in viability between low caffeine concentrations, a more sensitive method may be applied, such as fluorescent-activated cell sorting [304].

The growth and viability data presented in this study were measured in 1 biological replicate, with trypan blue counts repeated in 2 technical replicates, and resazurin/cyquant assays repeated in 3 technical replicates. As such, it is difficult to conclude the true effect of caffeine concentrations on cell growth and viability, and to determine any significant differences between DMSO control and caffeine exposed cells. These experiments require repeating in at least 2 biological replicates, with at least 3 technical replicates per experiment. Differences in growth and viability between assays may also be due to differences in assay measurement techniques. Trypan blue assays rely on manual counting from a small sample from the cell population, which can be subject to error and user variability, especially when few replicates are performed [305]. The Cyquant assay measures cell density by binding directly to nucleic acid content but does have detection limits based on the quantity of

binding agent used, and as such high nucleic acid content could be missed [305]. Conversely, resazurin assays measure the viability of cells through colour changes observed when oxidised resazurin is reduced by metabolically active cells [254]. Caffeine can influence cellular metabolism through its role as an adenosine receptor antagonist, phosphodiesterase inhibitor and by inducing calcium release [306]. As such, resazurin measurements could be disrupted by caffeine's impact on cell metabolism, since the assay measures viability via metabolic activity. Due to the limitations of trypan blue and resazurin assays, measuring cell growth and viability using the Cyquant assay alone would be suitable. Whilst further biological replicates are required to give a more accurate overview of cell growth and viability when exposed to caffeine, the results are sufficient to show that cells are actively proliferating and viable, and therefore a suitable model system to measure translocations.

The preliminary experiments observed two translocation events across all concentrations in the *TCF3-PBX1* translocation only, Table 4.1. A translocation event was observed at 10µM caffeine at 2 days exposure, equivalent to medium caffeine intake and average serum concentration in the first trimester [294]. A translocation was also observed at 80µM caffeine at 4 days exposure. This was representative of an extreme caffeine concentration, double the maximum serum concentration observed in 3rd trimester, expected to be equivalent to 16 cups of coffee [294, 296]. Studies have observed associations between increased risk of ALL and consumption of two or more cups of coffee per day, which is equivalent to the caffeine concentration of 10µM used in the current study where translocation events were observed [142, 143, 145]. The *TCF3-PBX1* translocation is commonly observed in ALL, however caffeine has not previously been associated with this specific subtype, although some associations have been observed for other subtypes, *MLL* and *ETV6-RUNX1,* which were not measured here [146].

The translocation *TCF3-PBX1* event was reproduced in frequency experiments with 13 out of 20 replicates observing positive RT-PCRs for 80µM caffeine exposed for 4 days. This gave an expected *TCF3-PBX1* translocation frequency of 65% in response to very high caffeine concentrations. Two translocation events were also observed in DMSO control cells, giving an expected translocation frequency of 10%. Although translocations were observed in the DMSO control cells, analysis by Fisher's exact test still suggested the rate of *TCF3-PBX1* translocations were significantly associated with caffeine exposure (P=0.0001)

When translocation events were measured across the various models a total of 11 translocations were observed in caffeine exposed cells, Table 4.3. Of these, 4 *TCF3-PBX1* translocations were observed at the highest caffeine exposure, 80µM, and 1 *TCF3-PBX1*

translocation was observed with 10 µM caffeine exposure, Table 4.3. Conversely, for *RUNX1-RUNX1T1*, translocations were more frequently observed at the lower 10µM caffeine exposure, with 5 translocation events, compared to the higher concentration exposure, 80µM, where only 1 translocation event was observed, Table 4.3. This suggests that the type of translocation induced may be dose dependent. Out of the 11 translocation events observed for caffeine exposed cells, 7 of these were observed in Model C. The expected translocation frequency of 65% calculated from previous experiments suggests that with 6 replicates, as in the model experiments, a total of 3-4 translocation events would be expected for each condition. Most of the model conditions however only observed 1-2 translocation events, with a single condition, 10µM caffeine, Model C day 4, observing 3 *RUNX1-RUNX1T1* translocations.

Due to its impact on many cellular processes, caffeine has the potential to contribute to translocation induction, which is supported by the findings in this study. Caffeine has been shown to be a topoisomerase inhibitor, a mechanism by which chemotherapy agents such as etoposide are known to induce DSBs and *MLL* translocations [157, 159]. Furthermore, through its involvement in ATM and ATR inhibition, caffeine can affect DNA repair by evading DNA checkpoint pathways, preventing any DSBs from being correctly repaired [148]. As such, caffeine has been shown to disrupt HR through interaction with RAD51 filaments which is required for normal DSB repair [153, 156, 307]. As HR is required for normal repair of DSBs, other repair mechanisms such as conventional and alternative NHEJ may become more prevalent, allowing DSBs to be repaired with little to no homology, increasing the likelihood of repair leading to translocation events.

When translocation induction was measured in DMSO control cells for each model, 2 events were observed in Model B for *TCF3-PBX1* and 4 events were observed in Model C for *TCF3-PBX1*, Table 4.3. The translocation events in DMSO control cells were observed at a much higher frequency than previously measured for DMSO control cells in preliminary and frequency experiments. Due to the presence of these translocation events, particularly the high frequency observed in Model C for both control and caffeine exposed cells, it is possible that the DMSO may be contributing to the induction of translocations in NALM6 cells, either alone or by priming cells to be susceptible to caffeine stressors. DMSO is commonly used as a cryoprotectant and is generally considered to be non-genotoxic at low concentrations, as evidenced by a lack of significant DNA damage detected by COMET assays when TK6 cells were exposed to up to 8% DMSO for 3 hour, however its ability to enter the cell and interact with DNA means its impact within cells may be underestimated [308-310]. DMSO can alter gene expression, as shown by Sumida *et al.* 2011 in a study of human and rat hepatocytes cultured in 0.1%-2% (v/v) DMSO for 24 hours [311]. Although microarray analysis did find

some increases in gene expression at 0.1% DMSO, the number of probe sets with altered expression was more prominent as the DMSO concentration increased [311]. DMSO has been shown to impact the epigenetic landscape, as investigated by Verheijen *et al.* 2019, where cardiac and hepatic microtissues were exposed to 0.1% DMSO for 2 to 336 hours followed by analysis of the transcriptome, proteome, and the methylation profile [312]. Verheijen *et al.* 2019, observed differences in gene expression and miRNAs in DMSO exposed microtissues, and observed differentially methylated regions in the cardiac microtissues, suggesting there may be tissue specific differences in DMSO effects [312]. DMSO has been shown to influence cellular processes such as cell cycle progression and mitochondrial integrity [313]. The impact of DMSO on DNA structure could play a major role in the possible influence of DMSO on translocation induction. DMSO can affect chromatin condensation and change DNA conformations with increased frequency of Z-DNA observed in cells exposed to low levels of DMSO (0.1%-1.5% v/v) [314]. As Z-DNA is a left-handed helix, it can be very fragile and lead to breaks which could result in translocations [123].

Alternatively, the translocation events observed in both DMSO and caffeine exposed cells could be due to spontaneous events occurring in NALM6 cells. NALM6 cells were chosen to be used in the translocation screening model due to deficient DNA repair mechanisms, which are expected to be required for chromosomal translocations [285]. With NALM6 cells being an immortalised cancer cell line, multiple replications and exogenous stressors can lead to heterogeneity and development of further mutations [315, 316]. This, along with a defective repair system that prime cells to develop translocations, suggests that spontaneous gain of mutations could account for the translocations observed in both the DMSO and caffeine exposed cells. To understand if spontaneous translocation events are occurring in this model, the experiments condition should be repeated without DMSO, as discussed below. Furthermore, in a real life scenario, foetal cells exposed *in utero* are unlikely to possess a defective DDR, or the same defective mismatch repair response found in NALM6 cells, so to more accurately reflect real life, the experiments should be repeated on a range of cell lines that contain a normal DDR or different defective DDRs, such as TK6 cells with induced HR and NHEJ defects [317].

To visually confirm the presence of the observed translocations using the gold standard method for their detection, which also can assess detection rates, a selection of translocation positive samples for both DMSO control and caffeine exposed cells, were analysed by *RUNX1* and *TCF3* breakapart FISH probes. In samples analysed with *TCF3* probes*,* a higher frequency of translocations were observed in both caffeine exposed and DMSO control cells than those observed in normal bone marrow cells, although this wasn't significant for DMSO control cells. Although the average percentage of translocations in

caffeine exposed cells was higher than in DMSO cells, the difference was not statistically significant. Additionally, the DMSO control with no expected translocations did not observe any *TCF3* translocations above the threshold of the normal bone marrow samples, Table 4.7. For samples analysed with *RUNX1* probes, all caffeine exposed cells observed translocations and abnormalities at a higher frequency than observed in normal bone marrow samples, which was statistically significant for overall abnormalities but only marginally significant for translocations alone. There was no significant difference in *RUNX1* translocations or overall abnormalities between DMSO control and bone marrow control cells. Out of the 3 DMSO controls with *RUNX1-RUNX1T1* events identified by RT-PCR, only 2 observed *RUNX1* translocations at a higher frequency than the normal bone marrow threshold, Table 4.6. The lack of translocations identified in RT-PCR positive sample DMSO control Model C Day 4 #3, may be due to false positive PCR results, or translocations may appear in less than 1% of cells, and so was not captured by FISH since only 100 cells are routinely counted, Table 4.6. *RUNX1* translocations were also observed in RT-PCR negative sample DMSO control Model C Day 2, Table 4.6. Although this translocation could be *RUNX1-RUNX1T1*, as the *RUNX1* breakapart probe identifies any translocation of the *RUNX1* gene, the breaks observed with FISH could be other *RUNX1* translocations, such as the common ALL translocation *ETV6-RUNX1*, and as such this may be why no translocations are observed in the RT-PCR. When the results for *TCF3* and *RUNX1* were combined a significant difference was observed for the percentage of translocations alone and overall abnormalities in caffeine exposed cells compared to both bone marrow control and DMSO control cells. This suggests that although translocations have been observed in DMSO cells, they are more frequent in caffeine exposed cells.

To further investigate the potential influence of DMSO on translocation induction, the conditions in Model C were repeated with the concentration of DMSO previously used, 0.1%, a low concentration of DMSO, 0.00002%, plus 10μM and 80μM caffeine with no DMSO, and NALM6 cells with no DMSO or caffeine added. Out of 6 replicates for each condition, only a single *TCF3-PBX1* translocation was observed at 10μM caffeine exposure, Table 4.8. Although this did not present any clear evidence to support or dismiss the contribution of DMSO on translocation induction, it does suggest that caffeine does contribute to translocation induction, and that previous translocations observed in caffeine exposed cells should not be completely disregarded. Repeating 0%, 0.00002% and 0.1% DMSO concentrations for a higher number of replicates, such as 20, would provide more accurate evidence for the frequency of spontaneous or DMSO induced translocation events.

The evidence presented in this chapter suggests that caffeine exposure contributes to translocation incidence, and this may be translocation type dependent, but given the rarity

and somewhat random nature of these events, more evidence is required to confirm and understand the relationship.

5 Benzene exposure and translocation induction in NALM6 cells

5.1 Introduction

Exposure to benzene, an organic chemical compound, can come from a variety of sources including cigarette smoking, air pollution, gasoline, and solvents [183]. Some e-cigarette systems have also shown formation of benzene vapour, with levels lower that conventional cigarette smoking but higher than found in air pollution [318]. Workplace exposure to benzene, such as petrol station workers, mechanics, and oil refinery workers, has been linked to increased incidence of AML in adults [183]. Some epidemiological evidence suggests that exposure to benzene, including *in utero*, may increase the risk of childhood leukemia [202-205]. However not all studies support this hypothesis, which may be due to limited power or detection methods [201].

Studies have shown that benzene can travel across the placenta, exposing offspring *in vivo* at concentrations equal to or higher than maternal concentrations [188]. Benzene has been shown to cause DNA damage, including breaks, fragmentation, and increased recombination both *in vitro* and in animal models through suggested mechanisms such as increased ROS production which can cause chromosomal instability once legions are repaired by BER and NER, and disruption of NF-κB which can delay cell cycle progression [212-214]. In adult AML, an increase in chromosomal translocations was observed for cases with chronic benzene exposure, with the observed translocations also commonly found in childhood leukaemia [220-222]. This evidence suggests that benzene exposure *in utero* may influence the development of initiating chromosomal translocations involved in childhood leukaemia.

As benzene exposure can come from numerous sources, there is a range of possible blood benzene concentrations that could be physiologically relevant, such as benzene exposed workers with high levels of benzene observed [184, 319, 320]. However, the minimal levels of exposure, such as those from air pollution, are largely unknown, but may still exert a biological effect so it is important to investigate a range of benzene concentrations for their contribution to chromosomal translocation induction, so that accurate advice on different exposure levels and their risk can be provided.

This chapter describes the use of an *in vitro* cell line to investigate the role of physiologically relevant benzene concentrations on the induction of *TCF3-PBX1* and *RUNX1-RUNX1T1* translocations associated with childhood leukaemia. The acute leukaemia cell line, NALM6, was exposed to various physiologically relevant concentrations of benzene to allow for the measurement of chromosomal translocation induction, cell growth and viability.

5.2 Methods

5.2.1 Experimental design of preliminary tests to examine the effects of benzene exposure on translocation events

To investigate the influence of benzene exposure on NALM6 cell viability and translocation induction, as described in Section 2.6.1, a range of benzene concentrations were chosen to reflect real life exposures, Figure 5.1. A study by Brugnone *et al.* 1998 measured blood benzene concentrations in the general population and in participants exposed to benzene in the workplace, with each population group further split by smokers and non-smokers [319]. Each concentration of benzene chosen for this chapter reflects blood benzene concentrations observed in the Brugnone study: 1.5nM represents non-smokers in the general population, 3nM represents general population smokers, 6nM represents the average level of all benzene workers in the study, 12nM represents the highest benzene level observed in the general population smokers in the study and 24nM represents the highest benzene level observed in benzene workers who smoke in the study [319]. A concentration of 48nM benzene was used as an extremely high exposure level for proof of principle to understand if concentrations above physiologically relevant levels of benzene could induce translocations in the event that the physiologically relevant levels did not.

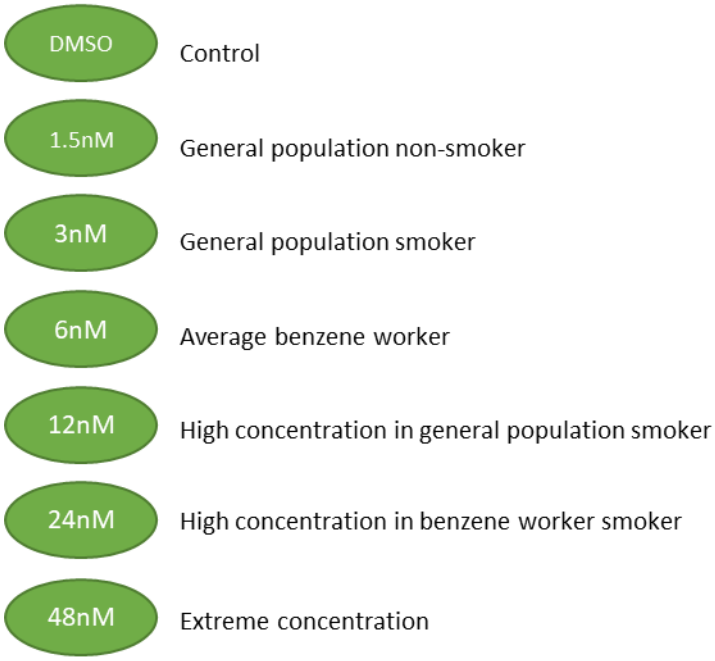


Figure 5.1. Benzene concentrations used in preliminary experiments and their corresponding physiological range.

5.2.2 Design of experiments to investigate the frequency of translocation events in response to benzene exposure

To investigate the frequency at which translocations may be expected to appear in NALM6 cells exposed to benzene, the preliminary experiment was repeated for 20 technical

replicates for 1.5nM and 48nM benzene exposed for 2 days. RNA was extracted from each of the 20 replicates and reverse transcribed into cDNA for use as a template in *RUNX1-RUNX1T1* and *TCF3-PBX1* translocation PCR assays, using HS Taq polymerase and 10pmol primer concentration for 40 cycles. Reported DMSO controls (section 4.2.2) are the same across all chapters 4-6.

5.2.3 Design of experiments to investigate translocation events in response to benzene exposure in NALM6 cells for different models

As described in Section 2.6.2, to optimise the preliminary NALM6 cell model, two variables were considered: i) the impact of treating cells either immediately or 24 hours after initial seeding, and ii) single vs daily dosage of benzene. Benzene exposure is likely to occur on a daily basis through many different routes, including cigarette smoking, second-hand cigarette smoke and air pollution. The benzene concentrations used to treat cells are outlined in Figure 5.2.

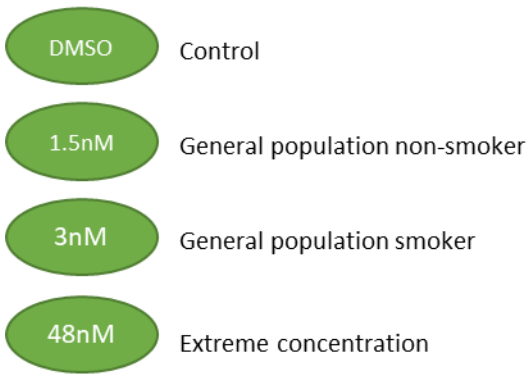


Figure 5.2. Benzene concentrations used in Model's A-D and their corresponding physiological range.

5.2.4 Statistical analysis

Described in section 4.2.4.

5.3 Results

5.3.1 Preliminary experiment measuring cell growth and translocation events in NALM6 cells exposed to benzene

To detect translocations through reverse transcription PCR, NALM6 cells need to be able to maintain adequate cell growth for RNA harvesting when exposed to benzene. Trypan blue results show increasing growth across all benzene concentrations between 24 and 72 hours after first exposure, Figure 5.3. At 96 hours, viable cell count decreases for all benzene concentrations, with the highest concentrations, 12nM, 24nM and 48nM having the lowest viable cell counts compared to control, Figure 5.3.

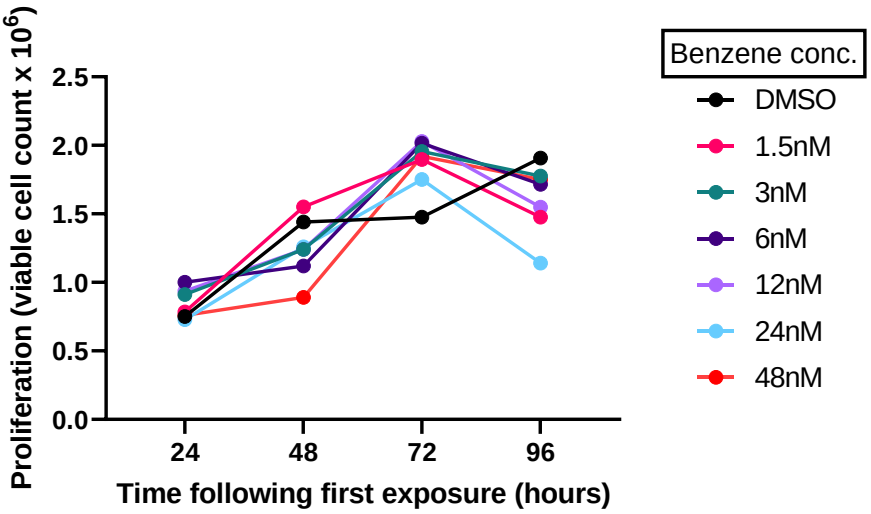


Figure 5.3. Trypan blue measurement of cell growth in NALM6 cells exposed to physiologically relevant concentrations of benzene. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of benzene or DMSO was added and then the cells were incubated at 37°C for up to 96 hours. Trypan blue counts occurred every 24 hours. The data shows 1 replicate.

A total of 8 translocation events were observed across all benzene concentrations, table 5.1. Out of 3 replicates, the highest frequency of translocation events was at 48nM benzene exposure for 2 days, with 2 *RUNX1-RUNX1T1* events in 3 replicates, table 5.1. A *RUNX1-RUNX1T1* event was also observed at 4 days exposure of 48nM and 24nM benzene, table 5.1. The final *RUNX1-RUNX1T1* event was observed at 3nM benzene exposure for 2 days, table 5.1. *TCF3-PBX1* translocation events were observed in the lower benzene concentrations, with events at 1.5nM benzene at both 2 and 4 days exposure, and at 12nM exposure for 4 days, table 5.1. Fisher exact tests show no significant difference between the frequency of *TCF3-PBX1* versus *RUNX1-RUNX1T1* events in benzene exposed and DMSO control cells (p=0.7126).

Table 5.1. Translocation events observed for NALM6 cells exposed to a range of physiologically relevant benzene concentrations. NALM6 cells were seeded into 2 12-well plates at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of benzene or DMSO was added into the wells, and then they were incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2). Cells were harvested for RNA extraction and treated with DNase then reverse transcribed using Precision nanoScript2 Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq and 10pmol primers. The data present shows 3 biological replicates. (DMSO control data for replicates 2 and 3 are common to chapters 4 and 6).

		Benzene conc.						
Translocation	Day cells harvested	DMSO Control	1.5nM	3nM	6nM	12nM	24nM	48nM
TCF3-PBX1 **(3 replicates)**	Day 2		1					
	Day 4		1			1		
RUNX1-RUNX1T1 **(3 replicates)**	Day 2			1				2
	Day 4						1	1

Sanger sequencing of purified PCR products from positive events are shown in table 5.2. The returned sequence from BLAST identified the correct expected translocated gene for each sequenced sample, allowing for the translocation junction to be identified. Due to low product concentration, sequencing was unable to be carried out for 4 of the translocations observed.

*Table 5.2. Sanger sequencing of purified PCR products from preliminary benzene experiments. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence.*

Sample	Sequence of translocation junction	Genes identified in BLAST
Benzene 1.5nM Day 2 rep 1 *(TCF3-PBX1)*	GGGCCTCCCGACTCCTACAGTGTTTTGAGTATCCGAGGAG	*TCF3 PBX1*
Benzene 1.5nM Day 4 rep 1 *(TCF3-PBX1)*	GGGCCTCCCGACTCCTACAGTGTTTTGAGTATCCGAGGAG	*TCF3 PBX1*
Benzene 48nM Day 2 rep 1 *(RUNX1-RUNX1T1)*	CCTTGAATGGCGCCCCCTCACCA	*RUNX1T1*
Benzene 3nM Day 2 rep 1 *(RUNX1-RUNX1T1)*	CCCCGAGAACCCTGAAAAAAAGAGGGAGAAG	*RUNX1 noise RUNX1T1*

5.3.2 Frequency of translocation events in response to benzene exposure

The translocation events observed in the preliminary experiments do not provide a clear pattern of translocation induction due to a limited number of replicates for each exposure. To investigate the frequency at which translocations may be expected to appear, the preliminary experiment was replicated 20 times for 1.5nM and 48nM benzene, alongside a DMSO control. These concentrations were chosen as two extremes of benzene exposure where a translocation was observed and confirmed via sequencing. For *RUNX1-RUNX1T1*, no translocation events were observed in the DMSO control experiments, whereas 5 translocation events were observed across 5/20 replicates treated with 1.5nM benzene and 6 translocation events were observed for 6/20 replicates treated with 48nM benzene. Analysis by Fisher exact test suggests that the relationship between benzene exposure and *RUNX1-RUNX1T1* translocations may be statistically significant, p= 0.047 for 1.5nM and p= 0.020 for 48nM benzene. For *TCF3-PBX1,* 2 translocation events were observed across 2/20 replicates treated with DMSO, and 1 translocation event was observed for 1/20 replicates treated with 1.5nM benzene. No significant relationship was observed between *TCF3-PBX1* translocations and 1.5nM benzene exposure when compared with the DMSO control using Fisher exact test (p= >0.9999). There was no statistical difference using Fisher

exact tests between the frequency of *TCF3-PBX1* vs *RUNX1-RUNX1T1* events in response to 1.5nM benzene exposure (p=0.1818)

5.3.3 Cell growth, viability and translocation events measured in response to benzene exposure in different models

To replicate the preliminary findings and address the issues of daily dosing and timing of those doses, further experiments were conducted using NALM6 cells treated with benzene or DMSO across 4 different models.

Trypan blue assays were performed every 24 hours following first exposure to benzene to assess the impact of each condition on NALM6 growth. In Model A (single exposure, 24 hours after seeding), cell growth increased each day across 96 hours with exposure to all benzene concentrations, but with a decrease in viable cell count at 96 hours with 1.5nM benzene, Figure 5.4. In Model B (single dose at seeding), growth is slow between 24-48 hours for all condition but then increases every 24 hours for the DMSO control and 1.5nM benzene, whereas for 3nM and 48nM benzene plateaus between 72 and 96 hours, Figure 5.4. Viable cell count in Models C (daily dose, 24 hours after seeding) and D (daily dose at seeding) plateaued for all benzene concentrations and DMSO control between 24 and 72 hours, with a sharp increase at 96 hours for Model C, and a gentle increase at 96 hours for Model D, Figure 5.4.

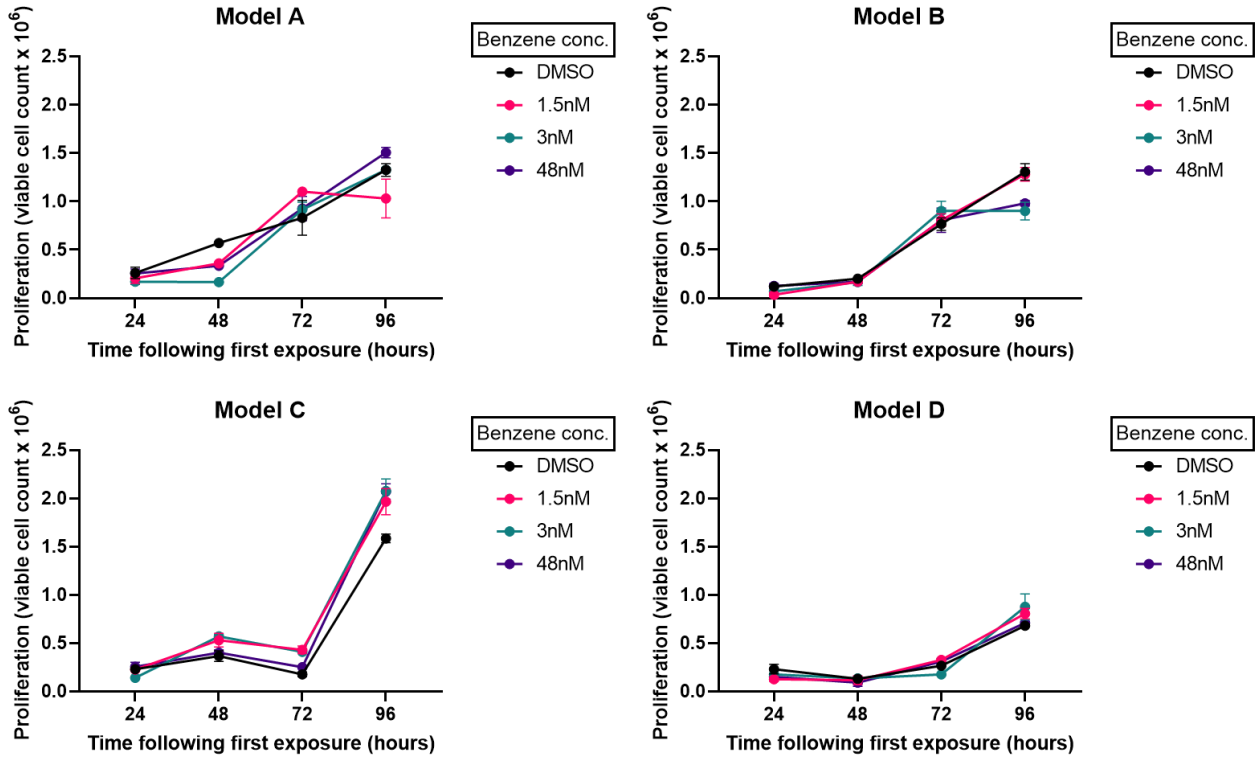


Figure 5.4. Trypan Blue Assay for each of the 4 benzene exposure models. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media and incubated at 37°C. Benzene concentrations 1.5nM, 3nM and 48nM were chosen to represent various physiologically relevant benzene exposures, with the equivalent DMSO (0.1% v/v) used as a negative control. Trypan blue counts occurred every 24 hours following the first dose of benzene. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

Cell viability assays were performed at day 2 and day 4 following first exposure using resazurin, a metabolic colorimetric assay, and Cyquant, a nucleic acid binding assay. No decrease in viability compared to control was observed for any benzene exposure in any model as measured by resazurin, Figure 5.5, indeed with increases in viability compared to controls were observed across all models.

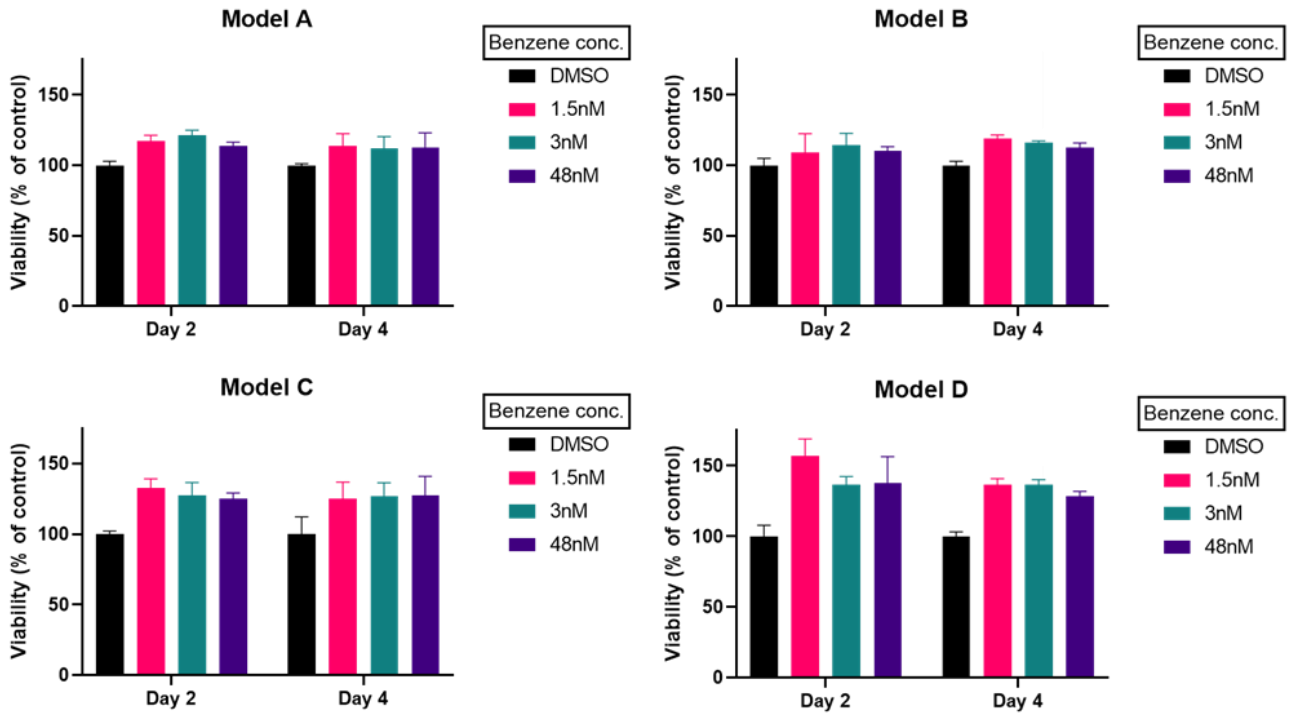


Figure 5.5. Resazurin assay measurements for each of the 4 benzene exposure models. On day 2 or day 4 of each model, resazurin was added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

Similarly, Cyquant viability assays show no reduction in cell viability compared to control in any model, with increases in viability observed across all models, Figure 5.6.

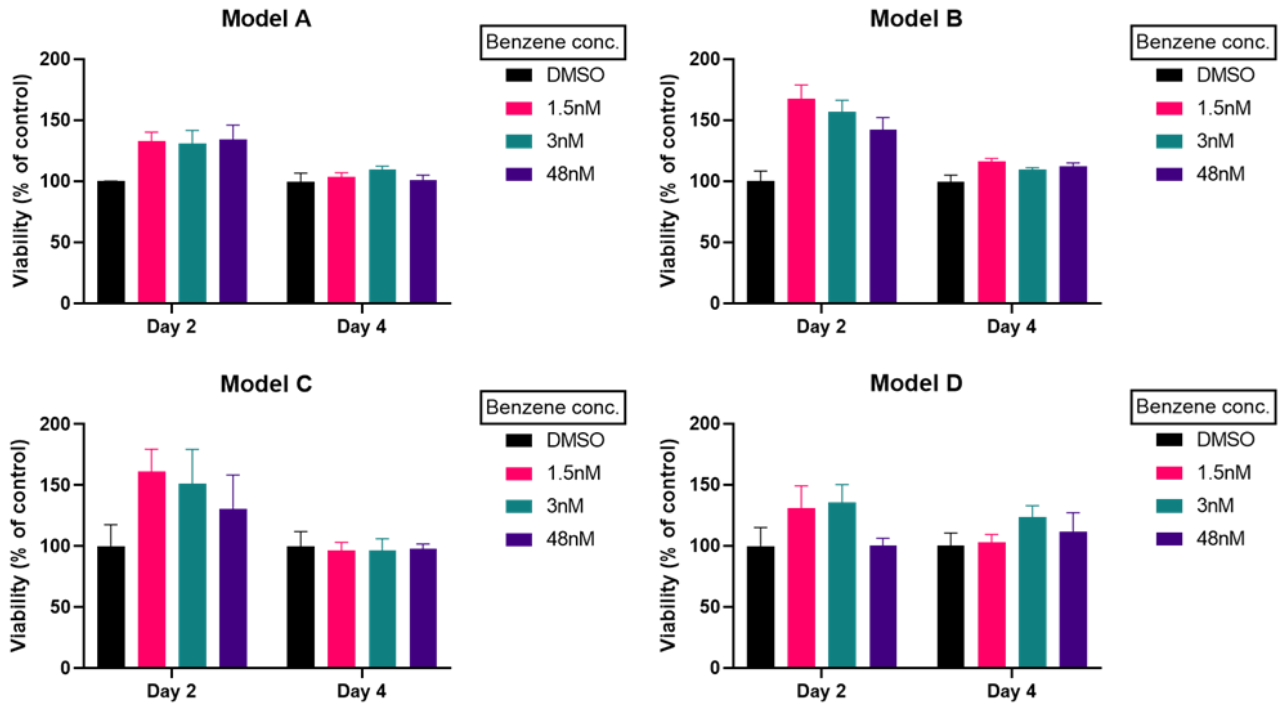


Figure 5.6. Cyquant assay measurement for each of the 4 benzene exposure models. On day 2 or day 4 of each model, samples were taken from each well and frozen at -80°C. Cyquant was added to each sample, incubated overnight at 4°C and fluorescence monitored at 480nm excitation and 520nm emission. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

A total of 12 translocation events were observed across 11/144 replicates treated with benzene, table 5.3. The DMSO control observed a further 6 translocation events across 48 replicates, table 5.3. The highest number of translocation events were observed in Model C, with a single *RUNX1-RUNX1* event across 6 replicates for each condition of cells treated with 1.5nM and 3nM benzene across 2 and 4 days, a single *TCF3-PBX1* event across 6 replicates for cells treated with 3nM and 48nM benzene for 2 days, and 4 *RUNX1-RUNX1* events across 12 replicates for cells treated with DMSO alone, table 5.3. Both translocations were observed in the same replicate for Model C, 3nM at 2 days exposure. Model A had a single replicate with a *RUNX1-RUNX1T1* event after 2 days exposure at each benzene concentration, table 5.3. Model D had a single *TCF3-PBX1* event overall at 1.5nM benzene for 4 days exposure, table 5.3. In Model B, *RUNX1-RUNX1T1* and *TCF3-PBX1* translocations were observed in different replicates of 3nM benzene exposure for 2 days, as well as 2 *TCF3-PBX1* events in the DMSO control table 5.3. Using Fisher exact tests, no significance was observed for translocation events in response to DMSO vs any benzene treatment (p=0.3776). No significance was observed between single dose vs daily dose (0.4585) and for treatment following 24 hours growth vs at seeding (p=0.0807) using Fisher exact tests. When considering any benzene exposure, no significant difference was

observed for the frequency of *TCF3-PBX1* and *RUNX1-RUNX1T1* events using Fisher exact tests (p=0.2508).

Table 5.3. Model optimisation translocation events in response to benzene. Each number represents the number of positive PCR bands seen during gel electrophoresis for 6 replicates of each condition. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media and incubated at 37°C. Model A: 24 hours growth, single dose of benzene or DMSO. Model B: single dose of benzene or DMSO at seeding. Model C: 24 hours growth, daily dose of benzene or DMSO. Model D: daily dose of benzene or DMSO starting at seeding. Benzene concentrations 1.5nM, 3nM and 48nM were chosen to represent various physiologically relevant benzene exposures, with the equivalent DMSO (0.1% v/v) used as a negative control. On day 2 and day 4 cells were also harvested for RNA extraction. The resultant RNA was treated with DNase and reverse transcribed using High-Capacity cDNA Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq with 10pmol primers. The data present shows 3 technical replicates and 2 biological replicates. (DMSO control data is common across caffeine (chapter 4) and cotinine (chapter 6)).

			Benzene conc.			
Translocation	Model type	Day cells harvested	DMSO Control	1.5nM	3nM	48nM
***TCF3-PBX1* (6 replicates)**	Model A	Day 2				
		Day 4				
	Model B	Day 2			1	
		Day 4	2			
	Model C	Day 2			1	1
		Day 4				
	Model D	Day 2				
		Day 4		1		
***RUNX1-RUNX1T1* (6 replicates)**	Model A	Day 2		1	1	1
		Day 4				
	Model B	Day 2			1	
		Day 4				
	Model C	Day 2	1	1	1	
		Day 4	3	1	1	
	Model D	Day 2				
		Day 4				

A selection of negative PCR samples were used as a template for PCR using nested primers to confirm the negative result, with no amplification observed. Sanger sequencing of the translocation breakpoints showed the expected fusion mRNA sequence, with the associated genes identified in BLAST, table 5.4. Sequencing results for translocations observed with DMSO treatment be found in chapter 4, table 4.5.

*Table 5.4. Sanger sequencing of PCR positive translocation events identified in each Model. Positive samples were used as a template for nested PCR using HS taq polymerase and 10pmol primers, 1-3 nested with original primers, 4-12 nested with nested primers. Samples were then purified for sequencing by DBS genomics. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence.*

Sample	Sequence of translocation junction	Genes identified in BLAST
Benzene 48nM Model C Day 2 (*TCF3-PBX1* Nested)	GGCCTCCCGACTCCTACAGTGTTTTGAGTATCCGAGGAG	*TCF3 PBX1*
Benzene 3nM Model B Day 2 (*TCF3-PBX1* Nested)	GGCCTCCCGACTCCTACAGTGTTTTGAGTATCCGAGGAG	*TCF3 PBX1*
Benzene 1.5nM Model D Day 4 (*TCF3-PBX1* Nested)	GGCCTCCCGACTCCTACAGTGTTTTGAGTATCCTAGGAG	*TCF3 PBX1*
Benzene 3nM Model C Day 2 (*TCF3-PBX1* Nested)	GGCCTCCCGACTCCTACAGTGTTTTGAGTATCCGAGGAG	*TCF3 PBX1*
Benzene 48nM Model A Day 2 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Benzene 3nM Model A Day 2 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAACCTCGAA*ATCGTACTGAGGAG	*RUNX1 RUNX1T1*
Benzene 3nM Model B Day 2 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAACCTCGAA*ATCGTACTGGAGAAG	*RUNX1 RUNX1T1*
Benzene 3nM Model C Day 2 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Benzene 3nM Model C Day 4 (*RUNX1-RUNX1T1* Nested)	GAGAACCTCGAA*ATCGTACTGGAGGAAG	*RUNX1 RUNX1T1*
Benzene 1.5nM Model A Day 2 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Benzene 1.5nM Model C Day 2 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAACCTCGAA*ATCGTACTGAGAAG	*RUNX1 RUNX1T1*
Benzene 1.5nM Model C Day 4 (*RUNX1-RUNX1T1* Nested)	CCCCGAGAACCTCGAA*ATCGTACTGAG-AG	*RUNX1 RUNX1T1*

5.4 Discussion

Concentrations of benzene in the blood can vary based on the environmental exposure levels, such as ambient benzene exposure (through air pollution, second hand smoke exposure), with low blood benzene levels of 3nM, or exposure to cigarette smoking and occupational benzene, with high blood benzene levels of 12nM [319, 320]. It is therefore

important to investigate a range of benzene concentrations to capture the effect that variation in benzene exposure may have on biological outcomes. However, as benzene exposure is associated with hematotoxicity [321], it is important to ensure that physiologically relevant concentrations were used in this study, described in section 5.2.1, did not cause significant cell death of NALM6 cells during treatment, so that biological end points of interest i.e., translocation events could be adequately addressed. Therefore, growth and viability were measured by trypan blue, resazurin and Cyquant assays to ensure a good level of cell viability was achieved to assess these outcomes.

In the preliminary experiments and Model B, a single dose of benzene added at seeding, trypan blue counts show increasing growth up to 72 hours, followed by a decrease or plateau in cell counts at 96 hours for benzene exposed cells compared to the DMSO control, Figures 5.3 and 5.4. This pattern of growth was also observed in Model A, a single dose following 24 hour growth period, with similar viable cell numbers between Model A and B, even with the extra 24 hours of growth, Figure 5.4. However, for Models C and D, where daily benzene doses were given after 24 hours growth or at seeding respectively, showed a different pattern of growth, both with a plateaued growth for the first 72 hours, Figure 5.4. This could be due to the daily addition of benzene, meaning that cells did not have time to recover between doses. Indeed, a study observing the impact on benzene exposed mice found that bone marrow cellularity decreased during daily exposure to benzene, then increased when exposure was postponed, including significant difference in white blood cell count compared to control [322], which is in agreement with the observations made here. At 96 hours, the viable cell count for both Model C and Model D increased, with a much steeper increase for Model C, which may be due to having a larger cell count available for recovery following the extra 24 hour growth before treatment.

Both resazurin and Cyquant cell viability assays did not show any reduction in cell viability across all models for benzene exposed cells compared to DMSO controls, Figures 5.5 and 5.6. However, increases in viability compared to the DMSO control were observed in both resazurin and Cyquant assays across all of the models, Figures 5.5 and 5.6. A study of benzene exposure on rat pituitary cells over 96 hours observed increased viability using the methylthiotetrazole (MTT) assay and increased cell energy content using the ATPlite assay, suggested to be due to benzenes activity as an endocrine disruptor, that can antagonise receptors and activate oncogenes [323]. This study observed these significant increases at benzene concentrations of 1.250nM, similar to the concentrations used in the current study, supporting the increased NALM6 viability observed [323]. Whilst decreases in cell viability have been observed in rat lymphocytes exposed to benzene and metabolised benzene, the concentration of benzene used was much higher, i.e 75µM, than the lower physiologically

relevant concentrations used in the current study [324]. This suggests that benzene may have a hormetic effect, in that lower concentrations may have some benefits to cell viability, but higher concentrations are toxic. As has been discussed previously (chapter 4) observations for growth and viability in this study are likely to be limited due to a single measurement, but also affected by the differences in assay measurement techniques used.

The preliminary experiment observed both *TCF3-PBX1* and *RUNX1-RUNX1T1* translocation events after exposure to physiologically relevant levels of benzene. Translocations were most frequently observed for *RUNX1-RUNX1T1* with 5 events observed across 36 replicates for any benzene concentration, which included translocations observed at 3nM (smoker), 12nM (benzene worker smoker) and 48nM (extreme) benzene exposure. *TCF3-PBX1* translocation events were found at lower benzene concentrations with 2 translocation events observed at 1.5nM benzene exposure (non-smoker) and 1 event at 12nM (high smoker). However, the difference in frequency was insignificant, so more replicates may be required to confirm this finding.

The more frequent *RUNX1-RUNX1T1* event was reproduced during the frequency test with 5 events across 20 replicates for 1.5nM benzene exposure and 6 events across 20 replicates for 48nM benzene. No *RUNX1-RUNX1T1* events were observed for the DMSO control in either the preliminary experiment or the frequency test. This gave an expected *RUNX1-RUNX1T1* translocation frequency of 25-30%, with Fisher exact tests suggesting this was significant for both concentrations. For *TCF3-PBX1* only 1 translocation event out of 20 was observed for 1.5nM benzene, a frequency of 5%, and 2 translocation events were observed in the DMSO control, a frequency of 10%, with Fisher exact tests analysis suggesting the *TCF3-PBX1* events were independent of benzene concentration. When directly comparing the frequency of events in 1.5nM benzene exposed cells using Fisher exact tests, no significant difference was observed for the type of translocation ($p=0.1818$). However, when all benzene exposures were combined from the frequency experiments (i.e., 1.5nM and 48nM benzene events), a significant difference was observed between the number of *TCF3-PBX1* and *RUNX1-RUNX1T1* events ($p= 0.0466$). These findings suggest that the *RUNX1-RUNX1T1* translocation is likely to be more susceptible to benzene exposure.

The *RUNX1-RUNX1T1* translocation is commonly associated with both childhood and adult AML [222]. One study of adult AML cases observed an increased number of *RUNX1-RUNX1T1* abnormalities in cases who had been chronically exposed to benzene compared to non-chronic benzene exposed cases [220]. This suggests the mechanism of benzene translocation induction may be targeted and more likely to appear in specific gene pairs and

is in line with our observations here of more *RUNX1-RUNX1T1* translocations than *TCF3-PBX1* induced in response to benzene exposure.

When translocations were measured across the various models, again, the *RUNX1-RUNX1T1* translocation was again more prevalent with 8 events across benzene exposed cells compared to 4 *TCF3-PBX1* events, table 5.3. With each benzene concentration repeated for 6 replicates in total, results of the frequency tests suggest translocations are expected at a frequency of 1-2 events. The maximum number of translocation events observed out of 6 replicates across all conditions was 1, in line with the frequency tests. Model C, daily benzene dose following 24 hour growth period, had the highest number of translocations with 4 *RUNX1-RUNX1T1* events in total, at 1.5nM and 3nM benzene at both 2 day and 4 day exposure, i.e. one per condition, and 2 *TCF3-PBX1* events in total, at 2 day exposure with treatment of 3nM and 48nM benzene. Model A, single benzene dose following 24 hour growth period, had the next highest frequency with 3 *RUNX1-RUNX1T1* events, 1 at each benzene concentration at 2 days exposure. Exposure to benzene has been shown to dysregulate proliferation and differentiation of haematopoietic cells, with *in utero* exposure in mice shown to supress the cell cycle [216, 325]. Both Model A and Model C were exposed to benzene during the exponential growth phase, suggesting that benzene may have the largest impact on actively dividing cells.

Benzene is proposed to contribute to leukemogenesis through various mechanisms including disruption of signalling pathways and the haematopoietic stem cell niche, mutations, epigenetic alterations, and oxidative stress [215, 216]. The mechanism by which benzene influences chromosomal translocations could be due to the production of reactive oxygen species (ROS) [326]. Mice studies of *in utero* benzene exposure have observed increased production of ROS in the foetal samples [189]. ROS has been shown to increase DSBs and impair DNA repair, specifically in myeloid malignancies, both of which are required for chromosomal translocation occurrence [326, 327]. In future work it would be useful to measure ROS production and DSB occurrence to further understand the mechanism by which benzene exposure may induce translocations involved in childhood leukaemia.

As previously discussed, (chapter 4, section 4.4), translocation events were observed in some DMSO control cells, which suggests that DMSO may be influencing translocation development alone, or priming cells to be more susceptible to translocation induction. Also, as DMSO is a free radical scavenger, it could be expected to provide protection against benzene induced ROS, suggesting that the true impact of benzene in translocation induction may also be masked by the presence of DMSO [308, 328]. As such, it would be beneficial to

repeat these experiments and remove the DMSO, using a different solvent, or diluting benzene directly into the cell culture medium.

The preliminary data in this chapter suggest that NALM6 cells exposed to physiologically relevant levels of benzene may influence translocation events, with a higher frequency for *RUNX1-RUNX1T1* translocations, that have been associated with benzene exposure in adult AML [220]. Statistical tests for translocations observed during frequency tests in response to benzene exposure suggest a significant difference between *RUNX1-RUNX1T1* and *TCF3-PBX1.* Although a higher frequency of *RUNX1-RUNX1T1* translocations were observed across preliminary or model experiments, analysis by Fisher's exact tests does not find any significance, suggesting more replicates are required to solidify this finding.

6 Cotinine exposure and translocation induction in NALM6 cells

6.1 Introduction

Cotinine, the major metabolite of nicotine, is a useful biomarker of nicotine intake [225]. As a key component of cigarette smoke, nicotine is in an important molecule to monitor due to the serious impact of smoking on health, including the association with multiple cancers [162]. This importance is further increased due to the rising popularity of e-cigarettes, allowing nicotine to be inhaled as vapour, seen as a "safer" alternative to traditional cigarette smoking [329, 330]. Much less is known about the impact of e-cigarettes on health, but some studies have already shown similar impacts to those found with traditional cigarette smoking, such as suppression of immune-related genes [331]. Maternal cigarette smoking is associated with adverse health outcomes in children [168, 169], but associations with childhood leukaemia through epidemiological studies of parental exposure have been varied [170].

With a longer half-life than nicotine, cotinine is often measured via saliva, urine and blood [332]. This means that the body is exposed to cotinine for a longer length of time and as such would make an interesting target to investigate the influence of this exposure on the induction of translocations, as a component of cigarette smoking and e-cigarettes. Exposure to cotinine varies between person to person, based on how many cigarettes they smoke, or via exposure to second hand smoke. The cotinine level ranges to be able to distinguish between smokers and non-smokers are often debated, and have been observed to fluctuate between different populations, such as race [332, 333]. When investigating the impact of cotinine exposure, it is important to consider the range of cotinine levels that people may be exposed to.

Studies have shown that offspring are exposed to cotinine *in utero*, with some levels matching those found in active smokers [87]. Although the carcinogenic role of cotinine is still inconclusive, *in vitro* studies have provided evidence of DNA damage and sperm alteration [236, 239, 242]. This suggests it is a plausible target to investigate if exposure may initiate translocations associated with childhood leukaemia. To our knowledge no *in vitro* studies have identified translocation induction on exposure to cotinine.

This chapter describes the development of an *in vitro* cell line model to explore the role of physiologically relevant levels of cotinine exposure on the induction of *TCF3-PBX1* and *RUNX1-RUNX1T1* translocations associated with childhood leukaemia. To do this, the chosen acute leukaemia cell line, NALM6, was exposed to a gradient of cotinine concentrations, and the effects on cell growth, viability and translocation incidence were measured.

6.2 Methods

6.2.1 Experimental design of preliminary tests to examine the effects of cotinine exposure on translocation events

To investigate the influence of cotinine exposure on NALM6 cell viability and translocation induction, a gradient of cotinine concentrations were chosen to reflect real life exposure based on their relevancy to physiological concentrations using serum cotinine levels reported in literature, Figure 6.1. The lowest concentration, 10nM cotinine, reflects levels identified by self-reported non-smokers, with 100nM chosen to represent second hand smoke exposure [332]. To reflect the serum cotinine levels of average smokers, a concentration of 1µM was selected [334]. A cotinine concentration of 5µM was chosen to represent very high smokers. The highest concentration, 10µM, was used as an extremely high proof of principle exposure level. An overview of the experiment can be found in section 2.6.1.

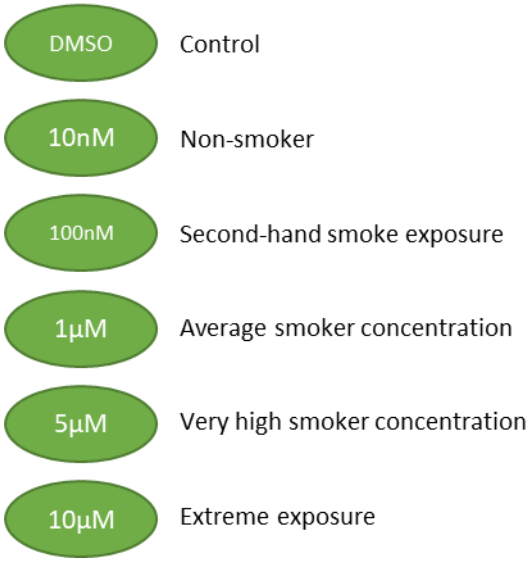


Figure 6.1. Cotinine concentrations used in preliminary experiments and their corresponding physiological range.

6.2.2 Design of experiments to investigate the frequency of translocation events in response to cotinine exposure

To investigate the frequency that translocations can be expected to appear in NALM6 cells exposed to cotinine, the preliminary experiment was repeated for 20 technical replicates using 5µM cotinine for 2 days exposure. RNA was extracted from all 20 replicates and reverse transcribed into cDNA for use as a template in *RUNX1-RUNX1T1* translocation RT-PCR assays using HS Taq polymerase with 10pmol primers for 40 cycles. Reported DMSO controls (section 4.2.2) are the same across all chapters 4-6.

6.2.3 Design of experiments to investigate translocation events in response to cotinine exposure in NALM6 cells for different models

As discussed in Section 2.6.2, to optimise the preliminary NALM6 cell model, two variables were considered: i) the impact of treating cells either immediately or 24 hours after initial seeding, and ii) single vs daily dosage of cotinine. As cotinine is exposed to the body through consumption of nicotine, it is unlikely that exposure is at a single time point, and like with consumption of cigarettes, exposure would be likely to occur frequently and daily. The cotinine concentrations used to treat cells are outlined in Figure 6.2

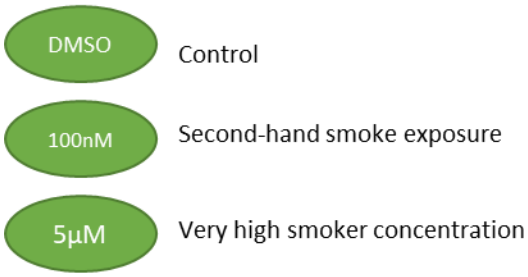


Figure 6.2. Cotinine concentrations used in Model's A-D and their corresponding physiological range.

6.2.4 Statistical analysis

Described in section 4.2.4.

6.3 Results

6.3.1 Preliminary experiment measuring cell growth and translocation events in NALM6 cells exposed to cotinine

It is important that upon exposure to cotinine, NALM6 cells are viable enough to continue proliferating, increasing the chance of translocations to occur and for fusion gene mRNA to be harvested for detection. Trypan blue assay results show increasing growth for all cotinine concentrations from 24 to 72 hours, with a reduction in viable cells at 96 hours, Figure 6.3.

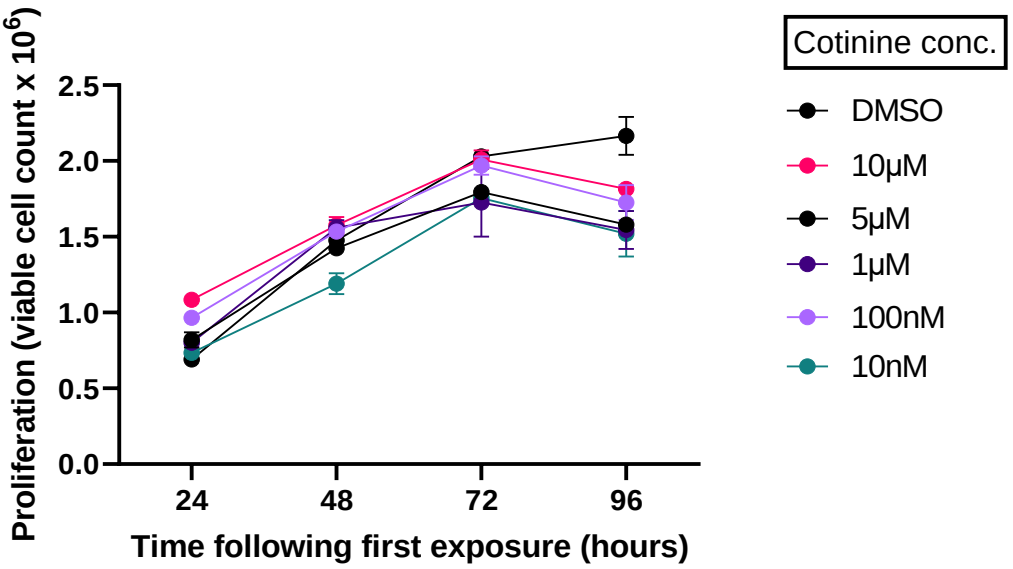


Figure 6.3. Trypan blue measurement of cell growth in NALM6 cells exposed to physiologically relevant concentrations of cotinine. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of cotinine or DMSO was added and then the cells were incubated at 37°C for up to 96 hours. Trypan blue counts occurred every 24 hours. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

Using the resazurin assay, it was observed that all cotinine concentrations had higher percentage viability than the DMSO control at both day 2 and day 4, Figure 6.4.

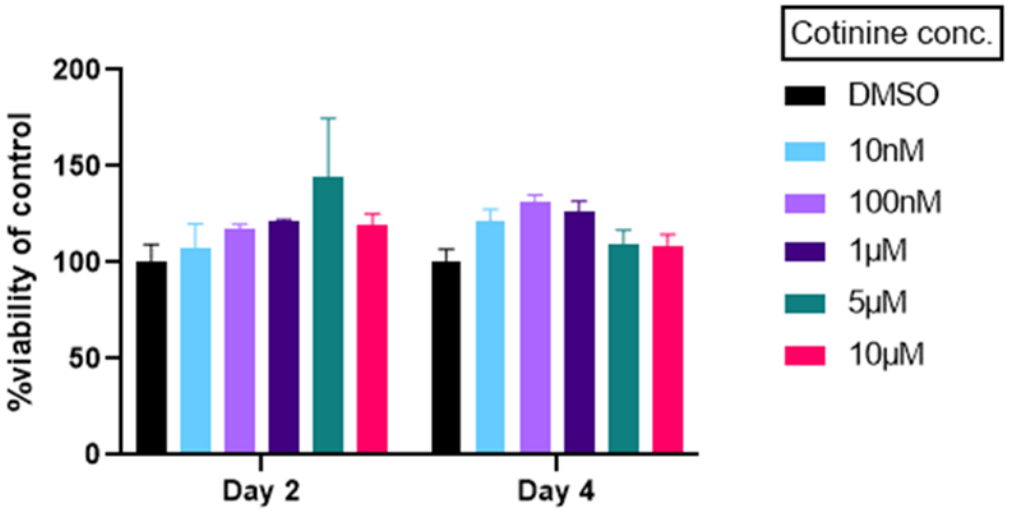


Figure 6.4. Resazurin assay measurement of cell viability of NALM6 cells exposed to physiologically relevant concentrations of cotinine at day 2 and day 4. NALM6 cells were seeded into 2 well plates at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of cotinine or DMSO was added into the wells, and then they were incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2). On day 2 or day 4, resazurin was added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

Two translocation events were observed for the *TCF3-PBX1* translocation, after 2 days exposure to 100nM and 10nM cotinine, Table 6.1. For the *RUNX1-RUNX1T1* translocation,

a positive event was observed for 2 days exposure to 5µM cotinine, Table 6.1. Fisher exact tests show no significant difference between the frequency of *TCF3-PBX1* versus *RUNX1-RUNX1T1* events in cotinine exposed and DMSO control cells (p=1).

Table 6.1. Translocation events observed for NALM6 cells exposed to a cotinine concentration gradient. NALM6 cells were seeded into two 12-well plates at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media. A single dose of cotinine or DMSO was added into the wells, and then they were incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2). Cells were harvested for RNA extraction and treated with DNase then reverse transcribed using Precision nanoScript2 Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq and 10pmol primers. The data present shows 3 biological replicates. (DMSO control data is common across chapters 4 and 5).

Translocation	Day cells harvested	DMSO Control	10nM Cotinine	100nM Cotinine	1µM Cotinine	5µM Cotinine	10µM Cotinine
TCF3-PBX1	Day 2		1	1			
	Day 4						
RUNX1-RUNX1T1	Day 2					1	
	Day 4						

Sanger sequencing of purified PCR products from positive events are shown in Table 6.2. The returned sequence from BLAST identified the correct expected translocated gene for each sample, allowing for the translocation junction to be identified.

*Table 6.2. Sanger sequencing of purified PCR products from cotinine concentration gradient. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence.*

Sample	Sequence of translocation junction	Genes identified in BLAST
Cotinine 100nM Day 2 rep 1 (*TCF3-PBX1*)	CCTCCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Cotinine 10nM Day 2 rep 1 (*TCF3-PBX1*)	CCTCCCGACTCCTACAGTGTTTTGAGTATCCGAG	*TCF3 PBX1*
Cotinine 5µM Day 2 rep 1 (*RUNX1-RUNX1T1*)	GGGCCCCGAGAACCTCGAA*ATCGTACTGAG	*RUNX1 RUNX1T1*

6.3.2 Frequency of translocation events in response to cotinine exposure

Although some translocation events were observed across the range of cotinine concentrations, there is no clear pattern of translocation induction over the 3 biological replicates. To determine the expected frequency of translocation events in response to cotinine, the preliminary experiments were repeated 20 times. Overall, 7 positive *RUNX1-RUNX1T1* events out of 20 replicates were observed when NALM6 cells were exposed to 5µM cotinine for 2 days. No *RUNX1-RUNX1T1* translocations were observed when NALM6 cells were exposed to DMSO. Analysis by Two-tailed Fisher's exact test suggests there is a relationship between *RUNX1-RUNX1T1* translocation events and cotinine exposure (p=0.0083).

6.3.3 Cell growth, viability and translocation events measured in response to cotinine exposure in different models

To investigate the impact of each model type on the growth of NALM6 cells, a trypan blue count was performed every 24 hours following the first cotinine exposure, Figure 6.5. In Model A (single exposure, 24 hours after seeding), cell growth increased across the 96 hours for all concentrations, with viable cell count higher for cotinine exposures than the DMSO control at hours 72 and 96. Cotinine exposed cells follow the same pattern of increasing growth every 24 hours as the DMSO control in Model B (single dose at seeding). In Model C (daily dose, 24 hours after seeding), cotinine exposed cells have increasing growth across 96 hours, with the DMSO control showing a reduction in viable cell count at 72 hours, followed by a steep increase at 96 hours. Model D cells (daily dose at seeding) have a slow growth for the first 72 hours compared to the other models. At 96 hours, the cotinine exposed cells in Model D have a sharp increase in viable cell count, with the DMSO control cells, showing a gentler increase.

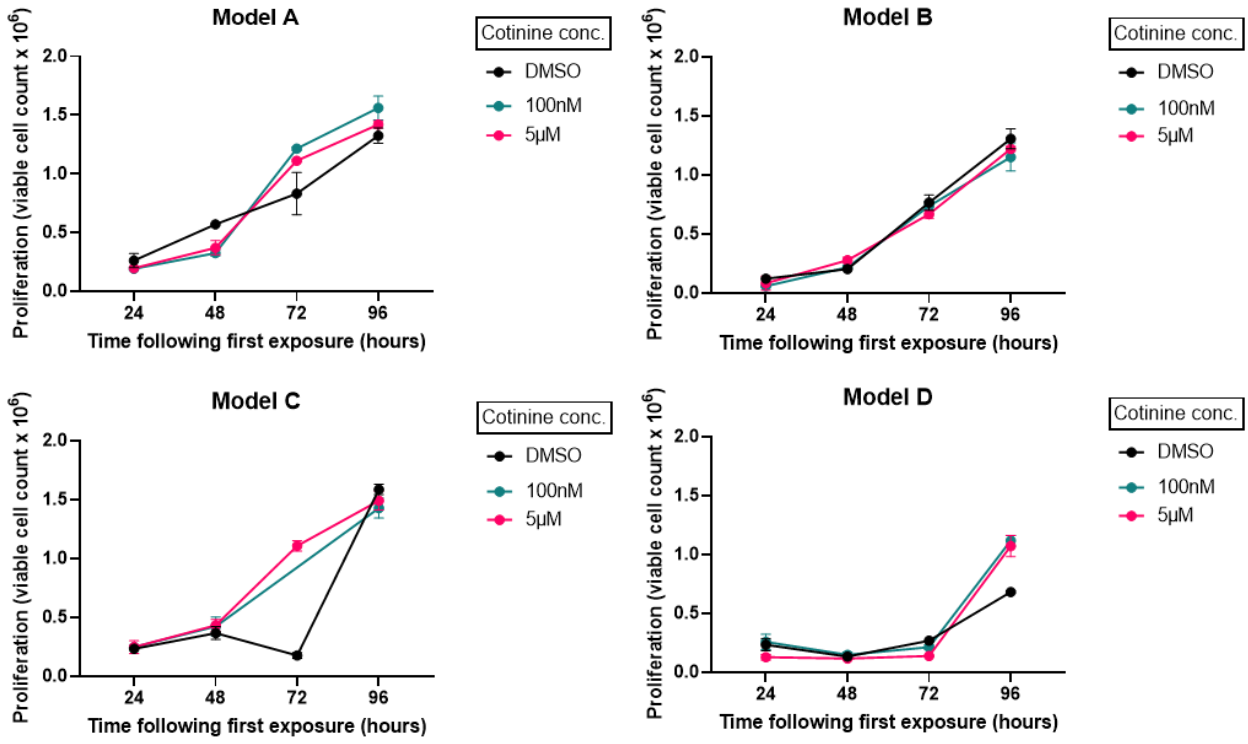


Figure 6.5. Trypan Blue Assay for each of the 4 cotinine exposure models. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media and incubated at 37°C. Cotinine concentrations 100nM and 5µM were chosen to represent various physiologically relevant cotinine exposures, with the equivalent DMSO (0.1% v/v) used as a negative control. Trypan blue counts occurred every 24 hours following the first dose of cotinine. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

Resazurin assays showed an increase in viability across all models for cotinine exposed cells compared to the DMSO control, Figure 6.6.

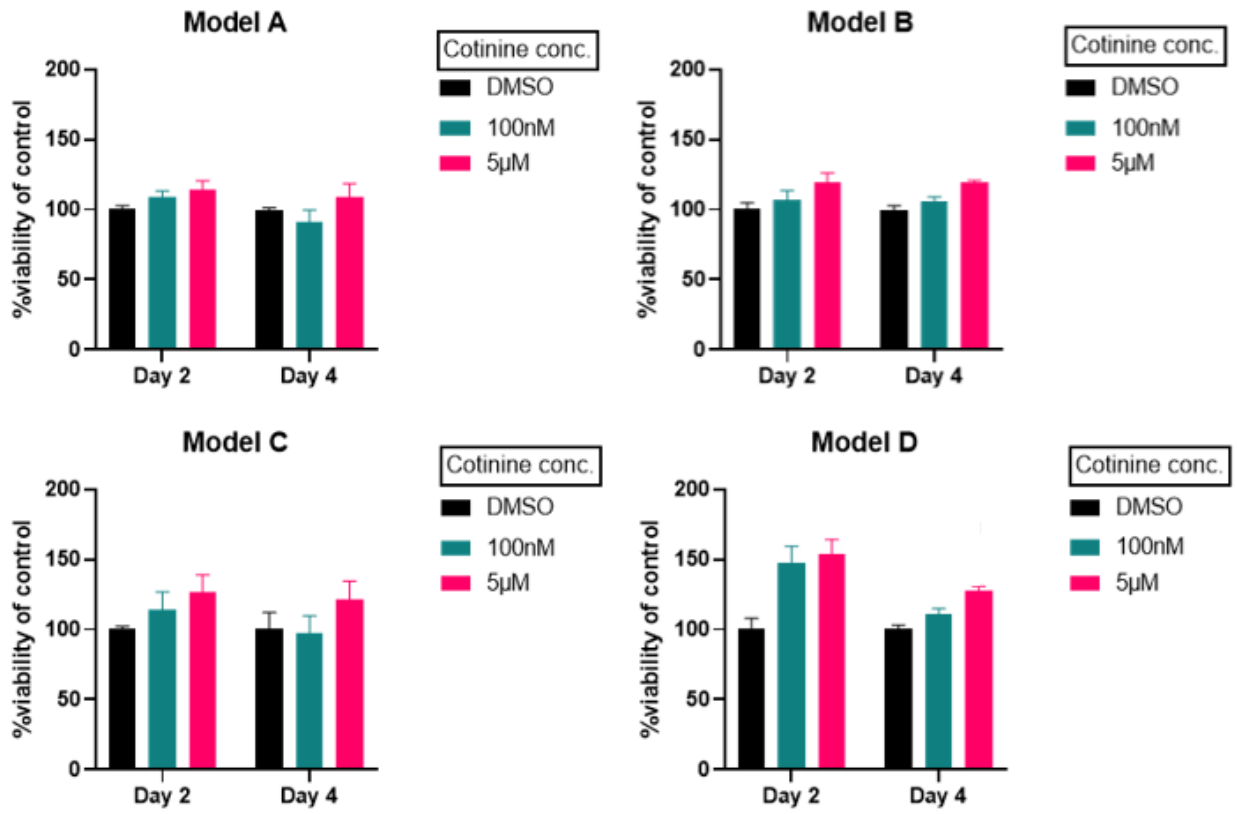


Figure 6.6. Resazurin assay measurements for each of the 4 cotinine exposure models. On day 2 or day 4 of each model, resazurin was added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

Cyquant viability analysis also showed an increase in cell viability across all models for cotinine exposed cells compared to DMSO controls, Figure 6.7.

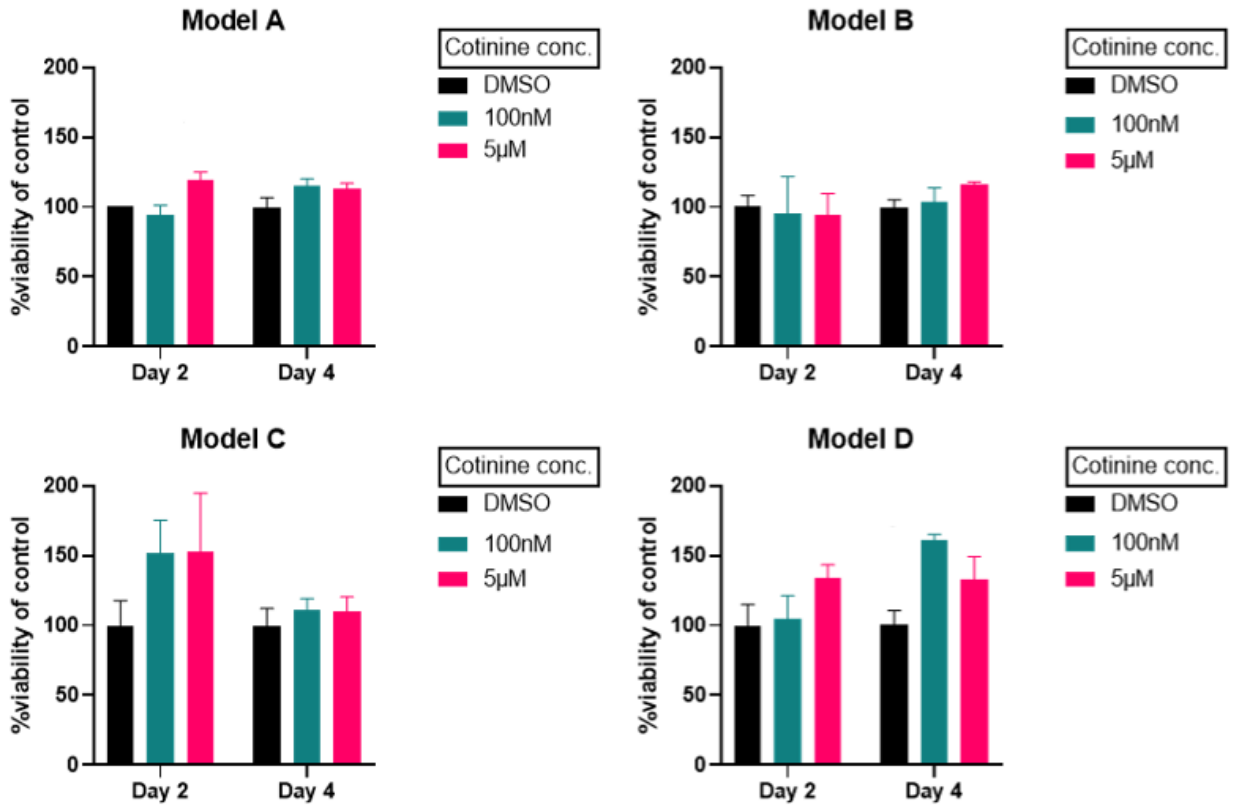


Figure 6.7. Cyquant assay measurement for each of the 4 cotinine exposure models. On day 2 or day 4 of each model, samples were taken from each well and frozen at -80°C. Cyquant was added to each sample, incubated overnight at 4°C and fluorescence monitored at 480nm excitation and 520nm emission. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

A total of 5 *TCF3-PBX1* and 11 *RUNX1-RUNX1T1* translocation events were observed across 96 replicates treated with cotinine, Table 6.3. The highest number of translocation events were observed in Model C, with a single *TCF3-PBX1* event observed across 6 replicates for each condition of cells treated with 100nM cotinine for 4 days and 5µM for 2 days, 2 *RUNX1-RUNX1T1* events were observed across 6 replicates for 100nM cotinine exposure for 2 days, and a single *RUNX1-RUNX1T1* event was observed across 6 replicates for 5µM cotinine at exposure of 2 and 4 days. A further 3 translocations were observed in Model D, with 2 *RUNX1-RUNX1T1* events across 6 replicates of 100nM cotinine exposure for 4 days plus a single *TCF3-PBX1* event across 6 replicates for the same exposure. Finally, a single *RUNX1-RUNX1T1* event was observed in Model A across 6 replicates of 100nM cotinine exposure for 2 days. Translocation events were also observed in DMSO controls for both *TCF3-PBX1*, in Model B, and *RUNX1-RUNX1T1*, in Model C, Table 6.3.

Using Fisher's exact test, no significance was observed for translocation events in response to DMSO vs any cotinine treatment (p=0.7805). When considering translocation events in response to single dose vs daily dose, a significant difference was observed (p= 0.0148),

however no significant difference was observed for treatment following 24 hours growth vs at seeding (p= 0.1837). When considering any cotinine exposure, no significant difference was observed for the frequency of *TCF3-PBX1* and *RUNX1-RUNX1T1* events using Fisher's exact test (p=0.3306).

Table 6.3. Model optimisation translocation events in response to cotinine. Each number represents the number of positive PCR bands seen during gel electrophoresis for 6 replicates of each condition. NALM6 cells were seeded into 2 identical well plates at a density of 0.5 x10^6 cells/ml in normal RPMI-1640 media and incubated at 37°C. Model A: 24 hours growth, single dose of cotinine or DMSO. Model B: single dose of cotinine or DMSO at seeding. Model C: 24 hours growth, daily dose of cotinine or DMSO. Model D: daily dose of cotinine or DMSO starting at seeding. Cotinine concentrations 100nM and 5µM were chosen to represent various physiologically relevant cotinine exposures, with the equivalent DMSO (0.1% v/v) used as a negative control. On day 2 and day 4 cells were also harvested for RNA extraction. The resultant RNA was treated with DNase and reverse transcribed using High-Capacity cDNA Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq with 10pmol primers. The data present shows 3 technical replicates and 2 biological replicates. (DMSO control data is common across chapters 4 and 5).

Translocation	Model type	Day cells harvested	DMSO Control	100nM Cotinine	5µM Cotinine
***TCF3-PBX1* (6 replicates)**	Model A	Day 2			
		Day 4			
	Model B	Day 2			
		Day 4	2		
	Model C	Day 2			1
		Day 4		1	
	Model D	Day 2			
		Day 4		1	
***RUNX1-RUNX1T1* (6 replicates)**	Model A	Day 2		1	
		Day 4			
	Model B	Day 2			
		Day 4			
	Model C	Day 2	1		1
		Day 4	3	2	1
	Model D	Day 2			
		Day 4		2	

As a confirmation of negative results, a selection of samples that did not contain visible bands during gel electrophoresis were used as a template for PCR using nested primers, and no amplification was observed. Sanger sequencing of the translocation breakpoints showed the expected fusion mRNA sequence, with the associated genes identified in BLAST, Table 6.4. Model D cotinine 100nM day 4 was not able to be sequenced due to low product concentration. Sequencing results for translocations observed with DMSO treatment be found in chapter 4, table 4.5.

*Table 6.4. Sanger sequencing of PCR positive translocation events identified in each Model. Positive samples were used as a template for nested PCR using HS taq polymerase and 10pmol nested primers. Samples were then purified for sequencing by DBS genomics. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence.*

Sample	Sequence of translocation junction	Genes identified in BLAST
Cotinine 5µM Model C Day 2 (*TCF3-PBX1* Nested)	CCTCCCGACTCCTACAGTGTTTT GAGTATCCGAG	*TCF3 PBX1*
Cotinine 100nM Model C Day 4 (*TCF3-PBX1* Nested)	CCTCCCGACTCCTACAGTGTTTT GAGTATCCGAG	*TCF3 PBX1*
Cotinine 5µM Model C Day 2 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAA*ATC GTACTGAG	*RUNX1 RUNX1T1*
Cotinine 5µM Model C Day 4 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAA*ATC GTACTGAG	*RUNX1 RUNX1T1*
Cotinine 100nM Model A Day 2 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAA*ATC GTACTGAG	*RUNX1 RUNX1T1*
Cotinine 100nM Model C Day 4 #1 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAA*ATC GTACTGAG	*RUNX1 RUNX1T1*
Cotinine 100nM Model C Day 4 #2 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAA*ATC GTACTGAG	*RUNX1 RUNX1T1*
Cotinine 100nM Model D Day 4 #1 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAACTC GTACTGAG	*RUNX1 RUNX1T1*
Cotinine 100nM Model D Day 4 #2 (*RUNX1-RUNX1T1* Nested)	GGGCCCCGAGAACCTCGAA*ATC GTACTGAG	*RUNX1 RUNX1T1*

6.4 Discussion

The results of physiologically relevant levels of cotinine exposure on NALM6 cell growth and viability showed no reduction in cell viability for any cotinine concentration compared with the DMSO control, Figures 6.3-4. This suggests these concentrations are suitable to be used for identifying chromosomal translocations, as the cells are capable of growth, and as such can go on to produce translocations that can be detected, although no reduction in survival could also suggest that the cells may not have obtained the damage necessary for producing translocations. The resazurin assay measurements of NALM6 cell viability show that the viability increases on exposure to cotinine compared to the DMSO control Figure 6.4. An *in vitro* study on human urothelial carcinoma cell lines also found that cotinine significantly increased cell proliferation using the WST-1 assay, although this was observed in cells exposed to concentrations of 100µM cotinine and above for 3 days, which is a much higher concentrations used in the current study [335]. As the current study only contains one biological replicate for the cotinine exposure gradient, it would be beneficial to repeat the experiment for further biological replicates to obtain a more accurate representation of cell viability at each physiologically relevant cotinine concentration, however the results here are in line with findings in the literature that cotinine exposure may increase proliferation.

Trypan Blue assays demonstrated that cell growth in Model A, single dose during exponential growth, Model B, single dose at seeding, and Model C, daily dose starting during exponential growth, increased every 24 hours, Figure 6.5. Cell growth in Model D, daily dose starting at seeding, is slow for the first 24 hours before increasing at 96 hours, yet cell count remains lower than in the other models, Figure 6.5. Model D conditions of daily dose of cotinine starting when cells are seeded, suggest that cotinine exposure initially slows cell growth, before dramatically increasing. The resazurin and Cyquant viability assays were generally in agreement, with no significant reductions in cell viability for cotinine exposure compared to the DMSO control for any model, Figures 6.6-7. All models show an increased cell viability for 5µM cotinine compared to 100nM cotinine, which is in line with results observed in a study showing increased proliferation of urothelial carcinoma cells exposed to higher concentrations of cotinine [335]. As has been discussed previously (chapter 4, section 4.4) observations for growth and viability in this study are likely to be limited due to a single measurement, but also affected by the differences in assay measurement techniques used.

In the preliminary experiments, a translocation was only observed for *RUNX1-RUNX1T1* at concentrations reflecting those of very high smokers, Table 6.1. Sanger sequencing of the PCR product confirmed that it was the expected *RUNX1-RUNX1T1* fusion product, with similar sequence pattern overlapping both genes across the translocation junction, Table 6.2. On the other hand, *TCF3-PBX1* translocations were observed in cells exposed to low levels of cotinine, 100nM and 10nM, associated with levels found in non-smokers and second-hand smoke exposure, Table 6.1. The PCR products from these positive events were confirmed to be the expected *TCF3-PBX1* fusion products, Table 6.2. The *RUNX1-RUNX1T1* translocation is found predominantly in childhood AML, which is much rarer than childhood ALL, and as such may require a higher level of exposure to induce translocations [3]. Epidemiological studies have found some association between smoking and specific ALL subtypes, however this was for *ETV6-RUNX1* which was not measured in the current study [176]. This suggests that there could be an association between level of exposure and childhood leukaemia subtype, however the difference was not statistically significant, and more replicates would be needed to understand a true relationship.

The highest frequency found of any of the translocations observed was 1 out of 3 biological replicates, making it difficult to discern any patterns of translocation induction. The 20 replicates of 5µM cotinine suggest translocations could be observed at a rate of 7 out of 20, with no translocations seen in the DMSO control. This aligns with observing a rate of 1 out of 3 translocations in the preliminary experiments. The number of positive translocations detected in the frequency experiment supports the preliminary results, suggesting that the translocation events observed were not random and at a rate of around 35%.

Considering the translocation frequency of 7/20 previously determined, 2 translocation events per 6 replicates were expected to be detected if exposure induced these events. In Model A, only 1 translocation was detected, suggesting that a single cotinine dose in the exponential phase is not the optimal condition for inducing translocations, Table 6.3. Model D observed 2 *RUNX1-RUNX1* translocations for 100nM cotinine at 4 days exposure, the expected frequency, with no other translocations observed, Table 6.3. Cell growth for Model D was slow in the first 72 hours, with daily doses of cotinine starting when NALM6 were seeded. This slow growth may be why translocation events are not seen at day 2 exposures. Model C had the highest number of translocations observed for cotinine exposed cells, a total of 6 events across both translocation types. Model C cells received a daily dose of cotinine or DMSO 24 hours after seeding, with the number of translocation events suggesting these conditions are the most likely to induce translocations. A total of 3 translocation events were observed for 5µM cotinine across all models, and a total of 7 events were seen for 100nM cotinine, which suggests that translocation induction in response to cotinine may be dose dependent. The preliminary results also suggested that translocation type may be dose dependent, with *TCF3-PBX1* events observed with lower concentrations and *RUNX1-RUNX1* events observed with higher concentrations of cotinine, however this pattern was not observed during model experiments, with no statistical difference in the type of translocation for cotinine exposed cells.

Cigarette smoking has well categorised carcinogens and known oxidative stress inducers, that are known to contribute towards cancer [336], but the impact of cotinine is less known. A study using the alkaline comet assay, used to detect DNA strand breaks, to assess cotinine exposure in neuroblastoma cells found significant differences in damage index and frequency when cells were exposed to 0.125mg/ml-2mg/ml cotinine for 3 hours [239]. The study also used a modified comet assay to evaluate oxidative damage and found increase damage in cells exposed to 0.125mg/ml-2mg/ml cotinine for 3 hours compared to control, but not at significant levels. Furthermore, using nasopharyngeal normal and carcinoma cells it has been demonstrated that oxidative stress led to cleavage of *AF9* gene, which is associated with leukaemia [337]. This suggests that cotinine may induce DSBs through oxidative damage, and as such could lead to translocation formation, in line with our preliminary findings. As discussed in section 4.4, the translocation events in DMSO control cells, however, suggest that DMSO may be contributing to the induction of translocations. As such, it would be beneficial to repeat these experiments using a different solvent for cotinine, such as ethanol or dimethyl formamide.

The preliminary data in this chapter suggests that cotinine may induce translocations, but subsequent data are unclear, and more evidence is required to confirm and understand the

relationship of cotinine exposure and translocation induction, such as the mechanism of cotinine on double strand break induction and DDR.

7 Influence of folate on translocation induction in NALM6 cells

7.1 Introduction

Folates are B-vitamins found throughout the diet in foods such as leafy green vegetables. The synthetic form of folate monoglutamate, folic acid, is used as a folate supplement and is used to fortify many food products across the world [338]. Folate plays an important role in DNA replication due to its role in one-carbon metabolism, required for the synthesis of purines and pyrimidines, and for methylation reactions [95]. When cells are folate deficient, the conversion of dUMP to dTMP can be impaired, leading to dUMP misincorporation during DNA synthesis, which can lead to DSBs [97, 98]. Folate status can also lead to altered methylation, which has an impact on many cellular processes [90].

Folate deficiency has been associated with an increased risk of cancers, Down's syndrome and adverse pregnancy outcomes and birth anomalies [90]. Supplementation with folic acid in pregnant women has been shown to reduce the formation of NTDs in offspring [105]. The recommended daily intake for pregnant women is 400µg/day [104]. Although some countries, including the USA and more recently the UK, have mandatory fortification programmes, folic acid fortification has been rejected by many countries due to the increased risk of colorectal cancer and progression of certain cancers, along with other adverse health outcomes due to high intake of folic acid [339]. High folic acid intake could lead to increased availability of nucleotides for proliferating cancers. The absorption of folic acid and metabolism into the active form is limited to >200–400 µg, after which unmetabolized folic acid is found in the serum, with could lead to accumulated high levels of folic acid [339]. It is important therefore, to understand the optimal concentration at which folate provides protection without adverse effects from high folate intake.

In vitro studies of fibroblast cells grown in folate deficient conditions have observed replication stress, with decreased fork progression rate, increased DSB formation and DNA repair [340]. *In vivo* studies using mouse prostate cells have observed DNA strand breaks and chromosomal rearrangements in folate deficient conditions [121]. As discussed in section 1.4, previous studies support the hypothesis that folate deficient conditions can make DNA more susceptible to instability and mutations, but to the best of our knowledge, no *in vitro* studies have observed the induction of childhood leukaemia translocations in response to folate deficient conditions.

This chapter describes the use of an *in vitro* model to investigate the influence of physiologically relevant folate concentrations on the induction of childhood leukaemia associated chromosomal translocations *TCF3-PBX1* and *RUNX1-RUNX1T1.* To achieve this, NALM6, an acute leukaemia cell line, was grown in a gradient of folate concentrations,

equivalent to those found in human serum, followed by measurements of cell growth, viability, and chromosomal translocation induction.

7.2 Methods

7.2.1 Experimental design of preliminary tests to examine the effects of folate levels on translocation events

To understand the role folate may play in translocation induction, NALM6 cells were exposed to a range of folic acid concentrations, chosen to mimic different physiological levels of folate status, Figure 7.1. Short term folate status is measured through serum folate, with red blood cell folate a reflection of long-term folate intake [341]. Serum folate levels were considered to reflect folate concentration in cell medium. As a control, 2µM folic acid was chosen as a supraphysiological level to reflect normal RMPI-1640 cell culture media [121, 342]. A normal human serum folate range is considered between 10nM and 50nM, with those boundary concentrations chosen for this study to represent a low and high “normal” folate level respectively [107, 343-345]. To represent a low serum folate status, 5nM folic acid was chosen, with 1nM folic acid considered folate depleted and 0.1nM folic acid considered folate deficient [345]. A high and very high physiological folate status was also investigated with folic acid concentrations of 100nM and 200nM, chosen to reflect high levels observed in colorectal cancer associated with increased proliferation and genetic instability [107, 121, 346].

An overview of the experiment can be found in Section 2.6.1, with NALM6 cells washed and resuspended in folic acid free RPMI-1640 media to remove any existing folic acid. RNA was extracted on day 4 which was selected as previous studies suggest it is sufficient to deplete endogenous folate whist maintaining cell viability [347].

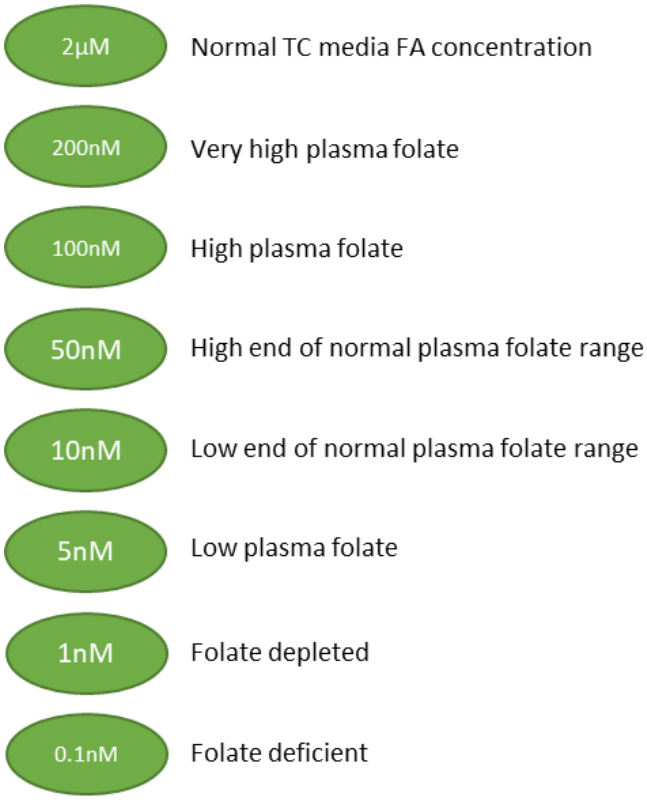


Figure 7.1. Folic acid concentrations used in preliminary experiments and their corresponding physiological range.

7.2.2 Design of experiments to investigate the frequency of translocation events in response to folate levels

To investigate the frequency at which translocations may be expected to appear in NALM6 cells grown in different folic acid concentrations, the preliminary experiment was repeated for 10 technical replicates using 1nM folic acid for 4 days. RNA was extracted from all 10 folic acid replicates and reverse transcribed into cDNA for use as a template in *TCF3-PBX1* translocation PCR assays, using HS Taq polymerase with 2.5pmol forward and 5pmol reverse primers for 40 cycles. Reported DMSO controls (section 4.2.2) are the same across all chapters 4-6.

7.2.3 Design of experiments to investigate translocation events in response to folate levels in NALM6 cells for different models

As discussed in Section 2.6.2, to optimise the preliminary NALM6 cell model, two variables were considered: i) the impact of treating cells either immediately or 24 hours after initial seeding, and ii) single vs daily dosage of folic acid. As in preliminary experiments, NALM6 cells were washed and resuspended in folic acid free RPMI-1640 media before first dose of

folic acid was added and RNA extraction and cell viability assays were performed on day 4 only. Folic acid concentrations used to treat cells are outlined in figure 7.2.

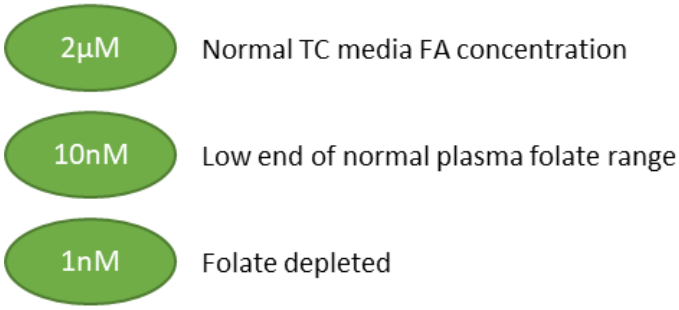


Figure 7.2. Folic acid concentrations used in Model's A-D and their corresponding physiological range.

7.2.4 Statistical analysis

Described in section 4.2.4.

7.3 Results

7.3.1 Preliminary experiments to investigate the influence of folate levels on translocation events

Standard RPMI-1640 media has supraphysiological levels of folic acid for normal cell growth and treatment. To induce and detect translocations, cells need to still be able to proliferate under folic acid deficient conditions. For all folic acid concentrations, NALM6 cells have increasing growth over 96 hours, although viable cell count at 96 hours for folate deficient cells (0.1nM, 1nM and 5nM folate) is less than the 2µM control, Figure 7.3.

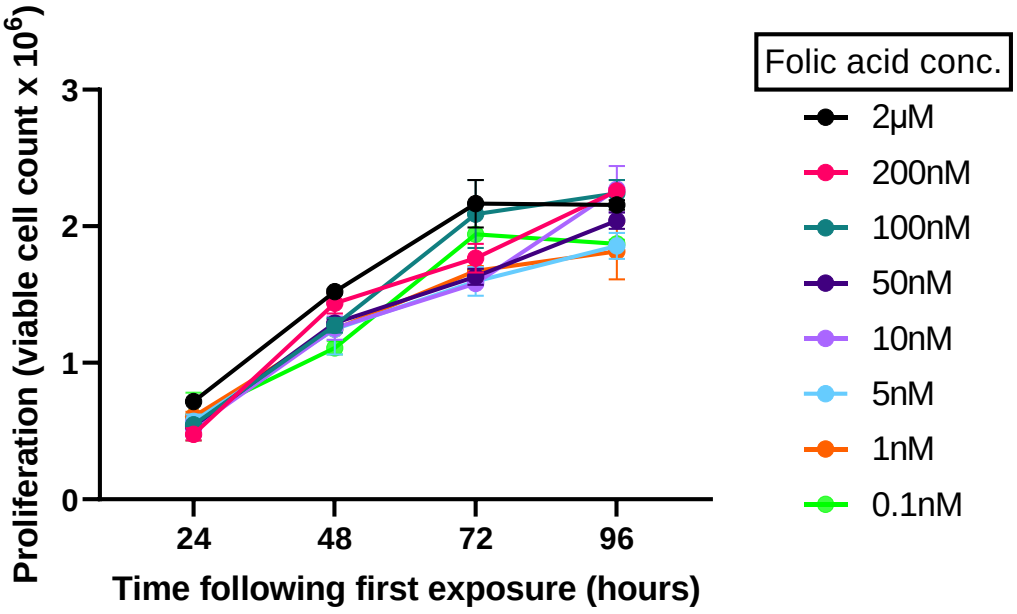


Figure 7.3. Trypan blue measurement of cell growth in NALM6 cells grown in varying folic acid. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml in folic acid free RPMI-1640 media. A single dose of folic acid was added and then cells were incubated at 37°C for up to 96 hours. Trypan blue counts occurred every 24 hours. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

The difference in viability between the 2µM control and decreasing folic acid concentrations was measured by the metabolic colorimetric resazurin assay. No decrease in viability was found for cells grown in different folic acid concentrations for 2 days, Figure 7.4. After 4 days growth, a decrease in viability compared to the 2µM control was observed for 0.1nM and 50nM folic acid, Figure 7.4.

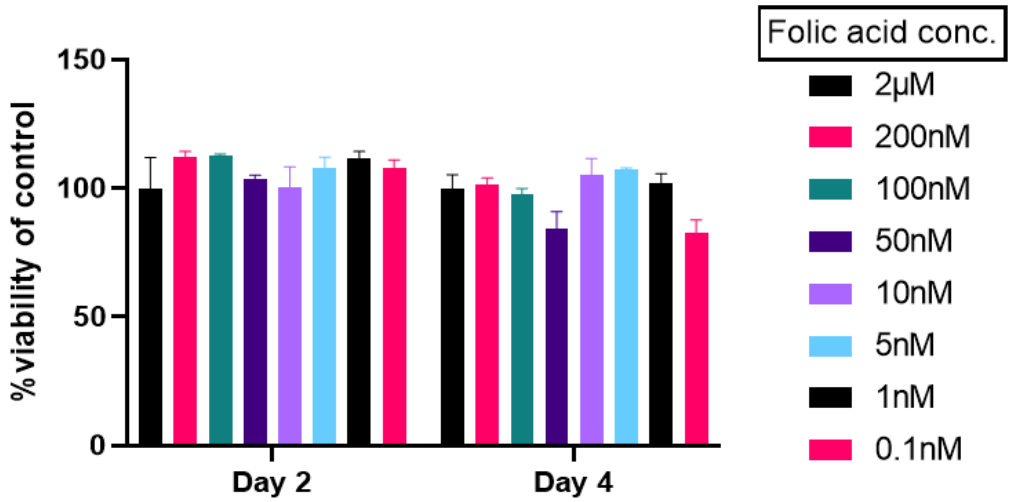


Figure 7.4. Resazurin assay measurement of cell viability of NALM6 cells grown in varying folic acid concentrations at day 2 and day 4. NALM6 cells were seeded into 2 well plates at a density of 0.5 x10^6 cells/ml in folic acid free RPMI-1640 media. A single dose of folic acid was added into the wells, and then they were incubated at 37°C for 48 hours (plate 1) or 96 hours (plate 2). On day 2 or day 4, resazurin was added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Media blanks were measured for each folic acid concentration. Results are reported as a percentage of the DMSO control + standard error.

A single *TCF3-PBX1* event was observed for 1/3 replicates for each condition of cells grown in 10nM and 1nM folic acid, Table 7.1. A single *RUNX1-RUNX1T1* event out of 3 replicates was observed at 100nM folic acid, Table 7.1. No translocation events were observed in the 2μM folic acid control cells, Table 7.1.

Table 7.1. Translocation events observed for NALM6 cells grown in varying folic acid concentrations. NALM6 cells were washed and resuspended in folic acid free RPMI-1640 media at a density of 0.5 x10^6 cells/ml in 12-well plates. A single dose of folic acid was added into the wells, and then incubated at 37°C for 96 hours. Cells were harvested for RNA extraction and treated with DNase then reverse transcribed using Precision nanoScript2 Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq and 10pmol primers. The data present shows 3 biological replicates.

Translocation	**2μM Folic acid Control**	**200nM Folic acid**	**100nM Folic acid**	**50nM Folic acid**	**10nM Folic acid**	**5nM Folic acid**	**1nM Folic acid**	**0.1nM Folic acid**
TCF3-PBX1 (3 replicates)					1		1	
RUNX1-RUNX1T1 (3 replicates)			1					

7.3.2 Frequency of translocation events in response to folate levels

As translocations are incredibly rare events it is difficult to predict how often they may be likely to appear if associated with a specific exposure. Although the results for preliminary investigation into various physiologically relevant levels of folate provide an indication that folate levels may influence these events, they do not provide a clear pattern of translocation

induction in NALM6 cells due to the limited number of biological replicates. To investigate the frequency of which translocations may be expected to be observed in NALM6 cells in response to folate, the preliminary experiment was repeated with 10 technical replicates using 1nM folic acid for 4 days. This concentration was chosen where a translocation event was observed in preliminary experiments. A total of 7 *TCF3-PBX1* events were observed across 10 replicates for cells grown in 1nM folic acid for 4 days. The DMSO control, whereby NALM6 cells were grown in standard RPMI-1640 media for 4 days, reported a total of 2 translocation events across 20 replicates. Analysis by two-tailed Fisher's exact test suggests there is a relationship between folate levels and *TCF3-PBX1* translocations ($p = 0.0017$).

7.3.3 Experiments to optimise a NALM6 cell line model to investigate translocation events in response to folate levels

To replicate the preliminary findings where translocations were observed in response to folate levels and to address the issues of daily dosing and timing of those doses, further experiments were conducted using NALM6 cells grown in various folic acid concentrations across 4 different models.

To understand how each model type will affect NALM6 growth a trypan blue count was performed every 24 hours following the first dose of folic acid, Figure 7.5. Model A (single exposure, 24 hours after seeding) and C (daily dose, 24 hours after seeding) show a similar pattern of growth, with cell growth increasing between 24 and 72 hours, Figure 7.5. At 96 hours, the viable cell count for the 2µM control is higher than the other folic acid concentrations. Model B (single dose at seeding) and D (daily dose at seeding) have a slow growth between 24 and 96 hours, with no increase in 2µM control growth compared to the folic acid deficient conditions in Model D, Figure 7.5.

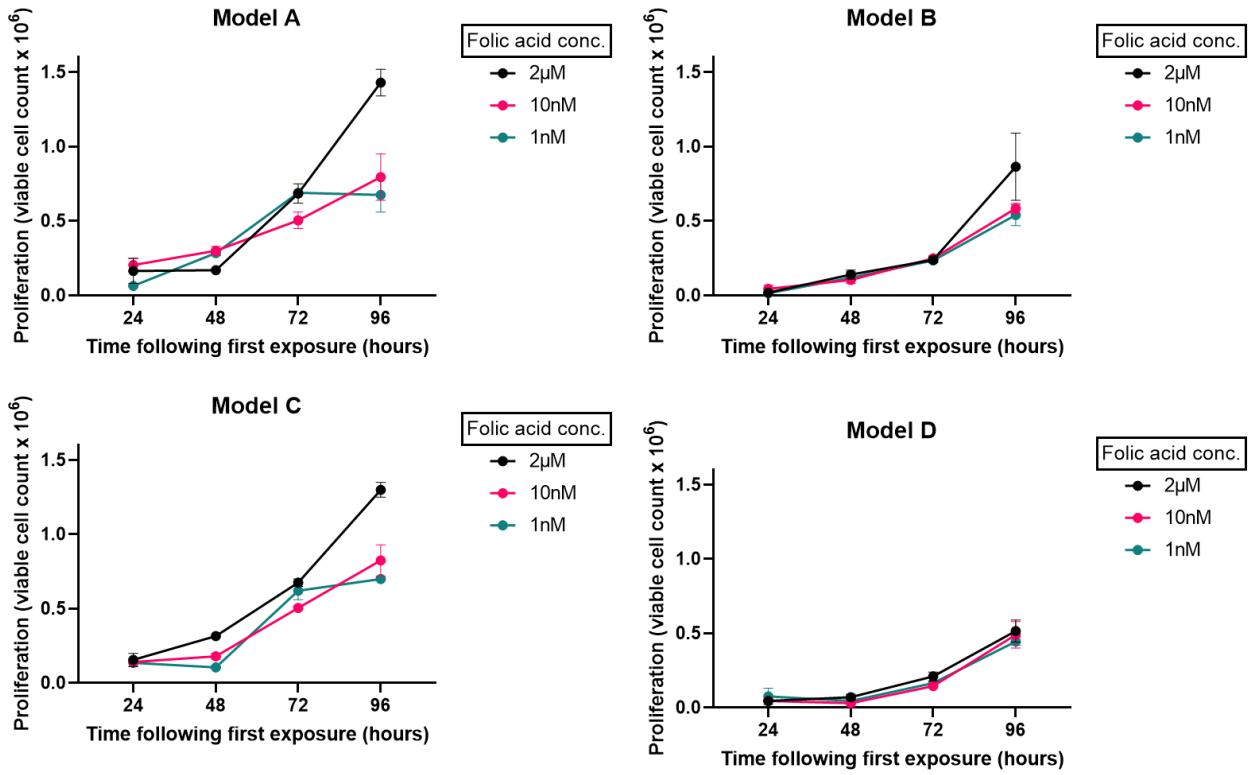


Figure 7.5. Trypan Blue Assay for each of the 4 folic acid deficient models. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml and incubated at 37°C. Folic acid concentrations 10nM and 1nM were chosen to represent various physiologically relevant folate deficiencies, with 2μM folic acid used as a negative control. Trypan blue counts occurred every 24 hours following the first dose of folic acid. The data shows the mean of 2 technical replicates for 1 biological replicate + standard error.

To understand the impact of each model on NALM6 cell viability a resazurin assay and Cyquant assay were performed on day 4. The resazurin assay shows a reduction in viability compared to control for all folic acid deficient conditions in each model, Figure 7.6.

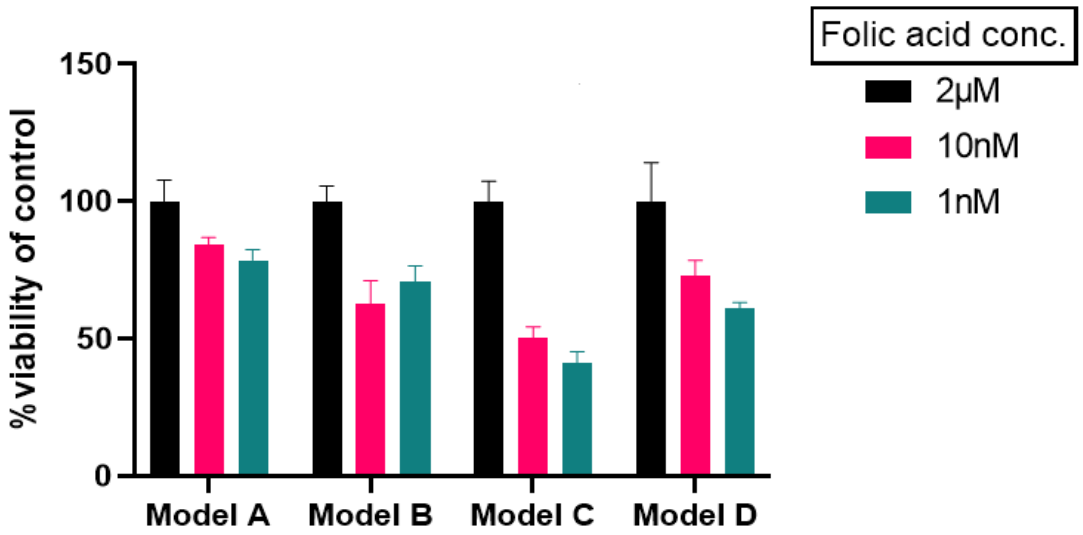


Figure 7.6. Resazurin assay measurements for each of the 4 folate deficient models. On day 4 of each model, samples were taken from each well and resazurin added to a final concentration of 0.02%, incubated for 4 hours at 37°C, and fluorescence monitored at an excitation wavelength of 560nm and an emission wavelength of 590nm. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the 2µM control + standard error.

The Cyquant viability analysis showed a reduction in cell viability compared to control for all folic acid deficient conditions across all models, Figure 7.7.

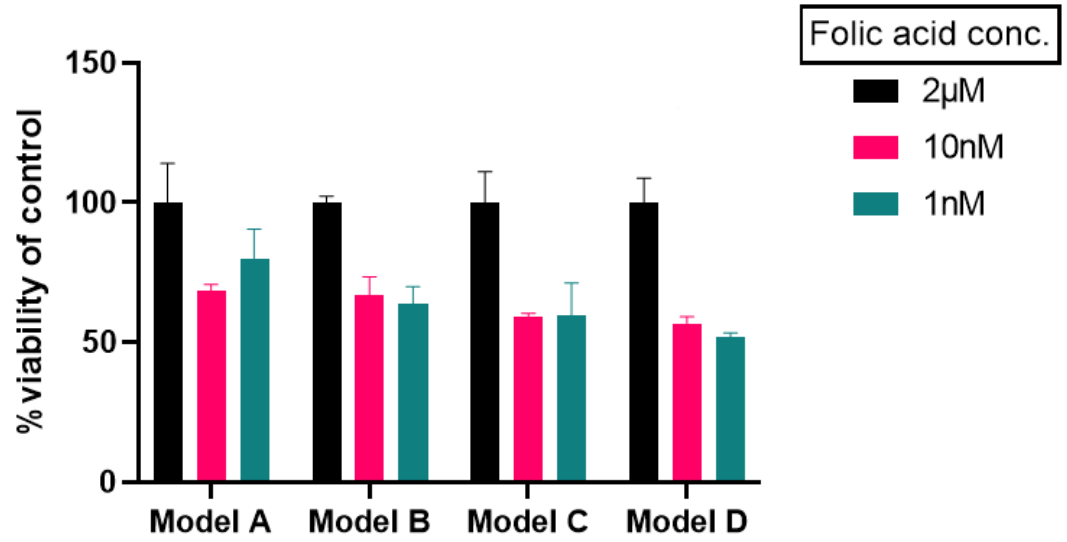


Figure 7.7. Cyquant assay measurement for each of the 4 folate deficient models. On day 4 of each model, samples were taken from each well and frozen at -80°C. Cyquant was added to each sample, incubated overnight at 4°C and fluorescence monitored at 480nm excitation and 520nm emission. The data presented represents 3 technical replicates for 1 biological replicate. Results are reported as a percentage of the DMSO control + standard error.

A total of 4 translocation events were observed across the 4 models, Table 7.2. Model A had a single *RUNX1-RUNX1T1* event across 6 replicates with cells grown in 1nM folic acid, Table 7.2. The remaining translocation events were observed in Model C, with a single *TCF3-PBX1* event across 6 replicates for cells grown in 10nM folic acid, and a single *RUNX1-RUNX1T1* event observed in 1/6 replicates for each condition of cells grown in 1nM

folic acid, and unexpectedly, the 2μM folic acid control, Table 7.2. Analysis by Fisher's exact tests observed no significant difference between the number of translocations in response to the folate control vs folate deficient conditions (p=1). No significant difference was observed between single dose vs daily dose (p= 0.6142) and for treatment following 24 hours growth vs at seeding (p= 0.1145) using Fisher's exact test.

Table 7.2. Model optimisation translocation events in response to folate levels. Each number represents the number of positive PCR bands seen during gel electrophoresis for 6 replicates of each condition. NALM6 cells were seeded at a density of 0.5 x10^6 cells/ml and incubated at 37°C. Model A: 24 hours growth, single dose of folic acid. Model B: single dose of folic acid at seeding. Model C: 24 hours growth, daily dose of folic acid or DMSO. Model D: daily dose of folic acid or DMSO starting at seeding. Folic acid concentrations 10nM and 1nM were chosen to represent folate deficient conditions, with standard medium folic acid concentration 2μM used as a control for the first exposure and followed by the equivalent DMSO (0.1% v/v) for any following dose in the control. On day 4 cells were harvested for RNA extraction. The resultant RNA was treated with DNase and reverse transcribed using High-Capacity cDNA Reverse Transcription Kit to use as a template in RT-PCR assays for TCF3-RUNX1 and RUNX1-RUNX1T1 translocations, using HS Taq with 10pmol primers. The data present shows 3 technical replicates and 2 biological replicates.

Translocation	Model type	2μM Folic acid Control	10nM Folic acid	1nM Folic acid
TCF3-PBX1 **(6 replicates)**	Model A			
	Model B			
	Model C		1	
	Model D			
RUNX1-RUNX1T1 **(6 replicates)**	Model A			1
	Model B			
	Model C	1		1
	Model D			

To confirm negative PCR results, a selection of negative samples were used as templates for a nested PCR, with no amplification observed. Following Sanger sequencing of purified positive PCR products using nested primers, translocation breakpoints were identified showing the expected fusion mRNA sequence, with the associated genes identified in BLAST, Table 7.3.

*Table 7.3. Sanger sequencing of PCR positive translocation events identified in each folate deficient model. Positive samples were used as a template for PCR using 10pmol of nested primers, and then purified for sequencing by DBS genomics. Sequences were entered into BLAST to identify gene matches, and translocation junction between 2 genes was identified. Underlined letters represent bases that were found to overlap in each gene identified. Letters with a * indicate that the base was correct in the RUNX1-RUNX1T1 fusion sequence but different in the original RUNX1T1 mRNA sequence.*

Sample	Sequence of translocation junction	Genes identified in BLAST
Folic acid 10nM Model C (*TCF3-PBX1*)	CCCGACTCCTACAGTGTTTT GAGTATCCTAGG	*TCF3 PBX1*
Folic acid 1nM Model A (*RUNX1-RUNX1T1*)	CCCCGAGAACCTCGAA*ATC GTACTGAGAAGCACT	*RUNX1 RUNX1T1*
Folic acid 1nM Model C (*RUNX1-RUNX1T1*)	CCCCGAGAACCTCGAA*ATC GTACTGAGAAGCACT	*RUNX1 RUNX1T1*
Folic acid 2μM Model C (*RUNX1-RUNX1T1*)	CCCCGTGAACCTGGAAGTG GCATCCTCCAGATCGTACT GAGAAGCACT	*RUNX1 RUNX1T1*

7.4 Discussion

Folate deficiency in humans can be monitored by serum folate concentrations with a normal physiological range between 10nM and 50nM [345]. Standard RPMI-1640 media has a supraphysiological concentration of folic acid added (2μM) to provide sufficient levels for growth of cultured cells [342]. To ensure that NALM6 cells grown in physiologically relevant folate deficient conditions were still able to proliferate under standard tissue culture conditions, cell growth and viability were measured by trypan blue, resazurin and Cyquant assays. Preliminary experiments show that NALM6 cells grown in folic acid concentrations between 200nM and 0.1nM were able to grow exponentially across 96 hours, following a similar pattern as the 2μM control, with cells grown in 5nM-0.1nM folic acid showing a decreased cell count compared to control, Figure 7.3. For the same concentrations cell viability did not decrease after 2 days growth, however some reductions in viability were observed after 4 days growth at 50nM and 0.1nM, Figure 7.4.

The reduction in NALM6 cell growth and viability at severely deficient folic acid concentrations was observed more prominently when multiple models were tested. Model A (24 hour growth, single dose) and Model C (24 hour growth, daily dose) had a similar growth pattern with cell count for all folic acid concentrations increasing over the first 72 hours, with a sharp increase in the 2μM control at 96 hours Figure 7.5. As this increase is not seen for the folic acid deficient cells at 96 hours, it suggests that folic acid deficiency is impacting cell growth at this point. The resazurin and Cyquant viability assays for Model A and C both show reduced viability in the folic acid deficient cells compared to the 2μM control on day 4, with significant decreases in Model C, Figures 7.6-7. In Model B (single dose at seeding), cell growth is slow for the first 72 hours, with cell count increasing at 96 hours for the 2μM control, Figure 5. The resazurin and Cyquant assays for Model B on day 4 both show reduction in cell viability for folic acid deficient cells compared to the 2μM control, Figures 7.6-7. In Model D (daily dose starting at seeding), cell growth is slow for all folic acid concentrations across the 96 hours. The resazurin and Cyquant viability assays both show a reduction in viability for folic acid deficient cells compared to the 2μM control.

As Model's A and C have an extra 24 hour growth period in standard medium to proliferate before being transferred to folate deficient medium, the viable cell count is, as expected, higher than Model's B and D, which are seeded directly into folate deficient medium, and as such have a shorter period of time to grow. This also suggests that Model A and C cells are already actively proliferating in the exponential phase when they are introduced into the new folic acid deficient medium and as such the change in environment does not appear to negatively affect growth for at least 72 hours. The change into a folic acid deficient environment for Model B and D appears to slow the rate of growth for at least 72 hours, with

Model B 2µM control recovering with increased growth at 96 hours. Model D, however, with daily doses of folic acid/DMSO, does not appear to recover, which suggests that the daily dose of folic acid/DMSO in the control cells may be affecting growth rate. All folic acid concentrations used are diluted in a vehicle of DMSO, added to cells at concentration of 0.1% (v/v). To eliminate the potential effect of DMSO, the experiments could be repeated with the folic acid free RPMI-1640 media supplemented with standard RPMI-1640 media to the desired folic acid concentration. An *in vitro* study which measured proliferation and viability of human placental choriocarcinoma (JEG-3) cells grown in physiologically relevant folate deficient conditions at 2nM (low), 20nM (normal) and 100nM (supplemented) folic acid, found that low levels of folic acid had significantly decreased proliferation and viability compared to the normal and supplemented cells [348]. This supports the findings observed in preliminary and model experiments, with a reduction in cell viability for low folate levels (1nM and below). As cell viability is reduced to 50% in many cases, it may make it more difficult for cells with translocations events to proliferate enough to allow for detection, which may be a reason as to why a low number of translocation events are observed compared to previous exposure chapters (4, 5 and 6). A higher number of replicates or cell density may be needed to increase the likelihood of detecting translocation events at low folate levels.

As has been discussed previously (section 4.4) observations for growth and viability in this study are likely to be limited due to a single measurement, but also affected by the differences in assay measurement techniques used. For instance, the reduction of resazurin requires electron donations from molecules such as NADPH and FADH [349]. The folate one-carbon metabolism pathway is known to produce NADPH, so deficiencies could lead to altered NADPH activity and affect the rate of resazurin reduction [350]. Due to this, using the Cyquant assay alone would be sufficient to measure cell viability in folate deficient cells.

The preliminary investigation into physiologically relevant folate levels observed translocation events for *TCF3-PBX1* at 10nM and 1nM folic acid, which are considered the low boundary of normal serum folate levels or folate depleted, respectively Table 7.1. A single *RUNX1-RUNX1T1* translocation event was observed at 100nM folic acid, considered a physiologically high folate status, Table 7.1. No translocation events were observed in the 2µM control during preliminary experiments. The preliminary results suggest that there may be some differences between folate status and types of translocations, however the low number of events make it difficult to determine a true pattern. As standard RPMI-1640 media is at a supraphysiological concentration, cells grown in folic acid deficient medium may have a more acute response to folate deficiency, and so translocations at 100nM folic acid could still be considered folate deficient in tissue culture settings compared to in human serum

folate, with some *in vitro* studies considering 100nM folic acid as mild folate depletion, with DNA damage observed [121].

Experiments investigating the frequency of *TCF3-PBX1* translocation events in NALM6 cells grown in 1nM folic acid suggest that translocation events can be reproduced and are expected to be observed 70% of the time, Table 7.2. This rate was much higher than translocation events seen in the preliminary experiment, with the highest rate of translocation events at 1 in 3 replicates (33%). DMSO control cells, although grown in standard medium with DMSO added, also observed translocation events in *TCF3-PBX1* at a rate of 10%, however Fisher's exact tests showed that the difference in translocation appearance was still significant between the groups.

During NALM6 cell model optimisation, translocation events were only observed in Model's A and C. No translocations were seen in Model B, which were grown in the same conditions as the preliminary and frequency experiments, where translocations were detected up to rates of 70%. For each model concentration, a total of 3 biological replicates and 2 technical replicates were performed giving 6 replicates overall. Considering the translocation frequency previously detected, 4 translocations per 6 replicates would be expected, however the highest frequency observed in any model was 1 out of 6 replicates (17%). Model A and C had an extra 24 hour growth period in standard medium which allowed for active growth and resulted in a higher cell count than the other models, which may have provided more opportunity for DSBs and repairs to form translocations. Model C observed the highest number of translocation events overall, including an event in the *RUNX1-RUNX1T1* 2µM control.

There are 2 main routes through which folate deficiency is likely to contribute to chromosomal translocation induction: reduced nucleotide biosynthesis and altered methylation. The folate cycle is important in the synthesis of both purines and pyrimidines, which are required for effective DNA replication and repair [95]. If dUMP is not converted to dTMP due to reduced folate metabolism, then excess uracil can build up and misincorporation into DNA can occur [95, 351]. The uracil is then removed through identification by uracil-DNA glycosylase which initiates base excision repair, leaving a single stand break that can lead to DSBs if uracil misincorporation frequency is high [95, 119, 351, 352]. Uracil misincorporation and subsequent DNA damage has been observed in cell lines, animal models and humans [97, 98, 119, 120]. Reduction in nucleotide availability through folate deficiency has also been shown to compromise DNA repair activity [98]. One carbon metabolism provides methyl donors for a range of cellular processes, including the epigenome. DNA methylation is important in regulating genes involved in DNA repair,

chromosome stability, and as such, altered methylation could contribute to translocation induction [77]. The translocation events observed in preliminary experiments support the hypothesis that folate deficient conditions may induce translocations.

However, a *RUNX1-RUNX1T1* translocation event was observed in the control cells for Model C, where a single dose of 2μM folic acid (dissolved in DMSO 0.1% v/v) was administered followed by further daily doses of DMSO (0.1% v/v) alone. As discussed in chapter 4, this suggests that DMSO may be contributing to the induction of translocations in NALM6 cells.

Repeating the folate deficient conditions without the presence of DMSO will allow further investigation of DMSO influence on translocation induction.

Further consideration should also be given to the potential contamination with excess folate in this model, such as through the foetal bovine serum used or through incomplete removal of folic acid from the medium during washes, which may make folic acid levels higher than expected. Furthermore, endogenous folate may not be depleted in NALM6 cells within 96 hours as expected. This may be why there is not a substantial effect observed on cell growth and viability with folic acid concentrations in the preliminary experiments, and why fewer translocations are observed overall in comparison to other exposures in chapters 4, 5 and 6. To optimise the folic acid exposure in this model, the endogenous folic acid status of the cells, and the exogenous folic acid status of the cell medium should be measured every 24 hours. This could be performed via U-HPLC or radio assays [353, 354].

The preliminary data in this chapter suggests that NALM6 cells grown in varying folic acid conditions may induce translocations, however subsequent data from frequency and model optimisation experiments are unclear and more evidence is required to increase inference to suggest that folate concentrations directly influence translocation events.

8 Exploring the potential role of environmentally-associated DNA methylation to contribute to risk of childhood acute lymphoblastic leukaemia: a subtype analysis.

8.1 Introduction

Acute lymphoblastic leukaemia (ALL) accounts for 25% of childhood cancers, with around 450 cases per year in the UK [1, 355]. The vast majority of childhood leukaemia development is thought to follow the "two-hit" hypothesis, where an initiating event, such as a chromosomal translocation occurs, but alone is not sufficient for disease progression and requires further events, such as altered DNA methylation, to advance [55]. Using neonatal blood spots taken a few days after birth, the *ETV6-RUNX1* chromosomal translocation has been identified at rates of 1-5% in the general population, which is 100-500 fold higher than leukaemia incidence [10, 56, 58]. This supports the two-hit hypothesis in that further spontaneous or environmentally induced events are required following initiating events for leukaemia onset.

Although childhood leukaemia survival rates are increasing, the incidence of childhood leukaemia is also increasing year on year [82]. Epidemiological studies have identified numerous environmental exposures that are associated with increased risk of childhood leukemia [14]. These include *in utero* exposures, such as parental smoking, maternal alcohol intake, maternal caffeine intake, maternal folate levels and radiation exposure, and post-natal exposures, such as nursery attendance and infection history [14]. The strength of associations for each risk exposure is varied and sometimes weak, which may be attributable to a lack of available data due to the rarity of childhood leukaemia and difficulty in accurately measuring/reporting environmental exposures retrospectively in case-control studies. Furthermore, the evidence for environmental contribution to different leukaemia cytogenetic subtypes is extremely limited, due to the inability to be able to stratify cases to detect robust associations between exposures and subtypes because of a lack of statistical power [356]. Therefore, the use of an intermediate biomarker linking risk exposure and leukaemia, including for cytogenetic subtypes, may provide stronger evidence of associations and the role risk exposures play in leukemogenesis.

The epigenetic landscape is a key factor in human health and disease, with modifications, such as DNA methylation, vital for regulation of gene expression [76, 77]. DNA methylation, the transfer of a methyl group onto cytosine nucleotides that immediately precede guanine, known as CpG sites, plays an important role in the developing embryo [73, 74]. In the early stages of development, the embryo undergoes global demethylation, allowing most genes to become active, with only some key paternally imprinted markers maintaining methylation throughout [357]. This demethylation is followed by a strict pattern of re-methylation throughout gestation to deactivate specific genes, in order to allow for tissue differentiation

during development. Early development is therefore a critical period for the establishment of DNA methylation, with any aberrations likely to influence gene expression and potential programming of offspring health in later life [357, 358]. Indeed, aberrant DNA methylation is associated with many adverse health outcomes including cancer and has been observed in childhood ALL [78, 80]. Altered methylation patterns have also been observed in response to environmental exposures [125, 359, 360]. Environmental exposure, especially in early life, given this is a critical time for epigenetic marks, is likely to play a key role in influencing DNA methylation patterns, suggesting that epigenetics are the interface between the environment and the genome. The mechanisms by which environmental exposures may contribute to leukemogenesis remain unclear, however as methylation is both affected by environment and is altered in leukaemia, it is plausible that DNA methylation may be a mediating factor.

Indeed previously, DNA methylation has been exploited as an intermediated biomarker using a “meet in the middle” approach, to provide evidence to support the hypothesis that DNA methylation is a mediating mechanism by which environmental exposure contributes to leukaemia [125]. This study compared altered DNA methylation patterns at gene loci in response to exposures associated with childhood leukaemia and in childhood leukaemia itself to identify common loci with altered methylation. For exposures expected to be associated with increased risk of childhood leukaemia, the direction of methylation, i.e. hypo or hyper methylation, was expected to be in the same direction in both the exposure and in ALL. Conversely, for those exposures expected to be preventative, the direction of methylation in the exposure was expected to be opposing to those observed in ALL. In the study, directionally concordant overlapping gene methylation was observed, and found not likely to be due to chance, between methylation in ALL and in response to maternal medical radiation exposure, maternal alcohol intake, maternal sugary caffeinated drinks intake, and nursery attendance at 8 months [125]. Therefore, for the current study, the use of this methodology was proposed to a) update the analysis for CpGs compared to previous gene level analysis and b) investigate specific ALL subtypes. By using a CpG level approach the analysis is more specific, pin-pointing the exact location of methylation changes. As there is limited epidemiological evidence for the relationship between environmental exposures and ALL subtypes due to small sample numbers available, the use of an intermediated biomarker, such as DNA methylation during this analysis, may help to understand if particular subtypes might link to specific exposures.

Altered DNA methylation patterns have been identified for 10 cytogenetic subtypes of ALL: T-ALL (T-cell lineage) and B-cell precursor (BCP)-ALL subtypes high hyperdiploidy (HeH), t(12;21) *ETV6-RUNX1*, t(1;19) *TCF3(E2A)-PBX1*, t(9;22) *BCR-ABL1*, dic(9;20), iAMP21, 11q23/*MLL(KMT2A*), undefined and non-recurrent subtypes [361]. Using the previously

established “meet in the middle” approach, differentially methylated CpGs (DMCs) associated with both individual ALL subtypes and environmental risk exposures were compared for significant overlaps, both overall and considering directionality of methylation. GO enrichment and KEGG pathway analysis were employed to investigate underlying mechanisms of environmental exposure in ALL subtypes.

8.2 Methods

8.2.1 Identification of CpGs with altered methylation in response to ALL-associated environmental risk factors

Previous literature has used epigenome wide association studies (EWAS) to identify DMCs that are associated with ALL environmental risk factors [125, 359, 360]. A study by Timms *et al.* 2019 used a subset of samples from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort where 1018 paired child and mother samples were analysed for genome-wide DNA methylation at birth and age 7, known as the Accessible Resource for Integrated Epigenomic Studies (ARIES) sub-population [125]. Environmental exposures were assessed using data from questionnaires carried out during the ALSPAC study. Methylation was measured in cord blood to assess methylation in relation to *in utero* environmental exposures, which include smoking, alcohol intake, sugary caffeinated drinks, radiation exposure, coffee intake and folic acid supplementation. Methylation was then measured in age 7 blood samples and used to assess the post-natal exposures nursery attendance and reported cold. Methylation data was determined by the Illumina Infinium® HumanMethylation450k BeadChip assay and normalised before multiple linear regression analysis to investigate associations between individual CpG sites and environmental exposures [125].

In addition to using data from the Timms *et al.* 2019 EWAS of maternal smoking, the most extensive meta-analysis of maternal smoking, conducted by Joubert *et al* was also assessed [360]. The meta-analysis incorporated 13 cohorts from the Pregnancy and Childhood Epigenetics (PACE) consortium, with a total of 6685 newborn participants [360]. Maternal smoking during pregnancy was assessed in relation to methylation. Maternal sustained and any smoking during pregnancy was determined using questionnaires carried out during each study, and methylation was measured in cord blood using the Illumina Infinium® HumanMethylation450k BeadChip assay [360]. The Joubert *et al.* 2016 meta-analysis of maternal smoking was performed with False Discovery Rate (FDR) at 5% and statistically significant DMCs were based on Bonferroni correction [360]. Only the data from sustained maternal smoking was considered for the current study.

In addition to the data from the Timms *et al.* 2019 EWAS of maternal folic acid supplementation, the most extensive meta-analysis of maternal plasma folate levels, conducted by Joubert *et al.* 2016 was also used [359]. The meta-analysis of maternal plasma folate levels used data from 2 study populations, 1275 participants in the Norwegian Mother and Child Cohort Study (MoBa) and 713 participants in the Generation R Study [359]. Maternal plasma folate levels were assessed in relation to methylation. Folate levels were determined using blood samples from mothers during pregnancy and methylation was measured in cord blood using the Illumina Infinium® HumanMethylation450k BeadChip assay [359]. Meta-analysis of maternal plasma folate levels was performed on CpGs common to both cohorts with FDR at 5% and statistically significant DMCs were based on Bonferroni correction [359].

The number of DMCs identified from each study is summarised in Table 8.1.

Table 8.1. Differentially methylated CpGs (DMCs) associated with environmental risk exposures associated with childhood ALL.

Environmental exposure	No. of associated DMCs
Maternal radiation exposure [125]	287
Maternal alcohol intake [125]	190
Maternal sugary caffeinated drinks [125]	65
Maternal coffee consumption [125]	15
Maternal smoking throughout pregnancy [125]	22
Maternal smoking at 3 months pregnant [125]	15
Maternal smoking pre-pregnancy [125]	7
Maternal sustained smoking in pregnancy (meta-analysis) [360]	6073
Day nursery attendance (at 8 months) [125]	11
Reported cold (at 6 months) [125]	75
Maternal folic acid supplementation [125]	9
Maternal plasma folate levels during pregnancy (meta-analysis) [359]	443

8.2.2 Identification of CpGs with altered methylation in ALL and across disease subtypes

The Nordlund *et al.* 2013 study has previously identified genome wide *de novo* methylation changes in subtypes specific to childhood ALL, and was chosen for use in the current “meet in the middle” approach due to having the largest data set available for ALL subtypes using

Illumina 450K arrays, which is key for providing like for like CpG analysis with the environmental exposures [361]. The study used bone marrow samples from 101 patients with T-ALL and 663 patients with B-cell precursor ALL and compared CpG methylation levels to reference cells including CD19+, CD3+, CD34+ of healthy donors and 86 bone marrow samples from patients in remission. DNA methylation levels were measured using Illumina Infinium® HumanMethylation450k BeadChip assay, and differential DNA methylation from 435,941 CpG sites was determined by the Wilcoxon rank-sum test. For T-ALL, remission bone marrow, CD3+ and CD43+ were used as reference cells. For B cell precursor ALL (BCP ALL), remission bone marrow, CD19+ and CD34+ were used as reference cells. The DMC's were split into ALL cytogenetic subtypes including T-ALL, high hyperdiploidy (HeH), 11q23/*MLL*, dic(9;20), t(1;19)*TCF3/PBX1*, t(12;21)*ETV6/RUNX1*, t(9;22)*BCR/ABL1*, and iAMP21, with BCP ALL samples with unknown cytogenetic aberrations described as "undefined", and those with no unique differential methylation described as non-recurrent. DMCs conserved across all the ALL subtypes were labelled as constitutive. Original DMC analysis data from all patient and control samples was obtained through correspondence with Dr Nordlund. The number of identified DMCs for each subtype is described in Table 8.2

Table 8.2. Differentially methylated CpG's (DMCs) found in childhood ALL cytogenetic subtypes [361]. Constituent DMCs are common across all subtypes.

Subtype	Total no. of DMCs	No. of non-constitutive DMCs	No. of DMCs unique to subtype
Constituent	9406	NA	NA
T-ALL	58157	48751	16841
***MLL*/11q23**	31403	21997	1736
dic(9;20)	53680	44274	2370
HeH	42779	33373	3014
t(1;19) *TCF3/PBX1*	21799	12393	1110
t(12;21) *ETV6/RUNX1*	45589	36183	2114
t(9;22) *BCR/ABL1*	23871	14465	271
iAMP21	44726	35320	2656
Undefined	39262	29856	56
Non-recurrent	42109	32703	27

8.2.3 Integration of DMCs associated with ALL and risk exposures

Bioinformatics and research computing (http://barc.wi.mit.edu/tools/compare/) tool was used to integrate the list of DMCs associated with each ALL subtype and constitutive groups, with the lists of DMCs associated with altered methylation in response to the ALL risk exposures. This was used to identify overlapping DMCs in any direction, i.e., hypo or hyper-methylated, between ALL methylation and risk exposure methylation. For each set of overlaps, hypergeometric tests were performed within *R*, to identify if observed methylation changes were due to chance based on a significance of $p \leq 0.05$. The total number of DMCs identified in the Nordlund *et al.* 2013 Illumina Infinium® HumanMethylation450k BeadChip assay, 435,941, was used as the population size. The "successes within the population" was designated as the number of DMCs associated with each environmental exposure, and the "sample size" was defined as the number of DMCs associated with each ALL subtype group. The "successes within the sample" was determined by the number of CpG overlaps between methylation changes in ALL subtypes and environmental exposures.

To ascertain the direction of the methylation change in DMCs associated with ALL subtypes, the difference between methylation values in Nordlund *et al.* 2013 control data and DMCs in subtypes was calculated to determine hypo or hyper methylation changes [361]. The methylation direction in ALL subtype DMCs was compared to risk exposure DMC direction (from literature) to identify which of the previously identified significant overlaps were in the same or opposing directions [125, 359, 360]. The sets of overlaps were then analysed by hypergeometric tests to determine if the directional overlaps in methylation change are due to chance.

An overview of the analysis is shown in Figure 8.1.

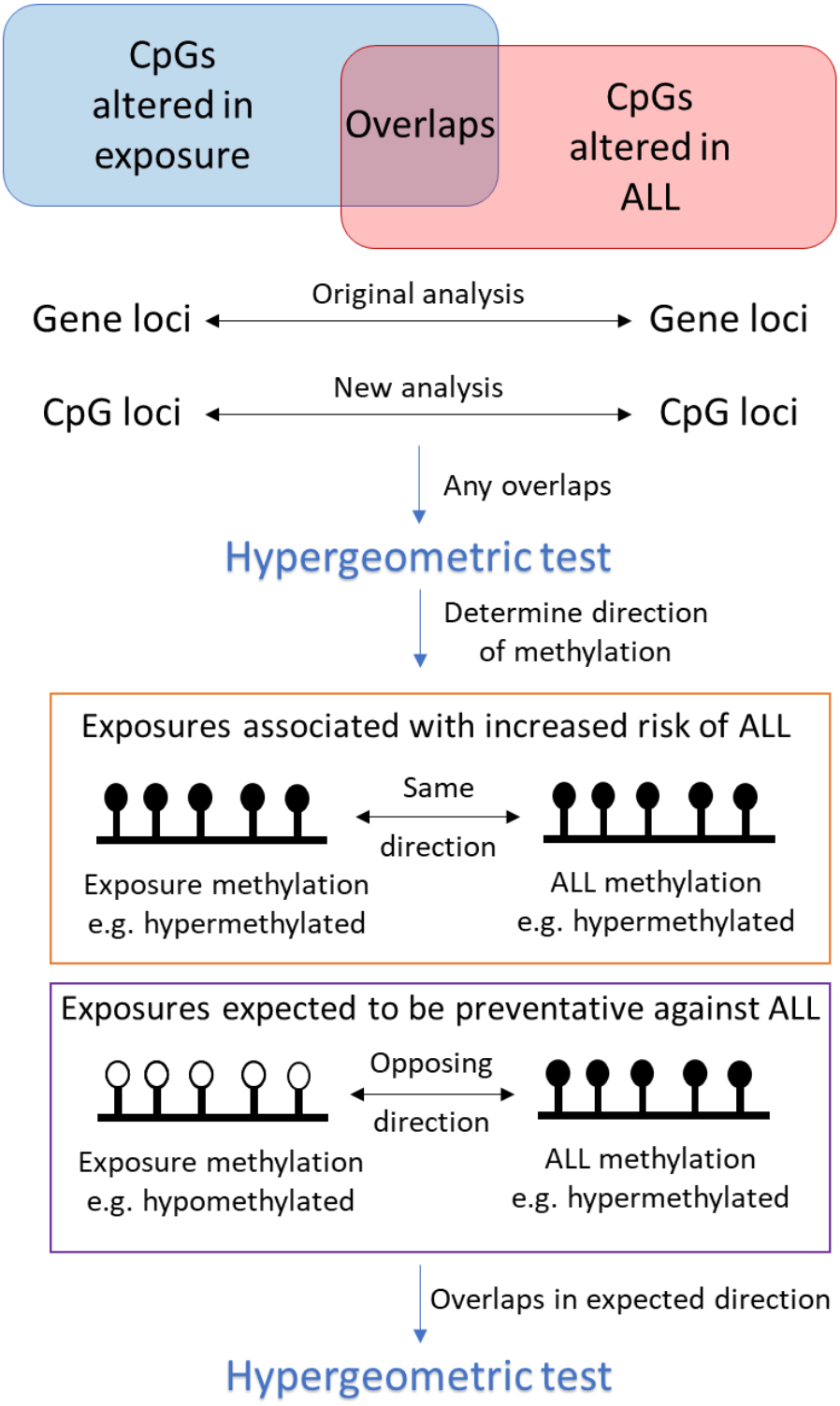


Figure 8.1. Overview of "meet in the middle" analysis for altered methylation changes in environment and ALL. For environmental exposures that are expected to increase the risk of childhood leukaemia, including maternal radiation exposure, maternal alcohol intake, maternal caffeinated drink intake, maternal coffee intake and maternal smoking, it would be anticipated that the direction of methylation change would be concordant, with the same direction observed in both exposure and ALL, i.e., hypermethylated in exposure and hypermethylated in ALL. Conversely, for environmental exposures that are expected to have a protective effect against ALL, including maternal folic acid supplementation/plasma folate level, day nursery attendance and reported colds, it would be anticipated that the direction of methylation change would be discordant, with opposing direction between exposure and ALL, i.e., hypermethylated in exposure and hypomethylated in ALL. The hypergeometric test calculates the likelihood that the overlaps identified within the population have occurred by chance.

8.2.4 Pathway and Gene Ontology enrichment analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analysis was performed via 2 platforms for comparison. Official gene identifiers mapped to DMCs were collated for each ALL subtype and each risk exposure using the published data available [125, 359-361]. The first platform utilised R studio with package Bioconductor packages, org.Hs.eg.db for mapping gene IDs and clusterProfiler for KEGG and GO analysis (August 2021) [362]. The second platform, used by Dr Jill McKay for the analysis, was DAVID (version DAVID 2021 released December 2021, access between June-August 2022) [363]. As DAVID is restricted to a maximum of 6,000 official gene identifiers, the gene identifiers were first converted from official gene symbols to entrez identifiers in DAVID, and entrez IDs used to carry out GO and KEGG analysis.

Significant threshold for GO enrichment analysis was set at $p<0.05$ (corrected for multiple testing), and at $p<0.05$ (uncorrected) for KEGG pathway enrichment analysis. Data from each subtype was compared using a Multiple List Comparator (https://molbiotools.com/listcompare.php) to assess common processes and pathways between subtypes.

8.3 Results

8.3.1 Constitutive CpG methylation changes across ALL subtypes and integration with potential ALL risk exposure associated methylation changes

A previous study by Timms *et al.* 2019 used the meet in the middle approach to identify overlaps at gene level associated with altered DNA methylation observed in risk-exposures and in constitutive ALL (DMCs common to all subtypes of ALL measured) [125]. Hypergeometric tests were used to assess the probability that these overlaps happened by chance. The overlaps were identified in overall methylation changes and in methylation changes occurring in the same direction (or opposite direction where exposures are expected to be protective). The same analysis was repeated in the current study at CpG level, as described in Figure 8.1, with results at gene level and CpG level presented in Table 8.3 for comparison.

The CpG level analysis found significant overlaps in overall methylation changes and with concordant directionality between constitutive ALL-associated DMCs and the risk-exposures maternal medical radiation and maternal alcohol intake during pregnancy, Table 8.3. This supports the findings observed at gene level, and although p values are more modest at CpG level, the percentage concordance for directionality of methylation changes is higher for CpGs than genes. For maternal sugary caffeinated drink intake during pregnancy, no significant overlaps were observed at CpG level, whereas at gene level, significant overlaps

were previously observed in both overall methylation changes and with concordant direction, Table 8.3. Similar to previous gene-level findings, maternal coffee consumption revealed no significant overlaps with ALL-associated methylation at CpG level, Table 8.3. Previous findings at gene level observed significant overlaps in 3 out of the 4 smoking risk-exposures, however this significance was lost when direction of methylation change was considered, Table 8.3. This outcome is supported at CpG level, with no significant directionally concordant overlaps observed in any of the smoking risk-exposures.

Opposing findings were observed at gene versus CpG level for both nursery attendance and reported colds at 6 months, Table 8.3. Previous gene level analysis observed significant overlaps for nursery attendance, however no overlaps were observed at CpG level, suggesting that different regions of those genes have altered methylation in ALL compared to those affected by nursery attendance. Conversely, for reported colds at 6 months, whilst no significant overlaps were observed at gene level, at CpG level, significance was observed in both overlaps in overall methylation changes, and when overlaps were discordant, as reported colds are expected to be a protective exposure. No significant overlaps were observed at both gene and CpG level for folic acid supplementation, however using data from a meta-analysis of maternal plasma folate levels during pregnancy, significant overlaps were observed at gene level and confirmed at CpG level, Table 8.3. Analysis of discordant DMCs associated with maternal plasma folate status at CpG level also observed significant overlaps, with 94% concordance for directionality.

*Table 8.3. Gene loci and CpG level comparison of overlapping DMCs found in constitutive ALL and risk-exposures. All p values are for hypergeometric tests. *Protective exposures assessed for overlapping methylation in opposing directions. (p value= <0.05 in bold)*

Risk exposure	Number of CpGs (genes) related to exposure	Number of overlapping genes with risk exposure DMCs and constitutive ALL-associated DMCs (P value)	Number of overlapping genes with the same direction of methylation change (% concordance; P Value)	Number of overlapping CpGs with risk exposure DMCs and constitutive ALL-associated DMCs (P value)	Number of overlapping CpGs with the same* direction of methylation change (% concordance; P Value)
Radiation exposure[125]	288 (239)	54 **(1.69 x 10^{-8})**	40 **(74%; 0.001)**	13 **(0.011)**	11 **(85%; 0.050)**
Alcohol intake[125]	192 (175)	33 **(3.98 x 10^{-4})**	29 **(88%; 0.006)**	9 **(0.025)**	9 **(100%; 0.025)**
Sugary caffeinated drinks[125]	66 (54)	12 **(7.42 x 10^{-3})**	10 **(83%; 0.045)**	4 (0.055)	4 (100%; 0.055)
Coffee consumption[125]	15 (13)	1 (0.754)	0	0	0
Smoking throughout pregnancy[125]	22 (13)	5 **(7.11 x 10^{-3})**	3 (60%; 0.140)	1 (0.381)	1 (100%; 0.381)
Smoking 3 months pregnant[125]	15 (7)	3 **(2.73 x 10^{-2})**	1 (33%; 0.530)	0	0
Smoking pre-pregnancy[125]	7 (4)	2 (5.45 x 10^{-2})	1 (50%; 0.350)	0	0
Sustained smoking in pregnancy (meta-analysis[360])	6073 (3176)	431 **(4.95 x 10^{-11})**	Not reported	99 (0.999)	65 (66%; 1)
Day nursery attendance (8 months) [125]	11 (11)	6 **(2.65 x 10^{-5})**	5 **(83%; 0.003)**	0	0
Reported cold (6 months) [125]	75 (60)	9 (0.156)	3 (33%; 0.952)	5 **(0.023)**	5 **(100%; 0.023)***
Folic acid supplementation[125]	9 (7)	2 (0.155)	1 (50%; 0.155)	0	0
Maternal plasma folate levels during pregnancy (meta-analysis[359])	443 (229)	62 **(4.32 x 10^{-13})**	Not reported	18 **(0.009)**	17 **(94%; 0.018)***

8.3.2 Subtype-specific DNA methylation changes in ALL and integration with potential ALL risk exposure associated changes in methylation

The same methodology was used to investigate the relationship between methylation changes associated with risk exposures and methylation changes associated with individual ALL subtypes. In line with findings for constitutive ALL, no significant overlapping DMCs

were observed between any ALL subtype and the following maternal risk exposures during pregnancy: sugary caffeinated drinks, coffee consumption, smoking 3 months pregnant, smoking pre-pregnancy and folate supplementation, Table 8.4.

Furthermore, in keeping with findings observed for constitutive ALL, maternal radiation exposure also had significant overlapping DMCs with the majority of ALL subtypes (7 out of 10), with overlapping methylation found between Heh, *BCR-ABL1* and iAMP21 and radiation exposure expected to be due to chance, Table 8.4. However, when the direction of methylation is considered (i.e., concordance as radiations is considered a risk factor), Table 8.5, only 1 subtype, 11q23/*MLL*, remains significant, although the percentage of concordance was lower than observed in constitutive ALL at 76% compared to 85%. Maternal alcohol exposure was also found to have significant overlapping methylation with constitutive ALL DMCs, but only 3 out of 10 subtypes, T-ALL, *ETV6-RUNX1* and undefined (Table 8.4) were found to have significant overlapping methylation with this exposure. However, when the direction of methylation was considered for concordant DMCs, only 2 subtypes remained significant, T-ALL and *ETV6-RUNX1*, suggesting these overlaps are not due to chance, Table 8.5. Although p values for these subtypes were modest compared to constitutive ALL results, the percentage concordance was still high, at 97%.

Unlike constitutive ALL findings, where no significant observations were found for any of the 4 smoking risk exposures, when subtype DMCs were compared, significant overlapping DMCs were observed between ALL subtypes and 2 of the smoking risk exposures, Table 8.4. DMCs associated with smoking throughout pregnancy, as analysed by Timms *et al.* 2019 had significant overlaps with 5 ALL subtypes: 11q23/*MLL*,dic(9;20), iAMP21, undefined and non-recurrent, Table 8.4. When the direction of methylation of the overlapping DMCs were assessed for concordance, as maternal smoking is expected to be a risk factor, 2 subtypes, 11q23/*MLL* and undefined, remained significant, suggesting the overlaps in these subtypes did not occur by chance, Table 8.5. When analysing the meta-analysis data from Joubert *et al.* 2016 for sustained smoking throughout pregnancy, significant overlapping DMCs were observed for all 10 ALL subtypes, Table 8.4. However, when directionality of methylation was considered, significance was lost, suggesting the overlapping methylation between smoking-associated methylation and subtype-associated methylation is likely to be due to chance, Table 8.5. The percentage concordance of overlaps for sustained smoking throughout pregnancy was low for all subtypes, at 61% or less.

DMCs associated with day nursery attendance were not found to be altered in constitutive ALL-associated DMCs, however during subtype analysis, 6 subtypes were found to have significant overlapping DMCs with nursery attendance including: T-ALL, 11q23/*MLL*,

dic(9;20), *ETV6-RUNX1,* undefined and non-recurrent, Table 8.4. As nursery attendance is expected to be protective against ALL, the overlapping DMCs were assessed for discordant methylation, with significance lost in all subtypes, Table 8.5.

DMCs associated with reported colds at 6 months and constitutive ALL-associated DMCs were found to have a significant level of overlapping methylation, however in the subtype analysis this relationship was only observed for 1 subtype out of 10 (Table 8.4). As reported colds at 6 months are expected to be protective, the significant overlapping DMCs observed with the 11q23/*MLL* subtype were assessed for discordant overlaps, Table 8.6, which remained significant at 100% discordance, suggesting this was not a chance finding. When analysing the meta-analysis data from Joubert *et al.* 2016 for maternal plasma folate levels during pregnancy significant overlapping DMCs were observed across all 10 ALL subtypes, Table 8.4. When considered for discordant overlapping DMCs, 8 out of 10 subtypes remained significant, Table 8.6. No significant overlaps were observed for Heh and *BCR-ABL1*, suggesting any overlapping DMCs between maternal plasma folate levels and these subtypes occurred by chance.

Table 8.4. Overlapping DMCs associated with both childhood ALL cytogenetic subtypes and environmental risk exposure. Number of overlaps (hypergeometric test p value).

	DMCs associated with ALL cytogenetic subtype (p value= <0.05 in bold)									
Risk exposure	T-ALL	11q23/*MLL*	dic(9;20)	Heh (High hyperdiploid)	*TCF3-PBX1*	*ETV6-RUNX1*	*BCR-ABL1*	iAMP21	undefined	non-recurrent
Radiation exposure[125]	60 **(0.0003)**	38 **(0.0002)**	49 **(0.0121)**	33 (0.1989)	22 **(0.0334)**	48 **(0.0008)**	19 (0.2337)	31 (0.4169)	40 **(0.0042)**	42 **(0.0048)**
Alcohol intake[125]	39 **(0.0047)**	15 (0.4109)	30 (0.1017)	24 (0.1309)	12 (0.2549)	29 **(0.0280)**	12 (0.3612)	22 (0.3249)	25 **(0.0398)**	25 (0.0771)
Sugary caffeinated drinks[125]	13 (0.0948)	7 (0.1962)	11 (0.1839)	9 (0.1964)	5 (0.2341)	9 (0.2501)	6 (0.1523)	10 (0.1354)	8 (0.2413)	8 (0.3042)
Coffee consumption[125]	3 (0.3232)	0	0	0	0	0	0	1 (0.8028)	0	0
Smoking throughout pregnancy[125]	4 (0.3367)	6 **(0.0038)**	7 **(0.0136)**	5 (0.0581)	3 (0.0948)	6 (0.2230)	2 (0.3411)	6 **(0.0205)**	6 **(0.0112)**	6 **(0.0155)**
Smoking 3 months pregnant[125]	1 (0.8833)	3 (0.0886)	2 (0.5678)	2 (0.4409)	1 (0.5367)	2 (0.4751)	1 (0.5703)	2 (0.4647)	2 (0.3968)	2 (0.4326)
Smoking pre-pregnancy[125]	1 (0.6330)	2 (0.0855)	1 (0.6014)	1 (0.5147)	1 (0.3017)	1 (0.5385)	0	1 (0.5313)	1 (0.4835)	1 (0.5089)
Sustained smoking in pregnancy (meta-analysis[360])	929 **(5.03×10^{-6})**	634 **(1.21×10^{-20})**	943 **(6.41×10^{-14})**	732 **(5.49×10^{-9})**	476 **(1.49×10^{-21})**	822 **(1.75×10^{-14})**	506 **(2.21×10^{-20})**	949 **(6.28×10^{-39})**	611 **(0.0023)**	682 **(2.41×10^{-5})**
Reported cold[125]	14 (0.1200)	10 **(0.0424)**	13 (0.1275)	10 (0.1981)	5 (0.3212)	12 (0.0890)	6 (0.2276)	7 (0.6601)	10 (0.1354)	10 (0.1852)
Day nursery attendance[125]	4 **(0.0481)**	4 **(0.0059)**	4 **(0.0371)**	3 (0.0856)	0	4 **(0.0216)**	2 (0.1188)	3 (0.0952)	4 **(0.0129)**	4 **(0.0165)**
Folic acid supplementation[125]	1 (0.7244)	1 (0.4898)	1 (0.6935)	1 (0.6053)	0	1 (0.6300)	1 (0.3976)	1 (0.6225)	1 (0.5723)	1 (0.5992)
Maternal plasma folate levels during pregnancy (meta-analysis[359])	123 **(9.43×10^{-16})**	57 **(1.78×10^{-5})**	92 **(3.58×10^{-7})**	55 **(0.0425)**	39 **(0.0005)**	71 **(0.0002)**	35 **(0.0203)**	99 **(9.99×10^{-14})**	63 **(0.0002)**	71 **(1.75×10^{-5})**

*Table 8.5. Concordant overlaps of DMCs associated with both childhood ALL cytogenetic subtypes and environmental risk exposure. Number of overlaps (% concordance; hypergeometric test p value). P = <0.05 in **bold.** N/A = not applicable, as no significance was observed prior to assessment of concordance.*

	Concordant DMCs overlapping between risk exposure and ALL cytogenetic subtype (% concordance; P Value)									
Risk exposure	**T-ALL**	**11q23/*MLL***	**dic(9;20)**	**Heh (High Hyperdiploidy)**	***TCF3-PBX1***	***ETV6-RUNX1***	***BCR-ABL1***	**iAMP21**	**undefined**	**non-recurrent**
Radiation exposure[125]	48 (80%; 0.0610)	29 (76%; 0.0436)	39 (80%; 0.2876)	N/A	19 (86%; 0.1351)	37 (77%; 0.1115)	N/A	N/A	33 (83%; 0.0915)	32 (76%; 0.2275)
Alcohol intake[125]	38 (97%; 0.0080)	N/A	N/A	N/A	N/A	28 (97%; 0.0448)	N/A	N/A	24 (96%; 0.0637)	N/A
Smoking throughout pregnancy[125]	N/A	5 (83%; 0.0181)	5 (71%; 0.1252)	N/A	N/A	N/A	N/A	4 (67%; 0.1835)	5 (83%; 0.0426)	5 (83%; 0.0549)
Sustained smoking in pregnancy (meta-analysis[360])	542 (58%; 1)	287 (45%; 1)	538 (57%; 1)	391 (53%; 1)	287 (60%; 0.8459)	465 (56%; 1)	285 (56%; 0.9973)	549 (58%; 0.9994)	366 (60%; 1)	415 (61%; 1)

*Table 8.6. Discordant overlaps of DMCs associated with both childhood ALL cytogenetic subtypes and environmental risk exposures expected to be protective. Number of overlaps (% concordance; hypergeometric test p value). P = <0.05 in **bold.** N/A = not applicable, as no significance was observed prior to assessment of discordance.*

	Discordant DMCs overlapping between risk exposure and ALL cytogenetic subtype (% discordance; P Value)									
Risk exposure	T-ALL	11q23/*MLL*	dic(9;20)	Heh (High Hyperdiploidy)	*TCF3-PBX1*	*ETV6-RUNX1*	*BCR-ABL1*	iAMP21	undefined	non-recurrent
Day nursery attendance[125]	0	0	0	N/A	N/A	0	N/A	N/A	0	0
Reported cold[125]	N/A	10 **(100%; 0.0424)**	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Maternal plasma folate levels during pregnancy (meta-analysis[359])	116 **(94%; $4.68x10^{-13}$)**	52 **(91%; 0.0004)**	82 **(89%; 0.0001)**	41 (75%; 0.6765)	33 **(85%; 0.0159)**	62 **(87%; 0.0114)**	28 (80%; 0.2437)	88 **(89%; $1.34x10^{-9}$)**	56 **(89%; 0.0066)**	63 **(89%; 0.0013)**

8.3.3 Comparison of KEGG Pathway and GO Analysis across Subtype Specific-Associated methylation

KEGG pathway and GO analysis were carried out for each of the individual ALL subtype specific-associated methylation patterns using 2 different platforms, DAVID and RStudio. Using DAVID, a range of 65-90 KEGG pathways were revealed across individual subtypes, with 49 pathways common to all subtypes, Table 8.7 and supplementary data S1. RStudio revealed similar findings, with a range of 63-102 KEGG pathways, and 49 pathways common to all subtypes, Table 8.7 and supplementary data S4. The 49 common pathways accounted for ~50% or more of the pathways identified in each subtype, Table 8.7. Of the 49 pathways common to all subtypes, 44 pathways were common to both DAVID and RStudio analysis (supplementary data Table S4k). These common pathways found associated with all ALL subtypes include signalling pathways (Hippo, MAPK, Wnt, Ras, stem cell pluripotency regulators) and cancer pathways (proteoglycans in cancer, melanoma, breast cancer), (supplementary data Table S4k). Most subtypes had at least 1 or more pathways that were unique to subtype, however these unique pathways accounted for 10% or less of overall pathways associated with each subtype, Table 8.7. DAVID analysis found no unique pathways for 11q23/*MLL*, dic(9;20) and non-recurrent subtypes, whereas RStudio analysis only identified the non-recurrent subtype as having no unique pathways, Table 8.7.

Analysis of GO biological processes with DAVID identified between 91-894 biological processes, with 63 common across all subtypes, whereas RStudio analysis identified between 153-215 biological processes, with 90 common across all subtypes Table 8.8 and supplementary data S2 and S5. The common biological processes shared across all subtypes identified by DAVID accounted for 10% or less of the overall GO processes in each subtype apart from dic(9;20) where they account for 69% of overall GO processes, Table 8.8. Conversely, the 90 GO process common to all subtypes found with RStudio, accounted for nearly half of overall processes for each subtype, Table 8.8. Of these common GO processes there is no direct overlap between DAVID and RStudio findings. However, many of the processes described in each platform are similar, such as transcription regulation, calcium regulation, signalling activity, neurotransmitter activity, transporter activity (supplementary data Tables S2k and S5k). DAVID analysis found at least 3 or more processes unique to subtype for each of the 10 ALL subtypes, however they mostly accounted for only 1-2% of overall processes, with 8% for T-ALL, Table 8.8. For 8 out of 10 subtypes, RStudio analysis found 5% or less unique processes, with 13% identified in T-ALL and no unique processes found for *BCR-ABL1*, Table 8.8.

Table 8.7. Comparison of KEGG pathways for individual ALL subtypes identified using DAVID and RStudio platforms.

	T-ALL	*11q23/MLL*	dic(9;20)	Heh (High Hyperdiploid)	*TCF3-PBX1*	*ETV6-RUNX1*	*BCR-ABL1*	iAMP21	Undefined	Non-recurrent
DAVID Analysis										
No. KEGG pathways	90	65	78	83	79	90	83	83	81	80
Unique to subtype (%)	9 (10%)	0	0	5 (6%)	2 (3%)	1 (1%)	1 (1%)	2 (2%)	1 (1%)	0
% Pathways shared by all subtypes (49 pathways)	54%	75%	63%	59%	62%	54%	59%	59%	60%	61%
RStudio Analysis										
No. KEGG pathways	99	84	85	100	89	95	102	98	91	63
Unique to subtype (%)	8 (8%)	3 (4%)	2 (2%)	4 (4%)	4 (5%)	3 (3%)	7 (7%)	2 (2%)	1 (1%)	0
% Pathways shared by all subtypes (49 pathways)	50%	58%	58%	49%	55%	52%	48%	50%	54%	77%

Table 8.8. Comparison of GO processes for individual ALL subtypes identified using DAVID and RStudio platforms.

	T-ALL	*11q23/MLL*	dic(9;20)	Heh (High Hyperdiploid)	*TCF3-PBX1*	*ETV6-RUNX1*	*BCR-ABL1*	iAMP21	Undefined	Non-recurrent
DAVID Analysis										
No. GO BPs	894	667	91	691	612	783	680	743	747	768
Unique to subtype (%)	71 (8%)	8 (1%)	4 (4%)	9 (1%)	4 (1%)	13 (2%)	7 (1%)	11 (2%)	3 (1%)	8 (1%)
% BPs shared by all subtypes (63 pathways)	7%	9%	69%	9%	10%	8%	9%	8%	8%	8%
RStudio Analysis										
No. GO BPs	215	153	226	183	178	196	173	188	187	201
Unique to subtype (%)	29 (13%)	5 (3%)	11 (5%)	4 (2%)	6 (3%)	8 (4%)	0	5 (3%)	2 (1%)	1 (0.5%)
% BPs shared by all subtypes (90 BPs)	42%	59%	40%	49%	51%	46%	52%	48%	48%	45%

8.3.4 Comparison of KEGG Pathway and GO Analysis between Subtype Specific-Associated methylation and risk exposure-associated methylation

Due to the small number of DMCs identified in the risk exposures, only the following, which had the highest number of DMCs of all risk exposures, were selected for analysis: Joubert *et al.* 2016 meta-analysis for sustained smoking in pregnancy, Joubert *et al.* 2016 meta-analysis for maternal plasma folate levels during pregnancy and Timms *et al.* 2019 data for radiation exposure. KEGG and GO analysis by DAVID identified significant pathways for smoking, folate, and radiation exposure, whereas RStudio analysis only identified significant pathways for smoking and folate, Table 8.9 and supplementary data S3 and S6.

Table 8.9. Comparison of KEGG and GO analysis for risk exposures identified using DAVID and Rstudio platforms.

	Sustained smoking in pregnancy	Maternal plasma folate levels during pregnancy	Radiation exposure
DAVID no. KEGG pathways	90	15	1
Rstudio no. KEGG pathways	110	16	0
DAVID no. GO processes	200	4	2
Rstudio no. GO processes	43	14	0

For sustained smoking in pregnancy, DAVID analysis identified 27 KEGG pathways which were also found to be in common to all ALL subtypes, Table 8.10. The overlaps for each subtype accounted for 41-53% of overall pathways identified as potentially affected by methylation changes associated with maternal smoking exposure. Twenty-one GO processes were observed in DAVID for all ALL subtypes, Table 8.11. The limiting subtype was dic(9;20), having only 11% of processes overlapping with those identified as potentially affected by methylation changes associated with sustained maternal smoking, compared to 82-94% of processes overlapping with the other 9 subtypes. Rstudio analysis identified 35 overlapping KEGG pathways that are potentially affected by methylation changes in response to sustained maternal smoking as well as all ALL subtypes, with overlaps for each individual subtype accounting for 37-59% of exposure-related pathways, Table 8.10. Rstudio analysis identified 21 GO processes which overlapped between the methylation analysis for sustained smoking and the methylation analysis for all ALL subtypes, with overlaps for individual subtypes accounting for between 56-67% of processes potentially altered due to smoking associated methylation, Table 8.11.

Comparison of DAVID analysis for maternal plasma folate associated methylation with each of the analyses for ALL subtypes identified 10 KEGG pathways and 1 GO process in common, Tables 8.10 and 8.11. The remaining 3 GO processes associated with maternal plasma folate associated methylation were also observed in the analysis of 9 of the 10

subtypes. Rstudio analysis of maternal plasma folate associated methylation with each of the analyses for ALL subtypes observed 11 KEGG pathways and 13 GO processes in common, Tables 8.10 and 8.11.

When comparing DAVID analysis for radiation exposure methylation with analyses for each ALL subtype, 1 KEGG pathway and 1 GO process was identified as overlapping for each subtype. These findings common to both radiation exposure and ALL subtypes included ‘MAPK signalling pathway’ and ‘nervous system development’, Tables 8.10, 8.11, 8.12 and 8.14.

Table 8.10. Comparison of overlaps between KEGG pathways observed in exposure and subtype identified using DAVID and Rstudio platforms. Number of overlaps (percentage of KEGG pathways found in exposure).

	KEGG pathways identified as overlapping between exposure and subtype (% of KEGG pathways found in exposure)										
Analysis platforms	ALL subtypes	T-ALL	11q23/*MLL*	dic(9;20)	Heh (High Hyperdiploidy)	*TCF3-PBX1*	*ETV6-RUNX1*	*BCR-ABL1*	iAMP21	undefined	non-recurrent
DAVID											
Sustained smoking in pregnancy (meta-analysis[360])	27 (30%)	47 (52%)	37 (41%)	38 (42%)	41 (46%)	40 (44%)	50 (56%)	48 (53%)	46 (51%)	39 (43%)	43 (48%)
Maternal plasma folate levels during pregnancy (meta-analysis[359])	10 (67%)	12 (80%)	11 (73%)	12 (80%)	12 (80%)	13 (87%)	13 (87%)	13 (87%)	11 (73%)	12 (80%)	12 (80%)
Radiation exposure[125]	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)	1 (100%)
Rstudio											
Sustained smoking in pregnancy (meta-analysis[360])	35 (32%)	64 (58%)	58 (53%)	53 (48%)	61 (55%)	52 (47%)	64 (58%)	65 (59%)	65 (59%)	58 (53%)	41 (37%)
Maternal plasma folate levels during pregnancy (meta-analysis[359])	11 (69%)	15 (94%)	14 (88%)	14 (88%)	15 (94%)	15 (94%)	15 (94%)	15 (94%)	13 (81%)	15 (94%)	13 (81%)

Table 8.11. Comparison of overlaps between GO processes observed in exposure and subtype identified using DAVID and Rstudio platforms. Number of overlaps (percentage of GO processes found in exposure).

	GO processes identified as overlapping between exposure and subtype (% of GO processes found in exposure)										
Analysis platforms	ALL subtypes	T-ALL	11q23/*MLL*	dic(9;20)	Heh (High Hyperdiploidy)	*TCF3-PBX1*	*ETV6-RUNX1*	*BCR-ABL1*	iAMP21	undefined	non-recurrent
DAVID											
Sustained smoking in pregnancy (meta-analysis[360])	21 (11%)	186 (93%)	186 (93%)	22 (11%)	165 (83%)	164 (82%)	187 (94%)	180 (90%)	184 (92%)	184 (92%)	182 (91%)
Maternal plasma folate levels during pregnancy (meta-analysis[359])	1 (25%)	4 (100%)	4 (100%)	1 (25%)	4 (100%)	4 (100%)	4 (100%)	4 (100%)	4 (100%)	4 (100%)	4 (100%)
Radiation exposure[125]	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)
Rstudio											
Sustained smoking in pregnancy (meta-analysis[360])	21 (49%)	27 (63%)	24 (56%)	29 (67%)	26 (60%)	25 (58%)	29 (67%)	25 (58%)	26 (60%)	25 (58%)	26 (60%)
Maternal plasma folate levels during pregnancy (meta-analysis[359])	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)	13 (93%)

Of the overlapping KEGG pathways identified by DAVID and Rstudio for sustained smoking in pregnancy and all ALL subtypes, 24 KEGG pathways were common between the analyses from both platforms. These include both signalling (MAPK, calcium, Hippo etc.) and cancer pathways (proteoglytans in cancer, gastric cancer) Tables 8.12 and 8.13. No overlaps were found between GO processes identified in DAVID and Rstudio, but similar processes can be observed such as transcription regulation, Table 8.14 and 8.15.

For maternal plasma folate during pregnancy, the 10 KEGG pathways identified as overlapping with all subtypes using DAVID were also found in Rstudio analysis, Table 8.12 and 8.13. These included a range of signalling pathways (cAMP, stem cell pluripotency) and signalling synapses (dopaminergic, glutamatergic, GABAergic). Rstudio also identified "nicotine addiction" pathway as common to all subtypes, Table 8.13. The only overlapping GO process identified by DAVID for maternal plasma folate associated methylation and all subtypes was 'cell-cell signalling', Table 8.14. Rstudio analysis on the other hand identified a range of processes involved in receptor and 171ctive171rtt 171ctiveity, Table 8.15.

Table 8.12. DAVID common KEGG pathways between risk exposures and all ALL subtypes. Pathways underlined are common to analysis performed by DAVID and Rstudio.

KEGG ID	KEGG Pathway	Number of genes	P value
Sustained smoking in pregnancy (meta-analysis[360])			
hsa01522	Endocrine resistance	27	0.001499
hsa04010	MAPK signaling pathway	62	0.002578
hsa04014	Ras signaling pathway	58	5.68E-05
hsa04015	Rap1 signaling pathway	57	3.34E-06
hsa04020	Calcium signaling pathway	49	0.013898
hsa04022	cGMP-PKG signaling pathway	35	0.027137
hsa04024	cAMP signaling pathway	46	0.012378
hsa04072	Phospholipase D signaling pathway	39	2.86E-04
hsa04151	PI3K-Akt signaling pathway	81	3.11E-05
hsa04360	Axon guidance	49	2.28E-05
hsa04390	Hippo signaling pathway	34	0.018943
hsa04510	Focal adhesion	47	0.001112
hsa04550	Signaling pathways regulating pluripotency of stem cells	33	0.008403
hsa04713	Circadian entrainment	23	0.022371
hsa04725	Cholinergic synapse	28	0.006041
hsa04727	GABAergic synapse	22	0.016531
hsa04728	Dopaminergic synapse	29	0.025937
hsa04750	Inflammatory mediator regulation of TRP channels	22	0.044468
hsa04810	Regulation of actin cytoskeleton	49	0.002079
hsa04921	Oxytocin signaling pathway	38	0.001316
hsa04925	Aldosterone synthesis and secretion	24	0.013336
hsa04928	Parathyroid hormone synthesis, secretion and action	30	4.78E-04
hsa05032	Morphine addiction	25	0.00244
hsa05165	Human papillomavirus infection	66	0.00724
hsa05200	Pathways in cancer	107	3.69E-04
hsa05205	Proteoglycans in cancer	45	0.005129
hsa05226	Gastric cancer	32	0.02563
Maternal plasma folate levels during pregnancy (meta-analysis[359])			
hsa04024	cAMP signaling pathway	8	0.005796
hsa04080	Neuroactive ligand-receptor interaction	9	0.024993
hsa04550	Signaling pathways regulating pluripotency of stem cells	6	0.012975
hsa04713	Circadian entrainment	6	0.002551
hsa04724	Glutamatergic synapse	6	0.005108
hsa04727	GABAergic synapse	5	0.011161
hsa04728	Dopaminergic synapse	5	0.040371
hsa04934	Cushing syndrome	6	0.017833
hsa05032	Morphine addiction	5	0.012039
hsa05217	Basal cell carcinoma	4	0.023681
Radiation Exposure[125]			
hsa04010	MAPK signaling pathway	8	0.033186

Table 8.13. Rstudio common KEGG pathways between risk exposures and all ALL subtypes. Pathways underlined are common to analysis performed by DAVID and Rstudio.

KEGG ID	KEGG Pathway	Number of genes	P value
Sustained smoking in pregnancy (meta-analysis[360])			
hsa04015	Rap1 signaling pathway	54	4.55E-07
hsa04360	Axon guidance	48	9.01E-07
hsa04151	PI3K-Akt signaling pathway	74	1.90E-05
hsa04928	Parathyroid hormone synthesis, secretion and action	29	6.29E-05
hsa04810	Regulation of actin cytoskeleton	48	0.000154
hsa04921	Oxytocin signaling pathway	36	0.000301
hsa04014	Ras signaling pathway	49	0.000376
hsa04072	Phospholipase D signaling pathway	34	0.000613
hsa04510	Focal adhesion	43	0.000637
hsa05032	Morphine addiction	23	0.001174
hsa04925	Aldosterone synthesis and secretion	24	0.001486
hsa05165	Human papillomavirus infection	62	0.001873
hsa04024	cAMP signaling pathway	44	0.002112
hsa04725	Cholinergic synapse	26	0.00251
hsa05030	Cocaine addiction	14	0.003147
hsa05205	Proteoglycans in cancer	41	0.00321
hsa04010	MAPK signaling pathway	55	0.003396
hsa04020	Calcium signaling pathway	46	0.004418
hsa04727	GABAergic synapse	21	0.004586
hsa04550	Signaling pathways regulating pluripotency of stem cells	30	0.005356
hsa04261	Adrenergic signaling in cardiomyocytes	31	0.005905
hsa05207	Chemical carcinogenesis – receptor activation	41	0.005933
hsa04713	Circadian entrainment	22	0.006222
hsa04390	Hippo signaling pathway	32	0.006458
hsa04934	Cushing syndrome	31	0.009594
hsa05226	Gastric cancer	30	0.009753
hsa04022	cGMP-PKG signaling pathway	32	0.015889
hsa04927	Cortisol synthesis and secretion	15	0.018582
hsa04728	Dopaminergic synapse	26	0.020111
hsa05218	Melanoma	16	0.021712
hsa04911	Insulin secretion	18	0.027636
hsa04270	Vascular smooth muscle contraction	25	0.037561
hsa05224	Breast cancer	27	0.040966
hsa04974	Protein digestion and absorption	20	0.04358
Maternal plasma folate levels during pregnancy (meta-analysis[359])			
hsa04713	Circadian entrainment	6	0.00035
hsa04724	Glutamatergic synapse	6	0.000829
hsa04024	cAMP signaling pathway	8	0.001273
hsa04727	GABAergic synapse	5	0.001714
hsa05032	Morphine addiction	5	0.001892
hsa04550	Signaling pathways regulating pluripotency of stem cells	6	0.002654
hsa05217	Basal cell carcinoma	4	0.003249
hsa04934	Cushing syndrome	6	0.003957
hsa04080	Neuroactive ligand-receptor interaction	9	0.005781
hsa05033	Nicotine addiction	3	0.006839
hsa04728	Dopaminergic synapse	5	0.009234

Table 8.14. DAVID common GO processes between risk exposures and all ALL subtypes.

GO ID	GO Biological Process	Number of genes	Bonferroni corrected P Value
Sustained smoking in pregnancy (meta-analysis[360])			
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	191	2.87E-04
GO:0001501	skeletal system development	120	1.75E-04
GO:0001503	Ossification	107	1.52E-06
GO:0006357	regulation of transcription from RNA polymerase II promoter	455	6.65E-04
GO:0007155	cell adhesion	287	3.98E-05
GO:0007267	cell-cell signaling	289	2.56E-04
GO:0007268	chemical synaptic transmission	138	0.017578
GO:0007399	nervous system development	485	1.46E-14
GO:0007417	central nervous system development	203	0.003712
GO:0007420	brain development	160	0.002503
GO:0008283	cell proliferation	378	1.78E-08
GO:0009887	animal organ morphogenesis	228	3.28E-09
GO:0010628	positive regulation of gene expression	400	4.84E-07
GO:0016477	cell migration	271	0.010563
GO:0030182	neuron differentiation	281	1.15E-08
GO:0030198	extracellular matrix organization	76	0.001939
GO:0035556	intracellular signal transduction	493	5.02E-10
GO:0045666	positive regulation of neuron differentiation	86	0.024979
GO:0045893	positive regulation of transcription, DNA-templated	316	7.37E-06
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	238	2.78E-04
GO:0098609	cell-cell adhesion	188	0.002854
Maternal plasma folate levels during pregnancy (meta-analysis[359])			
GO:0007267	cell-cell signaling	35	0.011735
Radiation Exposure[125]			
GO:0007399	nervous system development	50	0.016416

Table 8.15. Rstudio common GO processes between risk exposures and all ALL subtypes.

GO ID	GO Biological Process	Number of genes	Bonferroni corrected P Value
Sustained smoking in pregnancy (meta-analysis[360])			
GO:0005245	voltage-gated calcium channel activity	18	0.000255
GO:0022843	voltage-gated cation channel activity	36	0.001094
GO:0003779	actin binding	82	0.003437
GO:0005178	integrin binding	35	0.004054
GO:0015085	calcium ion transmembrane transporter activity	33	0.004106
GO:0005262	calcium channel activity	30	0.00438
GO:0005543	phospholipid binding	83	0.004693
GO:0005244	voltage-gated ion channel activity	42	0.008217
GO:0022832	voltage-gated channel activity	42	0.008217
GO:0046873	metal ion transmembrane transporter activity	77	0.009203
GO:0001227	DNA-binding transcription repressor activity, RNA polymerase II-specific	62	0.016325
GO:0001217	DNA-binding transcription repressor activity	62	0.016925
GO:0072509	divalent inorganic cation transmembrane transporter activity	35	0.017298
GO:0005544	calcium-dependent phospholipid binding	16	0.019444
GO:0005261	cation channel activity	61	0.020244
GO:0019838	growth factor binding	30	0.024417
GO:0004879	nuclear receptor activity	15	0.029
GO:0098531	ligand-activated transcription factor activity	15	0.029
GO:0022836	gated channel activity	60	0.032383
GO:0005539	glycosaminoglycan binding	44	0.037411
GO:0005201	extracellular matrix structural constituent	34	0.042672
Maternal plasma folate levels during pregnancy (meta-analysis[359])			
GO:0022843	voltage-gated cation channel activity	8	0.013739
GO:0022836	gated channel activity	12	0.013739
GO:0015267	channel activity	14	0.014634
GO:0022803	passive transmembrane transporter activity	14	0.014634
GO:0005244	voltage-gated ion channel activity	8	0.030355
GO:0022832	voltage-gated channel activity	8	0.030355
GO:0030594	neurotransmitter receptor activity	6	0.030355
GO:0005216	ion channel activity	12	0.031973
GO:0072509	divalent inorganic cation transmembrane transporter activity	7	0.031973
GO:0099604	ligand-gated calcium channel activity	3	0.049452
GO:0015276	ligand-gated ion channel activity	6	0.049452
GO:0022834	ligand-gated channel activity	6	0.049452
GO:0008066	glutamate receptor activity	3	0.049452

Supplementary data available to download using the link below:

https://livenorthumbriaac-my.sharepoint.com/:f:/g/personal/w18023919_northumbria_ac_uk/EgOKohFdIhxDklvxOs8YCgUBUZHpyzEfzDohATt99oVKiw?e=T3j5wo

8.4 Discussion

Alterations in DNA methylation have been observed in response to exogenous exposures and in disease, with some exposures, such as smoking, observing persistent changes in DNA methylation conserved from *in utero* exposure to 17 years of age [364, 365]. As such, these changes in response to exposure could be on the causal pathway and have been harnessed in previous studies to investigate DNA methylation as a mediating mechanism linking environmental risk exposures and ALL [125]. Timms *et al.* 2019 [125] used a "meet in the middle" approach to compare gene level DNA methylation changes associated with environmental risk exposures that have been suggested to be related to ALL and those methylation changes found constitutively across ALL subtypes. In this chapter this analysis was repeated using a more sensitive CpG level approach, which corroborated previous gene level findings that DNA methylation changes associated with maternal medical radiation exposure, maternal alcohol intake, and maternal plasma folate levels during pregnancy are also found in ALL at rates higher than expected by chance and may therefore contribute to ALL aetiology, Table 8.3. Similarly, no significant overlapping ALL-associated DNA methylation was observed at gene or CpG level with methylation associated with maternal smoking pre-pregnancy, or maternal coffee consumption or maternal folic acid supplementation during pregnancy. This may indicate that DNA methylation does not act as a mediating mechanism towards disease for these exposures but may also be indicative that these exposures themselves may not associate strongly with leukemia risk (Table 8.3). Maternal smoking associated DNA methylation changes significantly overlapped with ALL-associated methylation for 3/4 smoking risk exposures when measured at gene level, however at the more sensitive CpG level analysis, no significant overlaps were observed. Likewise, significant overlaps observed at gene level between ALL associated methylation and methylation associated with maternal sugary caffeinated drinks intake and day nursery attendance at 8 months, disappeared when analysed at the more sensitive CpG level. Conversely, where previously no significant overlapping methylation was observed at gene level for DNA methylation changes in response to reported colds and in ALL itself, significant overlaps in methylation were observed for the CpG level analysis, which when including consideration for directionality of methylation, remained significant. Overall, although significance levels for the CpG analysis were more modest than those observed at gene level, the percentage of concordant (or discordant for reported colds and maternal folate) overlaps for each risk exposure was generally higher, with many reaching 100% concordance. In general, the CpG analysis corroborated many of the more robust findings from the previous gene level analysis and therefore strengthens the evidence for DNA methylation having a potential mediating role in the causal pathway towards disease for maternal radiation, plasma folate and alcohol exposures during pregnancy. Whilst some

previous associations were lost and others opposing, this may be due the exposure-methylation relationship lacking robustness which could be attributable to a) the difficulty in assessing some exposures accurately b) inadequate power to detect associations for some exposures, discussed in detail below. However, taken together, these results suggest that CpG level analysis of DNA methylation is a more sensitive approach which precisely pin-points the exact genomic location where changes are occurring, which is likely to be useful in biomarker discovery. Conversely, this therefore may lack the advantage of the gene level analysis, in its potential to highlight functional changes that any altered methylation may have for gene expression.

Leukaemia is a heterogenous disease, arising from T-cell and B-cell lineages, and defined by cytogenetic subtypes based on the chromosomal alterations they possess [4, 10]. The previous findings were performed using constitutive ALL data, where altered DNA methylation was common to all 10 ALL subtypes investigated [125, 361]. To investigate any potential subtype specific influence of environmental risk exposures, the “meet in the middle” approach was utilised to compare the DNA methylation changes associated with risk exposures to DNA methylation patterns observed in 10 individual ALL subtypes. Supporting findings observed with methylation in constitutive ALL, no significant overlaps in DNA methylation changes were observed for any subtype with altered methylation in response to maternal sugary caffeinated drinks intake, coffee consumption, smoking in first 3 months of pregnancy, smoking pre-pregnancy, day nursery attendance and folic acid supplementation during pregnancy. However, methylation associated with maternal plasma folate during pregnancy, from Joubert *et al.* 2016 meta-analysis data*,* was found to significantly overlap with methylation in most (8/10) ALL subtypes in a discordant manner, as would be anticipated since maternal folate supplementation is thought to be protective. Folate is a key player in one-carbon metabolism, required for production of the universal methyl donor, *S*-adenosylmethionine [95]. If there is insufficient folate available for donation of methyl groups, then maintenance of correct methylation patterns may be affected. Indeed, folate consumption prior to conception and during pregnancy has been associated with variation in newborn methylation patterns, including sites associated with AML and ALL [124, 125, 366]. Taken together this evidence suggests that maternal plasma folate status plays a key role in most ALL subtypes, rather than specific subtypes, with DNA methylation a potential mediating mechanism of its influence.

Meanwhile, significantly overlapping methylation was observed for individual subtypes with those methylation changes associated with maternal radiation exposure, alcohol intake, smoking throughout pregnancy, sustained smoking in pregnancy from Joubert *et al.* 2016 meta-analysis data and reported colds, at rates higher than expected due to chance.

Methylation associated with maternal alcohol intake, significantly overlapped concordantly with methylation observed in 2 subtypes, T-ALL and *ETV6-RUNX1* suggesting that DNA methylation may be a mediating mechanism for the influence of alcohol on these subtypes. Whilst alcohol has been associated with increased risk of specific T-cell lymphomas [367], which may align with findings here, there has been no known association observed specifically between alcohol and *ETV6-RUNX1* subtype. Methylation associated with smoking throughout pregnancy from the EWAS carried out in Timms *et al.* 2016 significantly overlapped concordantly with methylation in the 11q23/*MLL* and undefined subtypes. However, analysis of methylation associated with sustained smoking during pregnancy from a larger meta-analysis, which could be considered more robust, did not find this association, and therefore this could be a chance finding. Indeed there is currently no evidence that maternal smoking is associated with increased risk of 11q23/*MLL* ALL subtypes, however foetuses from smoking mothers have been observed to have an increased frequency of 11q23 rearrangements compared to non-smoking mothers, so an association may be plausible [368]. Methylation found to be associated with reported colds at 6 months also significantly overlapped with methylation in the 11q23/*MLL* subtype, in a discordant direction, as would be anticipated for an exposure thought to be protective. There is currently no evidence associating infection history with any ALL subtypes, and therefore this may be a novel finding. The 11q23/*MLL* subtype also had significantly concordant overlapping methylation with that observed in response to maternal radiation exposure during pregnancy. One study of therapy related leukaemia, where a patient has developed leukaemia following cytotoxic or radiation treatment for a different malignancy, found that nearly 50% of therapy related ALL cases had 11q23 abnormalities [369], which may support this finding. A further study however, found that therapy related leukaemias that occurred following treatment with radiation therapy alone had higher rates of *BCR-ABL1* mutations compared to *MLL* [370]. The lack of an association between *BCR-ABL1* associated methylation and other subtypes, could suggest that DNA methylation is a mediating mechanism for radiation exposure in 11q23/*MLL* subtypes only and not in other subtypes.

The lack of significant overlaps in DMCs within every subtype for each risk exposure could be explained by the fact that DNA methylation is not the mediating mechanism by which the risk exposure influences leukemogenesis in ALL overall (in the case of sugary caffeinated drinks, coffee consumption, smoking 3 months pregnant, smoking pre-pregnancy and folic acid supplementation) or only has subtype specific influence (i.e. nursery attendance and smoking throughout pregnancy). However other factors may affect the findings observed due to the analysis method and data used. It is unlikely that people are exposed to only one exposure type, and as such findings may be influenced by compounding factors. The

measurement of exposures may be impacted by reporting bias as most of the risk exposures were assessed through self-reported questionnaires, [125, 360, 371, 372]. Misreporting could occur due to the perceived stigma surrounding smoking and caffeine intake during pregnancy which could lead to inaccuracies within exposure groups [86-88]. Many of the exposures had a very small number of associated DNA methylation changes, which may make it more difficult to observe any small changes, due to lack of power. The meta-analysis of maternal plasma folate levels used direct biomarker measurements for folate from 1,988 participants, with meta-analysis for sustained smoking having 6,685 participants, both much higher than 1,018 participants used for the remaining risk exposures, which may mean data from these larger studies were more robust [125, 359, 360]. It will therefore be important where possible to use data from larger scale studies in the future to refine the analysis in line with updated environmentally-associated methylation data.

Pathway enrichment analysis can be a helpful tool to identify causal pathways linking risk exposure and disease outcome. However, there are many analysis platforms available that use both common and unique databases and annotations to identify pathways that are unlikely to have occurred by chance [373]. As such it is beneficial to conduct analysis using multiple platforms to compare overlapping pathways, or identifying unique pathways that may be missed by only using a single platform [373]. Two such platforms, DAVID and Rstudio were used here to investigate pathway enrichment associated with the DMCs identified in ALL subtypes and risk exposures. As the number of DMCs for risk exposures was low, only the risk exposures with the largest number of DMCs were analysed due to statistical power, including radiation, sustained smoking from Joubert *et al.* 2016 meta-analysis, and maternal plasma folate from Joubert *et al.* 2016 meta-analysis. Variation was seen between the number of significant pathways/processes identified with each platform, Tables 8.7-9. If RStudio was the only platform used for analysis, the significant findings for radiation exposure would be missed, as these were only seen using DAVID, Table 8.9, supporting the need for multiple platforms during this analysis.

Comparison of KEGG pathways shared in all subtypes identified in both DAVID and Rstudio had 44 out of 49 pathways in common, supplementary data Table S4k, confirming the likelihood that these pathways play a role in ALL. The common pathways identified include signalling pathways (Hippo, MAPK, Wnt, Ras, stem cell pluripotency regulators) and cancer pathways (proteoglycans in cancer, melanoma, breast cancer), supplementary data Table S4k. Wnt signalling is important in the regulation of stem cell differentiation for normal haematopoietic development and is found up-regulated in both T-cell and B-cell leukaemia [374, 375]. This evidence supports the pathway enrichment findings for playing a role in leukemogenesis. When considering GO biological processes, no overlaps were observed

between DAVID and Rstudio analysis, however many of the overall processes described were common to both platforms (supplementary data Tables S2k and S5k). These included processes such as transcription regulation, calcium regulation, signalling activity, neurotransmitter activity and transporter activity. GO is a dynamic analysis that may vary due to gene annotations, incomplete data and platforms used, which could explain why no direct overlaps were observed between DAVID and Rstudio [376]. When considering KEGG and GO analysis for individual ALL subtypes, B-cell lineage subtypes had a very small percentage of pathways/processes unique to subtype using both platforms, with a large percentage common across all subtypes, compared to T-ALL, which had a larger percentage of pathways/processes unique to subtype, Tables 8.7-8. This may be due to the differences in haematopoietic lineages, with previous studies having identified differences in genetic alteration, signalling and differentially expressed proteins in T-ALL compared to B-ALL [377, 378]. Methylation may also just play a bigger role in the T-ALL subtype.

Pathway enrichment was then compared for DMCs associated with sustained smoking, maternal plasma folate and radiation with DMCs associated with each ALL subtype. Comparison of KEGG pathways and GO processes identified with DMCs associated with sustained smoking and common to all ALL subtypes, observed a range of common pathways and processes including a variety of signalling pathways (MAPK, Ras, Hippo, calcium) and cancer pathways (proteoglycans in cancer, gastric cancer) along with processes such as transcription regulation and transporter/receptor activity, Tables 8.12-15. As well as its known associations with many cancers, smoking has been associated with altered cell signalling and hormone production, along with differences in gene expression, supporting the pathways and processes identified in DAVID and Rstudio [379-381]. When KEGG pathways and GO processes identified for DMCs associated with maternal plasma folate levels were compared to those identified to be common across all ALL subtypes, signalling pathways and synapses were commonly observed, along with cell-cell interactions and transporter and receptor activity, Tables 8.12-15. Rats fed a methyl donor deficient diet during pregnancy had offspring with altered cAMP signalling pathways and production of neurosteroids, supporting the pathways and processes identified with DAVID and Rstudio [382]. Activation of the cAMP signalling pathway is associated with inhibition of apoptosis in response to DNA damage in leukemic cells, which could be a mechanism by which DNA methylation changes in response to folate levels could contribute to leukemogenesis [383]. GO and KEGG pathway enrichment only found significant results for radiation exposure using DAVID, which, when compared with all 10 ALL subtypes included MAPK signalling pathway and nervous system development, Tables 8.12 and 8.14. The MAPK signalling pathway plays a role in many cellular processes and is required for normal nervous system

development [384]. The MAPK pathway is also implicated in leukaemia, with epigenetically activated MAPK observed to promote cell survival, supporting leukaemia progression [385, 386]. This suggests that radiation exposure could potentially contribute to leukemogenesis through the regulation of MAPK signalling by DNA methylation.

The findings in this chapter support previous findings that DNA methylation may be a mediating mechanism by which risk exposures influence ALL disease development including when specific cytogenetic subtypes are considered. There has been some epidemiological evidence to support the association of risk exposures with specific ALL cytogenetic subtypes, however, due to the rarity of the disease and limited numbers available for analysis, subtype specific evidence is limited [356]. By using DNA methylation as a biomarker to conduct a “meet in the middle” approach to compare risk exposure and ALL subtype, it has been possible investigate the relationship between the 2 and identify potential causal pathways.

9 Discussion

9.1. Evaluation of chromosomal translocation induction in response to environmental risk exposures

Preliminary experiments (sections 4.3.1, 5.3.1, 6.3.1 and 7.3.1) to investigate the influence of physiologically relevant levels of the chosen exposures on translocation induction identified translocation events in NALM6 cells in response to caffeine, benzene, cotinine, and folate. These findings were further replicated during frequency experiments, with a range of frequencies observed between different exposures, as shown in the overview in Table 9.1. Analysis by Fisher's exact test suggested a relationship between translocation induction and each of the exposures compared to the DMSO control. This gave confidence in the model as a tool for detecting translocation events in response to environmental exposures.

Table 9.1. Frequency of translocation events across exposures. Data taken from sections 4.3.2, 5.3.2, 6.3.2, 7.3.2.

Exposure	***TCF3-PBX1* frequency**	***RUNX1-RUNX1T1* frequency**
DMSO control (4 days)	10%	0%
Caffeine 80μM (4 days)	65%*	NA
Benzene 1.5nM (2 days)	5%	25%*
Benzene 48nM (2 days)	NA	30%*
Cotinine 5μM (2 days)	NA	35%*
Folic acid 1nM (4 days)	70%*	NA

*** Fisher's exact test statistically significant <0.05 when comparing the exposure with DMSO control**

NA: RT-PCR was not performed

However, as the translocation events were rare and somewhat random across the exposures, further experiments were conducted in an attempt to optimise the likelihood of observing a translocation to further assess these events in response to the exposures. Two variables were considered to optimise the preliminary model: i) the impact of treating cells immediately or 24 hours after initial seeding, when more cells would be likely to be actively replicating, and ii) single vs daily dosage of exposure, to reflect real life conditions more closely.

When considering the effect of these two variables on cell growth and viability, the patterns were similar across most exposures, but differed across models, Figure 9.1.

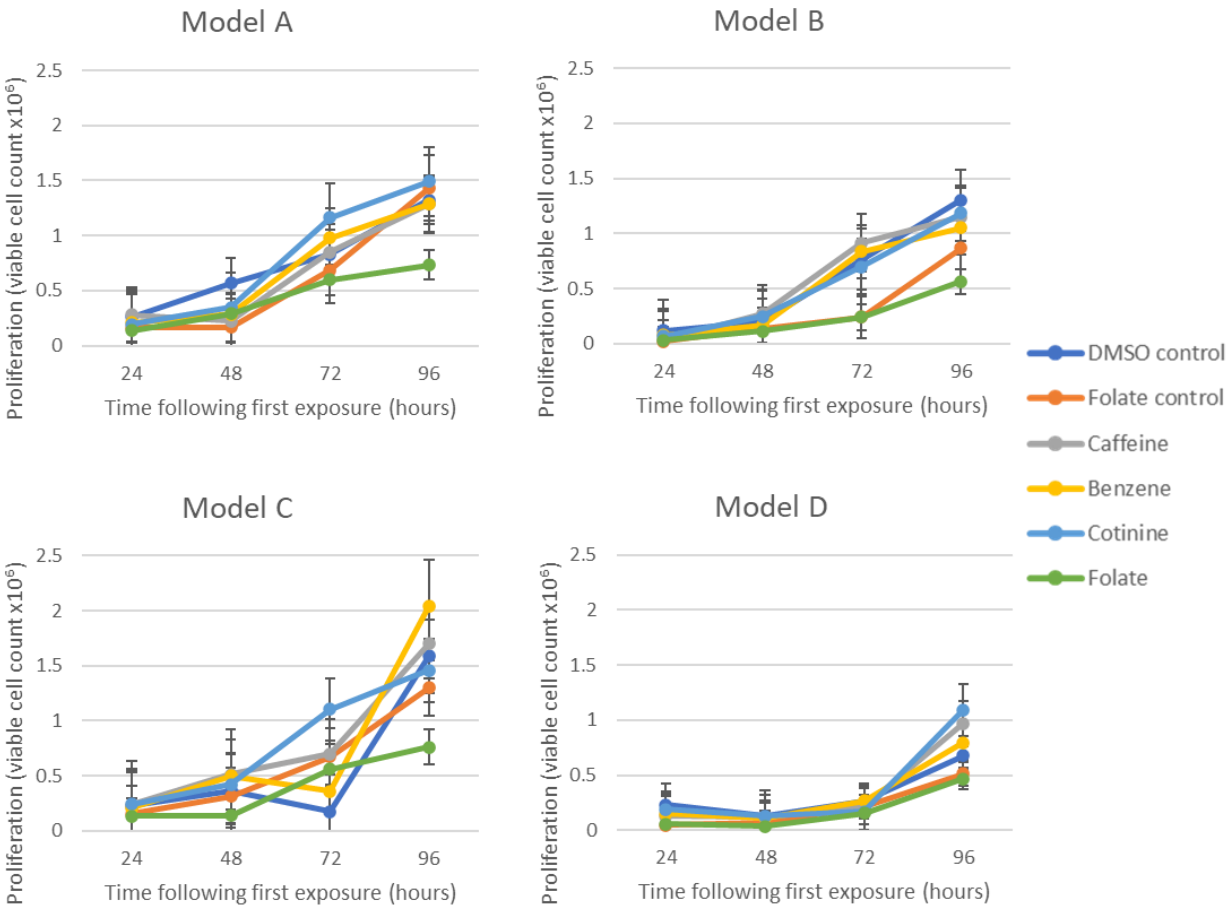


Figure 9.1. Overview of growth across different exposures for each model. Trypan blue counts were averaged for each different exposure (all concentrations included) from model experiments, presented with standard error.

Overall, in Model A (single exposure, 24 hours after seeding), cell growth increased every 24 hours, suggesting dosing after 24 hours allows cells to be more resilient to the exposure. In Model B (single dose at seeding), cells are slow to grow for the first 48 hours, but growth increases after 72 hours, suggesting cells are recovering from the slow growth following exposure. In Model C (daily dose, 24 hours after seeding) growth rates are generally slower for the first 72 hours, but then recover with a cell count higher than the other models after 96 hours, which suggests that although the daily dosing may be impacting growth, the timing of the first dosage after 24 hours growth, allows cells to recover better than if the first dose is administered at seeding. Conversely for Model D (daily dose at seeding), cells were also slow to grow for the first 72 hours across all exposures, with an increase in cell count at 96 hours, however overall cell count at this point was less than other models, which suggests that the daily dose in combination with first exposure at seeding hinders replication and cells are unable to recover as well as in other models. In terms of cell viability, for all exposures except folate, there were no significant decreases compared to the DMSO control observed in any of the models. Folate depletion however, led to a reduction in cell viability compared

to the control across all models, suggesting that, in contrast to cell growth, the type of exposure rather than any of the model conditions that may impact cell viability.

To understand the influence of each variable on translocation induction, Fisher's exact tests were performed for individual exposures to compare exposure at seeding vs. exposure after 24 hours and to compare single dose vs. daily dose of exposure, Table 9.2. The only significant relationship observed was between single dose versus daily dose for cotinine exposure, Table 9.2.

*Table 9.2. Comparison of model variables for each individual exposure. Fisher's exact tests were performed for individual exposures to compare exposure at seeding vs. exposure after 24 hours and to compare single dose vs. daily dose of exposure. P values for each test are presented. Statistically significant relationships are marked * = <0.05*

	Comparison of model variables for each individual exposure	
	Exposure at seeding vs. after 24 hours (p value)	Single dose vs. daily dose (p value)
DMSO controls	0.4414	0.4414
Caffeine (all concentrations)	0.1986	0.5235
Benzene (all concentrations)	0.7644	0.1289
Cotinine (all concentrations)	0.3167	0.0154*
Folate (all depleted concentrations)	0.234	1

*** Fisher's exact test statistically significant <0.05**

The lack of significance across all exposures for each variable may be due to a lack of power due to the rarity of translocations observed. To investigate if there may have been an interaction between the individual variables ANOVA tests were performed. Given the lack of effect observed on between model variables and translocation events for individual exposures, and to increase power to detect any effect results from all exposures were combined for this analysis. The number of translocation events when exposure was administered 24 hours after seeding was significantly higher (p=<0.001) compared to exposure at seeding. Translocation events were also more likely when exposure was administered daily compared with a single dose (p=0.007). The interaction between these 2 variables was analysed using a full factorial model, with a statistically significant relationship observed (p=<0.001). The mean highest number of events were observed in Model C (daily

dose, 24 hours after seeding). Bonferroni post-hoc tests confirmed a significant increase in the number of translocations events (across all exposures) for Model C compared to the other models: A (p=<0.001), B (p=<0.001) and D (p=<0.001), Figure 9.2. This suggests that Model C, with a combination of both variables, has the most optimal conditions required to induce translocation events in response to environmental exposures.

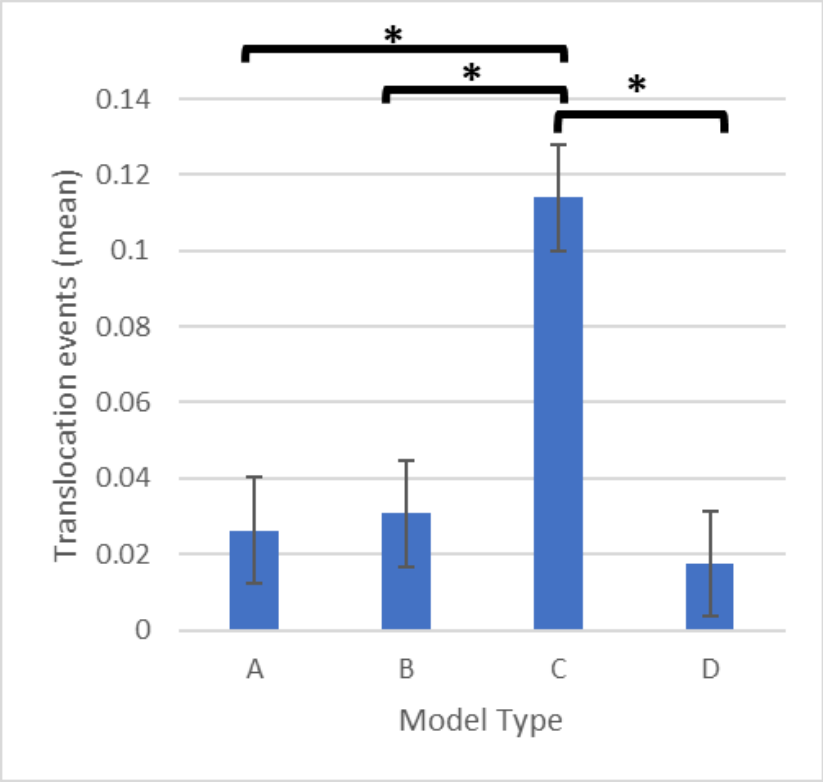


*Figure 9.2. Translocations events across all model types for combined exposures. A one-way ANOVA was performed across the different models, followed by Bonferroni's multiple comparisons test. Significant comparisons are marked * = <0.05.*

Whilst chromosomal translocations were successfully detected in NALM6 cells in response to caffeine, benzene, cotinine and folate, positive translocations were also unexpectedly observed in DMSO control cells, prompting questions regarding the causal relationship between chromosomal translocation observed in response to risk exposures. One potential concern could be the propensity for false positive results from the RT-PCR assay used to assess translocations. Therefore, to validate the capability of RT-PCR assays to detect translocations effectively, FISH analysis was conducted on a selection of caffeine exposed cells and DMSO control cells, using breakapart probes for a single gene within each translocation pair, *TCF3* and *RUNX1*. As such, any translocations in these genes would be detected, not just those involved in the translocation pairs investigated. Of 10 samples selected positive for translocations via RT-PCR, 9 had the same translocation detected via FISH, suggesting a low false positive rate using the RT-PCR method, and therefore that the resultant data were reliable. Of the 3 samples selected negative via RT-PCR for the

translocations measured with FISH probes, only 1 observed translocation events above the normal bone marrow threshold. Although this could be a false negative by RT-PCR, as the breakapart FISH probes can identify any translocations within *TCF3* or *PBX1*, the findings here may be a translocation pair that is not measured in the RT-PCR assay.

FISH analysis did allow for additional information to be generated regarding other abnormalities including translocations, which were found to be significantly higher in caffeine exposed cells compared to DMSO control cells for both *TCF3* and *RUNX1* probes. This may therefore suggest that exposed cells are more likely to display chromosomal damage compared to DMSO treated cells, but further investigation would be warranted for the individual exposures to explore this hypothesis.

A further concern may be that DMSO treatment itself may induce chromosomal translocations, which would impede accurate assessment of the exposure-translocation relationships. To assess this, further investigations exposing cells to caffeine without DMSO and including control cell population with no DMSO or caffeine observed a single translocation event via RT-PCR assays for cells exposed to caffeine with no DMSO present, but no translocations for either no or with DMSO controls, table 4.8. This may suggest that the translocation events observed for exposures in the presence of DMSO may not need to be discounted. However, to determine the true influence of the exposures on translocation induction it would be useful to repeat the experiments for each exposure with no DMSO present where possible, either using a different solvent or removing the solvent altogether, such as using caffeine dissolved in water, cotinine dissolved in ethanol, standard medium diluted in folic acid free medium to give specific folic acid concentration and benzene diluted directly in standard medium.

Given the presence of chromosomal translocations in the control cells exposed to DMSO alone, there is potential that the translocations are occurring spontaneously in the NALM6 cells. Heterogeneity within cell lines compared to primary cells is a concern due to their immortalised nature, as well as stressors from cell growth such as media conditions and physical handling which can cause changes within cell populations [315]. Cancers overall gain mutations to aid cancer progression, and as a leukaemia cell line, NALM6 could be spontaneously gaining genetic and epigenetic alterations [316].

However, given the presence of large numbers of translocation events observed in DMSO controls, particularly for Model C, (daily dose, 24 hours after seeding), it is likely that the presence of DMSO is contributing to translocation induction. Although DMSO, as an organic solvent, is generally considered to be non-genotoxic and protective against DNA damage during cryopreservation, it is not as inert within cells as previously assumed [308, 309, 312,

387]. Often when DMSO causes DNA damage it is at high concentrations, over 4%, which is much higher than the 0.1% v/v used in the current study [309, 312]. However, as a cryoprotectant, DMSO has the ability to enter the nucleus and impact the cell in a variety of ways. DMSO has been shown to effect cellular processes by altering gene expression, as well as disrupting DNA methylation mechanisms [312]. It has also been observed to induce cell differentiation [388]. Studies have also shown that DMSO can influence cell proliferation, cell cycle progression and DNA synthesis at low concentrations [314, 389, 390]. Potentially one of DMSOs most important effects in relation to translocation induction is its influence on DNA topology. DMSO is able to change DNA conformations, with increased frequency of Z-DNA at concentrations as low as 0.1% v/v [314]. Z-DNA is a left-handed helix, which are often fragile, and known to have breakpoint regions nearby, which could potentially lead to chromosomal translocations [123]. All of this evidence suggests that DMSO does have the ability to directly influence chromosomal translocations and is likely to be more frequent when in the presence of defective repair systems and/or exogenous stressors. The effect of DMSO on translocation induction may work in a dose dependent manner given that more translocations events are observed in Model C, where DMSO concentration was cumulatively higher over time, and in particular where 3/6 translocations were observed after 4 daily treatments with DMSO.

Further work is required to fully understand if translocations occur spontaneously in NALM6 cells or in response to DMSO or the other exposures examined here. Experiments using varying doses of DMSO, table 4.8, did not observe any translocations in DMSO or untreated NALM6 cells. As this was only conducted in 6 replicates, it would be beneficial to repeat this for a larger number of replicates to observe a more accurate representation of translocation frequency and the influence of DMSO. This would also be useful to determine if DMSO is dose dependent and at what concentration cells do not exhibit more frequent chromosomal translocations events than would be expected in cells not exposed to DMSO. Translocation events in NALM6 cells with and without DMSO could also be compared to other cell lines with different characteristics, such as a normal karyotype, to further understand the mechanism by which the translocations are occurring.

9.2. Evaluation of methods used to screen for chromosomal translocation induction in response to environmental risk exposures

Human immortalised cell lines are an important tool in research, allowing for the investigation of diseases and mutations at a cellular level [391]. Immortalised cell lines provide a more homogenous working material than primary human cells and they can be produced on a large scale to provide enough replicates to provide statistical power. This is a benefit over using animal models, which have more ethical implications in terms of the

number of subjects needed, especially when investigating a range of exposures and concentrations as with the current study. The NALM6 cell line used is easy to culture, and as a cancer cell line is already prone to mutations, which may provide a background level of susceptibility to induce chromosomal translocations, which are rare occurrences [315, 316]. Although translocations were observed in NALM6 cells, there are some limitations to using NALM6 as a model of translocation induction in response to environmental exposures.

As discussed in Section 4.4, NALM6 cells have a defective mismatch repair response, which may not increase the likelihood of translocation events, as suggested by Abdel-Rahman *et al.* 2005, where restored mismatch repair response in colon cancer cells did not prevent translocations from occurring, suggesting that other repair mechanisms may be more important in developing translocations [284]. This may have reduced the number of translocation events that occurred in the NALM6 cell line model, making cells less susceptible to translocation events. To understand the impact of specific DDRs on translocation events, the presence of translocation events could be measured in cell lines that contain normal and defective DDR, such as TK6 cells alone, or with induced defective HR and/or NHEJ repair responses [317].

Furthermore, as NALM6 is of B-cell lineage, it is interesting that as well as the *TCF3-PBX1* translocation that is common in B-cell ALL, induction of the *RUNX1-RUNX1T1* translocation is also observed, which is common in AML [3]. To understand if translocations can be induced in cell types they are not commonly associated with, the presence of translocation events could be measured in a variety of cell types, where respective translocations are common and uncommon, including: a myeloid leukaemia cell line, such as KG1; a T-ALL cell line, such as JURKAT; as well as other cancer types such as breast and colorectal cancers [392].

However, cell line models such as NALM6, lack physiological complexity and are not a true representation of real life exposure, and therefore other, more complex model systems may better reflect that, such as organoid models. Environmental exposure in animal models, such as mice, may provide a more realistic response in line with human physiology, however they require more intensive primary establishment, maintenance, and cost, which does not make them suitable as a screening tool [393]. Organoid models however may be a more useful system, with a more accurate representation of human exposure to environmental risk factors [393]. Organoids can be developed to mimic the bone marrow microenvironment in which haematopoietic cells would develop using induced pluripotent stem cells, representing true human physiology [393]. As they are quicker, cheaper, easier to scale and perhaps more ethically sound than animal models, organoids would be useful as a screening tool for

chromosomal translocation induction in response to environmental exposures [393]. Furthermore, analysis of some already established organoid models have provided evidence that some molecular characteristics such as DNA methylation are maintained from their equivalent primary cells, which would be beneficial given the potential role of epigenetics in the 2-hit hypothesis of leukaemia development [394]. However, development of organoids is still in its infancy, with a lack of robust protocols and heterogeneity across organoid models challenges that would need to be addressed [393].

During preliminary optimisation of the NALM6 cell line model, 2 cell densities were investigated for growth over 96 hours, with $0.5x10^6$ cells/ml maintaining the highest cell density at 96 hours. Experiments with caffeine, benzene and cotinine measured translocations events 48 and 96 hours after exposure, with folate experiments measuring translocation events at 96 hours only. However, by 96 hours, NALM6 cells at $0.5x10^6$ cells/ml were no longer growing exponentially. As the exposures have been shown to act most in proliferating cells, it may be useful to investigate a larger range of cell densities to find the optimum density that would allow for cells to be growing exponentially to measure translocation events [17, 148, 351]. The impact of the exposure on cell proliferation should also be considered, and a wider range of cell densities across different time points exposed to each concentration of caffeine, benzene, cotinine, and folic acid, would provide a more accurate representation of cell growth to inform what conditions would be optimum for RNA extraction during exponential growth for each exposure.

Furthermore, as discussed in Section 4.4, the cell growth and viability data presented in this study is only measured in 1 biological replicate, and to show any statistically significant changes, these experiments would need to be repeated in at least 2 biological replicates. Trypan blue assays are inexpensive and provide a visual result of live versus dead cells [254]. However, at high throughput trypan blue is time consuming and prone to human error due to manual counting and can only provide a small snapshot of the cell population. The resazurin assay on the other hand is also inexpensive, but quick to perform at high throughput [254]. However, the resazurin assay measures cell viability through the metabolic conversion of blue oxidised resazurin to red reduced resazurin, and as such, exposure to caffeine, benzene, cotinine and folic acid may affect metabolic activity, as discussed in each chapter, which could interfere with the assay. The Cyquant assay, although more expensive than trypan blue and resazurin, is quick to perform, and even allows for samples to be stored and measured later, offering more capability for high throughput experiments [395]. The Cyquant assay works through measuring a green fluorescent dye which binds to cellular nucleic acids, and so is not impacted by metabolic differences. Across all of the exposures, the resazurin and Cyquant assays showed similar results for cell viability, and give the

strengths and weaknesses of both assays, it would be satisfactory to just use the Cyquant assay for all measurements.

The RT-PCR assays used to detect *TCF3-PBX1* and *RUNX1-RUNX1T1* translocation events were successful at detecting translocations in at least 1% of cells. RT-PCR assays can be used to screen many samples with relative ease, and as the assay uses cDNA, the same sample can be analysed for multiple translocation events [396]. However, there were some limitations to using RT-PCR assays in this study. Only 2 of the 6 translocation pairs identified as common in childhood leukaemia were successfully developed into RT-PCR assays, chapter 3. This significantly reduced the ability to screen for the majority of chromosomal translocations that are observed in childhood leukaemia, and further optimisation would be required to expand the scope of the RT-PCR assays. RT-PCR was selected as a screening method due to the use of RNA, which can be converted to cDNA allowing for selective amplification of exons across translocation breakpoints, rather than trying to amplify large sections of DNA which are usually more difficult. However, RNA is much less stable than DNA and is degraded by Rnases, requiring careful handling during extraction, with degradation consequently leading to false negatives. Due to the multiple steps involved in RNA extraction and cDNA synthesis prior to RT-PCR, there are many opportunities where contamination can occur, which can lead to false positives. As such, Dnase was used to remove any DNA contamination and a PCR hood was used to reduce the possibility of contamination. Negative (H_2O template) controls were used during RT-PCR to check for contamination and positive cell line controls were used to ensure the assay was performed correctly. Another difficulty of using RT-PCR is the requirement that any fusion genes need to produce an mRNA transcript, so if breakpoints result in loss of transcription this will not be detected using RT-PCR. For this reason, FISH assays are utilised as a secondary confirmation of positive results, allowing for direct visualisation of chromosomal translocations, so would be able to confirm false positives or negatives. Indeed, FISH analysis confirmed that the majority of samples identified as positive by RT-PCR were positive. However, due to the cost and time required to analyse FISH assays, it would be unrealistic to use FISH as a screening tool in place of RT-PCR.

As an alternative to RT-PCR and FISH assays for screening cells to identify chromosomal translocations, a Genomic Inverse PCR for Exploration of Ligated Breakpoints (GIPFEL) method could be utilised [397]. This method uses genomic DNA, which is more stable template than RNA, which is digested and ligated to form enriched circular DNA. As translocation breakpoints are often within specific chromosomal regions, primers were designed to PCR amplify the regions where translocations form these known ligation joints, based on the restriction digest used for each translocation [397]. For positive samples, the

size of the product was confirmed by gel electrophoresis and then Sanger sequenced [57, 397]. This method was successfully used to detect *ETV6-RUNX1* translocation frequency in 1000 newborn samples, and could be applied for use in the current screening model [56]. However, the GIPFEL method is currently only available for the *ETV6-RUNX1* translocation, is a very time consuming method due to the numerous steps involved and has a high cost per sample of around €40.

Although translocation events were observed in the study, these are the end result of the damage caused and the mechanism by which the exposures contribute to translocation induction is still unclear. Further investigation is required to assess the intermediatory stages of DNA damage, such as the formation of DSBs, to understand the kinetics of DSB and repair when exposed to leukaemia risk factors, along with the appearance and persistence of these breaks. Methods such as the neutral COMET assay are used to detect DSBs *in vitro* through the visualisation of fragmented DNA in metaphase cells following agarose gel electrophoresis and staining [398]. The presence and absence of fragmented DNA could be measured at different time points following exposure to assess at which point DSBs occur and how long it takes for repair [398]. Chromosomal damage can also be measured using a micronuclei assay, which detects micronuclei caused by DNA breaks and instability in dividing cells [399]. The micronuclei assay is often used to investigate the cytogenetic damage caused by environmental exposures [399]. These methods assess damage at a single cell level, however they do have some limitations, such as a lack of information on the specific damage and location. Combination of these assays with FISH is suggested to overcome many of these limitations by providing deeper insight into origin and location of the damage [400]. A more recent method of detecting DSBs is through the measurement of γ-H2AX foci, which are generated through the phosphorylation of the histone protein H2AX following DSBs, which then recruit DNA repair proteins for repair by HR and NHEJ, and is then dephosphorylated once repair is complete [401]. The γ-H2AX assay can also be adapted for high throughput and sensitive analysis by using flow cytometry [402]. The current study also only took into consideration the impact of individual exposures, whereas in real life, these exposures are often compounding, such as benzene and cotinine, which would both be present in cigarette smoke. Further work could be performed to investigate the impact of combining exposures on translocation events. For example, a study on the impact of benzene exposure in mice that were folate deficient found an increased level of micronuclei, indicating chromosomal damage [126]. As such, combining exposures may lead to an increased likelihood of translocation events, and would better reflect real life exposure

9.3. Evaluation of DNA methylation as a mediating mechanism between risk exposure and ALL subtype

There is a lack of epidemiological evidence currently available to link environmental exposures with specifical leukaemia subtypes due to a lack of power because of the rarity of leukaemia, and more specifically individual subtypes. The “meet in the middle” approach, pioneered by Vineis and Perera [403], is beneficial when data is limited, by using a biomarker of susceptibility common to exposure and disease to identify potential associations. As DNA methylation is found to be altered in ALL and in response to environment it provides a suitable marker for investigation using the “meet in the middle” approach [78, 80, 125, 359-361]. The analysis of altered methylation at CpG level presented in chapter 8 supports previous findings at gene loci, with significant overlaps between methylation altered in constitutive ALL and maternal radiation exposure, maternal alcohol intake and maternal plasma folate levels during pregnancy, Table 8.3. A new finding here was a significant overlap between methylation altered in constitutive ALL and reported colds at 6 months, Table 8.3. This association was not found in the previous gene loci analysis, highlighting the potential utility of considering the CpG level approach taken here. However, the results for constitutive ALL are constrained by the requirement for DMCs to be present in all 10 subtypes reported in the Nordlund *et al.* 2013 study [361]. Likewise, many ALL subtypes are not categorised within the Nordlund *et al.* 2013 analysis, likely due to their rarity, which suggests they may be missing from analysis and that the constitutive ALL results might not be a true representation of all ALL subtypes [404]. Analysis could also be performed using EWAS data for AML to investigate the interactions between exposures and AML and its subtypes [405]. Leukaemia is a heterogenous disease, with many differences between subtypes, which may be masked by only looking at the constitutive DMCs across subtypes. This is especially important when considering T-ALL, which is derived from T-cell, a different cell type to the other 9 subtypes which are of B-cell origin, and as such, may have differing methylation profiles, which is suggested by the higher percentage of unique to subtype KEGG pathways/GO processes identified compared to B-cell subtypes, Tables 8.7-8 [377, 378]. Subtype specific associations with overlapping methylation were found for exposures previously identified as significant in constitutive ALL analysis, with maternal radiation exposure, maternal alcohol intake, maternal plasma folate levels and reported cold at 6 months, Tables 8.5-6. A significant subtype specific association was also observed for smoking throughout pregnancy, which was not observed for constitutive ALL, and would have been undetected if subtype analysis had not been performed.

As many of the exposures are measured through self-reporting or by using a proxy, this can lead to reporting bias [86-88]. Future work to unpin the contribution of DNA methylation as a mediating mechanism for each exposure could be to use direct measures of exposures, such as exposing cells to physiologically relevant levels of the exposure and measuring

DMCs so that results are more consistent and accurate. As more accurate assessments of exposure-methylation relationships are available, it would be beneficial for this analysis to be continuously updated to increase the robustness of associations. This is important given the findings reported here for folate, where no association was observed for the smaller EWAS of maternal folic acid supplementation, but a high degree of association was observed for the meta-analysis data of maternal plasma folate levels.

By identifying intermediate biomarkers of disease, it could prove useful in assisting disease prevention, whereby at risk individuals can be identified and monitored. DNA methylation analysis is becoming an important tool as an early biomarker of disease, with a randomised trial taking place in the UK investigating the use of the Galleri assay as a screening test for over 50 types of cancer [406]. The Galleri assay uses cell-free DNA obtained from patient blood samples to measure DNA methylation patterns, with the ability to identify cancer origins with a high specificity [406]. Screening methods lead to early detection, by which prevention methods can be put in place and treatment can have the biggest impact, which has been shown to reduce overall mortality [407-409].

9.4. Implication of this work in the wider context

Understanding the aetiology of childhood leukaemia is important for providing advice and guidance on the prevention of the disease. Although overall and 5 year survival rates are reaching over 90%, there are still high rates of relapse, and with intensive treatment regimes, side effects can include cardiovascular disease, neurocognitive impairment, fertility issues and secondary cancers [1, 410]. With incidence rates of childhood leukaemia increasing year on year it suggests that environment may be playing a role in disease development, and as many of these exposures are modifiable, focussing on prevention will work towards reducing the number of deaths and long term consequences of leukaemia treatment.

Epidemiological studies play an important role in determining risk factors for childhood leukaemia [14]. However, association are often conflicting and limited due to a lack of power, particularly for AML which is much less frequent than ALL. Furthermore, accurate measures of exposures can often be impacted by reporting bias, and as such “safe” levels of exposure are difficult to assess [86-88]. Many risk exposures associated with childhood leukaemia also have other adverse health impacts *in utero* which has resulted in specific guidance during pregnancy for caffeine and alcohol intake, folic acid supplementation and advice to cease smoking [100, 140, 167, 411]. However, there is little evidence that these guidelines are sufficient to prevent the development of childhood leukaemia. One of the main objectives

of this study was to understand the minimum level of exposure that is capable of inducing childhood leukaemia associated chromosomal translocations.

The current UK guidelines for maternal caffeine intake during pregnancy is 200mg caffeine per day, which is equivalent to 2 cups of coffee [140]. Findings in this study have shown translocation induction at caffeine exposures equivalent to 2 cups of coffee, which suggests that the current caffeine recommendation may be too high to protect against chromosomal translocations involved in childhood leukaemia. A recent meta-analysis of adverse pregnancy outcomes following maternal caffeine intake found no threshold for safe consumption of caffeine [412]. As pregnancy progresses, caffeine metabolism decreases which means that in later stages of pregnancy the same recommendation of caffeine leads to longer exposure at higher concentrations compared to earlier stages of pregnancy, so trimester based recommendations may be more appropriate [294]. Further investigation is required to provide advice on caffeine intake during pregnancy and pre-conception to prevent development of chromosomal translocations.

Normal serum folate levels are generally considered to be between 10nM and 50nM [107, 343-345]. Translocation events were observed at this concentration, which suggests that the lower boundary of “normal” folate levels may not be sufficient to protect against childhood leukaemia associated chromosomal translocations. Folic acid supplementation for pregnant women in the UK is recommended at 400µg per day. A study that measured the average serum folate levels of women in Colombia following supplementation with 400µg per day folic acid, found that average levels increased from 24.2nM to 45.8nM [413]. Mandatory folic acid fortification is implemented in many countries around the world, and has recently been introduced in the UK [414]. However, the proposed levels for the UK are 60-100µg, which is lower than that recommended supplementation in pregnancy [414]. In Brazil, folic acid fortification was introduced in 2004 with requirements of concentrations of 150µg [415]. Following fortification, average serum folate levels in women increased from 11.2nM to 20.3nM [415]. It is important to understand what folate levels are sufficient to provide protection against translocation induction, and further investigation would be beneficial to determine what concentration within the “normal” range is optimal for protection. Similarly, there is limited evidence on the impact of too much folic acid on childhood leukaemia. As the findings in this study did observe a translocation at 200nM, a very high folic acid concentration, this highlights the importance of finding a safe range of low and high folic acid for prevention of chromosomal translocations. Personalised recommendations would be beneficial by tracking plasma folate levels throughout pregnancy.

Smoking cessation is recommended in pregnancy, however nicotine dependence does make quitting difficult for many people [167, 416]. Cessation strategies include nicotine replacement therapies (i.e., nicotine gum), and the increasingly popular electronic nicotine delivery systems (i.e., e-cigarettes), which are viewed as safer alternatives to traditional cigarette smoking [417]. Although these systems contain less carcinogens than tobacco cigarettes, they all still contain nicotine, and some e-cigarettes also contain benzene [318]. In this study, translocation events have been observed in response to exposure to both cotinine (nicotine metabolite) and benzene at levels equivalent to non-smokers or second hand smoke exposure, which suggests exposure to any traditional cigarettes, nicotine replacement therapy, e-cigarettes and second hand smoke during pregnancy may lead to chromosomal translocations. Further evidence is required to support these findings and provide further advice and guidance on smoking/nicotine exposure in pregnancy and pre-conception. Translocation events were also observed at levels equivalent to those found in occupational benzene workers who smoke. Furthermore, benzene is also found in air pollution, with benzene concentrations in the air correlating with benzene levels in the blood (for non-smokers) [418]. Higher levels of benzene in both the air and blood are found in urban areas compared to rural areas [418]. However more evidence is required to understand the influence of air pollution on benzene concentration *in utero*, which could be provided by measuring benzene concentration in cord blood of non-smokers and measuring the corresponding ambient benzene concentration from individual locations. Further investigation is required to determine the benzene limit at which chromosomal translocations are induced, which could provide guidance on air pollution, emission, and workplace exposure policies. Benzene metabolites are also thought to play a large role in the toxicity of benzene; however, the exact mechanism and contribution of these metabolites is still to be determined [217]. Exposure to benzene metabolites *in vitro* has been shown to cause DNA damage, strand breaks and topoisomerase inhibition, along with an ability to reach the bone marrow [217]. Benzene metabolism can be affected by the level of exposure to benzene, with preference to different metabolites at different benzene doses, and can also be affected by polymorphisms in key metabolism enzymes [184]. Therefore, it would also be beneficial to screen for chromosomal translocations in response to benzene metabolites, which will provide further evidence on the mechanistic understanding of benzene contribution to translocation induction.

The use of NALM6 as a cell line model of environmental exposure allows for a mechanistic approach to understanding the impact of caffeine, benzene, cotinine and folic acid on chromosomal translocation induction. However, as NALM6 in not a true representation of human exposure *in utero*, the results presented aren't directly translatable to human subject

and further evidence is required from more relatable systems such as organoids or animal models to determine the threshold of translocation induction for each exposure. Nevertheless, the results presented in this thesis show that environmental exposures may play an important role in the development of childhood leukaemia, through induction of chromosomal translocations and changes to DNA methylation.

10 References

1. UK, C.w.C. *Childhood Cancer Facts & Figures*. 14/11/2018]; Available from: https://www.childrenwithcancer.org.uk/childhood-cancer-info/childhood-cancer-facts-figures/.
2. Allison, K. and G. Sledge, *Heterogeneity and Cancer* Oncology, 2014. **29**(9): p. 772-778.
3. Greaves, M.F. and J. Wiemels, *Origins of chromosome translocations in childhood leukaemia.* Nat Rev Cancer, 2003. **3**(9): p. 639-49.
4. Seth, R. and A. Singh, *Leukemias in Children.* Indian J Pediatr, 2015. **82**(9): p. 817-24.
5. Suguna, E., et al., *Acute Myeloid Leukemia: Diagnosis and Management Based on Current Molecular Genetics Approach.* Cardiovasc Hematol Disord Drug Targets, 2018. **18**(3): p. 199-207.
6. Malczewska, M., et al., *Recent Advances in Treatment Options for Childhood Acute Lymphoblastic Leukemia.* Cancers (Basel), 2022. **14**(8).
7. Pedrosa, F., et al., *Reduced-dose intensity therapy for pediatric lymphoblastic leukemia: long-term results of the Recife RELLA05 pilot study.* Blood, 2020. **135**(17): p. 1458-1466.
8. Care, L. *Childhood Acute Lymphoblastic Leukaemia (C-ALL)*. 14/12/2022]; Available from: https://www.leukaemiacare.org.uk/support-and-information/information-about-blood-cancer/blood-cancer-information/leukaemia/childhood-acute-lymphoblastic-leukaemia/.
9. Bhojwani, D., J.J. Yang, and C.H. Pui, *Biology of childhood acute lymphoblastic leukemia.* Pediatr Clin North Am, 2015. **62**(1): p. 47-60.
10. Greaves, M., *Childhood leukaemia.* BMJ, 2002. **324**: p. 283-287.
11. Rogers, R.L. and D.L. Hartl, *Chimeric genes as a source of rapid evolution in Drosophila melanogaster.* Mol Biol Evol, 2012. **29**(2): p. 517-29.
12. Malinowska-Ozdowy, K., et al., *KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia.* Leukemia, 2015. **29**(8): p. 1656-67.
13. Paulsson, K. and B. Johansson, *High hyperdiploid childhood acute lymphoblastic leukemia.* Genes, Chromosomes and Cancer, 2009. **48**(8): p. 637-660.
14. Timms, J.A., et al., *DNA methylation as a potential mediator of environmental risks in the development of childhood acute lymphoblastic leukemia.* Epigenomics, 2016. **8**(4): p. 519-536.
15. Hastings, P.J., et al., *Mechanisms of change in gene copy number.* Nat Rev Genet, 2009. **10**(8): p. 551-64.
16. Roukos, V. and M. Misteli, *The biogenesis of chromosome translocations.* Nature Cell Biology, 2014. **16**(4): p. 293-300.
17. So, A., et al., *Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells.* FEBS J, 2017. **284**(15): p. 2324-2344.
18. Roth, D.B., *V(D)J Recombination: Mechanism, Errors, and Fidelity.* Microbiol Spectr, 2014. **2**(6).
19. Heyer, W.D., K.T. Ehmsen, and J. Liu, *Regulation of homologous recombination in eukaryotes.* Annu Rev Genet, 2010. **44**: p. 113-39.
20. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annu Rev Biochem, 2010. **79**: p. 181-211.
21. Deriano, L. and D.B. Roth, *Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage.* Annu Rev Genet, 2013. **47**: p. 433-55.
22. Ramsden, D.A. and A. Nussenzweig, *Mechanisms driving chromosomal translocations: lost in time and space.* Oncogene, 2021. **40**(25): p. 4263-4270.
23. Zhang, Y., et al., *The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers.* Adv Immunol, 2010. **106**: p. 93-133.

24. Iarovaia, O.V., et al., *Dynamics of double strand breaks and chromosomal translocations.* Mol Cancer, 2014. **13**: p. 249.
25. Roukos, V., B. Burman, and T. Misteli, *The cellular etiology of chromosome translocations.* Curr Opin Cell Biol, 2013. **25**(3): p. 357-64.
26. Ezoe, S., *Secondary leukemia associated with the anti-cancer agent, etoposide, a topoisomerase II inhibitor.* Int J Environ Res Public Health, 2012. **9**(7): p. 2444-53.
27. Deweese, J.E. and N. Osheroff, *The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing.* Nucleic Acids Res, 2009. **37**(3): p. 738-48.
28. Montecucco, A., F. Zanetta, and G. Biamonti, *Molecular Mechanisms Of Etoposide.* EXCLI Journal, 2015. **14**: p. 95-108.
29. Bohlander, S.K., *ETV6: a versatile player in leukemogenesis.* Semin Cancer Biol, 2005. **15**(3): p. 162-74.
30. Sood, R., Y. Kamikubo, and P. Liu, *Role of RUNX1 in hematological malignancies.* Blood, 2017. **129**(15): p. 2070-2082.
31. Rodríguez-Hernández, G., et al., *Modeling the process of childhood ETV6-RUNX1 B-cell leukemias.* Oncotarget, 2017. **8**(60): p. 102674-102680.
32. Bach, C. and R.K. Slany, *Molecular pathology of mixed-lineage leukemia.* Future Oncology, 2009. **5**(8): p. 1271-1281.
33. McMahon, K.A., et al., *Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal.* Cell Stem Cell, 2007. **1**(3): p. 338-45.
34. Meyer, C., et al., *The MLL recombinome of acute leukemias in 2013.* Leukemia, 2013. **27**(11): p. 2165-76.
35. Zhou, C.C., et al., *AFF1 and AFF4 differentially regulate the osteogenic differentiation of human MSCs.* Bone Res, 2017. **5**: p. 17044.
36. Felice, M.S., et al., *Prognostic impact of t(1;19)/ TCF3-PBX1 in childhood acute lymphoblastic leukemia in the context of Berlin-Frankfurt-Münster-based protocols.* Leuk Lymphoma, 2011. **52**(7): p. 1215-21.
37. Zhou, B., et al., *The clinical outcomes and genomic landscapes of acute lymphoblastic leukemia patients with E2A-PBX1: A 10-year retrospective study.* Am J Hematol, 2021. **96**(11): p. 1461-1471.
38. Kęsy, J. and D. Januszkiewicz-Lewandowska, *Genes and childhood leukemia.* Postepy Hig Med Dosw (Online), 2015. **69**: p. 302-8.
39. Aspland, S.E., H.H. Bendall, and C. Murre, *The role of E2A-PBX1 in leukemogenesis.* Oncogene, 2001. **20**(40): p. 5708-17.
40. Bain, G. and C. Murre, *The role of E-proteins in B- and T-lymphocyte development.* Semin Immunol, 1998. **10**(2): p. 143-53.
41. Sanyal, M., et al., *B-cell development fails in the absence of the Pbx1 proto-oncogene.* Blood, 2007. **109**(10): p. 4191-9.
42. Haider, M.Z. and F. Anwer, *Genetics, Philadelphia Chromosome*, in *StatPearls*. 2022: Treasure Island (FL).
43. Inaba, H. and C.G. Mullighan, *Pediatric acute lymphoblastic leukemia.* Haematologica, 2020. **105**(11): p. 2524-2539.
44. Slayton, W.B., et al., *Dasatinib Plus Intensive Chemotherapy in Children, Adolescents, and Young Adults With Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia: Results of Children's Oncology Group Trial AALL0622.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2018. **36**(22): p. 2306-2314.
45. Martinez-Climent, J.A., *Molecular cytogenetics of childhood hematological malignancies.* Leukemia, 1997. **11**(12): p. 1999-2021.
46. Chowdhury, T. and H.J. Brady, *Insights from clinical studies into the role of the MLL gene in infant and childhood leukemia.* Blood Cells Mol Dis, 2008. **40**(2): p. 192-9.

47. Hess, J.L., *MLL: a histone methyltransferase disrupted in leukemia.* Trends Mol Med, 2004. **10**(10): p. 500-7.
48. Lin, C., et al., *AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia.* Mol Cell, 2010. **37**(3): p. 429-37.
49. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation.* Genes Dev, 2011. **25**(13): p. 1345-58.
50. Leblanc, T. and R. Berger, *Molecular cytogenetics of childhood acute myelogenous leukaemias.* Eur J Haematol, 1997. **59**(1): p. 1-13.
51. Kumar, P., et al., *Leukemia associated RUNX1T1 gene reduced proliferation and invasiveness of glioblastoma cells.* J Cell Biochem, 2021. **122**(11): p. 1737-1748.
52. Grinev, V.V., et al., *RUNX1/RUNX1T1 mediates alternative splicing and reorganises the transcriptional landscape in leukemia.* Nature communications, 2021. **12**(1): p. 520-520.
53. Liquori, A., et al., *Acute Promyelocytic Leukemia: A Constellation of Molecular Events around a Single PML-RARA Fusion Gene.* Cancers (Basel), 2020. **12**(3).
54. Lo-Coco, F. and S.K. Hasan, *Understanding the molecular pathogenesis of acute promyelocytic leukemia.* Best Practice & Research Clinical Haematology, 2014. **27**(1): p. 3-9.
55. Wiemels, J.L., et al., *Prenatal origin of acute lymphoblastic leukaemia in children.* The Lancet, 1999. **354**(9189): p. 1499-1503.
56. Schafer, D., et al., *Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening.* Blood, 2018. **131**(7): p. 821-826.
57. Hein, D., A. Borkhardt, and U. Fischer, *Genomic Inverse PCR for Screening of Preleukemic Cells in Newborns (GIPFEL Technology).* Methods Mol Biol, 2021. **2185**: p. 113-134.
58. Zuna, J., et al., *ETV6/RUNX1 (TEL/AML1) is a frequent prenatal first hit in childhood leukemia.* Blood, 2011. **117**(1): p. 368-369.
59. Greaves, M., *A causal mechanism for childhood acute lymphoblastic leukaemia.* Nat Rev Cancer, 2018. **18**(8): p. 471-484.
60. Greaves, M.F., *Speculations on the cause of childhood acute lymphoblastic leukemia.* Leukemia, 1988. **2**(2): p. 120-5.
61. Greaves, M.F. and L.C. Chan, *IS SPONTANEOUS MUTATION THE MAJOR 'CAUSE' OF CHILDHOOD ACUTE LYMPHOBLASTIC LEUKAEMIA? THE PAUCITY OF EVIDENCE FOR ENVIRONMENTAL AND GENETIC FACTORS IN ACUTE LYMPHOBLASTIC LEUKAEMIA.* British journal of haematology, 1986. **64**(1): p. 1-13.
62. Yuan, Y., et al., *AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10398-403.
63. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
64. Lipsick, J., *A History of Cancer Research: Tumor Suppressor Genes.* Cold Spring Harb Perspect Biol, 2020. **12**(2).
65. Gilliland, D.G. and J.D. Griffin, *The roles of FLT3 in hematopoiesis and leukemia.* Blood, 2002. **100**(5): p. 1532-42.
66. Higuchi, M., et al., *Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia.* Cancer Cell, 2002. **1**(1): p. 63-74.
67. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
68. Whitehead, T.P., et al., *Childhood Leukemia and Primary Prevention.* Curr Probl Pediatr Adolesc Health Care, 2016. **46**(10): p. 317-352.
69. Mori, H., et al., *Chromosome translocations and covert leukemic clones are generated during normal fetal development.* Proc Natl Acad Sci USA, 2002. **99**(12): p. 8242-8247.

70. Maia, A.T., et al., *Molecular tracking of leukemogenesis in a triplet pregnancy.* Blood, 2001. **98**(2): p. 478-482.
71. Greaves, M.F., et al., *Leukemia in twins: lessons in natural history.* Blood, 2003. **102**(7): p. 2321-33.
72. Zuna, J., et al., *Pre-natal, Clonal Origin of Acute Lymphoblastic Leukaemia in Triplets.* Leukemia & Lymphoma, 2003. **44**(12): p. 2099-2102.
73. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
74. Phillips, T., *The role of methylation in gene expression.* Nature Education, 2008. **1**(1): p. 116.
75. Lim, W.-J., et al., *Identification of DNA-Methylated CpG Islands Associated With Gene Silencing in the Adult Body Tissues of the Ogye Chicken Using RNA-Seq and Reduced Representation Bisulfite Sequencing.* Frontiers in Genetics, 2019. **10**.
76. Razin, A. and H. Cedar, *DNA methylation and gene expression.* Microbiol Rev, 1991. **55**(3): p. 451-8.
77. Meng, H., et al., *DNA methylation, its mediators and genome integrity.* Int J Biol Sci, 2015. **11**(5): p. 604-17.
78. Nishiyama, A. and M. Nakanishi, *Navigating the DNA methylation landscape of cancer.* Trends Genet, 2021. **37**(11): p. 1012-1027.
79. Davidsson, J., et al., *The DNA methylome of pediatric acute lymphoblastic leukemia.* Human Molecular Genetics, 2009. **18**(21): p. 4054-4065.
80. Chatterton, Z., et al., *Epigenetic deregulation in pediatric acute lymphoblastic leukemia.* Epigenetics, 2014. **9**(3): p. 459-67.
81. Yang, X., et al., *Secular trends in the incidence and survival of all leukemia types in the United States from 1975 to 2017.* J Cancer, 2021. **12**(8): p. 2326-2335.
82. WHO. *Incidence of childhood leukaemia, an ENHIS fact sheet*. 2009 11/08/2020]; Available from: https://www.euro.who.int/en/health-topics/noncommunicable-diseases/cancer/publications/2009/4.1-incidence-of-childhood-leukaemia,-an-enhis-fact-sheet.
83. Schuz, J. and F. Erdmann, *Environmental Exposure and Risk of Childhood Leukemia: An Overview.* Arch Med Res, 2016. **47**(8): p. 607-614.
84. Giddings, B.M., et al., *Childhood leukemia incidence in California: High and rising in the Hispanic population.* Cancer, 2016. **122**(18): p. 2867-75.
85. Whitehead, T.P., et al., *Polybrominated diphenyl ethers in residential dust: sources of variability.* Environ Int, 2013. **57-58**: p. 11-24.
86. Kesmodel, U.S., *Information bias in epidemiological studies with a special focus on obstetrics and gynecology.* Acta Obstet Gynecol Scand, 2018. **97**(4): p. 417-423.
87. Spector, L.G., et al., *Prenatal tobacco exposure and cotinine in newborn dried blood spots.* Pediatrics, 2014. **133**(6): p. e1632-8.
88. Taghavi, T., et al., *Cigarette consumption and biomarkers of nicotine exposure during pregnancy and postpartum.* Addiction, 2018. **113**(11): p. 2087-2096.
89. Zhao, R., L.H. Matherly, and I.D. Goldman, *Membrane transporters and folate homeostasis: intestinal absorption and transport into systemic compartments and tissues.* Expert reviews in molecular medicine, 2009. **11**: p. e4-e4.
90. Nazki, F.H., A.S. Sameer, and B.A. Ganaie, *Folate: metabolism, genes, polymorphisms and the associated diseases.* Gene, 2014. **533**(1): p. 11-20.
91. Zhao, R., et al., *Mechanisms of membrane transport of folates into cells and across epithelia.* Annual review of nutrition, 2011. **31**: p. 177-201.
92. Zhao, R. and I.D. Goldman, *Folate and thiamine transporters mediated by facilitative carriers (SLC19A1-3 and SLC46A1) and folate receptors.* Mol Aspects Med, 2013. **34**(2-3): p. 373-85.

93. Matherly, L.H., Z. Hou, and A. Gangjee, *The promise and challenges of exploiting the proton-coupled folate transporter for selective therapeutic targeting of cancer.* Cancer chemotherapy and pharmacology, 2018. **81**(1): p. 1-15.
94. Matherly, L.H., M.R. Wilson, and Z. Hou, *The major facilitative folate transporters solute carrier 19A1 and solute carrier 46A1: biology and role in antifolate chemotherapy of cancer.* Drug metabolism and disposition: the biological fate of chemicals, 2014. **42**(4): p. 632-649.
95. Clare, C.E., et al., *One-Carbon Metabolism: Linking Nutritional Biochemistry to Epigenetic Programming of Long-Term Development.* Annu Rev Anim Biosci, 2019. **7**: p. 263-287.
96. Cantarella, C.D., et al., *Folate deficiency as predisposing factor for childhood leukaemia: a review of the literature.* Genes Nutr, 2017. **12**: p. 14.
97. Duthie, S.J., G. Grant, and S. Narayanan, *Increased uracil misincorporation in lymphocytes from folate-deficient rats.* Br J Cancer, 2000. **83**(11): p. 1532-7.
98. Duthie, S.J. and A.R. Hawdon, *DNA instability (strand breakage, uracil misincorporation, and defective repair) is increased by folic acid depletion in human lymphocytes in vitro.* The FASEB Journal, 1998. **12**: p. 1491 - 1497.
99. Castillo-Acosta, V.M., et al., *Increased uracil insertion in DNA is cytotoxic and increases the frequency of mutation, double strand break formation and VSG switching in Trypanosoma brucei.* DNA Repair, 2012. **11**(12): p. 986-995.
100. *Institute of Medicine. Food and Nutrition Board. Dietary Reference Intakes: Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline*. 1998, Washington, DC: National Academy Press.
101. Allen, L.H., *Causes of Vitamin B12 and Folate Deficiency.* Food and Nutrition Bulletin, 2008. **29**(2_suppl1): p. S20-S34.
102. Ma, J., et al., *Methylenetetrahydrofolate reductase polymorphism, dietary interactions, and risk of colorectal cancer.* Cancer Res, 1997. **57**(6): p. 1098-102.
103. Molloy, A.M., et al., *Effects of Folate and Vitamin B12 Deficiencies During Pregnancy on Fetal, Infant, and Child Development.* Food and Nutrition Bulletin, 2008. **29**(2_suppl1): p. S101-S111.
104. WHO. *Periconceptional folic acid supplementation to prevent neural tube defects*. 2019 15/05/2020]; Available from: https://www.who.int/elena/titles/folate_periconceptional/en/.
105. Czeizel, A.E., et al., *Folate deficiency and folic acid supplementation: the prevention of neural-tube defects and congenital heart defects.* Nutrients, 2013. **5**(11): p. 4760-4775.
106. Li, K., M.L. Wahlqvist, and D. Li, *Nutrition, One-Carbon Metabolism and Neural Tube Defects: A Review.* Nutrients, 2016. **8**(11): p. 741.
107. Wang, X., et al., *Folate deficiency induces aneuploidy in human lymphocytes in vitro-evidence using cytokinesis-blocked cells and probes specific for chromosomes 17 and 21.* Mutat Res, 2004. **551**(1-2): p. 167-80.
108. Coppedè, F., *The genetics of folate metabolism and maternal risk of birth of a child with Down syndrome and associated congenital heart defects.* Frontiers in genetics, 2015. **6**: p. 223-223.
109. Safi, J., L. Joyeux, and G.E. Chalouhi, *Periconceptional folate deficiency and implications in neural tube defects.* Journal of pregnancy, 2012. **2012**: p. 295083-295083.
110. Amigou, A., et al., *Folic acid supplementation, MTHFR and MTRR polymorphisms, and the risk of childhood leukemia: the ESCALE study (SFCE).* Cancer Causes Control, 2012. **23**(8): p. 1265-77.
111. Ajrouche, R., et al., *Maternal reproductive history, fertility treatments and folic acid supplementation in the risk of childhood acute leukemia: the ESTELLE study.* Cancer Causes Control, 2014. **25**(10): p. 1283-93.

112. Milne, E., et al., *Maternal folate and other vitamin supplementation during pregnancy and risk of acute lymphoblastic leukemia in the offspring.* International Journal of Cancer, 2010. **126**(11): p. 2690-2699.
113. Thompson, J.R., et al., *Maternal folate supplementation in pregnancy and protection against acute lymphoblastic leukaemia in childhood: a case-control study.* The Lancet, 2001. **358**(9297): p. 1935-1940.
114. Metayer, C., et al., *Maternal supplementation with folic acid and other vitamins and risk of leukemia in offspring: a Childhood Leukemia International Consortium study.* Epidemiology, 2014. **25**(6): p. 811-22.
115. Singer, A.W., et al., *Maternal prenatal intake of one-carbon metabolism nutrients and risk of childhood leukemia.* Cancer Causes Control, 2016. **27**(7): p. 929-40.
116. Yang, Q.-H., et al., *Prevalence and effects of gene-gene and gene-nutrient interactions on serum folate and serum total homocysteine concentrations in the United States: findings from the third National Health and Nutrition Examination Survey DNA Bank.* The American journal of clinical nutrition, 2008. **88**(1): p. 232.
117. Lightfoot, T.J., et al., *Genetic variation in the folate metabolic pathway and risk of childhood leukemia.* Blood, 2010. **115**(19): p. 3923-3929.
118. Miller, J.W., et al., *Folate-deficiency-induced homocysteinaemia in rats: disruption of S-adenosylmethionine's co-ordinate regulation of homocysteine metabolism.* Biochem J, 1994. **298 (Pt 2)**(Pt 2): p. 415-9.
119. Blount, B.C., et al., *Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage.* Proc. Natl. Acad. Sci. USA, 1997. **94**(7): p. 3290-3295.
120. LeBlanc, D.P., et al., *Folate deficiency increases chromosomal damage and mutations in hematopoietic cells in the transgenic mutamouse model.* Environmental and Molecular Mutagenesis, 2018. **59**(5): p. 366-374.
121. Bistulfi, G., et al., *Mild folate deficiency induces genetic and epigenetic instability and phenotype changes in prostate cancer cells.* BMC Biol, 2010. **8**: p. 6.
122. Ni, J., et al., *Folate deficiency in human peripheral blood lymphocytes induces chromosome 8 aneuploidy but this effect is not modified by riboflavin.* Environ Mol Mutagen, 2010. **51**(1): p. 15-22.
123. Javadekar, S.M. and S.C. Raghavan, *Snaps and mends: DNA breaks and chromosomal translocations.* The FEBS Journal, 2015. **282**(14): p. 2627-2645.
124. Potter, C., et al., *Maternal Red Blood Cell Folate and Infant Vitamin B12 Status Influence Methylation of Genes Associated with Childhood Acute Lymphoblastic Leukemia.* Molecular Nutrition & Food Research, 2018. **62**(22): p. 1800411.
125. Timms, J.A., et al., *Exploring a potential mechanistic role of DNA methylation in the relationship between in utero and post-natal environmental exposures and risk of childhood acute lymphoblastic leukaemia.* Int J Cancer, 2019. **145**(11): p. 2933-2943.
126. Endoh, K., et al., *Low Folate Status Enhanced Benzene-Induced Cytogenetic Damage in Bone Marrow of Mice: A Relationship Between Dietary Intake and Tissue Levels of Folate.* Nutrition and Cancer, 2007. **59**(1): p. 99-105.
127. Padula, G., M.V. Ponzinibbio, and A.I. Seoane, *Possible radioprotective effect of folic acid supplementation on low dose ionizing radiation-induced genomic instability in vitro.* Indian journal of experimental biology, 2016. **54**(8): p. 537-543.
128. Reyes, C.M. and M.C. Cornelis, *Caffeine in the Diet: Country-Level Consumption and Guidelines.* Nutrients, 2018. **10**(11).
129. Cappelletti, S., et al., *Caffeine: cognitive and physical performance enhancer or psychoactive drug?* Curr Neuropharmacol, 2015. **13**(1): p. 71-88.
130. Nehlig, A., *Interindividual Differences in Caffeine Metabolism and Factors Driving Caffeine Consumption.* Pharmacol Rev, 2018. **70**(2): p. 384-411.

131. Zhou, S.F., et al., *Insights into the substrate specificity, inhibitors, regulation, and polymorphisms and the clinical impact of human cytochrome P450 1A2.* AAPS J, 2009. **11**(3): p. 481-94.
132. Lane, J.D., et al., *Menstrual cycle effects on caffeine elimination in the human female.* European Journal of Clinical Pharmacology, 1992. **43**(5): p. 543-546.
133. Rietveld, E.C., et al., *Rapid onset of an increase in caffeine residence time in young women due to oral contraceptive steroids.* European Journal of Clinical Pharmacology, 1984. **26**(3): p. 371-373.
134. Rodrigues, A.D., *Drug Interactions Involving 17α-Ethinylestradiol: Considerations Beyond Cytochrome P450 3A Induction and Inhibition.* Clin Pharmacol Ther, 2022. **111**(6): p. 1212-1221.
135. Grosso, G., et al., *Coffee, Caffeine, and Health Outcomes: An Umbrella Review.* Annu Rev Nutr, 2017. **37**: p. 131-156.
136. Qian, J., et al., *Impacts of Caffeine during Pregnancy.* Trends Endocrinol Metab, 2020. **31**(3): p. 218-227.
137. Grosso, L.M. and M.B. Bracken, *Caffeine metabolism, genetics, and perinatal outcomes: a review of exposure assessment considerations during pregnancy.* Ann Epidemiol, 2005. **15**(6): p. 460-6.
138. Brent, R.L., M.S. Christian, and R.M. Diener, *Evaluation of the reproductive and developmental risks of caffeine.* Birth Defects Res B Dev Reprod Toxicol, 2011. **92**(2): p. 152-87.
139. Group, C.S., *Maternal caffeine intake during pregnancy and risk of fetal growth restriction: a large prospective observational study.* BMJ, 2008. **337**: p. a2332.
140. NHS. *Should I limit caffeine during pregnancy?* 02/03/2020]; Available from: https://www.nhs.uk/common-health-questions/pregnancy/should-i-limit-caffeine-during-pregnancy/.
141. Milne, E., et al., *Maternal consumption of coffee and tea during pregnancy and risk of childhood ALL: a pooled analysis from the childhood Leukemia International Consortium.* Cancer Causes Control, 2018. **29**(6): p. 539-550.
142. Menegaux, F., et al., *Maternal coffee and alcohol consumption during pregnancy, parental smoking and risk of childhood acute leukaemia.* Cancer Detect Prev, 2005. **29**(6): p. 487-93.
143. Menegaux, F., et al., *Maternal alcohol and coffee drinking, parental smoking and childhood leukaemia: a French population-based case-control study.* Paediatr Perinat Epidemiol, 2007. **21**(4): p. 293-9.
144. Bonaventure, A., et al., *Childhood acute leukemia, maternal beverage intake during pregnancy, and metabolic polymorphisms.* Cancer Causes Control, 2013. **24**(4): p. 783-93.
145. Orsi, L., et al., *Parental smoking, maternal alcohol, coffee and tea consumption during pregnancy, and childhood acute leukemia: the ESTELLE study.* Cancer Causes Control, 2015. **26**(7): p. 1003-17.
146. Milne, E., et al., *Maternal consumption of coffee and tea during pregnancy and risk of childhood ALL: results from an Australian case-control study.* Cancer Causes Control, 2011. **22**(2): p. 207-18.
147. Ostertag, W., E. Duisberg, and M. StГjrmann, *The mutagenic activity of caffeine in man.* Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 1965. **2**(3): p. 293-296.
148. Porta, M., et al., *Coffee drinking the rationale for treating it as a potential effect modifier of carcinogenic exposures.* Eur J Epidemiol, 2003. **18**(4): p. 289-98.
149. Müller, W.U., et al., *Comet assay studies indicate that caffeine-mediated increase in radiation risk of embryos is due to inhibition of DNA repair.* Mutagenesis, 1996. **11**(1): p. 57-60.

150. Maréchal, A. and L. Zou, *DNA damage sensing by the ATM and ATR kinases.* Cold Spring Harb Perspect Biol, 2013. **5**(9).
151. Blackford, A.N. and S.P. Jackson, *ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response.* Mol Cell, 2017. **66**(6): p. 801-817.
152. Sarkaria, J., et al., *Inhibition of ATM and ATR Kinase Activities by the Radiosensitizing Agent, Caffeine.* Cancer Research, 1999. **59**: p. 4375-4382.
153. Oda, Y., M. Hidaka, and A. Suzuki, *Caffeine Has a Synergistic Anticancer Effect with Cisplatin via Inhibiting Fanconi Anemia Group D2 Protein Monoubiquitination in Hepatocellular Carcinoma Cells.* Biol. Pharm. Bull., 2017. **40**(10): p. 2005-209.
154. Clifford, B., et al., *G2 arrest in response to topoisomerase II inhibitors: the role of p53.* Cancer Res, 2003. **63**(14): p. 4074-81.
155. Wang, H., et al., *Effect of ATM and HDAC Inhibition on Etoposide-Induced DNA Damage in Porcine Early Preimplantation Embryos.* PLoS One, 2015. **10**(11): p. e0142561.
156. Tsabar, M., et al., *Caffeine impairs resection during DNA break repair by reducing the levels of nucleases Sae2 and Dna2.* Nucleic Acids Res, 2015. **43**(14): p. 6889-901.
157. Shin, C.G., et al., *Rapid evaluation of topoisomerase inhibitors: caffeine inhibition of topoisomerases in vivo.* Teratog Carcinog Mutagen, 1990. **10**(1): p. 41-52.
158. Ross, J.A., et al., *Maternal exposure to potential inhibitors of DNA topoisomerase II and infant leukemia (United States): a report from the Children's Cancer Group.* Cancer Causes Control, 1996. **7**(6): p. 581-90.
159. Cowell, I.G., et al., *Model for MLL translocations in therapy-related leukemia involving topoisomerase IIbeta-mediated DNA strand breaks and gene proximity.* Proc Natl Acad Sci U S A, 2012. **109**(23): p. 8989-94.
160. WHO. *Tobacco*. 08/06/2020]; Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco.
161. ASH. *Tobacco and the developing world*. 08/06/2020]; Available from: https://ash.org.uk/wp-content/uploads/2019/10/Tobacco-Developing-World.pdf.
162. West, R., *Tobacco smoking: Health impact, prevalence, correlates and interventions.* Psychology & health, 2017. **32**(8): p. 1018-1036.
163. Lichtman, M.A., *Cigarette smoking, cytogenetic abnormalities, and acute myelogenous leukemia.* Leukemia, 2007. **21**(6): p. 1137-40.
164. Moorman, A.V., et al., *Smoking and the risk of acute myeloid leukaemia in cytogenetic subgroups.* British journal of cancer, 2002. **86**(1): p. 60-62.
165. UK, C.R. *How does smoking cause cancer?* 08/06/2020]; Available from: https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/smoking-and-cancer/how-does-smoking-cause-cancer.
166. Hecht, S.S., *Tobacco carcinogens, their biomarkers and tobacco-induced cancer.* Nat Rev Cancer, 2003. **3**(10): p. 733-44.
167. NHS. *Stop smoking in pregnancy*. 08/06/2020]; Available from: https://www.nhs.uk/conditions/pregnancy-and-baby/smoking-pregnant/.
168. Cnattingius, S., *The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes.* Nicotine Tob Res, 2004. **6 Suppl 2**: p. S125-40.
169. Gould, G.S., et al., *Exposure to Tobacco, Environmental Tobacco Smoke and Nicotine in Pregnancy: A Pragmatic Overview of Reviews of Maternal and Child Outcomes, Effectiveness of Interventions and Barriers and Facilitators to Quitting.* International journal of environmental research and public health, 2020. **17**(6): p. 2034.
170. Chunxia, D., et al., *Tobacco smoke exposure and the risk of childhood acute lymphoblastic leukemia and acute myeloid leukemia: A meta-analysis.* Medicine (Baltimore), 2019. **98**(28): p. e16454.

171. Pereira, C.S., et al., *Smoking-induced chromosomal segregation anomalies identified by FISH analysis of sperm.* Molecular Cytogenetics, 2014. **7**(58).
172. Chang, J.S., et al., *Parental Smoking and the Risk of Childhood Leukemia.* American Journal of Epidemiology, 2006. **163**(12): p. 1091-1100.
173. de Smith, A.J., et al., *Correlates of Prenatal and Early-Life Tobacco Smoke Exposure and Frequency of Common Gene Deletions in Childhood Acute Lymphoblastic Leukemia.* Cancer Res, 2017. **77**(7): p. 1674-1683.
174. John, E.M., D.A. Savitz, and D.P. Sandler, *Prenatal exposure to parents' smoking and childhood cancer.* Am J Epidemiol, 1991. **133**(2): p. 123-32.
175. Milne, E., et al., *Parental prenatal smoking and risk of childhood acute lymphoblastic leukemia.* Am J Epidemiol, 2012. **175**(1): p. 43-53.
176. Metayer, C., et al., *Tobacco smoke exposure and the risk of childhood acute lymphoblastic and myeloid leukemias by cytogenetic subtype.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(9): p. 1600-11.
177. Metayer, C., et al., *Parental Tobacco Smoking and Acute Myeloid Leukemia: The Childhood Leukemia International Consortium.* Am J Epidemiol, 2016. **184**(4): p. 261-73.
178. de la Chica, R.A., et al., *Chromosomal instability in amniocytes from fetuses of mothers who smoke.* Jama, 2005. **293**(10): p. 1212-22.
179. Cárceles-Álvarez, A., et al., *Secondhand smoke: A new and modifiable prognostic factor in childhood acute lymphoblastic leukemias.* Environ Res, 2019. **178**: p. 108689.
180. Esakky, P., et al., *Paternal exposure to cigarette smoke condensate leads to reproductive sequelae and developmental abnormalities in the offspring of mice.* Reprod Toxicol, 2016. **65**: p. 283-294.
181. Pirini, F., et al., *Nuclear and mitochondrial DNA alterations in newborns with prenatal exposure to cigarette smoke.* Int J Environ Res Public Health, 2015. **12**(2): p. 1135-55.
182. Poli, P., et al., *DNA damage by tobacco smoke and some antiblastic drugs evaluated using the Comet assay.* Toxicology Letters, 1999. **108**: p. 267-276.
183. Snyder, R., *Leukemia and benzene.* Int J Environ Res Public Health, 2012. **9**(8): p. 2875-93.
184. Weisel, C.P., *Benzene exposure: an overview of monitoring methods and their findings.* Chem Biol Interact, 2010. **184**(1-2): p. 58-66.
185. Zhou, Y., et al., *Carcinogenic Air Toxics Exposure and Their Cancer-Related Health Impacts in the United States.* PLoS One, 2015. **10**(10): p. e0140013.
186. WHO. *Air Pollution*. 23/6/2020]; Available from: https://www.who.int/health-topics/air-pollution#tab=tab_1.
187. Jacobs, M., et al., *The association between ambient air pollution and selected adverse pregnancy outcomes in China: A systematic review.* Sci Total Environ, 2017. **579**: p. 1179-1192.
188. Dowty, B.J., J.L. Laseter, and J. Storer, *The transplacental migration and accumulation in blood of volatile organic constituents.* Pediatr Res, 1976. **10**(7): p. 696-701.
189. Badham, H.J. and L.M. Winn, *In utero exposure to benzene disrupts fetal hematopoietic progenitor cell growth via reactive oxygen species.* Toxicol Sci, 2010. **113**(1): p. 207-15.
190. Duarte-Davidson, R., et al., *Benzene in the environment: an assessment of the potential risks to the health of the population.* Occup Environ Med, 2001. **58**: p. 2-13.
191. Zahran, S., et al., *Maternal benzene exposure and low birth weight risk in the United States: a natural experiment in gasoline reformulation.* Environ Res, 2012. **112**: p. 139-46.
192. Xu, X., et al., *Association of petrochemical exposure with spontaneous abortion.* Occup Environ Med, 1998. **55**(1): p. 31-6.
193. Slama, R., et al., *Maternal personal exposure to airborne benzene and intrauterine growth.* Environ Health Perspect, 2009. **117**(8): p. 1313-21.
194. Chen, D., et al., *Exposure to benzene, occupational stress, and reduced birth weight.* Occup Environ Med, 2000. **57**(10): p. 661-7.

195. Lo Pumo, R., et al., *Long-lasting neurotoxicity of prenatal benzene acute exposure in rats.* Toxicology, 2006. **223**(3): p. 227-34.
196. Srám, R.J., et al., *Ambient air pollution and pregnancy outcomes: a review of the literature.* Environ Health Perspect, 2005. **113**(4): p. 375-82.
197. Llop, S., et al., *Preterm birth and exposure to air pollutants during pregnancy.* Environ Res, 2010. **110**(8): p. 778-85.
198. Ha, S., et al., *Prenatal and early life exposures to ambient air pollution and development.* Environ Res, 2019. **174**: p. 170-175.
199. Grippo, A., et al., *Air pollution exposure during pregnancy and spontaneous abortion and stillbirth.* Rev Environ Health, 2018. **33**(3): p. 247-264.
200. Estarlich, M., et al., *Exposure to ambient air pollution during pregnancy and preterm birth: A Spanish multicenter birth cohort study.* Environ Res, 2016. **147**: p. 50-8.
201. Pyatt, D. and S. Hays, *A review of the potential association between childhood leukemia and benzene.* Chem Biol Interact, 2010. **184**(1-2): p. 151-64.
202. Carlos-Wallace, F.M., et al., *Parental, In Utero, and Early-Life Exposure to Benzene and the Risk of Childhood Leukemia: A Meta-Analysis.* Am J Epidemiol, 2016. **183**(1): p. 1-14.
203. Spycher, B.D., et al., *Parental occupational exposure to benzene and the risk of childhood cancer: A census-based cohort study.* Environ Int, 2017. **108**: p. 84-91.
204. Kirkeleit, J., et al., *Maternal exposure to gasoline and exhaust increases the risk of childhood leukaemia in offspring - a prospective study in the Norwegian Mother and Child Cohort Study.* Br J Cancer, 2018. **119**(8): p. 1028-1035.
205. Heck, J.E., et al., *Parental occupational exposure to benzene and the risk of childhood and adolescent acute lymphoblastic leukaemia: a population-based study.* Occup Environ Med, 2019. **76**(8): p. 527-529.
206. Filippini, T., et al., *Association between Outdoor Air Pollution and Childhood Leukemia: A Systematic Review and Dose-Response Meta-Analysis.* Environ Health Perspect, 2019. **127**(4): p. 46002.
207. Sun, X.X., S.S. Zhang, and X.L. Ma, *No association between traffic density and risk of childhood leukemia: a meta-analysis.* Asian Pac J Cancer Prev, 2014. **15**(13): p. 5229-32.
208. Vinceti, M., et al., *Leukemia risk in children exposed to benzene and PM10 from vehicular traffic: a case-control study in an Italian population.* Eur J Epidemiol, 2012. **27**(10): p. 781-90.
209. Janitz, A.E., et al., *Traffic-related air pollution and childhood acute leukemia in Oklahoma.* Environ Res, 2016. **148**: p. 102-111.
210. Janitz, A.E., et al., *Benzene and childhood acute leukemia in Oklahoma.* Environ Res, 2017. **158**: p. 167-173.
211. Raaschou-Nielsen, O., et al., *Ambient benzene at the residence and risk for subtypes of childhood leukemia, lymphoma and CNS tumor.* Int J Cancer, 2018. **143**(6): p. 1367-1373.
212. Peng, D., et al., *Study on the cytogenetic changes induced by benzene and hydroquinone in human lymphocytes.* Hum Exp Toxicol, 2012. **31**(4): p. 322-35.
213. Lau, A., C.L. Belanger, and L.M. Winn, *In utero and acute exposure to benzene: investigation of DNA double-strand breaks and DNA recombination in mice.* Mutat Res, 2009. **676**(1-2): p. 74-82.
214. Pandey, A.K., et al., *In silico studies with human DNA topoisomerase-II alpha to unravel the mechanism of in vitro genotoxicity of benzene and its metabolites.* Mutat Res, 2009. **661**(1-2): p. 57-70.
215. Lu, P.C.W., S. Shahbaz, and L.M. Winn, *Benzene and its effects on cell signaling pathways related to hematopoiesis and leukemia.* J Appl Toxicol, 2020.
216. McHale, C.M., L. Zhang, and M.T. Smith, *Current understanding of the mechanism of benzene-induced leukemia in humans: implications for risk assessment.* Carcinogenesis, 2012. **33**(2): p. 240-52.

217. Ross, D., *The Role of Metabolism and Specific Metabolites in Benzene-Induced Toxicity: Evidence and Issues.* Journal of Toxicology and Environmental Health, Part A, 2000. **61**(5-6): p. 357-372.
218. Kumar, R., P.S. Godavarthy, and D.S. Krause, *The bone marrow microenvironment in health and disease at a glance.* J Cell Sci, 2018. **131**(4).
219. Son, M.Y., et al., *A mechanism for 1,4-Benzoquinone-induced genotoxicity.* Oncotarget, 2016. **7**(29): p. 46433-46447.
220. Irons, R.D., et al., *Acute myeloid leukemia following exposure to benzene more closely resembles de novo than therapy related-disease.* Genes Chromosomes Cancer, 2013. **52**(10): p. 887-94.
221. Kerzic, P.J. and R.D. Irons, *Distribution of chromosome breakpoints in benzene-exposed and unexposed AML patients.* Environ Toxicol Pharmacol, 2017. **55**: p. 212-216.
222. Manola, K.N., *Cytogenetics of pediatric acute myeloid leukemia.* Eur J Haematol, 2009. **83**(5): p. 391-405.
223. Hukkanen, J., P. Jacob, 3rd, and N.L. Benowitz, *Metabolism and disposition kinetics of nicotine.* Pharmacol Rev, 2005. **57**(1): p. 79-115.
224. Dani, J.A., *Overview of nicotinic receptors and their roles in the central nervous system.* Biol Psychiatry, 2001. **49**(3): p. 166-74.
225. Jarvis, M.J., et al., *Comparison of tests used to distinguish smokers from nonsmokers.* Am J Public Health, 1987. **77**(11): p. 1435-8.
226. Tsinisizeli, N., et al., *Determination of nicotine and cotinine in meconium from Greek neonates and correlation with birth weight and gestational age at birth.* Chemosphere, 2015. **119**: p. 1200-1207.
227. Holbrook, B.D., *The effects of nicotine on human fetal development.* Birth Defects Res C Embryo Today, 2016. **108**(2): p. 181-92.
228. Spindel, E.R. and C.T. McEvoy, *The Role of Nicotine in the Effects of Maternal Smoking during Pregnancy on Lung Development and Childhood Respiratory Disease. Implications for Dangers of E-Cigarettes.* Am J Respir Crit Care Med, 2016. **193**(5): p. 486-94.
229. Suter, M.A., et al., *Is there evidence for potential harm of electronic cigarette use in pregnancy?* Birth Defects Res A Clin Mol Teratol, 2015. **103**(3): p. 186-95.
230. Ginzkey, C., et al., *Analysis of nicotine-induced DNA damage in cells of the human respiratory tract.* Toxicol Lett, 2012. **208**(1): p. 23-9.
231. Heeschen, C., et al., *Nicotine stimulates angiogenesis and promotes tumor growth and atherosclerosis.* Nat Med, 2001. **7**(7): p. 833-9.
232. West, K.A., et al., *Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells.* J Clin Invest, 2003. **111**(1): p. 81-90.
233. Demirhan, O., et al., *The genotoxic effect of nicotine on chromosomes of human fetal cells: the first report described as an important study.* Inhal Toxicol, 2011. **23**(13): p. 829-34.
234. Chen, H., et al., *Maternal E-Cigarette Exposure in Mice Alters DNA Methylation and Lung Cytokine Expression in Offspring.* American Journal of Respiratory Cell and Molecular Biology (Online), 2018. **58**(3): p. 366-377.
235. Fox, A.M., F.H. Moonschi, and C.I. Richards, *The nicotine metabolite, cotinine, alters the assembly and trafficking of a subset of nicotinic acetylcholine receptors.* J Biol Chem, 2015. **290**(40): p. 24403-12.
236. Haussmann, H.J. and M.W. Fariss, *Comprehensive review of epidemiological and animal studies on the potential carcinogenic effects of nicotine per se.* Crit Rev Toxicol, 2016. **46**(8): p. 701-34.
237. Murray, R.P., J.E. Connett, and L.M. Zapawa, *Does nicotine replacement therapy cause cancer? Evidence from the Lung Health Study.* Nicotine & Tobacco Research, 2009. **11**(9): p. 1076-1082.

238. LaVoie, E.J., et al., *Evaluation of the Effects of Cotinine and Nicotine-NʙƄI-oxides on the Development of Tumors in Rats Initiated With N-[4-(5-Nitro-2-furyl)-2-thiazolyl]formamide.* JNCI : Journal of the National Cancer Institute, 1985. **75**(6): p. 1075-1081.
239. Dalberto, D., et al., *Cytotoxic and genotoxic evaluation of cotinine using human neuroblastoma cells (SH-SY5Y).* Genet Mol Biol, 2020. **43**(2): p. e20190123.
240. Conklin, B.S., et al., *Nicotine and Cotinine Up-Regulate Vascular Endothelial Growth Factor Expression in Endothelial Cells.* American Journal of Pathology, 2002. **160**(2): p. 413-418.
241. Arabi, M., *Nicotinic infertility: assessing DNA and plasma membrane integrity of human spermatozoa.* Andrologia, 2004. **36**(5): p. 305-10.
242. Sofikitis, N., et al., *Effects of cotinine on sperm motility, membrane function, and fertilizing capacity in vitro.* Urol Res, 2000. **28**(6): p. 370-5.
243. Wong, W.Y., et al., *Cigarette smoking and the risk of male factor subfertility: minor association between cotinine in seminal plasma and semen morphology.* Fertil Steril, 2000. **74**(5): p. 930-5.
244. Schrott, R., et al., *Sperm DNA methylation altered by THC and nicotine: Vulnerability of neurodevelopmental genes with bivalent chromatin.* Sci Rep, 2020. **10**(1): p. 16022.
245. Hurwitz, R., et al., *Characterization of a leukemic cell line of the pre-B phenotype.* Int J Cancer, 1979. **23**(2): p. 174-80.
246. Ueda, R., et al., *Serological analysis of cell surface antigens of null cell acute lymphocytic leukemia by mouse monoclonal antibodies.* Proc Natl Acad Sci U S A, 1982. **79**(14): p. 4386-90.
247. Quentmeier, H., et al., *The LL-100 panel: 100 cell lines for blood cancer studies.* Scientific reports, 2019. **9**(1): p. 8218-8218.
248. Findley, H.W., Jr., et al., *Two new acute lymphoblastic leukemia cell lines with early B-cell phenotypes.* Blood, 1982. **60**(6): p. 1305-9.
249. Asou, H., et al., *Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation.* Blood, 1991. **77**(9): p. 2031-6.
250. Rosenfeld, C., et al., *Phenotypic characterisation of a unique non-T, non-B acute lymphoblastic leukaemia cell line.* Nature, 1977. **267**(5614): p. 841-843.
251. Lanotte, M., et al., *NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3).* Blood, 1991. **77**(5): p. 1080-6.
252. Beverly, L., et al., *Growth Factor Requirements of Childhood Acute Leukemia: Establishment of GM-CSF-Dependent Cell Lines.* Blood, 1987. **70**(1): p. 192-199.
253. Gallagher, R., et al., *Characterization of the Continuous, Differentiating Myeloid Cell Line (HL-60) From a Patient With Acute Promyelocytic Leukemia.* Blood, 1979. **54**(3): p. 713-733.
254. Vega-Avila, E. and M.K. Pugsley, *An overview of colorimetric assay methods used to assess survival or proliferation of mammalian cells.* Proceedings of the Western Pharmacology Society, 2011. **54**: p. 10-14.
255. S Izraeli, et al., *Unexpected Heterogeneity in E2A/PBX1 Fusion Messenger RNA Detected by the Polymerase Chain Reaction in Pediatric Patients With Acute Lymphoblastic Leukemia.* Blood, 1992. **80**: p. 1413-1417.
256. Kozu, T., et al., *Junctions of the AMLl/MTG8(ETO) Fusion Are Constant in t(8;21) Acute Myeloid Leukemia Detected by Reverse Transcription Polymerase Chain Reaction.* Blood, 1993. **82**(4): p. 1270-1276.
257. S.P. Rornana, et al., *The t(12;21) of Acute Lymphoblastic Leukemia Results in a tel-AML1 Gene Fusion.* Blood, 1995. **85**: p. 3662-3670.
258. Pandolfi, P.P., et al., *Genomic variability and alternative splicing generate multiple PML/RARao transcripts that encode aberrant PML proteins and PML/RARao isoforms in acute promyelocytic leukaemia.* The EMBO Journal, 1992. **11**(4): p. 1397-1407.

259. Chen, Z., et al., *Development and Validation of a 3-Plex RTqPCR Assay for the Simultaneous Detection and Quantitation of the Three PML-RARa Fusion Transcripts in Acute Promyelocytic Leukemia.* PLOS ONE, 2015. **10**(3): p. p.e0122530.
260. Burmeister, T. and R. Reinhardt, *A multiplex PCR for improved detection of typical and atypical BCR-ABL fusion transcripts.* Leuk Res, 2008. **32**(4): p. 579-85.
261. van Dongen, J., et al., *Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease.* Leukemia, 1999. **13**: p. 1901-1928.
262. JR Downing, et al., *The der(11)-encoded MLL/AF-4 fusion transcript is consistently detected in t(4;11)(q21;q23)-containing acute lymphoblastic leukemia.* Blood, 1994. **83**: p. 330-335.
263. Jackson, K.A., et al., *Analysis of differential gene-regulatory responses to zinc in human intestinal and placental cell lines.* British Journal of Nutrition, 2009. **101**(10): p. 1474-1483.
264. Handschuh, L., et al., *Gene expression profiling of acute myeloid leukemia samples from adult patients with AML-M1 and -M2 through boutique microarrays, real-time PCR and droplet digital PCR.* International journal of oncology, 2018. **52**(3): p. 656-678.
265. Bustin, S.A., et al., *The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments.* Clinical Chemistry, 2009. **55**(4): p. 611-622.
266. Gonzales, P.R. and F.M. Mikhail, *Diagnostic and Prognostic Utility of Fluorescence In situ Hybridization (FISH) Analysis in Acute Myeloid Leukemia.* Curr Hematol Malig Rep, 2017. **12**(6): p. 568-573.
267. *Carnoy's fixative (L).* Cold Spring Harbor Protocols, 2009. **2009**(4): p. pdb.rec11742-pdb.rec11742.
268. Ami, D., et al., *Role of water in chromosome spreading and swelling induced by acetic acid treatment: a FTIR spectroscopy study.* European journal of histochemistry : EJH, 2014. **58**(1): p. 2330-2330.
269. Hliscs, R., P. Mühlig, and U. Claussen, *The spreading of metaphases is a slow process which leads to a stretching of chromosomes.* Cytogenetic and genome research, 1997. **76**(3-4): p. 167-171.
270. Liehr, T., *Fluorescence In Situ Hybridization (FISH) - Application Guide*. 2008: Springer Berlin Heidelberg.
271. Blake, R.D. and S.G. Delcourt, *Thermodynamic effects of formamide on DNA stability.* Nucleic acids research, 1996. **24**(11): p. 2095-2103.
272. Ostromohov, N., et al., *Real-Time Monitoring of Fluorescence in Situ Hybridization Kinetics.* Anal Chem, 2018. **90**(19): p. 11470-11477.
273. Kapuscinski, J., *DAPI: a DNA-specific fluorescent probe.* Biotech Histochem, 1995. **70**(5): p. 220-33.
274. Strick, R., et al., *Dietary bioflavonoids induce cleavage in the MLL gene and may contribute to infant leukemia.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4790-5.
275. van Waalwijk van Doorn-Khosrovani, S.B., et al., *Dietary flavonoids induce MLL translocations in primary human CD34+ cells.* Carcinogenesis, 2007. **28**(8): p. 1703-1709.
276. Bariar, B., et al., *Bioflavonoids promote stable translocations between MLL-AF9 breakpoint cluster regions independent of normal chromosomal context: Model system to screen environmental risks.* Environ Mol Mutagen, 2019. **60**(2): p. 154-167.
277. Pećina-Šlaus, N., et al., *Mismatch Repair Pathway, Genome Stability and Cancer.* Front Mol Biosci, 2020. **7**: p. 122.
278. Harfe, B.D. and S. Jinks-Robertson, *DNA mismatch repair and genetic instability.* Annu Rev Genet, 2000. **34**: p. 359-399.
279. Li, G.-M., *Mechanisms and functions of DNA mismatch repair.* Cell Research, 2008. **18**(1): p. 85-98.
280. Tham, K.-C., R. Kanaar, and J.H.G. Lebbink, *Mismatch repair and homeologous recombination.* DNA Repair, 2016. **38**: p. 75-83.

281. Bocker, T., et al., *Genomic instability in colorectal carcinomas: comparison of different evaluation methods and their biological significance.* J Pathol, 1996. **179**(1): p. 15-9.
282. Eshleman, J.R., et al., *Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53.* Oncogene, 1998. **17**(6): p. 719-725.
283. Abdel-Rahman, W.M., et al., *Spectral karyotyping suggests additional subsets of colorectal cancers characterized by pattern of chromosome rearrangement.* Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2538-43.
284. Abdel-Rahman, W.M., et al., *Restoring mismatch repair does not stop the formation of reciprocal translocations in the colon cancer cell line HCA7 but further destabilizes chromosome number.* Oncogene, 2005. **24**(4): p. 706-13.
285. Holt, S.M., et al., *Compromised repair of clustered DNA damage in the human acute lymphoblastic leukemia MSH2-deficient NALM-6 cells.* Mutat Res, 2009. **674**(1-2): p. 123-30.
286. Wlodarska, I., et al., *A new subtype of pre-B acute lymphoblastic leukemia with t(5;12)(q31q33;p12), molecularly and cytogenetically distinct from t(5;12) in chronic myelomonocytic leukemia.* Blood, 1997. **89**(5): p. 1716-22.
287. Smith, K.A., et al., *The role of topoisomerase II beta on breakage and proximity of RUNX1 to partner alleles RUNX1T1 and EVI1.* Genes Chromosomes Cancer, 2014. **53**(2): p. 117-28.
288. Epstein, F.H.M.D., et al., *The Molecular Genetics of Philadelphia Chromosome-Positive Leukemias.* The New England Journal of Medicine, 1988. **319**(15): p. 990-998.
289. Kozera, B. and M. Rapacz, *Reference genes in real-time PCR.* J Appl Genet, 2013. **54**(4): p. 391-406.
290. Taylor, S., et al., *A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines.* Methods, 2010. **50**(4): p. S1-5.
291. Drexler, H.G. and C.C. Uphoff, *Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention.* Cytotechnology, 2002. **39**(2): p. 75-90.
292. Schonn, I., J. Hennesen, and D.C. Dartsch, *Cellular responses to etoposide: cell death despite cell cycle arrest and repair of DNA damage.* Apoptosis, 2010. **15**(2): p. 162-72.
293. Solarczyk, K. and M. Kordon-Kiszala, *Let's not take DNA breaks for granted. The importance of direct detection of DNA breaks for the successful development of DDR inhibitors.* Front Cell Dev Biol, 2023. **11**: p. 1118716.
294. Yu, T., et al., *Pregnancy-induced changes in the pharmacokinetics of caffeine and its metabolites.* J Clin Pharmacol, 2016. **56**(5): p. 590-6.
295. Ginsberg, G., et al., *Physiologically Based Pharmacokinetic (PBPK) Modeling of Caffeine and Theophylline in Neonates and Adults: Implications for Assessing Children's Risks from Environmental Agents.* Journal of Toxicology and Environmental Health, Part A, 2004. **67**(4): p. 297-329.
296. Klebanoff, M.A., et al., *Serum caffeine and paraxanthine as markers for reported caffeine intake in pregnancy.* Ann Epidemiol, 1998. **8**(2): p. 107-11.
297. Zhong, M., et al., *Dual-color, break-apart FISH assay on paraffin-embedded tissues as an adjunct to diagnosis of Xp11 translocation renal cell carcinoma and alveolar soft part sarcoma.* Am J Surg Pathol, 2010. **34**(6): p. 757-66.
298. Ventura, R.A., et al., *FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue.* J Mol Diagn, 2006. **8**(2): p. 141-51.
299. Bavari, M., et al., *Neuroprotective, antiapoptotic and antioxidant effects of l-carnitine against caffeine-induced neurotoxicity in SH-SY5Y neuroblastoma cell line.* Drug and Chemical Toxicology, 2016. **39**(2): p. 157-166.
300. Lu, G.-Y., et al., *Caffeine induces tumor cytotoxicity via the regulation of alternative splicing in subsets of cancer-associated genes.* The International Journal of Biochemistry & Cell Biology, 2014. **47**: p. 83-92.

301. Lu, P.Z., C.Y. Lai, and W.H. Chan, *Caffeine induces cell death via activation of apoptotic signal and inactivation of survival signal in human osteoblasts.* Int J Mol Sci, 2008. **9**(5): p. 698-718.
302. Kaufmann, W.K., et al., *Caffeine and human DNA metabolism: the magic and the mystery.* Mutat Res, 2003. **532**(1-2): p. 85-102.
303. Guo, N., D.V. Faller, and C. Vaziri, *Carcinogen-induced S-phase arrest is Chk1 mediated and caffeine sensitive.* Cell Growth Differ, 2002. **13**(2): p. 77-86.
304. Kummrow, A., et al., *Quantitative assessment of cell viability based on flow cytometry and microscopy.* Cytometry A, 2013. **83**(2): p. 197-204.
305. Quent, V.M., et al., *Discrepancies between metabolic activity and DNA content as tool to assess cell proliferation in cancer research.* J Cell Mol Med, 2010. **14**(4): p. 1003-13.
306. Barcelos, R.P., et al., *Caffeine effects on systemic metabolism, oxidative-inflammatory pathways, and exercise performance.* Nutrition Research, 2020. **80**: p. 1-17.
307. Zelensky, A.N., et al., *Caffeine suppresses homologous recombination through interference with RAD51-mediated joint molecule formation.* Nucleic Acids Res, 2013. **41**(13): p. 6475-89.
308. Badisa, V.L., et al., *Mechanism of DNA damage by cadmium and interplay of antioxidant enzymes and agents.* Environ Toxicol, 2007. **22**(2): p. 144-51.
309. Muruzabal, D., et al., *Validation of the in vitro comet assay for DNA cross-links and altered bases detection.* Arch Toxicol, 2021. **95**(8): p. 2825-2838.
310. Falk, M., et al., *Chromatin architecture changes and DNA replication fork collapse are critical features in cryopreserved cells that are differentially controlled by cryoprotectants.* Sci Rep, 2018. **8**(1): p. 14694.
311. Sumida, K., et al., *Effects of DMSO on gene expression in human and rat hepatocytes.* Hum Exp Toxicol, 2011. **30**(10): p. 1701-9.
312. Verheijen, M., et al., *DMSO induces drastic changes in human cellular processes and epigenetic landscape in vitro.* Sci Rep, 2019. **9**(1): p. 4641.
313. Yuan, C., et al., *Dimethyl sulfoxide damages mitochondrial integrity and membrane potential in cultured astrocytes.* PLoS One, 2014. **9**(9): p. e107447.
314. Tunçer, S., et al., *Low dose dimethyl sulfoxide driven gross molecular changes have the potential to interfere with various cellular processes.* Sci Rep, 2018. **8**(1): p. 14828.
315. Hayashi, Y., K. Ohnuma, and M.K. Furue, *Pluripotent Stem Cell Heterogeneity*, in *Stem Cells Heterogeneity - Novel Concepts*, A. Birbrair, Editor. 2019, Springer International Publishing: Cham. p. 71-94.
316. Vendramin, R., K. Litchfield, and C. Swanton, *Cancer evolution: Darwin and beyond.* Embo j, 2021. **40**(18): p. e108389.
317. Hu, X., et al., *Genistein-induced DNA damage is repaired by nonhomologous end joining and homologous recombination in TK6 cells.* J Cell Physiol, 2019. **234**(3): p. 2683-2692.
318. Pankow, J.F., et al., *Benzene formation in electronic cigarettes.* PLoS One, 2017. **12**(3): p. e0173055.
319. Brugnone, F., et al., *Benzene in environmental air and human blood* Int Arch Occup Environ Health, 1998. **71**(8): p. 554-559.
320. Brugnone, F., et al., *Benzene in blood as a biomarker of low level occupational exposure* Science of the Total Environment, 1999. **235**(1): p. 247-252.
321. Qian, S., et al., *Benzene induces haematotoxicity by promoting deacetylation and autophagy.* J Cell Mol Med, 2019. **23**(2): p. 1022-1033.
322. Yoon, B.-I., et al., *Mechanism of action of benzene toxicity: Cell cycle suppression in hemopoietic progenitor cells (CFU-GM).* Experimental Hematology, 2001. **29**(3): p. 278-285.
323. Tapella, L., et al., *Benzene and 2-ethyl-phthalate induce proliferation in normal rat pituitary cells.* Pituitary, 2017. **20**(3): p. 311-318.
324. Martínez-Velázquez, M., et al., *Benzene metabolites induce apoptosis in lymphocytes.* Experimental and Toxicologic Pathology, 2006. **58**(1): p. 65-70.

325. Hirabayashi, Y., et al., *Mechanism of benzene-induced hematotoxicity and leukemogenicity: current review with implication of microarray analyses.* Toxicol Pathol, 2004. **32 Suppl 2**: p. 12-6.
326. Pannunzio, N.R. and M.R. Lieber, *Concept of DNA Lesion Longevity and Chromosomal Translocations.* Trends Biochem Sci, 2018. **43**(7): p. 490-498.
327. Sallmyr, A., J. Fan, and F.V. Rassool, *Genomic instability in myeloid malignancies: increased reactive oxygen species (ROS), DNA double strand breaks (DSBs) and error-prone repair.* Cancer Lett, 2008. **270**(1): p. 1-9.
328. Nishikawa, T., et al., *Benzene induces cytotoxicity without metabolic activation.* J Occup Health, 2011. **53**(2): p. 84-92.
329. Bozier, J., et al., *The Evolving Landscape of e-Cigarettes: A Systematic Review of Recent Evidence.* Chest, 2020. **157**(5): p. 1362-1390.
330. Breland, A., A. McCubbin, and K. Ashford, *Electronic nicotine delivery systems and pregnancy: Recent research on perceptions, cessation, and toxicant delivery.* Birth Defects Res, 2019. **111**(17): p. 1284-1293.
331. Martin, E.M., et al., *E-cigarette use results in suppression of immune and inflammatory-response genes in nasal epithelial cells similar to cigarette smoke.* Am J Physiol Lung Cell Mol Physiol, 2016. **311**(1): p. L135-44.
332. Kim, S., *Overview of Cotinine Cutoff Values for Smoking Status Classification.* Int J Environ Res Public Health, 2016. **13**(12).
333. Dempsey, D.A., et al., *Determination of tobacco smoke exposure by plasma cotinine levels in infants and children attending urban public hospital clinics.* Arch Pediatr Adolesc Med, 2012. **166**(9): p. 851-6.
334. Leenders, M., et al., *Plasma cotinine levels and pancreatic cancer in the EPIC cohort study.* Int J Cancer, 2012. **131**(4): p. 997-1002.
335. Suzuki, S., et al., *Cotinine, a major nicotine metabolite, induces cell proliferation on urothelium in vitro and in vivo.* Toxicology, 2020. **429**: p. 152325.
336. Caliri, A.W., S. Tommasi, and A. Besaratinia, *Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer.* Mutat Res Rev Mutat Res, 2021. **787**: p. 108365.
337. Tan, S.N., S.P. Sim, and A.S. Khoo, *Potential role of oxidative stress-induced apoptosis in mediating chromosomal rearrangements in nasopharyngeal carcinoma.* Cell Biosci, 2016. **6**: p. 35.
338. van Gool, J.D., et al., *Folic acid and primary prevention of neural tube defects: A review.* Reprod Toxicol, 2018. **80**: p. 73-84.
339. Patel, K.R. and A. Sobczyńska-Malefora, *The adverse effects of an excessive folic acid intake.* Eur J Clin Nutr, 2017. **71**(2): p. 159-163.
340. Lamm, N., et al., *Folate levels modulate oncogene-induced replication stress and tumorigenicity.* EMBO Mol Med, 2015. **7**(9): p. 1138-52.
341. Farrell, C.J., S.H. Kirsch, and M. Herrmann, *Red cell or serum folate: what to do in clinical practice?* Clin Chem Lab Med, 2013. **51**(3): p. 555-69.
342. Scientific, T.F. *Technical resources 21875 - RPMI 1640*. 23/2/22]; Available from: https://www.thermofisher.com/uk/en/home/technical-resources/media-formulation.187.html.
343. Courtemanche, C., et al., *Folate deficiency inhibits the proliferation of primary human CD8+ T lymphocytes in vitro.* J Immunol, 2004. **173**(5): p. 3186-92.
344. Fenech, M. and J. Rinaldi, *A comparison of lymphocyte micronuclei and plasma micronutrients in vegetarians and non-vegetarians.* Carcinogenesis (New York), 1995. **16**(2): p. 223-230.
345. WHO. *Serum and red blood cell folate concentrations for assessing folate status in populations*. Vitamin and Mineral Nutrition Information System 2015 23/2/22]; 1-5].

Available from: http://apps.who.int/iris/bitstream/10665/162114/1/WHO_NMH_NHD_EPG_15.01.pdf?ua=1.
346. Tomaszewski, J.J., et al., *Increased cancer cell proliferation in prostate cancer patients with high levels of serum folate.* Prostate, 2011. **71**(12): p. 1287-93.
347. Matsue, H., et al., *Folate receptor allows cells to grow in low concentrations of 5-methyltetrahydrofolate.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 6006-9.
348. Moussa, C., et al., *Altered folate metabolism modifies cell proliferation and progesterone secretion in human placental choriocarcinoma JEG-3 cells.* British Journal of Nutrition, 2015. **114**(6): p. 844-852.
349. Al-Nasiry, S., et al., *The use of Alamar Blue assay for quantitative analysis of viability, migration and invasion of choriocarcinoma cells.* Human Reproduction, 2007. **22**(5): p. 1304-1309.
350. Fan, J., et al., *Quantitative flux analysis reveals folate-dependent NADPH production.* Nature, 2014. **510**(7504): p. 298-302.
351. Duthie, S.J., *Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis.* J Inherit Metab Dis, 2011. **34**(1): p. 101-9.
352. Blount, B.C. and B.N. Ames, *DNA damage in folate deficiency.* Baillieres Clin Haematol, 1995. **8**(3): p. 461-78.
353. Deconinck, E., et al., *A validated ultra high pressure liquid chromatographic method for qualification and quantification of folic acid in pharmaceutical preparations.* J Pharm Biomed Anal, 2011. **54**(5): p. 995-1000.
354. García-Casal, M.N., et al., *High prevalence of folic acid and vitamin B12 deficiencies in infants, children, adolescents and pregnant women in Venezuela.* Eur J Clin Nutr, 2005. **59**(9): p. 1064-70.
355. UK, C.R. *What is childhood acute lymphoblastic leukaemia (ALL) ?* 2021 21/12/2022]; Available from: https://www.cancerresearchuk.org/about-cancer/childrens-cancer/acute-lymphoblastic-leukaemia/about.
356. Williams, L.A., et al., *Is There Etiologic Heterogeneity between Subtypes of Childhood Acute Lymphoblastic Leukemia? A Review of Variation in Risk by Subtype.* Cancer Epidemiol Biomarkers Prev, 2019. **28**(5): p. 846-856.
357. Koukoura, O., S. Sifakis, and D.A. Spandidos, *DNA methylation in the human placenta and fetal growth (review).* Molecular medicine reports, 2012. **5**(4): p. 883-889.
358. van Dijk, S.J., et al., *Recent developments on the role of epigenetics in obesity and metabolic disease.* Clin Epigenetics, 2015. **7**: p. 66.
359. Joubert, B.R., et al., *Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns.* Nat Commun, 2016. **7**: p. 10577.
360. Joubert, B.R., et al., *DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis.* Am J Hum Genet, 2016. **98**(4): p. 680-96.
361. Nordlund, J., et al., *Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia.* Genome Biol, 2013. **14**(9): p. r105.
362. Yu, G., et al., *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.* OMICS: A Journal of Integrative Biology, 2012. **16**(5): p. 284-287.
363. Sherman, B.T., et al., *DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis.* BMC Bioinformatics, 2007. **8**: p. 426.
364. Richmond, R.C., et al., *Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC).* Hum Mol Genet, 2015. **24**(8): p. 2201-17.
365. Yousefi, P.D., et al., *DNA methylation-based predictors of health: applications and statistical considerations.* Nat Rev Genet, 2022. **23**(6): p. 369-383.

366. Gonseth, S., et al., *Periconceptional folate consumption is associated with neonatal DNA methylation modifications in neural crest regulatory and cancer development genes.* Epigenetics, 2015. **10**(12): p. 1166-76.
367. Desmirean, M., et al., *Correlation between the prevalence of T-cell lymphomas and alcohol consumption.* Med Pharm Rep, 2021. **94**(3): p. 298-306.
368. de la Chica, R.A., et al., *Increased MLL gene rearrangements in amniocytes from fetuses of mothers who smoke.* Leuk Res, 2011. **35**(8): p. 1066-9.
369. Ishizawa, S., et al., *High frequency of pro-B acute lymphoblastic leukemia in adults with secondary leukemia with 11q23 abnormalities.* Leukemia, 2003. **17**(6): p. 1091-5.
370. Abdulwahab, A., et al., *Therapy-related acute lymphoblastic leukemia is more frequent than previously recognized and has a poor prognosis.* Cancer, 2012. **118**(16): p. 3962-3967.
371. Golding, J., M. Pembrey, and R. Jones, *ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology.* Paediatr Perinat Epidemiol, 2001. **15**(1): p. 74-87.
372. Relton, C.L., et al., *Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES).* Int J Epidemiol, 2015. **44**(4): p. 1181-90.
373. Chicco, D. and G. Agapito, *Nine quick tips for pathway enrichment analysis.* PLOS Computational Biology, 2022. **18**(8): p. e1010348.
374. Van Camp, J.K., et al., *Wnt signaling and the control of human stem cell fate.* Stem Cell Rev Rep, 2014. **10**(2): p. 207-29.
375. Ruan, Y., et al., *Wnt Signaling in Leukemia and Its Bone Marrow Microenvironment.* Int J Mol Sci, 2020. **21**(17).
376. Gaudet, P. and C. Dessimoz, *Gene Ontology: Pitfalls, Biases, and Remedies.* Methods Mol Biol, 2017. **1446**: p. 189-205.
377. Yu, R., et al., *iTRAQ-based quantitative protein expression profiling of biomarkers in childhood B-cell and T-cell acute lymphoblastic leukemia.* Cancer Manag Res, 2019. **11**: p. 7047-7063.
378. Teachey, D.T. and C.H. Pui, *Comparative features and outcomes between paediatric T-cell and B-cell acute lymphoblastic leukaemia.* Lancet Oncol, 2019. **20**(3): p. e142-e154.
379. Ion, R. and A.L. Bernal, *Smoking and Preterm Birth.* Reproductive Sciences, 2015. **22**(8): p. 918-926.
380. Miller, M.A., et al., *Gene and metabolite time-course response to cigarette smoking in mouse lung and plasma.* PLoS One, 2017. **12**(6): p. e0178281.
381. Sasco, A.J., M.B. Secretan, and K. Straif, *Tobacco smoking and cancer: a brief review of recent epidemiological evidence.* Lung Cancer, 2004. **45 Suppl 2**: p. S3-9.
382. El Hajj Chehadeh, S., et al., *Early methyl donor deficiency alters cAMP signaling pathway and neurosteroidogenesis in the cerebellum of female rat pups.* Am J Physiol Endocrinol Metab, 2014. **307**(11): p. E1009-19.
383. Skah, S., et al., *cAMP-mediated autophagy inhibits DNA damage-induced death of leukemia cells independent of p53.* Oncotarget, 2018. **9**(54): p. 30434-30449.
384. Iroegbu, J.D., et al., *ERK/MAPK signalling in the developing brain: Perturbations and consequences.* Neurosci Biobehav Rev, 2021. **131**: p. 792-805.
385. Pillai, P.M., et al., *Activation of the mitogen-activated protein kinase-extracellular signal-regulated kinase pathway in childhood B-cell acute lymphoblastic leukemia.* Pediatr Blood Cancer, 2022. **69**(10): p. e29771.
386. Wang, Y., et al., *P38 MAPK/AKT signalling is involved in IL-33-mediated anti-apoptosis in childhood acute lymphoblastic leukaemia blast cells.* Ann Med, 2021. **53**(1): p. 1461-1469.
387. Noda, M., et al., *A single-molecule assessment of the protective effect of DMSO against DNA double-strand breaks induced by photo-and γ-ray-irradiation, and freezing.* Sci Rep, 2017. **7**(1): p. 8557.
388. Pal, R., et al., *Diverse effects of dimethyl sulfoxide (DMSO) on the differentiation potential of human embryonic stem cells.* Arch Toxicol, 2012. **86**(4): p. 651-61.

389. Srinivas, S., T.A. Sironmani, and G. Shanmugam, *Dimethyl sulfoxide inhibits the expression of early growth-response genes and arrests fibroblasts at quiescence.* Experimental Cell Research, 1991. **196**(2): p. 279-286.
390. Guo, W., et al., *Low-concentration DMSO accelerates skin wound healing by Akt/mTOR-mediated cell proliferation and migration in diabetic mice.* Br J Pharmacol, 2020. **177**(14): p. 3327-3341.
391. Mirabelli, P., L. Coppola, and M. Salvatore, *Cancer Cell Lines Are Useful Model Systems for Medical Research.* Cancers (Basel), 2019. **11**(8).
392. Quentmeier, H., et al., *The LL-100 panel: 100 cell lines for blood cancer studies.* Scientific Reports, 2019. **9**(1): p. 8218.
393. Kim, J., B.-K. Koo, and J.A. Knoblich, *Human organoids: model systems for human biology and medicine.* Nature Reviews Molecular Cell Biology, 2020. **21**(10): p. 571-584.
394. Joshi, R., et al., *The DNA methylation landscape of human cancer organoids available at the American type culture collection.* Epigenetics, 2020. **15**(11): p. 1167-1177.
395. Jones, L.J., et al., *Sensitive determination of cell number using the CyQUANT cell proliferation assay.* J Immunol Methods, 2001. **254**(1-2): p. 85-98.
396. Spagnolo, D.V., et al., *The role of molecular studies in lymphoma diagnosis: a review.* Pathology, 2004. **36**(1): p. 19-44.
397. Fueller, E., et al., *Genomic inverse PCR for exploration of ligated breakpoints (GIPFEL), a new method to detect translocations in leukemia.* PLoS One, 2014. **9**(8): p. e104419.
398. Lu, Y., Y. Liu, and C. Yang, *Evaluating In Vitro DNA Damage Using Comet Assay.* J Vis Exp, 2017(128).
399. Sommer, S., I. Buraczewska, and M. Kruszewski, *Micronucleus Assay: The State of Art, and Future Directions.* Int J Mol Sci, 2020. **21**(4).
400. Hovhannisyan, G.G., *Fluorescence in situ hybridization in combination with the comet assay and micronucleus test in genetic toxicology.* Molecular Cytogenetics, 2010. **3**(1): p. 17.
401. Plappert-Helbig, U., et al., *Gamma-H2AX immunofluorescence for the detection of tissue-specific genotoxicity in vivo.* Environ Mol Mutagen, 2019. **60**(1): p. 4-16.
402. Motoyama, S., et al., *Advantages of evaluating γH2AX induction in non-clinical drug development.* Genes and Environment, 2018. **40**(1): p. 10.
403. Vineis, P. and F. Perera, *Molecular Epidemiology and Biomarkers in Etiologic Cancer Research: The New in Light of the Old.* Cancer Epidemiology, Biomarkers & Prevention, 2007. **16**(10): p. 1954-1965.
404. Roberts, K.G., *Genetics and prognosis of ALL in children vs adults.* Hematology Am Soc Hematol Educ Program, 2018. **2018**(1): p. 137-145.
405. Lin, W.Y., et al., *Genome-wide association study identifies susceptibility loci for acute myeloid leukemia.* Nat Commun, 2021. **12**(1): p. 6233.
406. Neal, R.D., et al. *Cell-Free DNA–Based Multi-Cancer Early Detection Test in an Asymptomatic Screening Population (NHS-Galleri): Design of a Pragmatic, Prospective Randomised Controlled Trial*. Cancers, 2022. **14**, DOI: 10.3390/cancers14194818.
407. Hackshaw, A., et al., *Estimating the population health impact of a multi-cancer early detection genomic blood test to complement existing screening in the US and UK.* Br J Cancer, 2021. **125**(10): p. 1432-1442.
408. Loud, J.T. and J. Murphy, *Cancer Screening and Early Detection in the 21(st) Century.* Semin Oncol Nurs, 2017. **33**(2): p. 121-128.
409. Ott, J.J., A. Ullrich, and A.B. Miller, *The importance of early symptom recognition in the context of early detection and cancer survival.* Eur J Cancer, 2009. **45**(16): p. 2743-8.
410. Gore, L., *What are the long-term complications of pediatric ALL treatments and how can they be mitigated? Perspectives on long-term consequences of curative treatment in childhood ALL.* Best Practice & Research Clinical Haematology, 2022. **35**(4): p. 101403.

411. Wilson, R.D. and D.L. O'Connor, *Maternal folic acid and multivitamin supplementation: International clinical evidence with considerations for the prevention of folate-sensitive birth defects.* Prev Med Rep, 2021. **24**: p. 101617.
412. James, J.E., *Maternal caffeine consumption and pregnancy outcomes: a narrative review with implications for advice to mothers and mothers-to-be.* BMJ Evid Based Med, 2021. **26**(3): p. 114-115.
413. Arias, L.D., et al., *Study Exploring the Effects of Daily Supplementation with 400 µg of Folic Acid on the Nutritional Status of Folate in Women of Reproductive Age.* Birth Defects Res, 2017. **109**(8): p. 564-573.
414. Morris, J.K. and N.J. Wald, *Importance of getting the right UK folic acid fortification policy.* Archives of Disease in Childhood, 2023. **108**(1): p. 74.
415. Chakraborty, H., et al., *Folic Acid Fortification and Women's Folate Levels in Selected Communities in Brazil - A First Look.* Int J Vitam Nutr Res, 2014. **84**(5-6): p. 286-94.
416. Roh, S., *Scientific Evidence for the Addictiveness of Tobacco and Smoking Cessation in Tobacco Litigation.* J Prev Med Public Health, 2018. **51**(1): p. 1-5.
417. Ioakeimidis, N., et al., *Smoking cessation strategies in pregnancy: Current concepts and controversies.* Hellenic J Cardiol, 2019. **60**(1): p. 11-15.
418. Manuela, C., et al., *Environmental and biological monitoring of benzene in traffic policemen, police drivers and rural outdoor male workers.* J Environ Monit, 2012. **14**(6): p. 1542-50.