

**Folate induced mutagenesis and identification
of folate-specific mutation spectra in
somatic cells**

By

Stephanie Diaz G

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in Partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University
Ottawa, Ontario

© 2019
Stephanie Diaz

Abstract

Folate is an essential water soluble B-vitamin required for the *de novo* synthesis of purines and dTMP and the synthesis of methionine. Disruption of any of the three folate-dependent biosynthetic pathways has been shown to have serious implications in a cell's survival and an individual's disease risk. Folate, generally folate deficiency, has been associated with colorectal cancer (CRC) and acute lymphocytic leukemia (ALL) in epidemiological studies. However, CRC risk has also been associated with high FA intake. Previous studies have shown that FA deficiency induces higher mutant frequency (MF) in bone marrow in addition to increasing chromosomal instability in RBC. Here, we used the MutaMouse model to determine the mutagenic potential of dietary FA in the colon, and to determine if there is a tissue- and diet-specific effect induced by FA intake in colon and bone marrow. Male mice were fed experimental FA defined diets: deficient (0 mg/kg), control (2 mg/kg) and supplemented (8 mg/kg) for 20 weeks from weaning. To determine the effects of FA intake on the induction of somatic mutations by a chemical mutagen, half of the mice from each diet group were gavaged with 50 mg/kg ENU and the other half with saline after 10 weeks on diet, followed by an additional 10 weeks on diet. NGS was used to sequence the lacZ reporter gene and we determined the FA-induced mutation spectra in both tissues. In the colon, mice in the saline treatment fed the FA supplemented diet had a higher MF compared to mice fed the FA control diet ($p = 2.3 \times 10^{-5}$) and FA deficient diet ($p = 2.9 \times 10^{-5}$). This effect was mirrored in the mice in the ENU treatment where the FA supplemented diet caused a higher MF compared to mice fed the FA deficient diet ($p = 0.045$). In contrast, in the bone marrow, the highest MF was observed in mice fed the FA deficient diet. However this was only observed in the saline treated mice. When comparing

the mutation spectra of the colon and the bone marrow, we identified a FA-diet-specific mutagenicity in the bone marrow, where the mutation spectra differed significantly for mice fed the FA deficient diet compared to those fed the control and supplemented diets. In the colon, however, we did not observe a difference in the mutation spectra across the diets, suggesting FA supplementation enhances the overall incidence of colon-specific mutation types as opposed to inducing a unique spectra. Furthermore, we show that the type of mutations observed in both tissues are characteristic of CRC and ALL. Ultimately, we demonstrated, for the first time, the mutagenic potential of FA-intake, the FA-induced mutation spectra by different FA intake levels and that the mutagenic potential of FA is tissue-specific.

Acknowledgments

First and foremost, I would like to thank the amazing Dr. Amanda MacFarlane who besides being an incredible scientist and my supervisor, she has also become a mentor and a role model inside and outside the lab. Also thank you for being so understanding, supporting and for embracing my uniqueness and culture ;) . Thank you, Dr. Alex Wong, for all your teachings and being my first formal introduction into programming. You and Dr. MacFarlane are not only two exemplary scientists but also individuals and parents. Also, a special thanks to my thesis supervisory committee, Dr. Francesco Marchetti and Dr. Nicolas Corradi, for their guidance and much appreciated input. I would also like to thank Danielle LeBlanc for the many hours of training, patience, willingness to help, and as a great drinking companion. A huge thank you to Remi Gagne for the many hours of coding, explaining, deciphering, and patience. Thank you for making my love for programming grow even more and for showing me things that blew my mind. Nathalie Behan thank you for always, and I mean always, being there. Thank you for your patience, help, guidance, for your stories, and for having a great taste in music (and for lending me your speaker ☺). A big thank you to the Canadian Institute of Health Research (CIHR) and Carleton University for their financial support, which allowed me to focus on classes and lab work throughout the two years. An immense thank you to my mom, dad and sister who have always, and will always, be there. Finally, thank you to the person that was there with me all the stressful afternoons and nights, supporting me and passing me a beer when necessary; thank you Kristianne Clark Galpin for always being there and for believing in me in those doubtful moments, especially while writing this thesis.

Table of Contents

<i>Abstract</i>	<i>ii</i>
<i>Acknowledgments</i>	<i>iv</i>
<i>Table of Contents</i>	<i>v</i>
<i>List of Abbreviations</i>	<i>vii</i>
<i>List of Figures</i>	<i>ix</i>
<i>List of Tables</i>	<i>x</i>
Chapter 1: Introduction	1
1.1 Folate and folate mediated one-carbon metabolism	1
1.1.1 The folate-mediated one-carbon metabolism	5
1.1.2 Folate transport.....	8
1.2 Mutagenic potential of folate	9
1.2.1 Purine and thymidylate synthesis	9
1.2.2 Methylation	11
1.2.3 Single nucleotide polymorphisms	12
1.3 Folate intake and health considerations	13
1.3.1 FA and health implications	14
1.3.2 FA and cancer	17
1.4 Model for studying mutations <i>in vivo</i>	22
1.5 Hypothesis	23
1.6 Objectives	24
Chapter 2: General overview and rationale of methods	25
2.1 Animal Model: The MutaMouse	25
2.2 Animal study design	28
2.3 Necropsy and tissue collection	31
2.4 Colon <i>lacZ</i> assay	31
2.5 Preparation for next-generation sequencing	32
2.6 Illumina sequencing	33
2.7 Bioinformatics	34
2.8 Previously reported results relevant to the study	35
2.8.1 Body Weights.....	35
2.8.2 Plasma and RBC folate.....	37
2.8.3 Bone marrow <i>lacZ</i> mutant frequency.....	39
Chapter 3: Investigating the folate-induced mutant frequency and mutation spectrum in the colonic epithelium of the male MutaMouse	41
3.1 Abstract	41
3.2 Introduction	42

3.3 Methods	48
3.3.1 Mouse study and colon collection.....	48
3.3.2 Mouse colon DNA extraction	49
3.3.3 Colon <i>lacZ</i> mutant frequency analysis.....	50
3.3.4 Statistical analysis.....	50
3.3.5 Library preparation and next-generation sequencing.....	51
3.3.6 Bioinformatics	52
3.4 Results	54
3.4.1 Colon mutant frequency.....	54
3.4.2 Sequencing of the colon mutant plaques.....	58
3.4.3 Proportion of mutation types of the colon	60
3.4.4 Distribution and hotspots of base substitutions and indels along the <i>lacZ</i> gene in the colon.....	63
3.5 Discussion	69
3.6 Conclusion	74
<i>Chapter 4: Investigating the folate-induced mutation spectra and its tissue-specific mutagenicity potential in the colon epithelium and bone marrow of the male MutaMouse</i>	76
4.1 Abstract	76
4.2 Introduction	77
4.3 Methods	83
4.3.1 Animal study	83
4.3.2 <i>LacZ</i> assay and preparation for sequencing	84
4.3.3 Illumina sequencing	85
4.3.4 Bioinformatics	86
4.4 Results	87
4.4.1 Comparison of mutant frequencies across tissues.....	87
4.4.2 Sequencing of the <i>lacZ</i> reporter gene in somatic tissues.....	89
4.4.3 Somatic tissue mutation spectra comparison	91
4.4.4 Distribution and hotspots of base substitutions and indels along the <i>lacZ</i> gene in bone marrow and comparison to hotspots in colon.....	95
4.5 Discussion	101
4.6 Conclusion	108
<i>Chapter 5: Discussion</i>	110
5.1 Context for the study	110
5.1.1 Mutagenic potential of folate	110
5.1.2 FA and CRC	111
5.1.3 FA-induced mutation spectra and somatic tissue specificity.....	115
5.2 Study strengths and limitations	120
5.3 Future directions	121
5.4 Conclusions	121
<i>References</i>	123

List of Abbreviations

AdoMet	S-Adenosylmethionine
ALL	Acute Lymphocytic Leukemia
BER	Base Excision Repair
CRC	Colorectal Cancer
DHF	Dihydrofolate
DHFR	Dihydrofolate Reductase
DSB	Double-strand Breaks
DRI	Dietary Reference Intake
dUMP	Deoxyuridine Monophosphate
dTMP	Thymidine Monophosphate
ENU	N-ethyl-N-nitrosurea
FA	Folic Acid
FR	Folate Receptors (FR α / FR β)
MTHFR	MethyleneTHF Reductase
MF	Mutant Frequency
MS	Methionine Synthase
NGS	Next Generation Sequencing
NTDs	Neural Tube Defects
PCFT	Proton-coupled Folate Transporters
PFUs	Plaque Forming Units
RBC	Red Blood Cell
RDA	Recommended Dietary Allowance

RFC	Reduced Folate Carrier
ROS	Reactive Oxygen Species
SBS	Single Base Substitution
SHMT	Serine Hydroxymethyltransferase
SNP	Single Nucleotide Polymorphism
THF	Tetrahydrofolate
TS	Thymidylate Synthase
UL	Upper Intake Level

List of Figures

Figure 1.1: Structure of tetrahydrofolate triglutamate.

Figure 1.2. Structures of folate compounds

Figure 1.3. Folate-mediated one carbon metabolism

Figure 2.1: Overview of the lacZ mutation assay

Figure 2.2: Animal study design

Figure 2.3: Body weights of male mice fed three FA diets

Figure 2.4: (A) Plasma Folate and (B) RBC folate

Figure 2.5: Bone marrow lacZ MF of male mice fed FA- defined diets and exposed to saline or ENU

Figure 3.1: LacZ mutant frequency in the colon of male mice fed FA defined diets

Figure 3.2: Proportional representation of mutation spectra of the colon of male mice on the saline treatment fed one of three FA diets

Figure 3.3: LacZ mutation spectra of the colon in mice fed one of three FA diets

Figure 3.4: Distribution of base substitutions across the lacZ gene in the colon in mice fed one of three folic acid diets

Figure 3.5: Distribution of indels across the lacZ gene in the colon in mice fed one of three folic acid diets

Figure 4.1: Comparison of lacZ mutant frequency in somatic tissues

Figure 4.2: Proportional representation of mutation spectra of the bone marrow and colon of male mice on the saline treatment fed one of three FA diets

Figure 4.3: LacZ mutation spectra of the colon and bone marrow in mice fed one of three FA diets

Figure 4.4: Distribution of base substitutions across the lacZ gene in the bone marrow in mice fed one of three folic acid diets

Figure 4.5: Distribution of indels across the lacZ gene in the bone marrow in mice fed one of three folic acid diets

List of Tables

Table 3.1: LacZ mutant frequency in colon

Table 3.2: Total of samples sequenced, and number of independent mutations identified in the lacZ gene from the colon

Table 3.2: Position of possible base substitution hotspots

Table 3.3: Indel hotspots along the lacZ gene in the colon

Table 4.1: Samples sequenced and number of independent mutations identified in the lacZ gene from the colon and bone marrow

Table 4.2: Position of candidate base substitution hotspots

Table 4.3: Indel hotspots along the lacZ gene in the bone marrow

Chapter 1: Introduction

1.1 Folate and folate mediated one-carbon metabolism

Folate is an essential water-soluble B vitamin that is present in all cells. By transferring and activating one-carbon-units, it acts as an enzyme cofactor for the synthesis of nucleotides (purines and thymidine) and the synthesis of methionine from homocysteine¹. Fresh fruits and vegetables are natural food sources of folate. Folic acid (FA) is a synthetic, oxidized and stable form of folate often found in fortified foods (e.g. white flour and cereal) and dietary supplements.

Folate contains a reduced pteridine moiety, with A and B rings linked to *p*-aminobenzoylglutamate by a methylene group at carbon 6. An oxidative and irreversible cleavage of the C⁹-N¹⁰ bond can degrade folate to pteridine and *p*-aminobenzoylglutamate (Fig.1.1). Once inside the cell, folates are found as polyglutamates, however, in order for folates to be transported into the cells they need to be in the form of monoglutamate vitamins. Therefore, dietary polyglutamate folate forms must be converted to their monoglutamate forms prior to intestinal absorption, by glutamate carboxypeptidase II in the human intestine and by γ -glutamyl hydrolase in rodent intestine^{2,3}. After folate monoglutamate is generated, it is then transported into the intestinal epithelial cells by proton-coupled folate transporters (PCFT). Once in the cells, the folate monoglutamates are processed into functional metabolic cofactors by restoring the poly-glutamate chain by the polyglutamyl synthetase enzyme, which progressively adds γ -linked polyglutamate peptides. The intracellular folates have polyglutamate tails varying in length, up to 9 glutamate residues. The glutamate polypeptide allows the retention of folate in the cell due to increased affinity for intracellular folate-binding enzymes⁴. It is these

tetrahydrofolate (THF) polyglutamates at various oxidation levels of formate that serve to carry and activate one-carbon groups for biosynthetic reactions ^{1,3}. As such, THF is the bioactive form of folate.

Folate from natural food sources is predominately found in the form of 5-methyl-tetrahydrofolate (5-methyl-THF) which is converted to THF through the transfer of a methyl group to homocysteine, via methionine synthase, to form methionine, a precursor for the synthesis of S-Adenosylmethionine (AdoMet), a reaction catalyzed by the vitamin B12 dependent enzyme methionine synthase (Fig. 1.2).

FA, the oxidized form of folate, contains a single glutamate and itself is not bioactive. Once it is transported into the cell by the PCFTs, it is first reduced to dihydrofolate (DHF) at the B ring and then to THF by dihydrofolate reductase (DHFR), making it chemically identical to natural food folates ^{1,3}(Fig. 1.2).

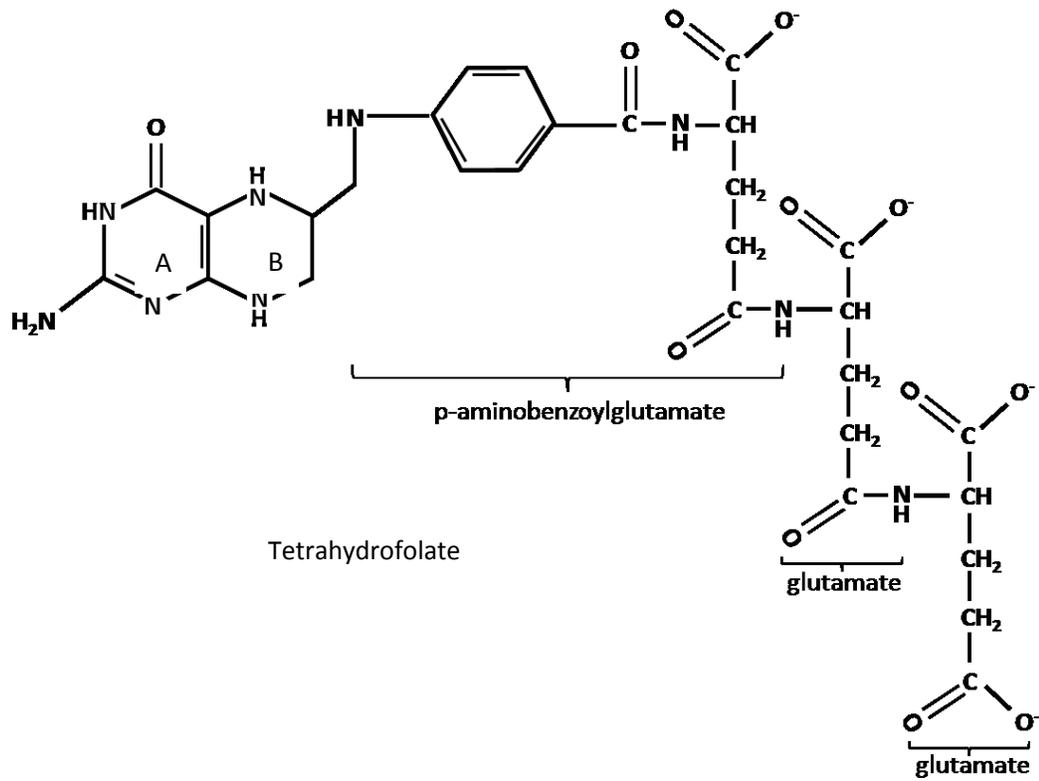


Figure 1.1: Structure of tetrahydrofolate triglutamate.

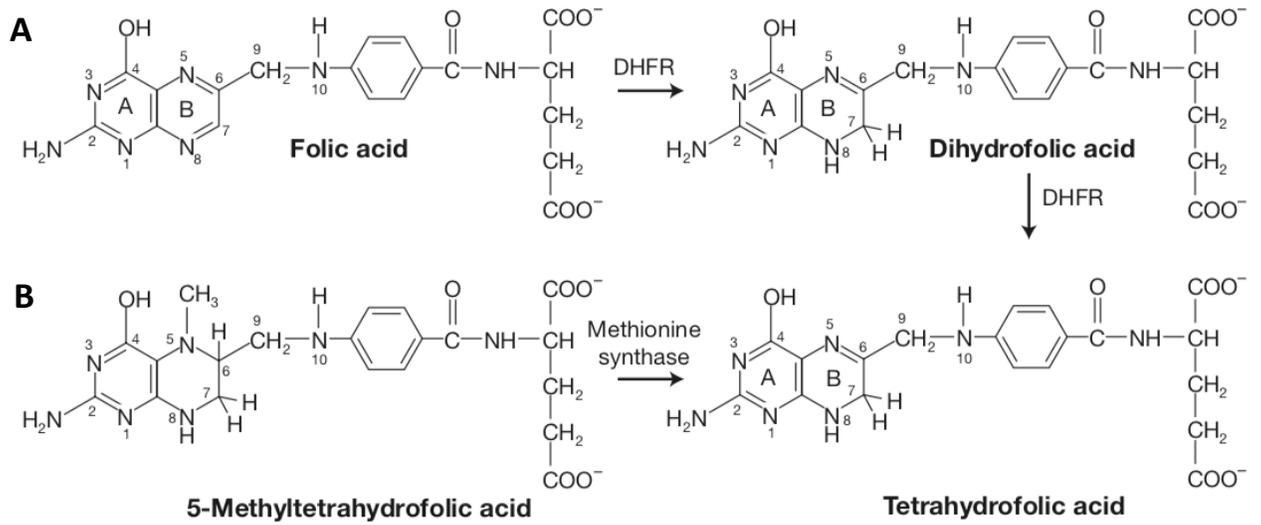


Figure 1.2. Structures of folate compounds: (A) Structure of folic acid and (B) Structure of folate from natural food sources and their conversions to tetrahydrofolic acid (THF), their bioactive form. DHFR, dihydrofolate reductase.

1.1.1 The folate-mediated one-carbon metabolism

The folate-mediated one-carbon metabolism encompasses many independent metabolic pathways which require THF as a cofactor to carry and activate one-carbon units in cellular biosynthetic reactions. Folate metabolism in the cell is compartmentalized in the nucleus, cytoplasm and mitochondria, each with different endpoints ¹. One-carbon metabolism in the mitochondria results in the conversion of serine to formate, which is required for the synthesis of glycine and formylmethinyl-tRNA, which feeds back to be used in mitochondrial protein synthesis ^{1,4,5}. One-carbon metabolism in the nucleus is involved in dTMPs biosynthesis ⁴.

One-carbon metabolism in the cytoplasm includes two biosynthetic pathways, the *novo* biosynthesis of purines and the remethylation of homocysteine to form methionine ⁴. Initially, formate derived from the mitochondria serves as a formyl group donor to THF to form 10-formylTHF in an ATP-dependent reaction catalyzed by 10-formylTHF synthetase ⁶ (Figure 1.3). Here, 10-formylTHF is required for the *de novo* biosynthesis of purines by supplying the #2 and #8 carbons to the purine ring through the activity of the purinosome, a multi-complex consisting of six enzymes found in the cytoplasm ⁴. Purines can also be synthesized by a single-step salvage pathway where hypoxanthine and phosphoribosyl are condensed into inosine monophosphate, which is a precursor for AMP and GMP synthesis. Regulation of purine synthesis is accomplished through feedback inhibition of the *de novo* pathway and the formation of purinosome by purine nucleotides synthesized by the salvage pathway ^{4,7}.

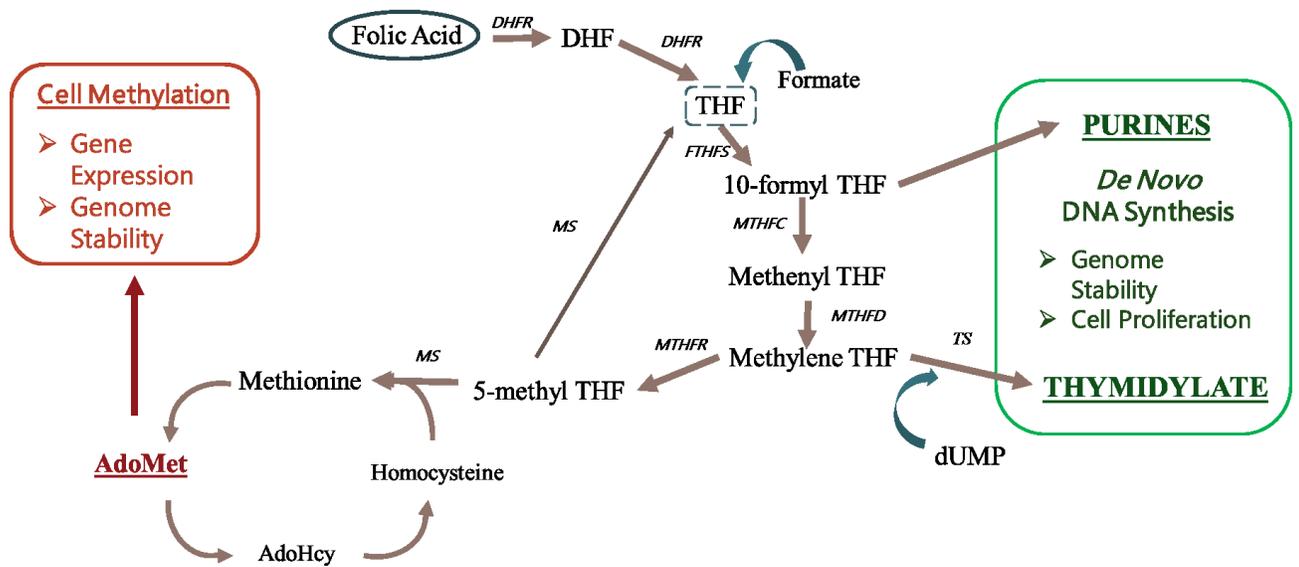


Figure 1.3. Folate-mediated one carbon metabolism. AdoHcy, adenosylhomocysteine; AdoMet, adenosylmethionine; DHF, dihydrofolate; DHFR, dihydrofolate reductase; FTHFS, 10-formylTHF synthetase; MTHFC, methenylTHF cyclohydrolase; MTHFD, methyleneTHF dehydrogenase; MTHFR, methyleneTHF reductase; dUMP, deoxyuridine monophosphate; MS, methionine synthase; THF, tetrahydrofolate; TS, thymidylate synthase.

Alternatively, 10-formylTHF can be reduced by methenyl cyclohydrolase to 5,10-methenylTHF and again by methyleneTHF dehydrogenase 1 (MTFHD1) in a NADPH dependent reaction to form 5,10-methyleneTHF. Here, 5,10-methyleneTHF is used by thymidylate synthase (TS) as a cofactor for the methylation of deoxyuridine monophosphate (dUMP) to form thymidylate (thymidine monophosphate [dTMP]). This is the only folate-dependent reaction where folate cofactors serve as donors of one-carbons and as donors of 2 electrons by the oxidation of THF ^{4,6}. This biosynthetic pathway takes place in the nucleus where the enzymes involved in *de novo* dTMP biosynthesis (serine hydroxymethyltransferase [SHMT], TS and DHFR) translocate following sumoylation during S-phase ⁸.

Alternatively, 5,10-methyleneTHF can be further reduced to 5-methylTHF by methyleneTHF reductase (MTHFR) in an NADPH-dependent irreversible reaction. 5-methylTHF donates a one-carbon to allow for the re-methylation of homocysteine to form methionine, and enter the methionine/homocysteine re-methylation cycle ⁴. This reaction is catalyzed by methionine synthase (MS), a B-12-dependent enzyme. Methionine can be adenylated to form AdoMet, which serves as the universal methyl donor of many methylation reactions in the cell, subsequently controlling gene transcription and protein expression through the methylation of DNA and histones ⁹. Once methyltransferase enzymes have that are dependent on AdoMet have donated their methyl group, S-adenosylhomocysteine (AdoHcy) is formed as a byproduct, which is hydrolyzed to homocysteine, thus, continuing the methionine/homocysteine cycle ⁶.

The methionine cycle is tightly regulated to ensure the maintenance of methionine and AdoMet synthesis, and to prevent the accumulation of other metabolites. Because the

reduction of 5,10-methyleneTHF to 5-methylTHF is irreversible, any disruption to the methionine cycle can result in the accumulation of 5-methylTHF, known as the “methyl trap”, and affects folate-derivative-dependent reactions due to the lost regeneration of THF. The methyl trap can be a direct result of vitamin B-12 deficiency since B-12 is a cofactor of MS and is needed to convert 5-methyl-THF to THF by the re-methylation of homocysteine⁶.

Because folate mediated one-carbon metabolism is essential for the *de novo* synthesis of purines, thymidylate, and synthesis of AdoMet, folate plays an important role in the regulation of DNA biosynthesis, repair, and cellular methylation.

1.1.2 Folate transport

Mammalian cells cannot synthesize folates; therefore, folate requirements must be met from dietary sources, natural food sources, or fortified foods. Specific transporters mediate the intestinal absorption of folate due to the hydrophilic nature of the molecule³. PCFT is a high-affinity folate transporter that allows the low-pH folate transport into many tissues; this is the primary folate transporter in the small intestine. Transport of folates at a neutral to mildly acidic pH is accomplished by receptor-mediated endocytosis by folate receptors (FR α / FR β) localized in the cell membrane and both high affinities for FA. At physiological pH, the reduced folate carrier (RFC), an anionic exchanger, is the major deliverer of folates to systemic tissues and has very low affinity for FA. RFC has an important role in folate homeostasis given its ubiquitous expression in tissues¹⁰.

Once folate monoglutamate is generated, it can be transported across the proximal jejunum by PCFT found in the apical brush-border membrane of epithelial cells. The high concentration of folate in the enterocytes facilitates its efflux into the periserosal space

across the basolateral membrane, a process believed to be mediated by a member of the multidrug-resistance-associated protein. From here, folates enter the vascular system. Folates that have been absorbed by the intestine can also be delivered to the hepatic sinusoids next to the sinusoidal membrane where PCFT is also localized. Once in the liver, folates can either (1) be converted into polyglutamate storage forms, (2) they can be secreted into the bile and be reabsorbed in the jejunum, or (3) they can enter the hepatic portal vein and reach systemic circulation³. Finally, folate can be reabsorbed in the kidney where high concentrations of folate accumulate, and renal reabsorption is mediated by FR α . Blood folates that are not bound to serum proteins are filtered at the glomerulus and excreted in the urine. However, clearance of folate is low due to its highly efficient reabsorption by FR α at the luminal brush-border membrane of proximal tubules^{3,11,12}.

1.2 Mutagenic potential of folate

1.2.1 Purine and thymidylate synthesis

Folate mediated one-carbon metabolism plays a critical role in *de novo* synthesis of purines and dTMP, encompassing three of the four nucleotides required for DNA synthesis. *De novo* nucleotide synthesis is important for cell proliferation and genome stability. Any disruption to the one carbon metabolic pathways involved in the *de novo* synthesis of these nucleotides can result in nucleotide pool imbalances, impaired DNA synthesis and cell death^{13,14}. The effect of insufficient purine synthesis was studied in Ade-C cells, a Chinese hamster ovary cell line which is auxotrophic for purines. The inability to synthesize *de novo* purines resulted in a significant reduction in DNA replication, abnormal repair of UV-induced DNA damage and increase in cell mutagenicity and lethality¹³.

De novo biosynthesis of dTMP requires the TS-mediated transfer of the methyl group from 5,10-methyleneTHF to dUMP to form dTMP. Low cytosolic concentration of 5,10-methyleneTHF decreases synthesis of dTMP which in turn increases the cellular dUMP/dTMP ratio. Therefore, folate deficiency disrupts the thymidylate pathway resulting in an increased concentration of dUMP, thus leading to the misincorporation of uracil into DNA ^{15,16}.

Removal of misincorporated uracil in DNA is completed via the base excision repair (BER) pathway in mammalian cells. BER is initiated by cleaving a glycosidic bond to release the incorrect base via uracil DNA glycosylase (UDG). Subsequently, it requires a DNA re-synthesis step, by polymerase activity in addition to ligase, to introduce the proper base and seal the nick ¹⁵. However, the DNA re-synthesis step of BER can encounter problems due to the low cellular concentrations of dTMP due to folate deficiency. The high concentration of dUMP can result in re-introduction of uracil into the repaired DNA strand, thus continuing a “futile cycling” of BER, and resulting in persistent DNA breaks and gaps ¹⁵. If DNA breaks occur on opposite strands in the same region of DNA undergoing repair, it can result in double-strand DNA breaks (DSB). Hence, disruption of *de novo* thymidylate synthesis leads to DNA instability and increases mutagenesis.

Because the conversion of dUMP to dTMP is entirely folate dependent, uracil accumulation in DNA is considered a specific and functional biomarker of folate status ^{17,18}. Inverse association was observed between uracil misincorporation in lymphocyte DNA and the folate status in red blood cells ¹⁸. This inverse association between folate status and uracil content has also been seen in mice ¹⁹.

1.2.2 Methylation

DNA methylation is an epigenetic modification that is mitotically stable and heritable, and which can change gene expression not caused by changes in the DNA coding sequence. Epigenetic regulation of gene expression is critical for all cellular processes including cellular differentiation, embryonic development, organogenesis, etc. Silencing and activation of genes, DNA repair, imprinting, and chromatin remodeling are all regulated by epigenetic mechanisms. Therefore, any dysregulation of these epigenetic processes can result in altered risk for chronic diseases and developmental disorders²⁰⁻²².

5-methylTHF is required to remethylate homocysteine to form methionine, in order to be metabolized into AdoMet, an essential methyl donor needed for the induction and maintenance of cytosine methylation^{20,23}. DNA methylation patterns reflect the frequency at which cytosine is methylated at specific sites along the DNA. CpG islands are typically unmethylated and are usually associated with the 5' end of housekeeping genes and tissue-specific genes, as well as in the some 3' ends and exons; therefore methylation of these segments regulate gene expression directly^{20,24}. Hypermethylation of CpG islands in promoter, exons and shore regions are correlated with transcriptional silencing of the corresponding gene^{25,26}. Silencing by hypermethylation is essential in the case of X-chromosome inactivation, where one of the two X chromosomes in the somatic cells of females is randomly inactivated in order to compensate for the unequal dose of x-linked genes between females and males. This inactivation happens due to the hypermethylation of CpG islands in the promoters along the chromosome^{20,27,28}. On the other hand, methylation can also activate transcription, as is the case for the telomerase gene (*hTERT*), where the DNA polymerase adds TTAGGG repeats to the chromosome to maintain

telomere length ²⁹. Methylation of the promoter of p53 protein is needed to regulate its expression and it is directly influenced by nutrient status of folic acid and vitamin B12 ³⁰. Female and male patients with MA have low FA, which directly correlates with lower methylation in the p53 promoter, leading to its higher expression in these patients ³⁰.

Folate deficiency lessens the formation of AdoMet, which could result in gene-specific or genome wide hypomethylation of DNA. Therefore it can potentially increase the expression of proto-oncogenes, decrease the expression of tumor-suppressor genes, and induce chromatin changes ^{21,23}.

1.2.3 Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs), and other mutations, in genes coding for enzymes and other proteins associated with folate metabolism and transport, can have causal role in the development of NTDs and other diseases.

MTHFR is a key enzyme in folate-mediated one-carbon metabolism that converts 5,10-methyleneTHF to 5-methylTHF, the latter of which is required to remethylate homocysteine to methionine. The two most common SNPs in MTHFR are 677C>T and 1298A>C. Mutations in this gene results in a reduction of the enzyme activity, lowering the concentration of circulating folate (5-methylTHF) and increasing plasma homocysteine concentration ³¹. In addition, effects of the MTHFR 677C>T on DNA methylation correlate with folate status where low global DNA methylation is seen only under conditions of low folate ³². Mothers who carried the same SNP in the MTHFR gene were more likely to have children with Down Syndrome ³³. MTHFR 677C>T homozygosity in infants and MTHFR 677C>T allele in combination with the 1298A>C allele are associated in an increased risk of NTDs ³⁴. MTHFR 677C>T and 1298A>C polymorphisms have been associated with an

increased risk of pancreatic cancer and they are associated with low RBC folate in people with the disease ³⁵.

The competition for folate precursors for synthesis of DNA nucleotides and methionine synthesis is mediated by SHMT, an enzyme that catalyzes the reversible conversion of serine and THF to 5,10-methyleneTHF. Individuals with the SHMT1 L474F (1420C>T) SNP exhibited higher serum folate than those with the LL allele, a polymorphism that impairs the nuclear transport of SHMT1 and results in its cytosolic accumulation, thus disrupting dTMP synthesis ³⁵.

1.3 Folate intake and health considerations

Adequate consumption of folate is key to avoid inadequacy given its important role in ensuring normal development, as well as growth and maintenance of overall health. The World Health Organization (WHO) adopted the Estimated Adequate Requirement (EAR) and Recommended Dietary Allowance (RDA) set by the US Institute of Medicine ¹⁷. These Dietary Reference Intake (DRIs) recommendations for folate are based on the dietary folate intakes required to maintain a red blood cell (RBC) folate concentration, a primary biomarker of folate status, that reduces the risk for developing macrocytic anemia ³⁶⁻³⁸. Serum folate is also a folate status biomarker that reflects more short-term intakes and homocysteine concentration is a functional biomarker of folate status but is also influenced by other B vitamins including vitamins B6 and B12.

The bioavailability of folic acid when compared to food folate is 85% and 50%, respectively. Therefore, the DRIs for folate are based on Dietary Folate Equivalent (DFE) values where 1 unit of natural folate is equivalent to 1.7 units of FA. The intake recommendation for children 4-8 years old is 200 ug/day (DFE), for youth 9-13 years old

is 300 ug/day (DFE), and for those over 14 years old (adults) is 400 ug/day (DFE), for males and females. These recommendations increase to 500-600 ug DFE/day during pregnancy and lactation. The DRIs also include the tolerable upper intake level (ULs), which is the highest usual daily intake likely to pose no risk of adverse health effects. The UL for FA for adults is 1000 ug/day. In the case of folate, the UL is unique in that it only applies to the synthetic form of folate, folic acid found in fortified foods and supplements

36-38

1.3.1 FA and health implications

Inadequate intake of folate is the major cause of folate deficiency. However, in addition to low dietary intake, folate deficiency can be attributed to a number of different causes including, but not limited to, an increased requirement for the vitamin due to neoplastic diseases or pregnancy, metabolic inhibitors, or malabsorption conditions, and various polymorphisms in genes involved in the one-carbon metabolism¹⁷. Some malabsorptive conditions that can cause deficiency are tropical sprue, extensive inflammatory bowel disease, and celiac disease³⁹. Chronic alcoholism is also associated with folate deficiency, due to increased renal excretion and intestinal malabsorption^{40,41}. Folate deficiency has also been linked to various chronic diseases, including cancer and vascular disease¹⁷.

1.3.1.1 FA deficiency and megaloblastic anemia

Megaloblastic anemia (MA), anemia characterized by macrocytic red blood cells, results from folate deficiency. Folate is an important cofactor required for the maturation of all cells, including marrow erythroblasts. Folate is needed for the *de novo* synthesis of

three of the four nucleotides required for DNA synthesis, thus low folate status disrupts DNA synthesis in RBC precursors, thus interfering with cell maturation and proliferation. Since RNA synthesis and cytoplasmatic elements remain relatively unaffected, the resulting erythroblasts, which are identified as megaloblasts, are large and oval shaped, with an immature nucleus^{30,42}.

1.3.1.2 FA deficiency and neural tube defects

Folate is important for development. During embryogenesis the completion of the neural tube closure is important for proper brain and nervous system development and it requires cell proliferation, differentiation, survival and migration. Any of these events can be disrupted by a disturbance in folate metabolism. Neural tube defects (NTDs), one of the most common types of congenital anomalies, arise when the embryonic neural tube fails to close completely during development early in the first trimester of pregnancy^{6,43}. There are two main forms of NTDs; anencephaly, which is the failure of the cranial end of the neural tube to close, and spina bifida, which is the failure of the caudal end of the neural tube to close. Anencephaly is lethal, resulting in death before or right after birth, and accounts for one-third of NTDs. Spina bifida accounts for the remaining two-thirds of NTDs, making it the most common NTD. Spina bifida causes paraplegia, paralysis of the lower extremities and impairments in bowel and bladder functions. However, unlike anencephaly, spina bifida is usually not fatal⁴³.

An association between folate metabolism and NTDs was first established in the early 1960's by Hibbard (1964). He analyzed the folate status of 1,484 pregnant women in Liverpool, using the formiminoglutamic acid (FIGLU) excretion test and bone marrow biopsies. The FIGLU excretion test works by assessing the relationship between folate and

histidine. Histidine needs FIGLU to be broken down to glutamic acid, a reaction that is dependent on folate and its derivatives, therefore in a deficient state the metabolism of histidine cannot continue beyond this step. This leads to an accumulation of FIGLU and its excretion in urine. The study found a higher excretion of FIGLU in mothers of infants born with congenital birth defects ⁴⁴. This was further demonstrated when mothers of infants with NTDs had significantly lower RBC folate in the first trimester compared to controls ⁴⁵. These findings motivated further studies to determine the preventative effect of FA towards NTDs ^{46,47}. A randomized double-blind prevention trial was launched in 1983 in seven countries by the Medical Research Council (MRC). This trial recruited a total of 1817 women at high risk of having a pregnancy with NTD, i.e. those with a previously affected pregnancy, who were then randomized into groups given supplementation of FA (4 mg FA), other vitamins (A, D, B_{1,2,3}, C, and nicotinamide), both or neither. The study concluded that FA supplementation confers a protective effect towards NTDs, showing that groups on the supplemented FA had a 72% protective effect on the recurrence of NTDs when compared to those in groups with no FA supplementation ⁴⁷. A subsequent study determined the effects of periconceptual vitamin supplementation and its efficacy in reducing the primary occurrence of NTDs, i.e. women with no history of NTDs in previous pregnancies. This was a randomized control trial where women planning a pregnancy were randomly assigned into a group given a multivitamin containing 0.8 mg FA, or a trace-element supplement, daily for a month prior to conception and up to the end of the third month of gestation. This study showed a 100% reduction of NTD risk in women in the FA supplemented group, compared to the trace-element supplemented group where some NTD-affected pregnancies were observed ⁴⁸. These studies demonstrated the protective

effect of FA supplementation against NTDs and led to the recommendation to consume a FA supplement before pregnancy and during the first trimester.

These results led to a series of public health initiatives promoting the use of FA supplements among women of child-bearing age during the early 1990s in order to reduce the risk for NTDs. The recommended doses varied between supplements of 400 ug FA daily for all low risk women and 4 mg FA/day for women with a previous NTD-affected pregnancy ⁶. In 1996 the Food and Drug Administration (FDA) mandated the fortification with FA of enriched grain products in the United States which started in 1998. Canada followed, and by the end of 1998, fortification of white wheat flour, as well as enriched cornmeal and pasta became mandatory ^{6,49}. The prevalence of NTDs fell by 46% in Canada and its prevention was ~25% in the United States following FA fortification ^{49,50}. Consistent with recommendations for NTD prevention, the IOM recommended that women of child-bearing age consume 400 ug FA/day (in addition to food folate), which should be acquired from fortified foods and/or supplements. However, data from a 2004 Canadian Community Health Survey found that, despite fortification, only ~18% of women of reproductive-age consumed the recommended 400 ug FA/day from diet and supplements ⁵¹.

1.3.2 FA and cancer

Canadian Cancer Statistics estimated that a total of 206,200 cancer diagnoses and 80,800 deaths from cancer occurred in 2017. The top four diagnosed cancers in the country were lung and bronchus, colorectal, breast, and prostate cancers, in that order, and they account for about half of all cancer diagnoses and deaths. Cancer is a disease that results from clonal expansion of an abnormal cell which escapes the cell's built-in mechanisms

and exogenous restraints of cell proliferation. Cancer is an example of Darwinian evolution in that the hallmarks of the disease are acquired through somatic mutations and selection⁵. Mutations which confer a selective advantage, leading to enhanced proliferation and survival of cells, can arise from replication errors or instances where DNA damage was either not repaired, or was repaired incorrectly. DNA damage can be caused by exogenous factors such as chemicals and UV light. Endogenous factors include mitotic errors and aldehydes, or by aberrant enzymes involved in genome editing and DNA repair⁵².

Many studies have shown the effect of folate deficiency on DNA instability, mutagenesis, and its involvement in the development of various epithelial cell malignancies. In a case-controlled study, patients with chronic ulcerative pancolitis with depressed RBC folate had a higher risk of dysplasia and cancer⁵³. A reduction in *de novo* nucleotide synthesis was observed in spleen from rats fed a folate deficient diet, resulting in a nucleotide pool imbalance which could promote base misincorporation mutations⁵⁴. DNA methylation in the liver of rats was found to progressively decrease in those on a folate deficient diet⁵⁵. Furthermore, FA deficiency has been observed to impair DNA repair in neurons from hippocampal cultures of transgenic mice, making them more sensitive to damage induced by amyloid β -peptide⁵⁶. Therefore, folate deficiency can affect all tissues and promote mutations by deregulating the one-carbon metabolism.

Numerous studies on colorectal cancer (CRC) and folate have shown an association between folate deficiency and an increase in CRC risk, where in some cases the risk is reduced by increasing the FA intake. Multiple published case-controlled studies have shown that individuals with the highest dietary folate intake have a 35% reduction in the risk of CRC development compared to those with the lowest intake⁵⁷. This is further

supported by a prospective cohort study in which after 8.5 years of post-fortification follow-up it was found that those with higher folate intake (≥ 900 ug/day) had a ~30% reduced risk of developing CRC than those that had a lower folate intake (≤ 200 ug/day) ⁵⁸. In *in vitro* studies, immortalized normal human colonocytes cultured in a FA deficient medium showed a decrease in DNA stability in addition to altered activity and expression of proteins involved in cell proliferation, apoptosis, and malignant transformation ^{59,60}. Overexpression of *SHMT1* has been associated with colonic tumors in human and rats ⁶¹ and reduced dTMP synthesis induced by *Shmt1* hemizyosity has been associated with increased risk of intestinal tumorigenesis in mice ⁶². Meta-analysis of multiple association studies determined an association of MTHFR 677C>T and 1298A>C polymorphisms with a decreased risk of CRC in different populations ^{63,64}.

Folate intake has also been associated with acute lymphocytic leukemia (ALL) risk and treatment outcome. Folic acid deficiency was associated with complications and adverse outcomes in children in India with ALL during maintenance chemotherapy ⁶⁵. *In vitro* studies have shown that human lymphocytes cultured in low FA concentrations have increased DNA instability, strand breakage, and uracil misincorporation ⁶⁶. The human SNP SHMT1 1420C>T has been associated with reduced risk of ALL ⁶⁷ and malignant lymphoma ⁶⁸. MTHFR 677C>T and 1298A>C SNPs were associated with lower risk of adult ALL, which suggests that folate inadequacy may be involved in the development of ALL ⁶⁹. SNPs in folate carrier proteins have also been observed. The RFC 80G>A SNP was found to impact the membrane transport of dietary folates and increased the risk of pediatric ALL ^{70,71}.

While the effects of folate deficiency have been extensively studied, the effect of FA supplementation, beyond its relation with NTDs, has not been given as much attention. Given the increase of FA intake in the Canadian population since mandatory fortification in the late 1990's, it is important to investigate whether high FA intake in the population may have potential beneficial or detrimental effects beyond NTDs. As previously mentioned, the UL of 1 mg/day for folate applies only to FA and not naturally occurring folate ³⁸. Among individuals that consume both fortified foods and FA containing supplements, there is a meaningful proportion who surpass the UL ⁵¹. While inconsistent, studies that have examined the association between high FA intake and adverse effects, including diseases such as cancer, suggest that further investigation is warranted.

Adequate intake and availability of folate enhances DNA stability, by supporting constant availability of *de novo* synthesized nucleotides ⁷². However, tissues with high rates of proliferation might need higher *de novo* nucleotide synthesis to maintain, repair, and maximize DNA synthesis. A large cohort study following fortification found an inverse correlation between folate intake and risk of colorectal cancer ⁵⁸. An analysis of dietary data collected post-FA fortification from 1998-2005 in the United States found an association between higher FA intake and decreased risk of pancreatic cancer in women only. This study also found that FA supplements were not associated with pancreatic cancer ⁷³. The long-term use of high-dose FA supplements is associated with a 75% reduction in colorectal cancer in women who started the supplementation as healthy women ⁷⁴. FA supplementation is also recommended when individuals take drugs that inhibit folate absorption in order to prevent complications that might lead to cancer. Sulfasalazine is a drug taken by patients with ulcerative colitis and it is also a competitive inhibitor of folate

absorption. In these patients, FA supplementation is suggested to prevent dysplasia or cancer ⁷⁵.

In vitro studies in U937 cells, a human leukemia cell line, have shown that FA supplementation alleviates oxidative stress, as well as apoptosis and mitochondria-associated death signaling which is induced by an oxidized cholesterol, 7-ketocholesterol ⁷⁶. *In vitro* experiments in the human liver cell line, L02, showed that FA supplementation reduced the Benzo(a)pyrene – induced mutagenicity and genotoxicity ⁷⁷. This protective effect of FA supplementation was also seen *in vivo*, where a significant decrease in micronucleus frequency was observed in the bone marrow of mice in the FA supplemented intervention when compared to control ⁷⁷.

Despite the research that shows the beneficial effects of FA supplementation, there is a public concern regarding its possible detrimental effects, especially given that some of the population's folate status suggests that they are surpassing the FA UL. It has been suggested that high-FA intake could potentially facilitate the proliferation of neoplastic cells by maintaining a high pace of *de novo* nucleotide synthesis in individuals who harbor an existing neoplasm ⁷². A double-blind randomized trial examining the effect of folic acid on colorectal adenoma risk in men and women with history of colorectal adenomas found that high-dose FA intake did not confer an overall protective effect, and instead was associated with increased risk of having 3 or more adenomas when compared to the control group that were taking a placebo; a finding that requires further investigation ⁷⁸. In mice, FA supplementation has been observed to suppress the development of ileal polyps in the small intestine and colon. However, if the polyps were already established, the effect of FA supplementation was the opposite ⁷⁹. This effect has also been observed for breast

cancer. A prospective cohort study of ~25,000 postmenopausal women found that the risk of developing breast cancer was significantly increased in the group taking FA supplements⁸⁰. Another prospective study found that high concentrations of plasma folate was associated with increased risk for breast cancer in women with the BRCA1/2 mutation⁸¹.

Given the variety of results from studies on folate and cancer risk of different tissues, and the since the mechanism of how folate affects tissues differently remains unknown, it is clear that further research is required.

1.4 Model for studying mutations *in vivo*

Mutations play a role in carcinogenesis, birth defects, and other diseases. Due to the complexities of *in vivo* cellular metabolism, the active mutagen may arise as a metabolic product in a target tissue which might not be the same as where it was activated. Therefore, an *in vivo* mutation assay is more suitable and effective method to detect, quantify, and sequence mutations in somatic and germ cells.

The MutaMouse model was used in this project to detect and measure the mutagenic potential of FA deficient, adequate and supplemented intake. The MutaMouse is a transgenic mouse model that allows the *in vivo* study of induced gene mutations in different organs and tissues⁸². These transgenic mice were produced by injecting the lambda bacteriophage vector (λ gt10lacZ) into the male pro-nucleus of fertilized eggs of CD2 (BALB/c x DBA/2) F₁ mice, which were later implanted into pseudo-pregnant BCBA mice. This resulted in a transgenic mouse model containing 29 ± 4 copies of the λ gt10lacZ in all of its cells. This non-endogenous reporter gene permits the detection of loss of

function mutations and allows the measurement of the DNA mutant frequency (MF) in any tissue following *in vivo* exposure to a mutation-inducing substance ^{82,83}.

A mutation analysis can be carried out by extracting genomic DNA, containing the vector, from the tissue of interest, packaging of the *lacZ* transgene into a bacteriophage, which is then used to infect an *Escherichia coli* (*E. coli*) host. The *E. coli* is plated in a selective and a non-selective medium followed by overnight incubation to allow plaque formation. The selective media contains phenyl- β -D-galactopyranoside (P-gal) that allows the positive selection of plaques formed by *E. coli* containing a loss-of-function mutated copy of the *lacZ* gene, which can be used to calculate mutant frequency (MF) ^{84,85}. Mutant plaques can be collected and mutations sequenced using next generation sequencing (NGS) which can provide insight into the type of mutations caused by particular substances and inform mode of action of the mutagen ⁸⁴. The MutaMouse has been routinely used in *in vivo* genotoxicity studies since 1989 to determine and analyze the risk of numerous environmental exposure-induced mutations on various tissues. In fact, in 2011 the OECD released the test guidelines and protocol for the use of the MutaMouse for *in vivo* mutation assay to detect gene mutations ⁸⁶, noting the assay's relevance for assessing *in vivo* mutagenic effects that are dependent on cellular events such as metabolism and DNA synthesis and repair.

1.5 Hypothesis

Dietary FA will affect somatic mutations in the male MutaMouse. Specifically, FA deficiency will cause a higher mutant frequency in colon epithelial cells in mice treated

with saline and the known mutagen, N-ethyl-N-nitrosurea (ENU). FA supplementation will have a protective effect on mutagen-induced mutations.

1.6 Objectives

- 1.** To investigate the effect of folic acid consumption on somatic DNA mutations in the colon of male mice.
- 2.** To determine the effect of folic acid deficient and supplemented diets on ENU-induced mutations in somatic cells (colon).
- 3.** To compare the effect of FA consumption on somatic cells, bone marrow and colon, in male mice.
- 4.** To identify folate- and tissue-specific mutation profiles of colon and bone marrow.

Chapter 2: General overview and rationale of methods

This chapter outlines the background of the methods of the study, including the animal model used and the animal study design. A rationale for the methods and the results obtained from a previous analysis in the lab, but that pertain to the study presented here, are also discussed.

2.1 Animal Model: The MutaMouse

The MutaMouse transgenic mouse model was used in this study. The MutaMouse genome contains multiple copies of the lambda bacteriophage vector (λ gt10*lacZ*) in all of its cells, both somatic and germ cells. There are approximately 29 ± 4 copies of the vector in tandem repeats at a single site on chromosome 3 of the transgenic mouse⁸³. The MutaMouse allows for testing the *in vivo* mutagenicity of extrinsic factors such as chemicals and environmental factors⁸⁶. Mutations are assessed by extracting high molecular DNA from the tissue of interest. The lambda vector is then packaged *in vitro* into phages and these are used to infect an appropriate strain of *E. coli* (*galE*⁻ *lacZ*). Mutant identification is achieved by an overnight incubation of *E. coli* on P-Gal medium. P-gal is toxic to *galE*⁻ strains that express a functional *lacZ* gene; therefore only phages that have a mutated *lacZ* gene will be able to form plaques in the P-gal selective medium⁸⁵. Non-selective titre plates serve as a control to determine the number of background plaque forming units (PFUs) (Fig. 2.1). The *lacZ* MF is determined by calculating the proportion of plaques containing a *lacZ* mutation in the phage population as follows:

$$\text{PFUs}/\mu\text{l} = \left(\frac{\text{\# of titer plaques}}{\text{volume of cells plated (15 }\mu\text{l)}} \right)$$

Total PFUs = (*PFUs/μl* × (*volume of cells + volume of packaged phage particles – volume of cells plated on titer plates*))

$$\mathbf{Mutant\ frequency\ (MF)} = \left(\frac{\mathit{Total\ mutant\ plaques}}{\mathit{Total\ PFUs}} \right)$$

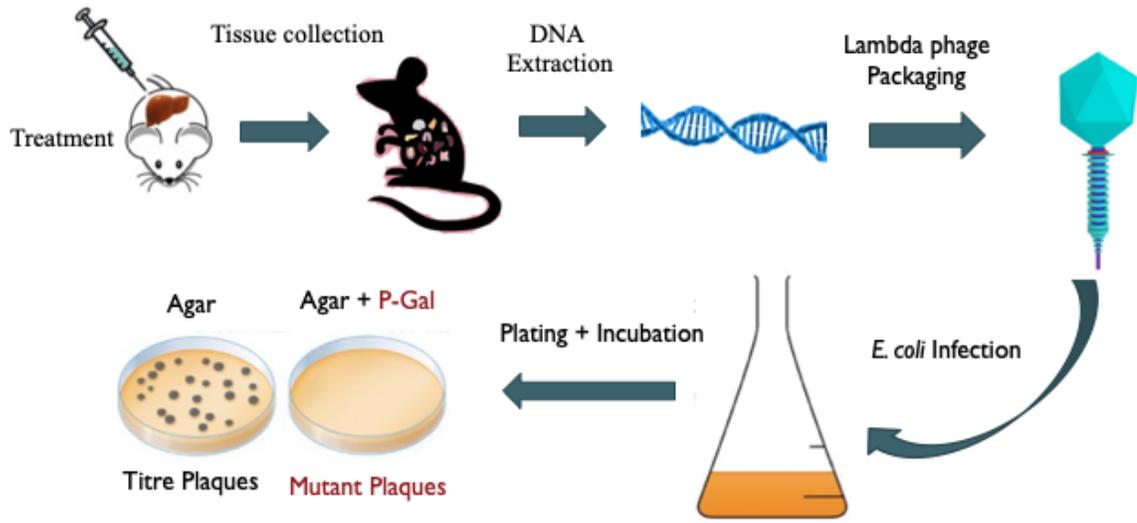


Figure 2.1: Overview of the lacZ mutation assay

2.2 Animal study design

Mice were cared for in accordance with the Guidelines of the Canadian Council on Animal Care (CACC), described in the CACC Guide to the Care and Use of Experimental Animals ⁸⁷. The study was approved by the Health Canada Ottawa Animal Care Committee. A total of 60 weanling male mice derived from the Health Canada in-house MutaMouse (BALB/c x DBA/2 CD2F1) colony were used. The mice were housed at standard humidity and temperature with a 12-hour light cycle. The mice had *ad libitum* access to food and water.

At 5 weeks of age, mice were fed one of three FA-defined diets based on the AIN-93G formula and kept on diet for 20 weeks with *ad libitum* access (Fig. 2.2) ^{88,89}. The diets contained either 0 mg FA/kg (deficient), 2 mg FA/kg (control) or 8 mg FA/kg (supplemented) (Dyets, Inc.; Bethlehem, PA) ⁸⁸.

These diets have been used routinely in our lab and represent physiologically relevant levels of FA. The deficient diet (0 mg FA/kg) is used to represent an inadequate dietary FA intake. Intake of this diet depletes tissue folate status and results in higher circulating homocysteine, however, the tissues do not become completely folate depleted given that folate can be synthesized by intestinal bacteria ^{90,91}. The control diet (2 mg FA/kg) is representative of an adequate dietary FA intake for rodents, as recommended by the American Institute of Nutrition and corresponds with the RDA for adult humans, which is 0.4 mg folate per day ⁸⁹. The supplemented diet (8 mg FA/kg) is 4-fold the control diet and represents a dietary FA intake of 1.6 mg per day in adult humans. This intake is over the FA UL, but represents an intake that can be achieved through the combined intake of folate

from natural dietary food sources, FA fortified foods, and over-the-counter vitamin supplements that are available without prescription in Canada.

Following 10 weeks on the FA defined diets, half of the mice from each diet group (n = 10 per diet) were given a 50 mg/kg dose of ENU by gavage and the other half (n = 10 per diet) were given saline by gavage. The mice were kept on diet for another 10 weeks (Fig. 2.2). In mice, highly proliferative tissues such as colon and circulating folate in plasma can become folate depleted within five weeks on a FA deficient diet. However, other tissues (eg. the liver) can take up to 10 weeks on a FA deficient diet to demonstrate evidence of depletion ⁶². Therefore, this study design ensures that folate status in the majority of tissues would represent the experimental diet at the time of ENU exposure.

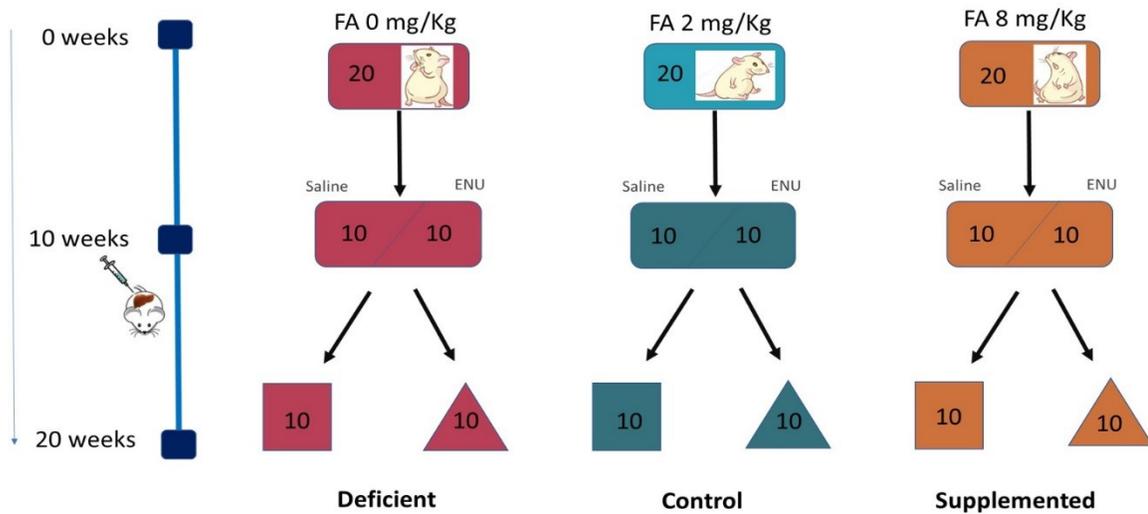


Figure 2.2: Animal study design. A total of 60 male mice were fed one of three FA diets (deficient, control and supplemented [n=20 each]) from 5 weeks of age for 20 weeks, *ad libitum*. After 10 weeks on diet, half of the mice from each diet (n=10 each diet) were given a 5uL/g dose of ENU (50 mg/Kg) or saline by gavage.

2.3 Necropsy and tissue collection

Each mouse was weighed weekly from weaning until the week prior to necropsy for a total of 19 weeks. After 20 weeks on diet, the mice were killed under isoflurane anesthesia by cardiac puncture followed by cervical dislocation. Among the tissues collected, were bone marrow, cauda epididymis, testes and the colonic epithelium, which will be referred to as “colon” hereafter. These tissues were chosen for analysis because they are associated with folate-dependent pathologies, including megaloblastic anemia, decreased cauda sperm counts, and colon cancer. Colon is a highly proliferative tissue and it is sensitive to FA deficiency given its high demand for *de novo* DNA synthesis.

At the time of necropsy, the colon was dissected, fecal pellets were squeezed out gently and then flushed with 5 mL of cold phosphate buffered saline (PBS). It was then cut open and laid flat with the lumen side facing up on a glass plate. The epithelial layer was lightly scraped away from the underlying colon tissue using two glass slides and collected into 1.5 mL Eppendorf tubes. All tissues collected were flash frozen in liquid nitrogen and stored at -80 °C until use. Blood was drawn via cardiac puncture into a heparin coated syringe (5:1 ratio of blood to heparin) and ~300 uL were transferred to K₂EDTA collection tubes. The remaining blood (~400uL) was separated into RBCs and plasma by centrifugation within an hour of collection, flash frozen in liquid nitrogen, and stored at -80 °C until later use.

2.4 Colon *lacZ* assay

The MutaMouse model provides the ability to detect and quantify mutations induced in a neutral transgene in all somatic tissues and germ cells. Mutagenic events that

occurred in any tissue can be scored by recovery of the vector and analysis of the phenotype of the reporter gene in *E.coli*. In this study, the *lacZ* assay was performed in the colon epithelium of the male MutaMouse.

The colon epithelium is highly proliferative and needs to maintain a dynamic equilibrium between proliferation, differentiation, and apoptosis. Given that both folate deficiency and supplementation have been associated with an increased risk of colon carcinogenesis depending on the model used, *in vivo* studies of FA-induced mutations are of high importance ⁹. The MutaMouse has been used previously to measure the spontaneous mutant frequency in many tissues, including the colon, when exposed to a variety of chemicals or environmental contaminants ^{85,88}. However, the mutagenic effects of FA have been studied only in the context of maternal folate exposure during pregnancy and its effect on the developing colon of F1 offspring from SWR females and MutaMouse males ⁹². Here, the direct FA-induced mutagenic effect on the colon of the male MutaMouse was quantified. Furthermore, collection of the mutant plaques from the *lacZ* assay allows for the sequencing of the mutant transgenes permitting identification of the FA-induced mutation spectrum ^{84,85}. Identification of the mutation spectrum can be indicative of the mechanisms through which FA induces mutations.

2.5 Preparation for next-generation sequencing

The mutant transgenes (plaques) collected from the colon *lacZ* assay, and those previously collected from the bone marrow of the same mice, were sequenced, allowing the comparison of FA-induced mutation spectra in two highly proliferative somatic tissues.

Mutant plaques from both tissues were pooled by sample (mouse). Sequencing was performed on samples for all three FA diets from the saline treated group only for each tissue. Plaques were also collected from titer plates, which acted as wild type/background MF plaque controls. Tubes containing the pooled plaques were boiled to melt the agar and an aliquot of 10 uL of the suspension was transferred to a well in a PCR plate. To control for PCR errors, each sample was amplified and sequenced in duplicate as a technical replicate (two separate reactions). The final 50 uL PCR mastermix per well contained the 10 uL aliquot from the plaques suspension, 10 uL of 5X Q5 Reaction Buffer, 0.5 uL of Q5 High-Fidelity DNA Polymerase, 1 uL of dNTPs (10 mM) (NEB Inc.), 2.5 uL of the forward primer (10 uM), 2.5 uL of the reverse primer (10 uM) and 23.5 uL of nuclease-free water. The PCR primers were designed using flanking DNA outside of the *lacZ* gene [GeneBank: J01636.1]. The forward primer was GGCTTTACACTTTATGCTTC and reverse primer was ACATAATGGATTTCCTTACG. The *lacZ* gene was amplified using the following thermocycle program: 95 °C for 3 min; 30 cycles of 95 °C for 45 s, 50 °C for 1 min, 72 °C for 4 min; final extension at 72 °C for 7 min. The PCR products were purified using the QIAquick PCR purification kit (Qiagen). DNA from purified PCR products was quantified using a Qubit™ dsDNA HS Assay kit (Invitrogen™) in a Qubit 2.0 Fluorometer. Samples were diluted with nuclease-free water to a final DNA concentration of 0.2 ng/uL and stored in LoBind tubes at -20°C until library preparation.

2.6 Illumina sequencing

Sequencing libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina). The Nextera XT protocol to create the sequencing libraries was followed except for the step where the libraries are normalized. Briefly, the Nextera transposome is used to

“tagment” the genomic DNA (gDNA) of each sample, this fragments the gDNA and tags the sequences. The “tagmented” DNA is amplified in the next step using a 12 cycle PCR program. During this amplification, the Index 1 (N7XX) and Index 2 (S5XX) and full adapter sequences are added to the “tagmented” DNA. The SPRIselect reagent beads (Beckman Coulter, Inc.) replaced the AMPure XP beads, used in the protocol, to purify the libraries and remove shorter fragments. Here, the normalization steps from the protocol were replaced by assessing the quality of the libraries using a High Sensitivity DNA Screen Tape in an Agilent 2200 TapeStation. The region was set to 150-1250 bp (base pair) and its molarity (pM/L) was calculated. Then, the libraries were diluted to a final molarity of 500 pM. Finally, 4 uL aliquots from the 500 pM solutions of each library were pooled together.

Next-generation sequencing of the *lacZ* mutant plaques was performed in-house using the Illumina NextSeq 500 System. We used the NextSeq 500/550 High Output Kit 2.5 (150 cycles) to sequence the libraries. Ultimately, the total samples sequenced were as follows: 28 samples from colon, 29 samples from bone marrow, each sample with its duplicate, and the titer plaques from the control diet of each tissue.

2.7 Bioinformatics

Bioinformatic analysis of sequence data was conducted using the pipeline described in Beal et al., 2015⁸⁴. Raw sequence data was converted into FASTQ format using the bcl2fastq conversion software v2.17.1.14 (Illumina, Inc). Reads trimmed and aligned to the reference *lacZ* sequence from the MutaMouse [GeneBank: J01636.1] using bowtie2⁹³ Settings: “- very-sensitive-local”). Alignment pileup for each sample was done using SAMtools⁹⁴. From the pileups, the frequency of each mutation type, including base

substitutions, deletions and insertions (indels), relative to the reference sequence was determined. Then, these were normalized by subtracting the false mutation frequency (estimated PCR artifacts and sequencing noise). To reduce false indel calls produced by homopolymer sequencing errors, indels with a false mutation frequency higher than the highest false mutation frequency for base substitution were ignored. The mutation-calling threshold was set based on the number of plaques sequenced for each sample (mutation-calling threshold = $1/\#$ plaques sequenced of sample). For a mutation to be counted and be considered a true mutation, it had to be present in both technical replicates at frequencies above the threshold. The mean true mutation frequency provided an estimate of total number of each type of mutation. Identical mutations that occurred more than once at a position within a sample were considered recurrent mutations and the result of clonal expansion from a single independent event.

To determine mutation spectrum, which is the distribution of the type of mutations in the *lacZ* gene for each FA diet, only independent mutations were used. Significant differences in mutation spectra within tissues was determined by one-way ANOVA followed by Tukey's Honest Significant Difference (Tukey HSD).

2.8 Previously reported results relevant to the study

The following sections summarize previously published results that are relevant to the current study, including body weight, plasma, and RBC folate at time of necropsy⁸⁸.

2.8.1 Body Weights

Neither the body weights measured throughout the study, or overall weight gain, differed significantly among mice in the different diet groups (Fig. 2.3)⁸⁸.

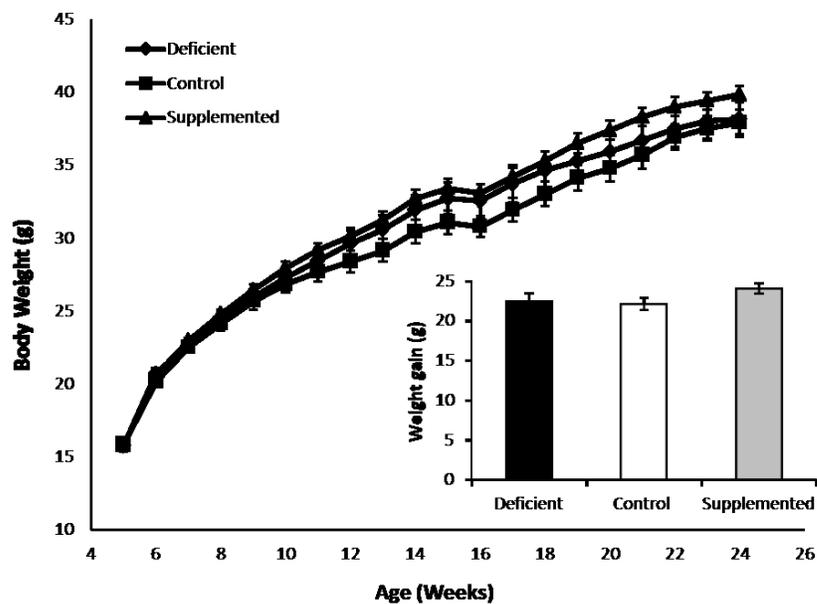


Figure 2.3: Body weights of male mice fed three FA diets. 0 mg FA/kg (deficient), 2 mg FA/kg (control) or 8 mg FA/kg (supplemented). *Inset*, total body weight gain from weaning. Data are presented as mean \pm SEM. Differences were assessed by one-way ANOVA, Holm–Sidak post- hoc analysis. Figure from LeBlanc et al., 2018.

2.8.2 Plasma and RBC folate

The plasma and RBC folate were measured in a subset of mice (n=5) from each of the six experimental groups. Folate was measured using the *Lactobacillus casei* (*L.casei*) microbiological assay as described by A. Molloy and J. Scott^{88,95}. Plasma folate serves as an indicator of recent folate intake, whereas RBC folate reflects the long-term folate status in tissues⁹⁶.

Mice fed the FA deficient diet had a 6.6-fold lower plasma folate compared to mice fed the FA control diet. In contrast, mice fed the FA supplemented diet had a 2-fold higher plasma folate when compared to those fed the FA control diet (Fig. 2.4 A). These results indicate that the folate status in the male mice reflected the dietary FA intake.

RBC folate concentrations were 8.7-fold lower in mice fed the FA deficient diet compared to those fed the FA control diet. Those fed the FA supplemented diet had RBC folate concentrations 1.3-fold higher compared to those fed the FA control diet. Therefore, folate status in the tissues of the male mice was reflective of their FA dietary intake (Fig. 2.4 B)⁸⁸.

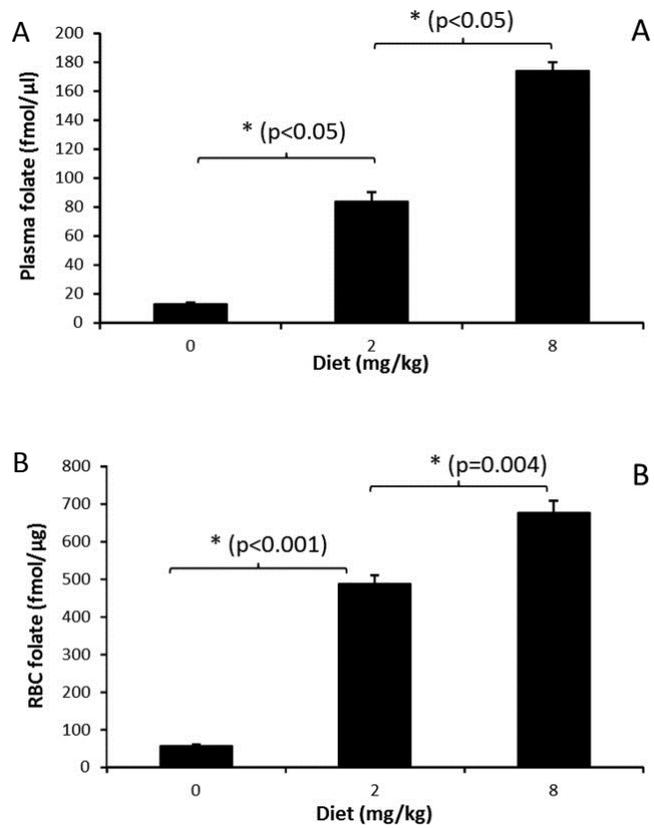


Figure 2.4: (A) Plasma Folate and (B) RBC folate. *, significantly different than control, one-way ANOVA, Tukey's post- hoc analysis ($P < 0.05$). Data are presented as mean \pm SEM. Figure from LeBlanc et al., 2018.

2.8.3 Bone marrow *lacZ* mutant frequency

DNA was extracted from the bone marrow of the left and right femur of the male mice using the phenol chloroform method. The *lacZ* assay was performed as originally described by Gingerich et al. [2014]^{88,97}.

Mice fed the FA deficient diet had a significantly higher MF, a 1.9-fold increase, compared to those fed the FA control diet ($p=0.035$, Fig. 2.5A). All ENU-treated samples regardless of FA diet had a significantly higher MF compared to those exposed to the saline treatment; however, FA diet had no effect on MF in the ENU-treated mice (Fig. 2.5B)⁸⁸.

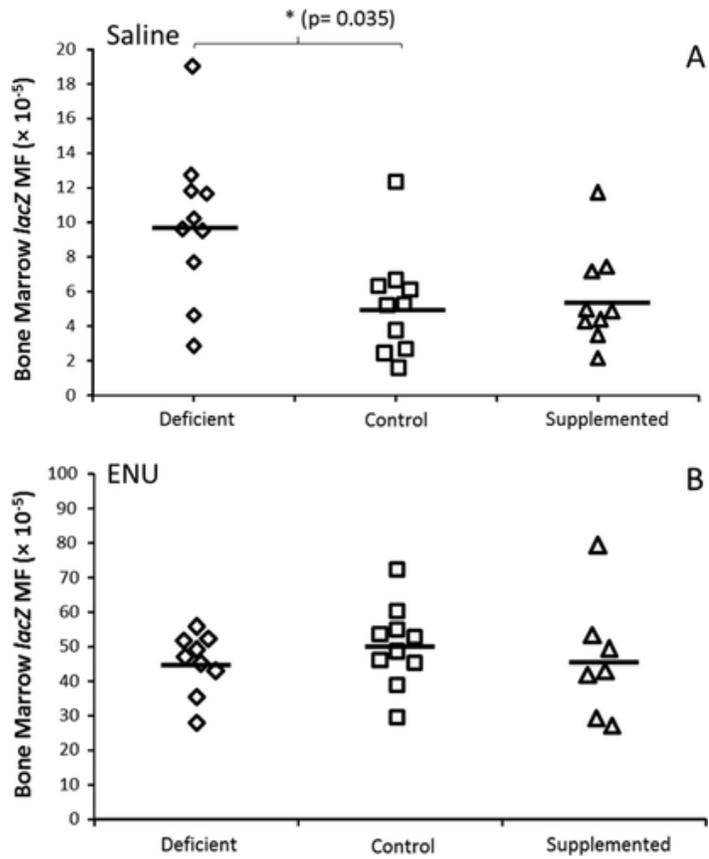


Figure 2.5: Bone marrow *lacZ* MF of male mice fed FA- defined diets and exposed to saline or ENU. (A) Male mice exposed to saline, (B) male mice exposed to ENU. *, statistically different based on two- factor binomial regression with a Bonferroni–Holm post- hoc analysis. Note the difference in y- axis between (A) and (B). Figure from LeBlanc et al., 2018.

Chapter 3: Investigating the folate-induced mutant frequency and mutation spectrum in the colonic epithelium of the male MutaMouse

3.1 Abstract

Colorectal cancer (CRC) is the third most common cancer globally and second most common in Canada. Folate has been suggested to have a dual effect on the risk of CRC. Some epidemiological studies have indicated an association between folate deficiency and an increased risk of developing colon cancer. Others have suggested an association between high intake of FA and an increased risk of developing the disease. Here, we used the MutaMouse to evaluate the effect of dietary FA on DNA mutagenesis *in vivo* in the colon. Male mice were fed experimental FA diets (deficient, control, or supplemented) for 20 weeks from weaning. In addition, to determine the effect of FA intake on the induction of mutations by a chemical mutagen, after 10 weeks on the diet half of the mice from each diet group were gavaged with 50 mg/kg ENU and the other half with saline after which they remained on the diet for an additional 10 weeks. After determining the FA-induced MF, we used Next Generation Sequencing to identify the FA-induced mutation spectra in the colon. We observed that mice in the saline treatment fed the FA supplemented diet had a higher MF compared to mice fed the FA control diet ($p = 2.3 \times 10^{-5}$) and FA deficient diet ($p = 2.9 \times 10^{-5}$). Moreover, FA supplementation did not have a protective effect on the colon in ENU exposed mice. Instead, the FA supplemented diet caused a higher MF compared to mice fed the FA deficient diet ($p = 0.045$). By identifying the FA-induced MF, we show for the first time the capacity for supplemental FA intakes to induce mutations. In addition, by identifying the FA-induced mutation spectra we show that the type of mutations observed are characteristic of CRC but not necessarily unique to the

supplemented diet. These findings demonstrate the capacity of FA supplementation to induce mutations in the colon of this mouse model and support the association of FA-intake and CRC.

3.2 Introduction

The number of Canadians diagnosed with cancer every year is increasing⁹⁸. Almost 1 in 2 Canadians are expected to be diagnosed with cancer during their lifetime⁹⁸. Colorectal cancer (CRC) is the third most common cancer globally⁹⁹. In Canada, colon cancer is the second most common type of cancer, just below lung cancer, accounting for 13% (~26,800) of all new cancer cases in 2017. Similarly, in the United States, colon cancer is the third most common cancer, estimated to account for 9% and 7% of new cancer cases (~101,420), and 9% and 8% of estimated deaths (~51,020) in males and females, respectively¹⁰⁰. Survival from colon cancer is strongly associated with stage at diagnosis, where those diagnosed at stage I have a ~92% chance of survival compared to just 12% for those diagnosed at stage IV. Currently, in Canada, all 10 provinces and Yukon have implemented colorectal cancer screening programs in the hopes that detection of this cancer can be made as early as possible. However, diagnoses at all four stages are roughly equally distributed, with half the diagnoses being at stage III and IV despite the screening programs for both men and women⁹⁸. Therefore, it is important to investigate possible risk factors of the disease to improve preventive and diagnostic efforts.

Adenocarcinoma that arises from colon and rectal epithelium accounts for 90% of CRC cases¹⁰¹. The tumor can then breach the intestinal wall and spread via the lymphatic system to lymph nodes and other organs through the blood stream. The epithelial layer of the colon consists of a single sheet of columnar epithelial cells folded to form finger-like

invaginations called crypts of Lieberkühn¹⁰². The differentiated cells are found in the top-third of the crypt and they are derived from multipotent stem cells, the crypt base columnar cells. These cells undergo self-renewal and give rise to a population of transit-amplifying cells that become one of the epithelial lineages of the colon^{102,103}. Cellular transformation in the colon has been observed to start in the crypts in a bottom-up fashion and also in a top-down fashion with carcinogenesis starting at the surface of the epithelium^{104,105}. Cancer stem cells, in which the capacity of self-renewal is deregulated; have been identified in colon cancer and are hypothesized to represent the cells from which cancer originates^{106,107}.

Adult stem cells of organs with different incidences of cancer accumulate mutations through life at similar rates. It was estimated that the colon accumulates ~36 mutations per year in its adult stem cells with a resulting correlation between the number of somatic point mutations and age¹⁰⁸. While the mutation profiles arise from intrinsic mutational processes, differences in cancer incidence among various tissues may be attributed to tissue-specific clonality of stem cells or the interaction and impact of external agents¹⁰⁸. Differences in cancer risks among tissues can be explained by the total number of stem cell divisions; for the colon, environmental and inherited factors play an important role in the risk of CRC¹⁰⁹.

Folate is a water-soluble B vitamin found in foods such as fresh fruits and vegetables, legumes and organ meats. FA is the synthetic oxidized form of folate found in fortified foods such as enriched white wheat flour and other enriched cereals. Folate is required for the activation and transfer of one-carbon-units needed for the *de novo* synthesis of purines and thymidine (dTMP), and the synthesis of methionine¹. *De novo* synthesis of

purines and dTMP is essential for cell proliferation and genome stability. Methionine can be adenylated to form AdoMet, the universal methyl donor in the cell, which is required for methylation of various molecules in the cell including DNA and histones. In turn, DNA and histone methylation patterns regulate gene expression, cellular differentiation, and organogenesis, among other processes.

Disruption to any of the three folate-dependent biosynthetic pathways can have serious implications in a cell's survival, ultimately affecting organs and tissues as a whole. Impairments in purine synthesis can result in a reduction of DNA replication and abnormal DNA repair^{13,14}. Insufficient dTMP synthesis may cause a nucleotide imbalance which leads to the accumulation of uracil and its mis-incorporation into DNA, resulting in increased DNA double strand breaks due to the "futile cycle" of DNA repair^{15,16}. Reduction in the production of AdoMet reduces the availability of methyl groups and the cellular methylation capacity, and can thus alter methylation patterns in the genome. Genome wide hypomethylation can result in genome instability, disrupt gene expression, and alter risk for diseases such as cancer^{20,110}.

Folate is an extrinsic factor that has a dual effect on the risk of CRC. The protective or promoting effect in CRC depends on the folate status and the normal, not malignant, status of the tissue. A number of human studies suggest an association between high folate intake and/or status and a reduced risk of developing colon cancer. A meta-analysis of prospective studies showed an association between high dietary folate intake and a reduction of CRC risk¹¹¹. Long-term FA consumption as part of a daily multivitamin (Vit A, C, D, E) intake was associated with a significantly reduced risk of colon cancer in a large prospective cohort study of 88,756 women⁷⁴. Animal studies have shown that folate

deficiency increases the risk of CRC by increasing neoplastic lesions in rats with an underlying predisposition ¹¹² and that high dietary folate intake reduced the evolution of macroscopic neoplasms in the colorectum of rats ¹¹³. The effects of folate deficiency have also been studied *in vitro*. Low folate concentrations induce global and region-specific (e.g., p53 tumor suppressor gene) hypomethylation in cultured human-derived colon carcinoma cells ¹¹⁴. Immortalized normal human colonocytes in a folate deficient medium exhibited an increase in DNA strand breakage and uracil misincorporation ⁵⁹.

Genetic polymorphisms in genes related to folate metabolism have also been associated with CRC risk. Methylenetetrahydrofolate reductase (MTHFR) is a key enzyme needed to irreversibly convert 5,10-methyleneTHF to 5-methylTHF, which is required to remethylate homocysteine to methionine, ensuring the production of AdoMet. AdoMet is crucial for methylation reactions and epigenetic regulation of gene expression. Common polymorphisms in this gene alter folate-mediated one-carbon metabolism by decreasing the activity of the enzyme and they have been shown to modulate the risk of CRC ^{63,110}. A meta-analysis of multiple case-control studies found an association between the *MTHFR* 677C>T and 1298A>C single nucleotide polymorphisms (SNPs) and a decreased risk of CRC in different populations ^{63,64}. A case-control study that lasted 12 years, found the *MTHFR* 677C>T to have a protective effect towards colon cancer risk; however, this is an effect that is negated by a low folate intake ¹¹⁵. Recently, a genome-wide association study determined the *MTHFR* 677C>T variant to be the major genetic modifier of folate biomarkers ¹¹⁶. The *MTHFR* 677C>T SNP is associated with lower serum and red blood cell folate and higher plasma homocysteine, all indicative of impaired folate status and metabolism in tissues ¹¹⁶. *In vitro*, human colonocytes homozygous for the *MTHFR*

677C>T SNP demonstrated global DNA hypomethylation when cultured in a medium with low FA and hypermethylated when cultured in a medium with an adequate or high folate supply¹¹⁷. The *MTR* 2756A>G polymorphism of the methionine synthase gene, also important in the homocysteine-methionine pathway, decreases the enzyme activity and has also been associated with a decreased risk of CRC^{31,118}. In mice, reduced dTMP synthesis induced by *Shmt1* hemizyosity was associated with increased risk of intestinal tumorigenesis⁶². Together the data suggest that folate deficiency or aberrant folate metabolism, specifically that which affects nucleotide synthesis rather than methylation capacity, may increase the risk for developing CRC.

In contrast to these studies that suggest that a higher folate status has a beneficial effect on CRC risk, there are a number of studies that suggest FA fortification may increase it. The introduction of FA fortification in North America and Chile was correlated with a transient increase in colon cancer risk immediately following implementation^{119,120}. A double-blind randomized trial in participants with recent history of colorectal adenomas found that long-term (>3 years) exposure to FA supplementation did not increase overall risk for CRC recurrence, but it did increase the risk of developing more advanced and higher number of recurrent adenomas⁷⁸. It has been proposed that the timing of FA intervention may be key to these observations such that FA may promote the growth of existing lesions.

Timing of FA supplementation was studied in a genetic murine model predisposed to intestinal tumorigenesis that can develop numerous dysplastic colonic aberrant crypt foci, an early precursor of CRC. This model, *Apc*^{+/-} *Msh*^{-/-}, is genotypically similar to human CRC and allows the study of environmental and genetic factors in familial and

sporadic carcinogenesis. FA supplementation started before the establishment of neoplastic foci provided a protective effect against the development of small intestinal adenomas and colonic aberrant crypt foci ¹²¹. However, if neoplastic foci were present, FA supplementation had the opposite effect and increased the number of neoplastic foci. These contrasting results suggest that the protective or promoting effect of FA may depend on the folate status of the tissue during specific times in cellular transformation ¹²². A protective or rescuing effect may be seen in normal non-cancerous tissues where FA supplementation is used to counteract folate deficiency whereas in tissues with established neoplasms FA supplementation may promote the progression of the malignant lesion.

Given the association of folate status with CRC and the potential of folate to induce mutations by disrupting one-carbon metabolism it is of interest to measure the potential mutagenicity of FA in the colon. We used the MutaMouse to measure and characterize the folate-induced mutations in the colon. The MutaMouse is a transgenic mouse model used to detect and measure the mutation-inducing potential of many substances. The MutaMouse carries multiple copies of the *lacZ* transgene contained in the lambda bacteriophage vector (λ gt10*lacZ*) incorporated into the genome of all of its cells, which allows for the identification of mutations in this non-endogenous gene ⁸³. Since the DNA can be isolated from any tissue, positive selection of the mutated *lacZ* gene allows for the calculation of mutant frequency (MF) in response to an environmental intervention in all tissues ^{84,85}. Here, we used the MutaMouse to evaluate the effect of dietary FA on the MF in the colonic epithelium to determine its mutagenic potential and identify potential mechanisms of action that may link folate metabolism with CRC risk.

3.3 Methods

3.3.1 Mouse study and colon collection

For this experiment, all tissues were in-hand. They were collected in a previous study and stored for future research at -80°C. The animal study was performed in-house at Health Canada and followed the guidelines described in the Guide to the Care and Use of Experimental Animals¹²³. The study was approved by the Health Canada Ottawa Animal Care Committee. The study included 60 weanling male mice (~21 days) from the in-house MutaMouse colony that were housed with *ad libitum* access to water and diets and maintained at standard humidity. At 5 weeks of age, the male mice were fed one of the three FA diets and kept on diet for 20 weeks: 0 mg FA/kg (deficient), 2 mg FA/kg (control) or 8 mg FA/kg (supplemented). The deficient diet represents inadequate dietary FA, the control diet approximates the RDA for adult humans, which is 0.4 mg per day, and the supplemented FA is 4-fold the control FA diet, representing 1.6 mg per day, which would be above the FA UL.

After 10 weeks on the FA defined diets, half of the mice in each diet group (n=10 per diet) were given 50 mg/kg of ENU treated by gavage and the other half (n=10 per diet) were given saline by gavage. The mice were kept on diet for another 10 weeks before necropsy and tissue collection.

At the time of necropsy, the colon was dissected, fecal pellets were lightly squeezed out using very little pressure, not to disturb the epithelium. The unobstructed colon was then flushed slowly with 5 mL of cold PBS. The colon was then cut open lengthwise and laid out on a glass plate. Using a glass microscope slide to hold down one end of the colon, another glass slide was used to lightly scrape the epithelium off and it was collected into a

1.5 mL Eppendorf tube. The contents were then flash frozen in liquid nitrogen and kept at -80 °C until later use.

3.3.2 Mouse colon DNA extraction

The DNA extraction from frozen colon epithelium was performed using the phenol/chloroform extraction procedure as described⁹⁷. The steps for ‘Small intestine and colon’ extraction were followed, skipping the cracking step. The colon was thawed and digested in 5 mL of lysis buffer and incubated at 37°C overnight, with gentle shaking. Following digestion, RNase A was added to a final concentration of 100 ug/mL and incubated for 1 h at 37°C. 5 M NaCl was added to a final concentration of ~1.5 M. The sample was then mixed and centrifuged at 2,000 x g for 20 min. Only the aqueous sample was transferred to a new 15 mL tube. Hydrated phenol was first equilibrated with equal part of 1 M Tris-HCl, pH 8, and shaken, and the upper aqueous phase was discarded; this was repeated with 0.1 M Tris-HCl, pH 8. An equal volume of chloroform:isoamyl alcohol (24:1) was added to the phenol, mixed and centrifuged for 10 min at 1500 x g. An equal volume of the bottom organic layer was added to each sample, and samples were inverted for 20-30 min and then centrifuged at 1500 x g for 10 min. The upper aqueous layer was transferred to a new 15 mL tube leaving behind any white precipitate. Enough 5 M NaCl was added to each sample to bring the concentration to 200 mM (1/50 volume) and mixed. An equal volume of chloroform/isoamyl alcohol was added to the samples, the samples were inverted and centrifuged as previously described. The aqueous fraction was transferred to a new 15 mL tube followed by the addition of ethanol at a 2:1 ratio to start DNA precipitation. DNA was spooled and washed with 70% ethanol. Once dried, the

sample was dissolved in TE buffer (10 mM Tris pH 7.6, 1 mM EDTA). Isolated DNA was kept at 4°C until it was used for the lac Z assay.

3.3.3 Colon *lacZ* mutant frequency analysis

The *lacZ* assay was performed using the DNA extracted from the colon. The recovery and packaging of the lambda vector into a lambda phage from the extracted DNA was accomplished by using the Transpack Packaging extract (Agilent Technologies, Santa Clara, CA). These particles then infect the recombinase A deficient (*recA*-) and galactose epimerase deficient (*galE*-) *E. coli*. The *E. coli* was then incubated overnight permitting the growth of a bacterial lawn allowing the visibility and quantification of plaques. A minimum of four non-selective control plates (allowing growth of non-mutant and mutant plaques) and four selective mutant plates were counted for each sample. An average of more than 110,000 total PFUs were scored for each treatment. The mutant frequency (MF) was calculated per sample as the number of mutant plaques divided by the total number of PFU.

Once the mutant plaques were counted, they were collected using a transfer pipet into sterile microtubes containing autoclaved milliQ sterile water (0.3 mutants/uL; 1 sample per tube). The mutant plaques were pooled by sample (mouse) and stored at -80°C until sequencing.

3.3.4 Statistical analysis

The colon *lacZ* MF data were fit to a generalized linear model with a Poisson distribution using the *glm* function in R with the quasi-Poisson error distribution to account for over-dispersion (variability) of the data. One of the colon DNA samples did not produce any PFUs; therefore, it was removed from the analysis. Replicate plates that were outside

the error distribution of the diet group were considered outliers and removed from the analysis. Based on these removals, four of the original 60 samples were removed from the analysis. Ultimately, there was a total of 10 samples for both deficient and control FA diets for both treatments and eight samples for the supplemented diet for both treatments (Table 3.1). The mean PFUs counted were 146,189 and 138,633 in saline and ENU treatments, respectively. Significant differences among the FA diets and treatments were identified by a two-way ANOVA followed by Tukey's Honest Significant Difference (Tukey HSD).

3.3.5 Library preparation and next-generation sequencing

Mutant plaques were collected and pooled by sample for sequencing. Sequencing was performed on samples for all three FA diets from the saline treated group only. Plaques from titer plates were also collected to act as wild type/background mutation controls. Microtubes containing the pooled plaques for each mouse were boiled to melt the agar and an aliquot of 10 uL of the suspension was transferred to a well in a PCR plate. Amplification and sequencing of the *lacZ* gene for each sample was done in two technical replicates to control for PCR errors. A PCR mastermix was prepared for each sample and its replicate resulting in a final volume of 50 uL. The PCR mastermix contained a 10 uL aliquot from the plaque suspension, 10 uL of 5X Q5 Reaction Buffer, 0.5 uL of Q5 High-Fidelity DNA Polymerase, 1 uL of dNTPs (10 mM) (NEB Inc.), 2.5 uL each of the forward and reverse primers (10 uM), and 23.5 uL of nuclease-free water. A 30-cycle PCR amplification of the *lacZ* gene was performed using the following thermocycle program: 95 °C for 3 min; 30 cycles of 95 °C for 45 s, 50 °C for 1 min, 72 °C for 4 min; final extension at 72 °C for 7 min. The PCR products were purified using the QIAquick PCR purification kit (Qiagen). The QubitTM dsDNA HS Assay kit (InvitrogenTM) and the Qubit

2.0 Fluorometer were used to quantify the purified PCR product. Samples were diluted with nuclease-free water to a final DNA concentration of 0.2 ng/uL and stored in LoBind tubes at -20°C until library preparation.

Sequencing libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina) following the Nextera XT protocol except for the last step where the libraries are normalized. Briefly, the genomic DNA (gDNA) of each sample is “tagmented” and tagged by the Nextera transposome. Index 1 (N7XX), Index 2 (S5XX) and full adapter sequences are added to the “tagmented” DNA and it is then amplified in a 12 cycle PCR. We used the SPRIselect reagent beads (Beckman Coulter, Inc.) to purify the libraries and remove shorter fragments instead of the AMPure XP beads, used in the protocol. We replaced the normalization steps from the protocol by assessing the quality of the libraries using a High Sensitivity DNA Screen Tape in an Agilent 2200 TapeStation. The region was set to 150-1250 bp (base pair) and its molarity (pM/L) was calculated. The libraries were then diluted to a final molarity of 500pM. Libraries were pooled together containing 4 uL aliquots from the 500 pM solutions of each library.

The Illumina NextSeq 500 System and the NextSeq 500/550 High Output Kit 2.5 (150 cycles) were used to sequence the *lacZ* libraries. Ultimately, plaques from a total of 28 samples were sequenced, each sample with its duplicate, and a pool of 105 titer plaques from the control diet (Table 3.2).

3.3.6 Bioinformatics

Bioinformatic analysis of sequence data was conducted using the pipeline described in Beal et al., 2015⁸⁴. Briefly, raw sequence data was converted into FASTQ format using the bcl2fastq conversion software v2.17.1.14 (Illumina, Inc). Reads were trimmed and

aligned to the reference *lacZ* sequence from the MutaMouse using bowtie2⁹³. Pileups were created using SAMtools⁹⁴. From the pileups, a frequency of each mutation type, which included base substitutions and indels, relative to the reference sequence (*lacZ* gene) were determined and adjusted for the false mutation proportion (PCR artifacts and sequencing noise). Indels with a false mutation proportion higher than the highest false mutation proportion for base substitutions were ignored to reduce false indel calls. “True mutations” were called when both technical replicate values were above the mutation calling threshold (1/# plaques of sample). Recurrent mutations, specifically the same mutation that occurred at the same location more than once per animal, were considered to be the result of clonal expansion and were thus counted as one mutation. Independent, unique mutations were used to determine mutation spectra (type of mutations) and distributions of mutations across the *lacZ* gene. One-way ANOVA followed by Tukey’s Honest Significant Difference (Tukey HSD) were performed to identify any significant differences on mutation types among the FA diets.

3.4 Results

3.4.1 Colon mutant frequency

Table 3.1 presents the mean colon MF and standard deviation (SD) by diet and treatment of the 56 samples included in the analysis after removal of outliers from replicates (described in methods). There was a total of 10 samples for deficient and control diets for both treatment groups and eight samples for the supplemented diet for both treatment groups.

Table 3.1: LacZ mutant frequency in colon. MF: mutant frequency. SD: standard deviation.

Treatment	FA Diet	# of Samples	Mean MF x 10⁻⁵ ± SD
Saline	Deficient	10	11.1 ± 1.5
	Control	10	11 ± 1.5
	Supplemented	8	15.5 ± 2.4
ENU	Deficient	10	59.7 ± 8.4
	Control	10	61.3 ± 4.1
	Supplemented	8	70.1 ± 11.7

Among saline treated mice, those fed the FA supplemented diet had a significantly 1.4-fold higher MF compared to mice fed the FA control diet ($p = 2.3 \times 10^{-5}$) and those fed the FA deficient diet ($p = 2.9 \times 10^{-5}$) (Fig 3.1 A). In ENU treated mice, mice fed the FA supplemented diet also had a significantly 1.2-fold higher MF than mice fed the FA deficient diet ($p = 0.045$) (Fig. 3.1 B).

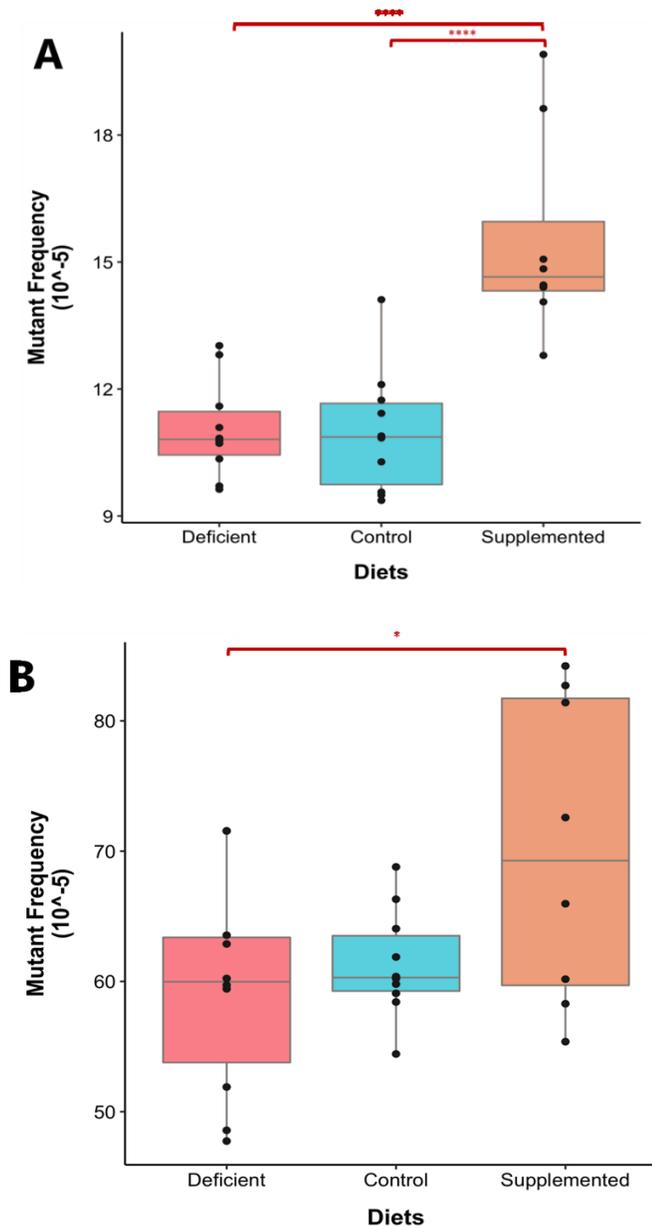


Figure 3.1: LacZ mutant frequency in the colon of male mice fed FA defined diets. (A) Male mice exposed to saline, and (B) male mice exposed to ENU. The points represent the mean mutant frequency for each mouse. The horizontal line represents the group median, the box represents the upper and lower quartiles and the whiskers extend across the full range of observed mutant frequencies. Mice fed the supplemented FA diet and treated with either saline or ENU had a significantly higher mutant frequency in the colon compared to mice fed a FA deficient or control diet. Note the Y-axis scale of saline and ENU groups. (Two-way ANOVA, Tukey HSD, * $p \leq 0.05$, *** $p \leq 0.005$, ***** $p \leq 0.00005$). Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg. Treatment: Saline or N-ethyl-N-nitrosurea (ENU). Groups included $n=8-10$ /diet/treatment.

3.4.2 Sequencing of the colon mutant plaques

Mutations were measured and characterized in the colon by sequencing the lacZ reporter transgene of all samples in the saline treatment, using the Illumina NextSeq 500 System. A total of 28 samples (452 plaques) were sequenced. Each sample was sequenced with a technical replicate (Table 3.2). Sequencing generated a total of 18,493,058,083 base reads (nucleotide reads).

After correcting for false mutation proportion and applying the mutation calling threshold on the colon sequencing data, “true mutations” were called in nine mice fed the FA deficient and FA control diets, and eight mice fed the FA supplemented diet. There was a total of 360 recurrent mutations from which 176 independent mutations (including indels) were identified in all samples. Only independent mutations were used in the analysis.

Table 3.2: Total of samples sequenced, and number of independent mutations identified in the lacZ gene from the colon. The total number of mice (samples) per FA diet that were sequenced, and the number of base substitutions, deletions and insertions identified per FA diet.

FA Diet	# of Samples	Total Samples	Total number of Libraries (with duplicates)	# of Samples used in Analysis ¹	Independent mutations		
					Base substitutions	Deletions	Insertions
Deficient	10			9	44	7	3
Control	10	28	56	9	39	12	3
Supplemented	8			8	52	5	-
				Total	135	24	6

¹ Number of samples used in the analysis may differ from the original number of samples sequenced due to the fact that a sample may have been dropped if none of its mutation calls remained after correcting for false mutation proportion and applying the mutation calling threshold.

3.4.3 Proportion of mutation types in the colon

Proportions of the colon mutation spectra among the three FA diets can be seen in figure 3.2. Some qualitative differences were observed in the mutation type proportions among the FA diets, but they were not statistically significant. A:T → G:C transitions were present in mice fed the FA control and supplemented diet but not in those fed the FA deficient diet. Insertions were only observed in mice fed the FA deficient and control diets. The greatest proportion of A:T → C:G transversions was observed in mice fed the FA supplemented diet. All four types of possible transversions (G:C → C:G, G:C → T:A, A:T → T:A, A:T → C:G) and deletions were seen in mice fed all three diets.

Overall, the proportions of transitions, transversion and indels, respectively, in mice fed the FA diets were: 0.48, 0.34 and 0.18 in those fed the FA deficient diet; 0.50, 0.20 and 0.30 in those fed the FA control diet and 0.55, 0.35 and 0.10 in those fed the FA supplemented diet, respectively. Therefore, while proportions of transitions were similar in mice across all FA diets, the mice fed the FA control diet had the lowest proportion of transversions and the mice fed the supplemented diet had the lowest proportion of indels.

The *lacZ* mutation spectrum of the colon was adjusted to the MF obtained from the *lacZ* assay (Fig. 3.3).

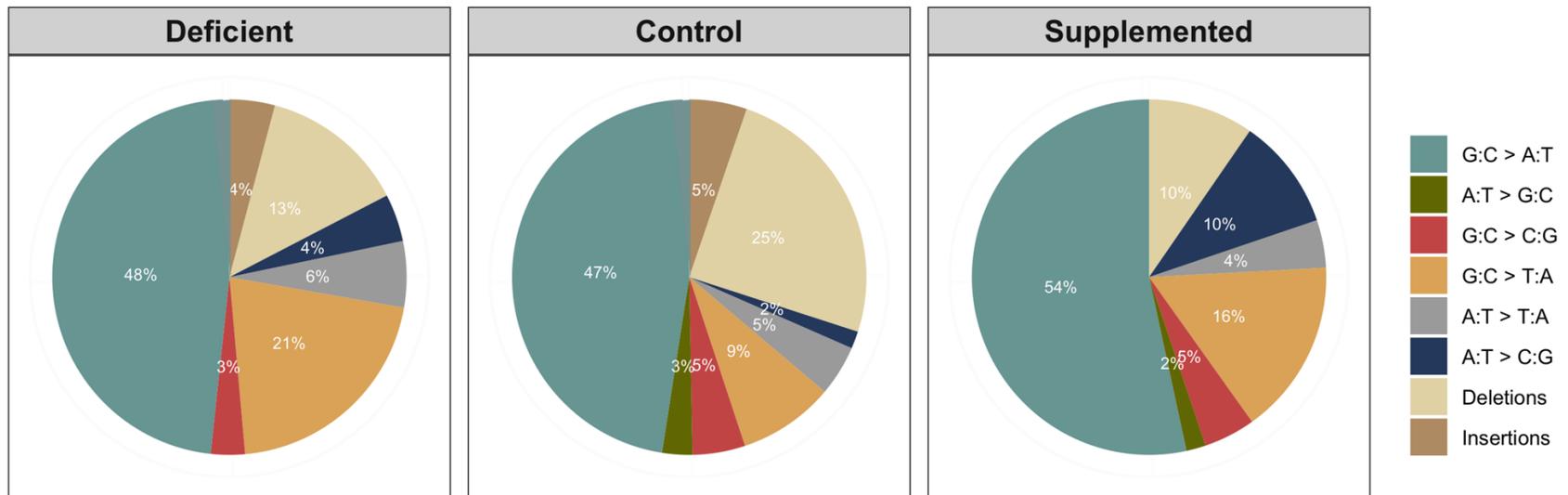


Figure 3.2: Proportional representation of mutation spectra of the colon of male mice on the saline treatment fed one of three FA diets. No significant differences were identified by one-way ANOVA, Tukey HSD. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

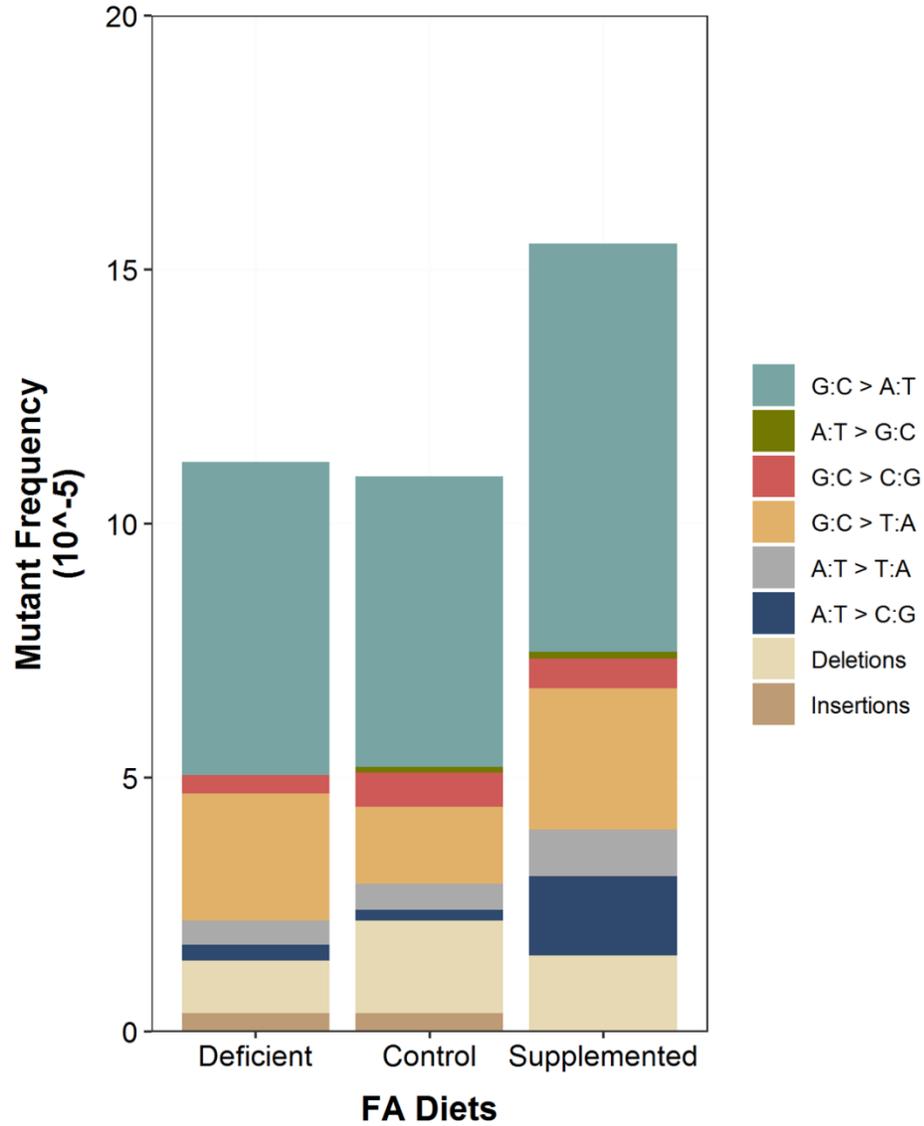


Figure 3.3: LacZ mutation spectra of the colon in mice fed one of three FA diets. Proportions of all type of mutations adjusted to MF of colon. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

3.4.4 Distribution and hotspots of base substitutions and indels along the *lacZ* gene in the colon

Plotting the distribution of the independent FA-induced mutations in colon along the *lacZ* gene allowed the identification of candidate mutational hotspots. A position was considered a possible hotspot if a mutation occurred at the same position in more than one sample per diet.

A total of six, three and nine possible hotspots were identified in mice fed the FA deficient, control and supplemented diets, respectively (Table 3.2). All of the hotspots in mice fed the FA deficient and control diets occurred at C:G base pairs. In mice fed the FA supplemented diet one (~11%) of the hotspots occurred at a T:A base pair and the rest occurred at a G:C base pair. The majority of the hot spot mutations were G:C→A:T transitions among all diets. Five hotspots were common to more than one diet, however there were samples with no hotspots observed in all diet groups. All five shared hotspots occurred at C:G base pairs, with four being G:C→A:T transitions and one a G:C→C:G transversion. A visual representation of the distribution of the base substitutions along the *lacZ* gene can be observed in figure 3.4.

Table 3.2: Position of possible base substitution hotspots. Independent base substitutions which occurred at the same position in more than one sample per diet in the colon. The numbers denote the number of mice with the mutation per diet. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg. * Hotspot present in more than one diet

Position	FA Diet			Type of Mutation
	Deficient	Control	Supplemented	
50			2	T > A
204 *		2	2	G > C
436		2		C > T
928	3			G > A
1030	2			G > T
1187 *	2		2	C > T
1196			2	C > T
1399	2			G > A
2374			2	C > T
2659 *	3		2	C > T
2713 *		4	2	G > A
2743			2	C > A / C > T
2744 *	2		2	G > A

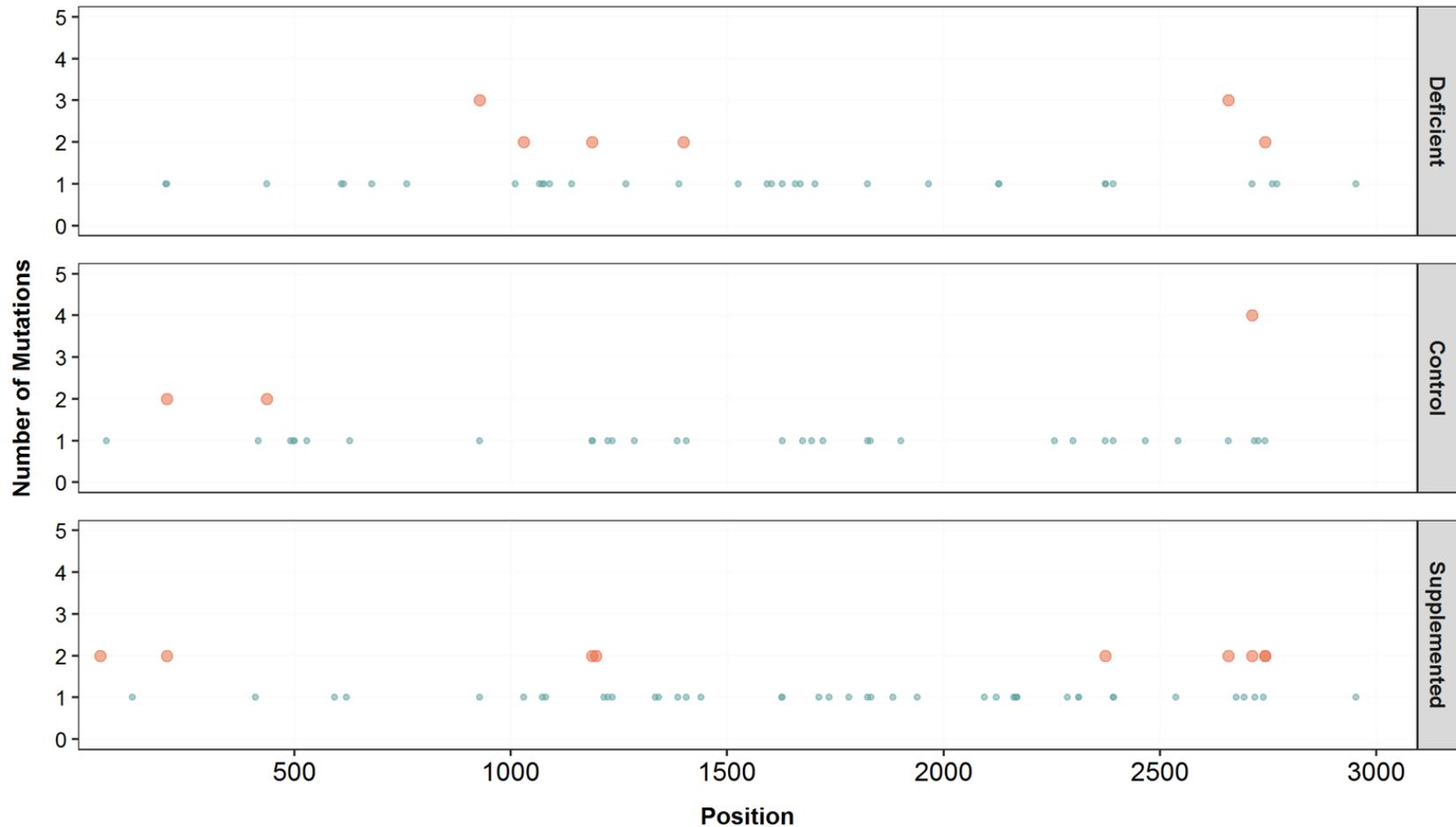


Figure 3.4: Distribution of base substitutions across the lacZ gene in the colon in mice fed one of three folic acid diets. Dots represent a base-substitution. Red dots represent hotspots, mutation that occurred at the same position in more than one sample on the same diet. Number of mutations show how many samples share the base substitution at the same position. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

Candidate indel hotspots in the colon are shown in table 3.3. Only deletion hotspots were identified, and only in the group fed the FA control diet.

A total three deletions hotspots were identified in mice fed the FA control diet. None of the deletion hotspots occurred at the same position as a base substitution hot spot. Deletion hotspot at position 1578 occurred in three mice.

A visual representation of the distribution of indels across the *lacZ* gene can be observed in figure 3.5.

Table 3.3: Indel hotspots along the lacZ gene in the colon. Independent indels that occurred at the same position in more than one mouse per diet in the colon. The numbers denote the number of mice with the mutation per diet. Note that no insertion hotspots were observed in the colon. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

		FA Diets		
		Deficient	Control	Supplemented
Position	Deletions			
442	2			
1578	3			
2556	2			

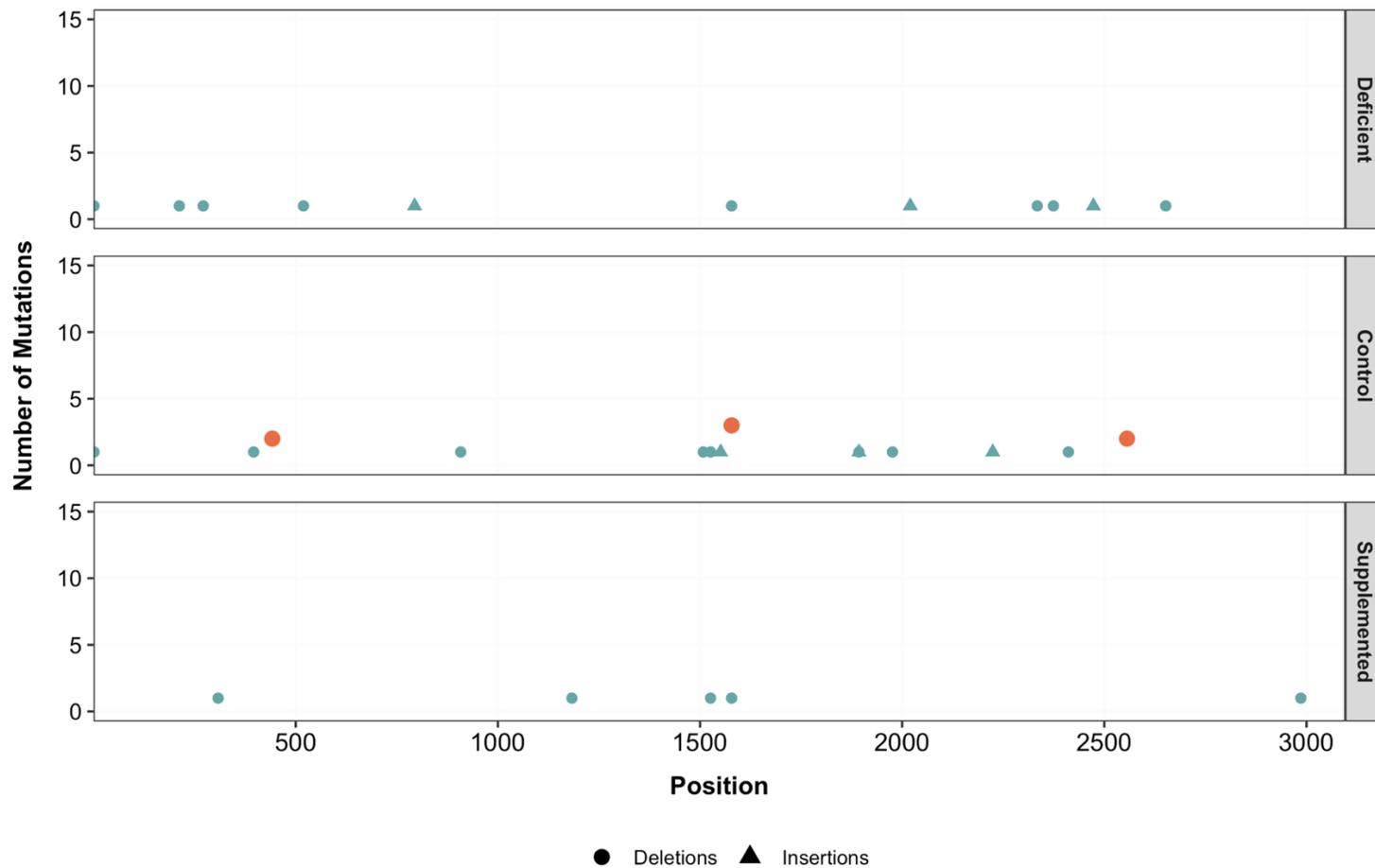


Figure 3.5: Distribution of indels across the lacZ gene in the colon in mice fed one of three folic acid diets. Dots represent deletions and triangles represent insertions. Red dots represent the deletion hotspots, deletions that occurred at the same position in more than one sample on the same diet. Number of mutations show how many samples share the indel at the same position. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

3.5 Discussion

Folate plays an important role in the maintenance of genome stability through its role in the one-carbon metabolism. Folate is required for the *de novo* synthesis of DNA nucleotides, purines and dTMP, and the maintenance of AdoMet. Altering folate status can lead to disturbances in one-carbon metabolism, thus causing genome instability and increasing risk of diseases. Folate deficiency has been associated with many diseases, including cancers^{17,47}. Folate supplementation has generally been associated with reducing the risk of cancer and rescuing hypomethylation caused by folate deficiency^{74,111,114}.

The colon is a highly proliferative tissue and folate-dependent one-carbon metabolism has been associated with the development and prevention of CRC, where both folate deficiency and supplementation have been associated with a dual effect towards CRC. These studies have looked at the epidemiological, physiological, and methylation effects of folate intake in the colon. However, the underlying mechanism of mutation induction by folate in the tissue remains unclear. Here, we studied the effect of dietary FA on the induction of genetic mutations in the colon of the male MutaMouse by means of the *lacZ* assay, which allows the detection of point mutations, small deletions, and insertions. We show that the colon in the MutaMouse strain is susceptible to FA-induced DNA mutations, where mice fed a FA supplemented diet had a higher colon MF than those fed a control or deficient diet.

Our results show that MF in the colon was significantly higher in mice fed the FA supplemented diet when compared to those fed either the FA control and deficient diets in both treatments, saline and ENU. Studies have suggested that high FA intakes may alter one-carbon metabolism by saturating the enzymatic activity of DHFR¹²⁴. DHFR is

required for the sequential reduction of FA to 7,8-DHF and then to THF, the bioactive form of folate. Saturation of DHFR can result in the appearance of unmetabolized FA in plasma and tissues, which has been proposed to cause a functional folate deficiency. At high concentrations of FA, the activity of rat and human DHFR isolated from liver is reduced 850 and 1300 times, respectively¹²⁴. This would decrease the generation of THF from DHF limiting the folate cofactors available for the *de novo* nucleotide synthesis and for the conversion of homocysteine to methionine. FA absorption happens in the intestine^{3,4}, making this tissue potentially more susceptible to saturation of DHFR. Other studies have suggested that high concentrations of FA could potentially act as a folate antagonist by increasing DHF concentrations which may also inhibit some of the enzymes required for dTMP, purine and methionine synthesis. It has been shown that thymidylate synthase, the enzyme responsible for the *de novo* synthesis of dTMP, can be inhibited by the accumulation of DHF in a noncompetitive manner¹²⁵. DHF and its polyglutamate derivatives have also been found to inhibit a key intermediate enzyme (AICAR transformylase) involved in the *de novo* synthesis of purines¹²⁶. In addition, it has also been observed that DHF can function as a competitor inhibitor of MTHFR which could result in inhibition of the formation of 5-methylTHF leading to a decrease in methionine synthesis^{127,128}. Ultimately, high levels of FA could potentially result in DNA instability and alteration of methylation patterns due to a *de facto* functional folate deficiency. This could explain why mice fed the FA supplemented diet had a higher MF in the colon compared to those fed the FA control and deficient diets in the saline treatment.

Alternatively, it has been proposed that folate deficiency may increase DNA instability and malignant transformation. Folate deficiency can induce DNA double-strand-

breaks (DSBs) by disrupting dTMP synthesis and increasing misincorporation of uracil in DNA and/or reducing methionine synthesis and cellular methylation capacity resulting in genome wide hypomethylation; either can result in chromosomal damage ^{1,129}. Folate deficiency can also result in a nucleotide pool imbalance due to impaired *de novo* purine synthesis, which affects DNA repair mechanisms, therefore increasing mutations and DNA instability ^{16,130}.

One explanation for why we did not observe an increase in MF in the colon of FA deficient mice could possibly be the microbiota in the colon. Bifidobacteria is naturally found in the gastrointestinal tract of humans and mice and it is known to have beneficial effects ^{131,132}. Bifidobacteria has also been found to be a folate producing bacteria ¹³². The uptake of microbiota-produced vitamins occurs predominantly in the colon and these vitamins contribute to the homeostasis of colonic epithelial cells ^{133–135}. In our study, mice fed a FA deficient diet had a similar MF to those fed the control diet. We could rationalize that folate produced by the gut microbiota could compensate for the depletion in those mice fed the FA deficient diet; hence, no disruption to the one-carbon metabolism, which results in a low MF.

Characterization of FA-induced mutations was accomplished by recovery of the reporter *lacZ* gene from the MutaMouse mice. The mutation spectra of the colon varied among FA diets, suggesting a FA-intake dependent mutation spectrum. Base substitutions, transitions and transversions, comprised the majority the types of mutations in all FA diets. Spontaneous somatic mutations resulting in base-substitutions can be induced by an extrinsic environmental factor, which may confer cells to advantageous characteristics which can result in cancer ^{5,136,137}. Mutation patterns of these somatic mutations are called

mutation signatures. These signatures are unique combinations of mutation types generated from different mutational processes. The Catalog of Somatic Mutations in Cancer (COSMIC) is a large genome-wide database that describes over 4 million coding mutations, from genome-wide sequencing results and curation of publications^{138,139}. The database contains information for each type of signature, including proposed etiology for the mutational processes underlying a particular mutation type and the type of cancers in which the signature has been found^{138,140}. In the COSMIC database, we identified the following single base substitution (SBS) mutational signatures associated with CRC: SBS [1, 2, 3, 5, 9, 13, 15, 18, 20, 26, 40, 44]^{139,140}. We then compared the FA mutational spectra from our results to the characteristic mutation types found for the CRC SBS to identify the possible etiology underpinning specific types of mutations.

The G:C→A:T transition was the predominant mutation observed in mice fed all of the FA diets, representing ~50% of the mutation spectra of each diet. These transitions are characteristic of CRC-associated SBS [1, 2, 3, 5, 15, 26, 40, 44]^{139,140}. The proposed etiologies of this type of transition is age, spontaneous deamination of 5-methyl-cytosine, activation of AID/APOBEC enzymes, defective DNA mismatch repair (MM), and errors during replication by polymerase¹³⁹⁻¹⁴¹.

SBS [1, 15, 26, 44] are associated MM repair which results in a G:C→A:T transition^{139,140}. MM is a DNA repair mechanism that detects and corrects DNA damage, and is involved in initiating the cellular response to such damage. MM repair is a primary repair mechanism activated in the colon. Given that the colon is a highly proliferative tissue, it relies on MM to repair base substitution mismatches during replication¹⁴². Due to MM being the main DNA repair mechanism of this tissue, any mutation or mechanism

that disrupts its function can lead to an increase in this particular base substitution. Although, there are other possible etiologies that also result in a G:C→A:T transition, it would be of interest to determine the association and effect, if any, of FA on this repair mechanism.

The SBS2 signature is also associated with a G:C→A:T transition and its proposed etiology is the activation of AID/APOBEC^{139,140}. The AID/APOBEC enzyme family, were originally identified as DNA mutators and antiviral factors^{143,144}. They are involved in the DNA cytosine deamination (C→U) with the purpose of inhibiting viruses and retrotransposons by disrupting their genome^{145,146}. APOBEC can also target the host's genome to create genome diversity, in order to help with antibody actions and to drive evolution. However, if deregulated, these enzymes can also disrupt genome stability and lead to cancer^{143,147}. The APOBEC enzymes are considered one of the most significant carcinogens¹³⁸. Its potent mutagenicity may explain the similar high proportions of G:C→A:T mutations in the *lacZ* gene seen across FA diets.

In addition, APOBEC is also the etiology suggested for G:C→C:G transversion (SBS13). However, from our results we saw that the G:C→A:T transitions (45-52%) are preferential over the G:C→G:C transitions (3-5%) in the MutaMouse. This preference for a mutation type is a direct result of DNA polymerases inserting an adenine opposite to the deaminated cytosine, causing a C>A transition. In contrast, the action of REV1, a deoxycytidyl transferase, is required in order to incorporate a cytosine opposite a deaminated cytosine resulting in a G:C→C:G transversion^{145,148}. Because of APOBEC's role as a potent mutagen it is possible that it masks some of the mutagenic effects of FA. Therefore, it would be of interest in future studies to find the sole effect of dietary FA on

this type of mutation and the relationship between APOBEC and FA's role in nucleotide synthesis, particularly since APOBEC's mutations involve purines.

The G:C→T:A was the most common transversion observed in the colon across all FA diets. CRC signatures in which this transversion was characteristic were SBS [18, 20, 26, 44]. The etiology of CRC-associated SBS signatures [20, 26, 44] is defective MM repair, as described above. However, SBS18 has been associated with damage caused by reactive oxygen species (ROS) ^{139,140}. Production of ROS has been associated with FA deficiency as a result of homocysteine accumulation by depletion of folate coenzymes ^{149,150}. ROS puts the cell in a pro-oxidant state and leads to DNA hypomethylation. Although not significant, the proportion of G:C→T:A transversion in our results were 2.2-fold higher in the group fed the FA deficient diet compared to the control diet and FA control and 1.5-fold higher when compared to the supplemented diets. Thus, ROS production may be the etiology driving this type of transversion in the colon. Moreover, this suggests the possible role of FA intake involvement in CRC by increasing G:C→T:A transversions, which are characteristic of the disease.

3.6 Conclusion

In this study, we showed FA intake has an effect on the MF of the colon in the male MutaMouse. We observed a significantly higher MF in mice fed the FA supplemented diet compared to those fed the FA control and deficient diets. MF in mice fed the FA supplemented diet was significantly higher in mice exposed to both saline and ENU treatments. We hypothesized that this increase in MF may be a consequence of the saturation of DHFR by FA supplementation ultimately disrupting the formation of THF and one-carbon metabolism. We also proposed that since gut microbiota can produce

folate, they could potentially compensate in the colon of mice fed the FA deficient diet, therefore mitigating disruption to one-carbon metabolism resulting in a low MF. However, this over overcompensation might be localized to the colon. Moreover, we show that FA supplementation does not have a protective effect in the colon when exposed to a mutagen (ENU) and in fact results in a higher MF than observed in FA deficient mice.

Identification of the FA-induced mutation spectra in the colon was achieved by NGS of the mutant *lacZ* gene obtained from the *lacZ* assay. We observed a somewhat distinct FA-induced mutation spectra per FA diet. Although, differences among mutation types across diets was not found to be significant, this is the first instance where the direct effect and the capacity to induce mutations in DNA of different FA-intake is shown.

Moreover, by comparing the FA-induced mutation spectra to mutation signatures of CRC from the COSMIC database, we found that mutations signatures associated with CRC were also observed in our mice. The etiology of these mutations included age, spontaneous deamination of 5-methyl-cytosine, activation of AID/APOBEC enzymes, defective DNA mismatch repair (MM) and damage from ROS. ROS production, the suggesting underlying etiology of SBS13 (G:C→T:A), has previously been associated with FA deficiency. Presence of this transversion in our results supports the association between ROS production and FA, while also supporting the association between FA intake and CRC by inducing mutations characteristic of the disease.

Taken together, we propose that FA supplementation enhances the type of mutations commonly associated with CRC suggesting it increases the overall susceptibility of the tissue to the disease. In addition, future studies could focus on the relationship between APOBEC, specific repair mechanisms, and FA.

Chapter 4: Investigating the folate-induced mutation spectra and its tissue-specific mutagenicity potential in the colon epithelium and bone marrow of the male MutaMouse

4.1 Abstract

Variation in cancer incidence among tissues is attributable in part to exposure to environmental risk factors. A tissue-specific mutagenicity, where certain mutations are preferentially selected in different tissues, results in tissues-specific mutation profiles. Folic acid is an extrinsic factor which has been associated with colorectal cancer and acute lymphocytic leukemia. Here, we used the MutaMouse model to determine the mutagenic potential of FA *in vivo* in the colon and bone marrow, and to determine if there is a tissue- and diet-specific effect induced by FA intake. Male mice were fed experimental FA diets (deficient, control, or supplemented) for 20 weeks from weaning. NGS was used to sequence the *lacZ* reporter gene and we determined the mutation spectra of both tissues. The tissues presented with contrasting responses to FA both in MF and mutation spectra. In the colon, the highest MF was observed in mice fed the FA supplemented diet in mice treated with either saline or ENU. However, in the bone marrow, the highest MF was observed in mice fed the FA deficient diet, and only in the saline treated mice. When comparing the mutation spectra of the tissues, the bone marrow had a FA-diet-specific mutation spectra that differed significantly for mice fed the FA deficient diet compared to the control and supplemented diets. However, in colon, the FA supplemented diet did not cause a unique mutation spectra but rather enhanced the overall incidence of colon-specific mutation types. Furthermore, we show that the types of mutations we observed are characteristic of CRC and ALL. We demonstrate, for the first time, the mutagenic potential

of FA-intake, the mutation spectra induced by different FA intake levels and that the mutagenic potential of FA is tissue-specific.

4.2 Introduction

Spontaneous mutations that occur in somatic cells accumulate steadily through life. These mutations can arise from endogenous and exogenous processes ⁵. A mutation from one cell can be inherited by its descendants (clones) and those mutations that confer a selective advantage will have preferential growth and survival. One possible outcome of a somatic mutation, or the accumulation of many mutations, is cancer. Cancer is a disease in which a mutation confers a clone the ability to divide and grow uncontrollably by escaping the built-in and exogenous restraints of proliferation of normal somatic cells ^{5,136}. About 95% of the somatic mutations observed in solid tumors that result in altered protein products are the result of single-base substitutions ¹³⁷. The remainder consists of deletions and insertions of a few bases. Variability in the number of mutations found in tumors is attributed to defects in DNA repair, mutations in the proofreading domain of DNA polymerases (POLE and POLD1) and exposure to an exogenous mutagen ^{137,151}.

The variation in cancer incidence among tissues is well known and it is attributable to exposure to environmental risk factors, inherited genetic variation and recently, and to the lifetime number of cell divisions of the tissue's stem cells ¹⁰⁹. Mutation profiles are tissue-specific even when the accumulation of lifetime somatic mutations is similar among tissues ¹⁰⁸. While some genes are associated with cancers in numerous tissues, there are also tissue-specific cancer genes that are mutated in selected types of cancers. The BRCA1 is one tissue-specific cancer gene which, if it has a loss-of-function mutation, increases the risk of breast and ovarian cancer by up to 80% but not the risk for other cancers ^{142,152}.

Moreover, the type and location of mutations in a particular gene can vary among cancers from various tissues. The p53 tumor suppressor gene mutational spectrum differs among cancers, where in colon cancer transitions at CpG islands predominate, whereas G>T transversions along the gene are most frequent in lung and breast cancers¹⁵³. In addition, some exogenous mutagens, such as environmental mutagens and toxins, have a preferential tissue target, at times producing cancer in one tissue but having no effect in many others¹⁵⁴⁻¹⁵⁷. This suggests a tissue-specific mutagenicity where different mutations are preferentially selected in different tissues since the advantages given by a mutation to a cancer cell can vary by cell type¹⁴².

Sequencing has shed light on possible mechanisms and processes involved in tumor evolution. Cancer genomes provide a history of the mutagenic processes that have occurred before and after the development of a neoplastic transformation^{158,159}. Cancer genome sequencing has also revealed mutational patterns, signatures, which may reflect the effect of extrinsic factors. Mutational signatures induced by mutagenic agents could provide new insights on the origins and causes of human tumors and cancers. Signatures of many known and suspected carcinogenic agents, including drugs, have been determined¹⁶⁰.

Folic acid is an exogenous factor that has been associated with CRC and acute lymphocytic leukemia (ALL). Folate is a water-soluble B vitamin found in fresh fruits and vegetables, legumes and organ meats. FA is the synthetic oxidized form of folate found in fortified foods such as enriched white wheat flour and other enriched cereals. Folate is required for the *de novo* synthesis of purines and thymidine (dTMP), and the synthesis of methionine¹. *De novo* synthesis of purines and dTMP is essential for genome stability and cell proliferation. Methionine can be adenylated to form AdoMet, which is the universal

methyl donor in the cell, required for the methylation of various molecules in the cell including DNA and histones. DNA and histone methylation patterns are essential for the regulation of gene expression, cellular differentiation, and organogenesis, among other processes.

Disruption to any of the three folate-dependent biosynthetic pathways can have serious implications in a cell's survival and tissues as a whole, and similarly can impact the propensity for a cells to undergo transformation. Impairment in purine synthesis can result in a reduction of DNA replication and abnormal DNA repair and has been shown to induce mutations in Chinese hamster ovary cells ^{13,14}. Insufficient dTMP synthesis may cause a nucleotide imbalance resulting in the accumulation of uracil and its misincorporation into DNA which could increase DNA double strand breaks due to the “futile cycle” of DNA repair ^{15,16}. In addition, reduction in the production of AdoMet lowers the availability of methyl groups and the cellular methylation capacity. This can alter methylation patterns in the genome and lead to genome wide hypomethylation resulting in genome instability, disruption of gene expression, and alter risk for diseases such as cancer ^{20,110}. However, while altered folate metabolism has been implicated in cancer development, the study of folate from the perspective of it acting as a mutagen is lacking, and the folate-induced mutational signature(s), if they exist, are unknown.

CRC is one of the most common malignancies worldwide ⁹⁹. About 90% of CRC cases derive from the colon and rectal epithelium where the tumor can breach the intestinal wall and then spread to lymph nodes and other organs ¹⁰¹. Tumorigenesis in the colon has been observed to start at the surface of the epithelium and also in the crypts ^{104,105}. Cancer stem cells have been identified in colon cancer, these stem cells have a deregulated capacity

of self-renewal ^{106,107}. Environmental, including food components, and inherited factors have been found to play an important role in the risk of CRC ^{109,142}.

Folate is an extrinsic factor that has been shown to have a dual effect on the risk of CRC. FA supplementation has been associated with a protective effect towards a high risk of CRC derived from folate deficiency. A large prospective cohort study of 88,756 women showed that long-term FA consumption as part of a daily multivitamin (Vit A, C, D, E) intake was associated with a significant reduced risk of colon cancer ⁷⁴. Animal studies have shown that folate deficiency increases the risk of CRC by increasing neoplastic lesions in rats with an underlying predisposition ¹¹² and that high dietary folate intake reduced the evolution of macroscopic neoplasms in the colorectum of rats ¹¹³. On the other hand, FA supplementation has been associated with promoting the progression of malignant lesion in CRC in the presence of established neoplasms. A study in a murine model (*Apc*^{+/-} *Msh*^{-/-}), predisposed intestinal tumorigenesis mice, showed a protective effect provided by FA supplementation against the development of small intestinal adenomas and colon aberrant crypt foci in when supplementation started before the establishment of neoplastic foci ¹²¹. However, once neoplastic foci are present, FA supplementation had the opposite effect and increased the number of neoplastic foci.

Leukemia and multiple myeloma are cancers that originate in the bone marrow. In Canada, leukemia has a higher incidence in males than in females with a mortality rate of ~50% for both sexes. ALL is a neoplastic disease caused by somatic mutations to a single lymphoid progenitor cell at one of the many stages of development ¹⁶¹. Incidence of ALL has its first peak in childhood and a second peak in adults of ~50 years of age ¹⁶². FA intake and genetic polymorphisms in genes related to folate transport and metabolism have also

been associated with ALL risk and treatment outcome, as well as other bone marrow diseases. A previous study from our lab showed that bone marrow from mice was susceptible to the induction of DNA mutations and chromosomal instability in conditions of folate deficiency. Additionally, we did not observe a protective effect from FA supplementation on these endpoints when mice were exposed to a mutagen (ENU)⁸⁸. In contrast, *in vitro* experiments in the human liver cell line, L02, showed that FA supplementation reduced Benzo(a)pyrene-induced mutagenicity and genotoxicity⁷⁷. This protective effect of FA supplementation was also seen *in vivo*, where a significant decrease in micronucleus frequency was observed in the bone marrow of mice in the FA supplemented intervention when compared to control⁷⁷. In regards to polymorphisms, the human SNP SHMT1 1420C>T has been associated with reduced risk for ALL⁶⁷, whereas MTHFR 677C>T and 1298A>C SNPs were associated with lower risk of adult ALL. These suggest that altered folate metabolism may be involved in the development of ALL⁶⁹.

Finally, folic acid deficiency was associated with complications and adverse outcomes in children in India with ALL during maintenance chemotherapy⁶⁵. SNP 80G>A in the *RFC1* gene, the reduced folate carrier, is associated with low plasma folate and high homocysteine concentrations in healthy persons¹⁶³. An association between the RFC1 80G>A polymorphism and increased ALL susceptibility was found in a study where DNA was isolated from pediatric ALL patients and compared to controls⁷¹. Additionally, RFC1 is also a major transporter of methotrexate, an antifolate chemotherapeutic that is a key treatment of childhood ALL. A study examining the plasma concentrations of methotrexate and disease outcome in children with ALL found that patients with the SNP had worse prognosis and that homozygosity for the 80G>A polymorphism had higher methotrexate

plasma concentrations ¹⁶⁴; the poorer outcomes could have been due to lower cellular uptake of methotrexate.

Given the association of folate status with CRC and ALL cancers and the potential of folate to induce mutations by disrupting one-carbon metabolism, it is of interest to measure the potential mutagenicity of FA in the colon and bone marrow. The MutaMouse was used to measure and characterize folate-induced mutations in both somatic tissues. The transgenic rodent mutation assay allows us to detect and measure the mutation-inducing potential of many substances and to determine mutational spectra *in vivo*. The *lacZ* reporter gene is incorporated into the MutaMouse genome and can be recovered to detect *de novo* mutations ⁸³. Positive selection of mutated *lacZ* genes allows for the calculation of mutant frequency (MF) in response to an environmental intervention in all tissues and sequencing of the *lacZ* gene provides data which enables a mutation spectrum to be compiled for the given mutagenic agent ^{84,85,159}. Here, we used the MutaMouse to evaluate the effect of dietary FA on the MF and mutational spectra in the colonic epithelium and bone marrow to determine its tissue-specific mutagenic potential and identify potential mechanisms of action that may link folate metabolism with CRC and ALL risk.

4.3 Methods

4.3.1 Animal study

The animal study was performed in-house at Health Canada. The study followed the guidelines described in the Guide to the Care and Use of Experimental Animals ¹²³ and it was approved by the Health Canada Ottawa Animal Care Committee. Sixty weanling males derived from the in-house MutaMouse colony that were housed with *ad libitum* access to water and diets and maintained at standard humidity. At 5 weeks of age the male mice were fed one of the three FA diets, which have been routinely used in the lab, and continued the diet for 20 weeks : 0 mg FA/kg (deficient), 2 mg FA/kg (control) or 8 mg FA/kg (supplemented) (Dyets, Inc.; Bethlehem, PA) ⁸⁹. The control diet (2 mg/kg) represents an adequate dietary FA intake for rodents and corresponds with the RDA for adult humans, which is 0.4 mg folate per day. The deficient diet (0 mg/kg) was used to represent an inadequate dietary FA intake, in our hands it has been shown to decrease both tissue and plasma folate and increase homocysteine but not render the rodents completely folate depleted. The supplemented diet (8 mg/kg) is 4-fold the control diet and is used to represent a dietary FA intake of 1.6 mg per day in adult humans, representing a folate intake that could be achieved through the combined consumption of folic acid fortified foods and over-the-counter supplements sold in Canada. After 10 weeks on diet, half of the mice in each diet group (n=10) were given a 50 mg/kg dose of ENU dissolved in phosphate buffer by gavage (n = 10) and the other half in each diet group (n = 10) were given saline by gavage. The mice were fed the diet for another 10 weeks. At the time of necropsy, the bone marrow and the colon epithelium, were collected and flash frozen in liquid nitrogen.

4.3.2 *LacZ* assay and preparation for sequencing

The mutant plaques from the bone marrow used for sequencing were isolated in a previous study; the bone marrow came from the same mice from which the colon was collected⁸⁸. Briefly, DNA was extracted from the bone marrow (collected from both the right and left femurs) using a phenol chloroform method and used in the *lacZ* assay⁸⁸. Similar to that described for the colon, recovery and packaging of the λ gt10*lacZ* copies was accomplished using the Transpack Packaging extract (Agilent Technologies, Santa Clara, CA). Infection of the host bacteria (*galE*⁻ *recA*⁻ *E.coli*) with the package phages was followed overnight incubation of the plated bacteria to allow visibility and quantification of plaques the following day. Plaques from plates containing agar and P-gal, selective medium, were used to estimate MF and plaques from plate containing only agar, non-selective medium, were used to calculate PFUs. MF was calculated per sample as the number of mutant plaques divided by the total number of PFU. The mutant plaques from a single mouse were pooled and stored at -80°C until sequencing. After statistical analysis outliers were removed from the MF analysis.

Mutant plaques from remaining mice representing all three FA diets from the saline treatment, only, of the bone marrow were prepared for sequencing (Table 4.1). Plaques from titer plates (~200) were collected to act as wild type/baseline mutation plaque controls. Tubes containing the pooled plaques were boiled to melt the agar. An aliquot of 10 uL of the suspension was transferred to a well in a PCR plate. Amplification and sequencing of the *lacZ* gene for each mouse was done in duplicate as a technical replicate (two separate reactions) to control for PCR errors. The PCR mastermix final volume was 50 uL per well and contained: 10 uL aliquot from the plaques suspension, 10 uL of 5X Q5

Reaction Buffer, 0.5 uL of Q5 High-Fidelity DNA Polymerase, 1 uL of dNTPs (10 mM) (NEB Inc.), 2.5 uL of each primer (10uM), forward and reverse, and 23.5 uL of nuclease-free water. The *lacZ* gene was amplified in a 30-cycle PCR as follows: 95 °C for 3 min; 30 cycles of 95 °C for 45 s, 50 °C for 1 min, 72 °C for 4 min; final extension at 72 °C for 7 min. The PCR products were purified using the QIAquick PCR purification kit (Qiagen). DNA was quantified using a Qubit™ dsDNA HS Assay kit (Invitrogen™) in a Qubit 2.0 Fluorometer. Samples were diluted with nuclease-free water to a final DNA concentration of 0.2 ng/uL and stored in LoBind tubes at -20°C until library preparation.

4.3.3 Illumina sequencing

Libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina). The Nextera XT protocol was followed except for the last step where libraries are normalized. Briefly, the Nextera transposome was used to “tagment”, fragment, and tag the genomic DNA (gDNA) of each sample. The “tagmented” DNA was then amplified using a 12 cycle PCR program. During this amplification, the indexes (N7XX and S5XX) and full adapter sequences were added to the “tagmented” DNA. To purify the libraries and remove shorter fragments the SPRIselect reagent beads (Beckman Coulter, Inc.) were used instead of the AMPure XP beads, used in the protocol. Evaluation of the quality of the libraries was done using a High Sensitivity DNA Screen Tape in an Agilent 2200 TapeStation. This replaced the normalization steps from the Nextera XT protocol. On the TapeStation the region was set to 150-1250 base pairs and its molarity (pM/L) was calculated. The libraries were then diluted to a final molarity of 500pM. Finally, 4 uL aliquots from the 500 pM solutions of each library were pooled together.

Sequencing of the *lacZ* libraries was performed in-house using the Illumina NextSeq 500 System with the NextSeq 500/550 High Output Kit 2.5 (150 cycles). A total of 29 samples (mice) from bone marrow were sequenced, only those from the saline diet excluding the outliers, as mentioned before. Each sample was sequenced with its duplicate, plus the titer plaques from the control diet (Table 4.1).

4.3.4 Bioinformatics

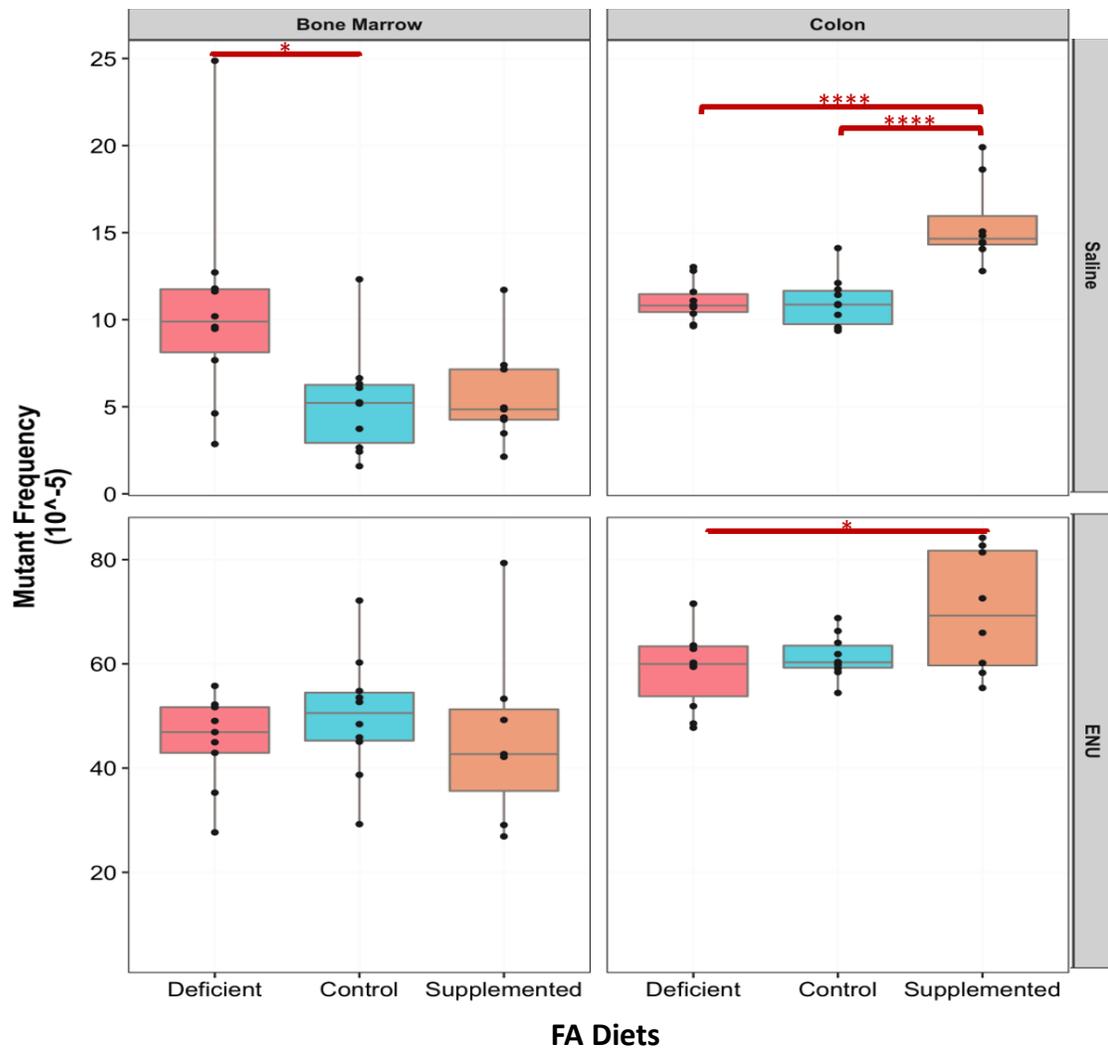
Bioinformatic analysis of sequence data was conducted using the pipeline described in Beal et al., 2015⁸⁴. In summary, the raw sequence data was converted into FASTQ format using the bcl2fastq conversion software v2.17.1.14 (Illumina, Inc). Reads were trimmed and aligned to the reference *lacZ* sequence from the MutaMouse using bowtie2⁹³. Alignment pileups were created using SAMtools⁹⁴. From the pileups, the proportions of base substitutions and indels were determined for each library relative to the reference sequence. The mutation proportions were normalized by subtracting the false mutation proportion (estimates PCR artifacts and sequencing errors) from the total mutation proportion. To reduce false indel calls, indels with false mutation proportions higher than the highest false mutation proportion from the base substitutions were ignored. Mutation calls in which both replicates were above the mutation calling threshold (1/# plaques of sample) were considered true mutations. Identical mutations which occurred more than once at the same location were considered recurrent mutations, a result from clonal expansion from a single independent event. For our analysis we considered only independent mutations, unique mutations. These were used to determine mutation spectra and to map the distribution of the mutations across the *lacZ* gene.

4.4 Results

4.4.1 Comparison of mutant frequencies across tissues

The *lacZ* assay was previously performed in bone marrow⁸⁸ from the same mice used in the colon study. A comparison of the results from the two somatic tissues is presented in Figure 4.1.

The bone marrow showed significant differences in MF such that the FA deficient diet resulted in a 1.9-fold higher MF than the control diet in the saline treatment group ($p = 0.035$). As expected, ENU treatment induced a higher MF compared to saline treatment in both tissues. However, in the bone marrow there were no significant differences based on diet⁸⁸. In contrast, in colon, as stated previously in chapter 3, there was a significant difference in MF among the FA diets in both treatment groups, such that FA supplementation resulted in a 1.4-fold and 1.2-fold higher MF than the control and deficient diets in the saline and ENU treatments, respectively. Given that these results are from the same mice, we show that there are tissue specific responses to FA intake.



4.4.2 Sequencing of the *lacZ* reporter gene in somatic tissues

A total of 28 mice (samples) (452 plaques) from colon and 29 samples (463 plaques) from bone marrow were sequenced. (Table 4.1). Sequencing generated a combined total of 53,881,611,015 base reads (nucleotide reads) for both tissues; 18,493,058,083 and 35,388,552,392 from the colon and the bone marrow, respectively.

After correcting for false mutation proportion and applying the mutation calling threshold on the sequencing data, “true mutations” in the colon were called from nine samples each from the FA deficient and control diets and eight samples from the FA supplemented diet. In the bone marrow, “true mutations” were called from 10 samples from the FA deficient diet, 9 samples from the FA control diet and 8 samples from the FA supplemented diet. There was a combined total of 360 and 399 recurrent mutations identified in all samples, from which 165 and 149 were independent mutations (including indels) in the colon and bone marrow, respectively. Only independent mutations were used in the analysis.

Table 4.1: Samples sequenced, and number of independent mutations identified in the lacZ gene from the colon and bone marrow. The total number of mice (samples) per FA diet that were sequenced, and the number of base substitutions, deletions and insertions identified per FA diet

Tissue	FA Diet	# of Samples	Total	Total number of Libraries (with duplicates)	# of Samples used in Analysis ²	Independent mutation types		
						Base substitutions	Deletions	Insertions
Colon	Deficient	10	28	56	9	44	7	3
	Control	10			9	39	12	3
	Supplemented	8			8	52	5	-
					Total	135	24	6
Bone Marrow	Deficient	10	29	58	10	53	11	1
	Control	10			9	40	11	2
	Supplemented	9			8	28	3	-
					Total	121	25	3

² Number of samples used in the analysis may differ from the original number of samples sequenced due to the fact that a sample may have been dropped if none of its mutation calls remained after correcting for false mutation proportion and applying the mutation calling threshold.

4.4.3 Somatic tissue mutation spectra comparison

Some qualitative differences in mutation types between colon and bone marrow were observed. The bone marrow proportions of transitions, transversions and indels, respectively, in mice fed the FA diets were: 0.62, 0.23 and 0.15 in those fed the FA deficient diet; 0.62, 0.20 and 0.18 in those fed the FA control diet and 0.70, 0.19 and 0.10 in those fed the FA supplemented diet, respectively. As reported in chapter 3, the proportions in colon were 0.48, 0.34 and 0.18 in those fed the FA deficient diet; 0.50, 0.20 and 0.30 in those fed the FA control diet and 0.55, 0.35 and 0.10 in those fed the FA supplemented diet, respectively.

Overall, the proportions of transitions were higher in mice across all FA diets in the bone marrow than in colon, while the opposite was observed for transversions. In regard to indels, in bone marrow the proportion of mutations that were indels were lower in mice fed the FA control and deficient diet and the same in mice fed the FA supplemented diet when compared to the proportion from the colon in mice fed the same diets.

Proportions of mutation types in the bone marrow among the three FA diets can be seen in figure 4.2 (upper). Bone marrow had a significant difference among the proportions of the type of mutations. In bone marrow a significantly higher proportion of G:C→C:G transversions was observed in the mice fed the FA deficient diet compared to those fed the FA control diet ($p=0.016$) and the FA supplemented diet ($p=0.02$). The A:T → G:C transitions were present in mice fed the FA control in bone marrow and FA control and supplemented diet in colon but not in those fed the FA deficient diet in both tissues. However, these observations were not significant for either tissue. The mutation spectra observed in mice fed the FA supplemented diet in the bone marrow was made up of only

three type of mutations, G:C→A:T, G:C→T:A and deletions. All four types of possible transversions (G:C→C:G, G:C→T:A, A:T→T:A, A:T→C:G) were observed only in the group fed the FA deficient diet in the bone marrow as opposed to the colon where the four transversion types were observed in mice fed all three FA diets. Finally, insertions were only observed in the group fed the FA control and deficient diets in both tissues.

Figure 4.3 shows the proportion of types of mutation adjusted to the MF obtained from the lacZ assay of each tissue.

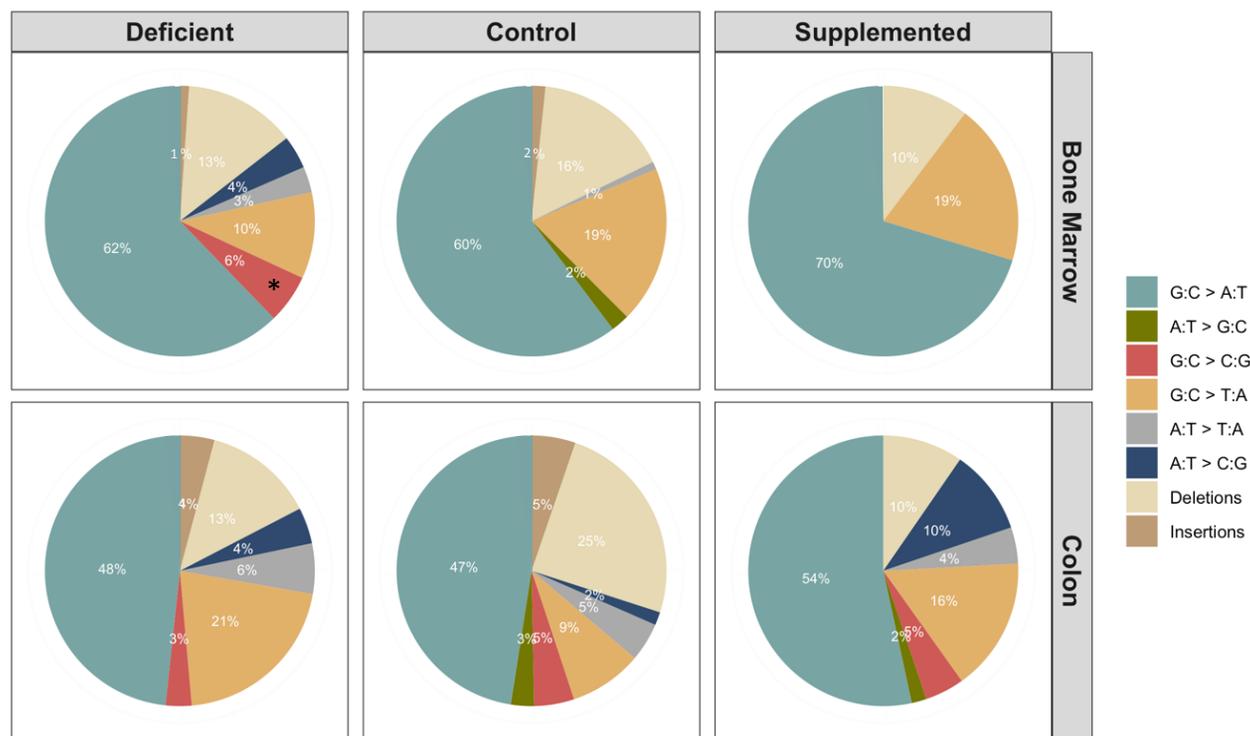


Figure 4.2: Proportional representation of mutation spectra of the bone marrow and colon of male mice on the saline treatment fed one of three FA diets. In the bone marrow the G:C → C:G transversion is significantly higher in mice fed the FA deficient diet when compared to mice fed the FA control or supplemented diets (One way ANOVA, Tukey HSD, * $p \leq 0.05$). No significant differences were identified among the mutation proportions in the colon. Diets: Control = 2 mg FA/kg; Deficient = 0 mg FA/kg; Supplemented = 8 mg FA/kg

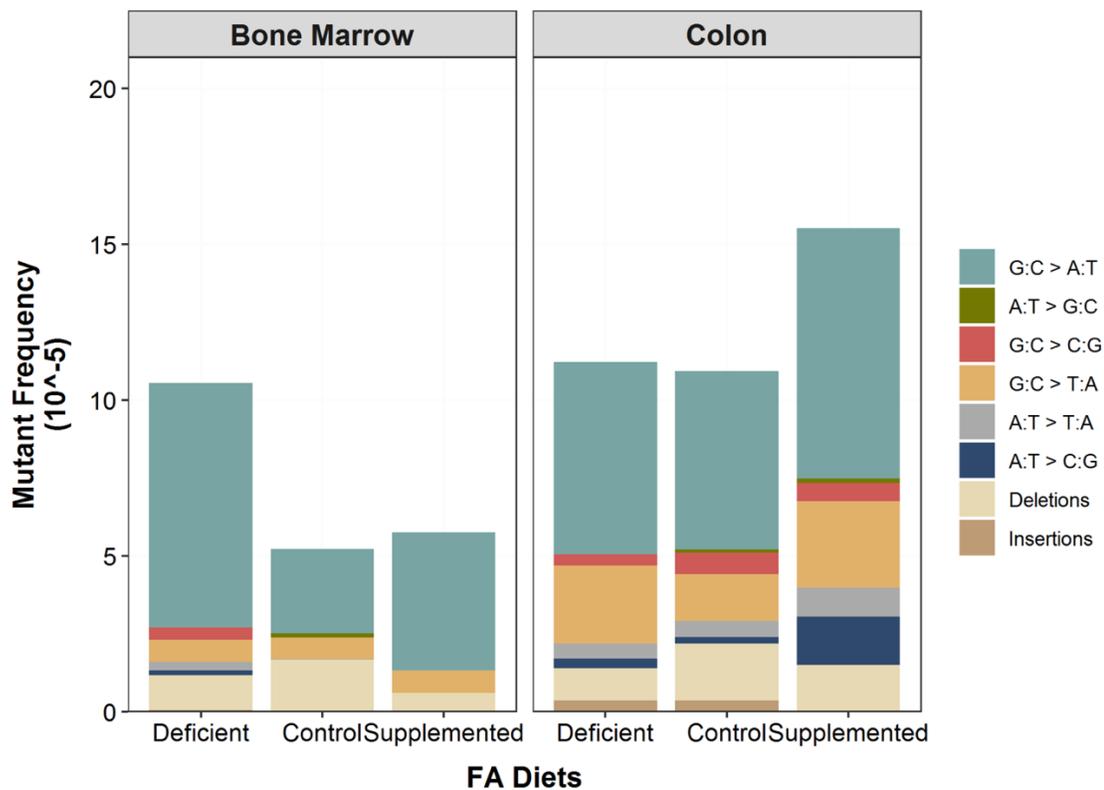


Figure 4.3: LacZ mutation spectra of the colon and bone marrow in mice fed one of three FA diets. Proportions of all type of mutations adjusted to MF of each tissue respectively. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

4.4.4 Distribution and hotspots of base substitutions and indels along the *lacZ* gene in bone marrow and comparison to hotspots in colon

Plotting the distribution of the independent FA-induced mutations in the bone marrow along the *lacZ* gene allowed for the identification of candidate mutational hotspots. A position was considered a candidate hotspot if a mutation occurred at the same position in more than one sample (mouse) per diet.

A total of seven hotspots were identified in the bone marrow of mice fed the FA deficient and five hotspots in mice fed the FA control and supplemented diets (Table 4.2). All of the hotspots identified among FA diets occurred at C:G base pairs. All of the hotspot mutations were G:C→A:T transitions. Six hotspots common in more than one diet and two hotspots were common in all diets. A visual representation of the distribution of the base substitutions along the *lacZ* gene can be observed in figure 4.4.

Overall, there were 30% more hotspots identified in bone marrow than in colon. Five hotspots were observed in both tissues although the FA diet group in which the hotspots were identified were not always the same. The type of mutations observed at the hotspots also differed between both tissues. While, in bone marrow all hotspots occurred at C:G base, in colon ~11% of the hotspots from the supplemented diet occurred at a T:A base pair.

Table 4.2: Position of candidate base substitution hotspots. Independent base substitutions that occurred at the same position in more than one sample per diet in the bone marrow. The numbers denote the number of mice with the mutation per diet. Positions in red color are hotspots also identified in colon. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg. * Hotspot present in more than one diet

Position	FA Diet			Type of Mutation
	Deficient	Control	Supplemented	
928 *	3		2	G > A
1072 *	5	4	2	C > T
1090 *	3		2	G > A
1187 *	7	3	4	C > T
1627 *	4	2	3	G > A
2374		2		C > T
2659	2			C > T
2713 *	2	2		G > A

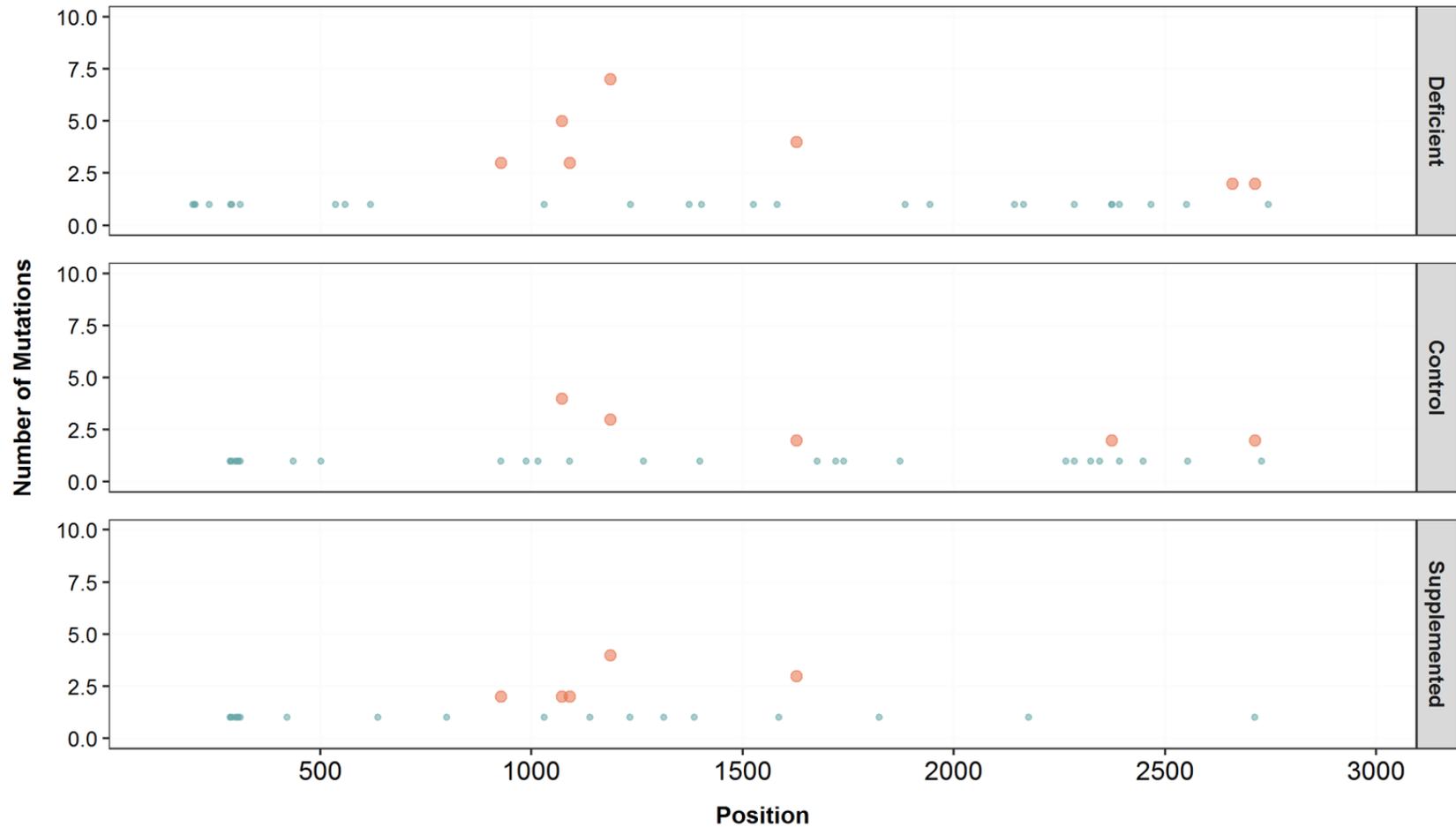


Figure 4.4: Distribution of base substitutions across the lacZ gene in the bone marrow in mice fed one of three folic acid diets. Dots represent a base-substitution. Red dots represent hotspots, representing a mutation that occurred at the same position in more than one mouse from the same diet. The y-axis represents the number of mice that have a base substitution at the same position. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

Candidate indel hotspots in bone marrow are shown in table 4.3. Only one deletion hotspot and no insertion hotspots were identified among the FA diets.

The only deletion hotspot was identified in the bone marrow of mice fed the FA deficient and control diets. The deletion hotspot did not occur at the same position as base substitution hot spot. The deletion hotspot occurred in both diet groups at position 1578. A visual representation of the distribution of indels across the *lacZ* gene in the bone marrow can be observed in figure 4.5.

Overall, we identified more deletion hotspots in colon than in bone marrow, a 3:1 ratio. However, the deletion hotspot in bone marrow was shared between two FA diets. The only deletion hot spot in bone marrow was also identified as a deletion hotspot in colon. Lastly, no insertion hotspots were identified for either tissue.

Table 4.3: Indel hotspots along the lacZ gene in the bone marrow. Independent indels that occurred at the same position in more than one mouse per diet in the bone marrow. The numbers denote the number of mice with the mutation per diet. Position is in red font to denote that the hotspot was also identified in colon. Note that no insertion hotspots were observed in the bone marrow. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg. * Hotspot position present in more than diet.

FA Diets			
	Deficient	Control	Supplemented
Position	Deletions		
1578 *	2	2	

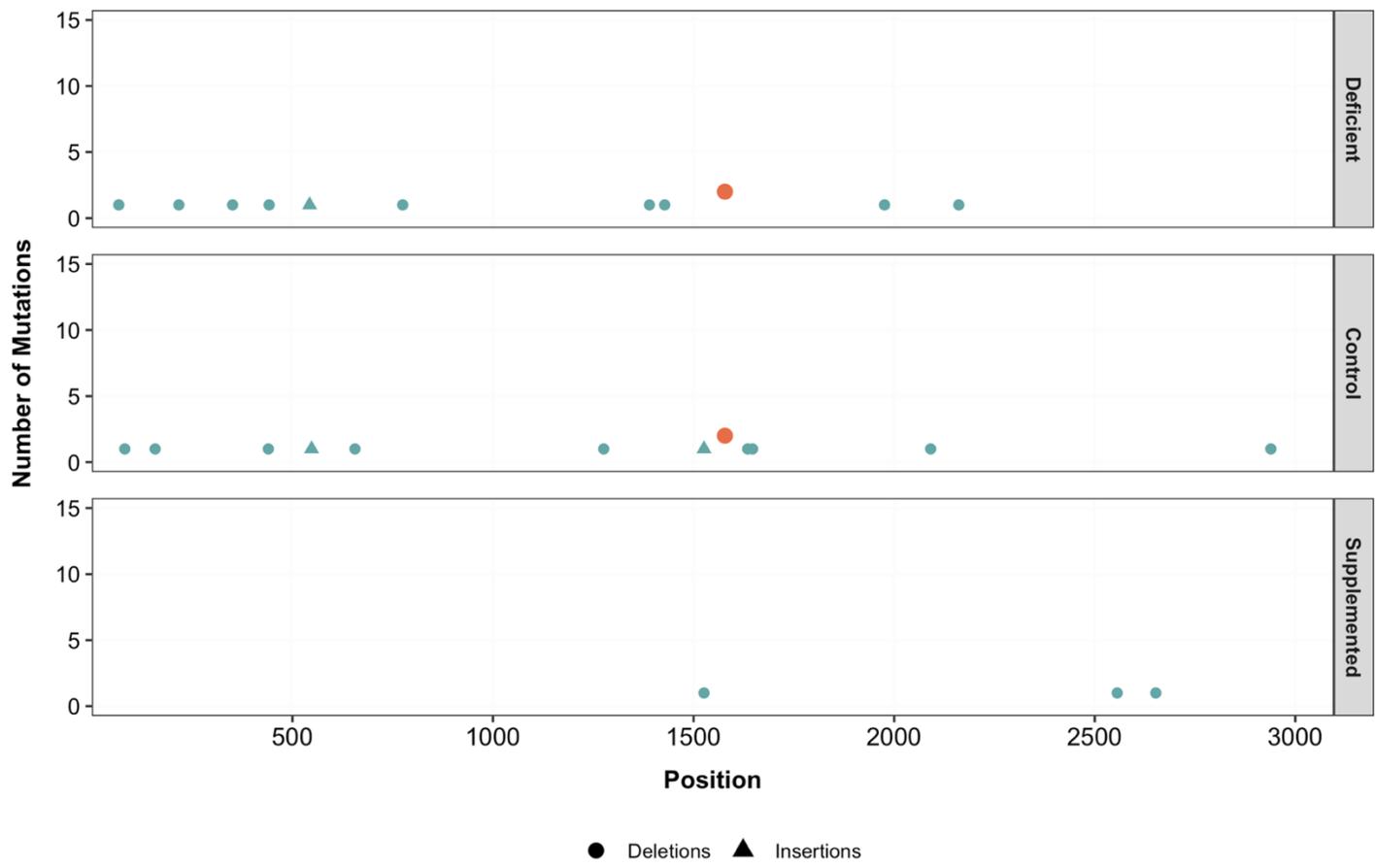


Figure 4.5: Distribution of indels across the lacZ gene in the bone marrow in mice fed one of three folic acid diets. Dots represent a deletions and triangles represent insertions. Red dots represent deletion hotspots, deletions that occurred at the same position in more than one mouse in the same diet group. The y-axis represents the number of mice that share an indel at the same position. Diets: Control = 2 mg FA/kg; Deficient= 0 mg FA/kg; Supplemented = 8 mg FA/kg.

4.5 Discussion

Folate-mediated one-carbon metabolism is required for the *de novo* synthesis of DNA nucleotides and maintenance of AdoMet, thus it is important for DNA stability, gene expression, cell proliferation and methylation maintenance. Disruption of one-carbon metabolism by altering folate intake or metabolism can result in DNA instability and an increased risk of disease. Folate deficiency has been associated with many diseases, including cancers^{17,47}. Folate supplementation has generally been associated with reducing the risk of cancer and rescuing hypomethylation caused by folate deficiency^{74,111,114}.

Bone marrow is a highly proliferative tissue and the origin of ALL, a cancer initiated in white blood cell precursors (lymphocytes) in the bone marrow. The risk of developing ALL and its response to treatment have been associated with folate status and SNPs in proteins involved in one-carbon metabolism. Folate status has also been associated with CRC, where folate deficiency and supplementation have been observed to have a dual effect on cancer initiation and growth. Because mutagenicity has been suggested to be tissue-specific where different mutations are preferentially selected in different tissues¹⁴², we compared the FA-induced MF of the colon with previous MF results from the bone marrow from the same mice⁸⁸, which were consistent with the hypothesis that folate deficiency causes genetic mutations and genome instability. An increased induction of chromosomal damage, as indicated by micronucleated red blood cells, and increased MF of the bone marrow was observed in mice fed the FA deficient diet compared to mice fed the control and supplemented FA diets⁸⁸. Results which were not mirrored in the colon, where MF was significantly higher in those fed the FA supplemented diet when compared to the other diets. Therefore, we demonstrate that the FA-induced DNA mutations are tissue

specific where in bone marrow FA deficiency results in higher MF and in colon FA supplementation results in higher MF. We also compared the FA-induced mutation spectra from both tissues which further confirms the tissue-specific mutagenicity induced by FA intake.

The differences in MF and response to FA doses suggest that the intrinsic metabolic and/or DNA repair mechanisms differ between these two somatic tissues. One difference between these two somatic tissues is the type of folate to which they are exposed. FA consumed from fortified foods and vitamin supplements as well as reduced folates (mostly 5-methylTHF) from natural food sources are mostly absorbed across the intestinal epithelium primarily by the PCFT in the small intestine. In the case of our mouse study, the casein-based AIN-93G diet has low levels of natural folate, therefore it is enriched with a FA-containing vitamin mix. As such, FA is the primary folate consumed. To be bioactive, FA needs to be reduced to THF. Serum folates circulate as monoglutamate derivatives, with 5-methylTHF being the predominant form found in blood ^{3,4}. Thus, bone marrow has access primarily to 5-methylTHF which is ready to enter the one-carbon metabolism as opposed to the colon where FA has to be reduced. In addition, the intracellular concentration of folate, which depends on the intracellular bound folate and the kinetics of long-term turnover of folate, differs among tissues ¹⁶⁵. This may explain the different rates at which tissues can become folate depleted and consequently influence the mutagenic potential of altered folate metabolism.

Nearly all intracellular folate is bound by enzymes that catalyze the folate-dependent reactions. Unbound folate is susceptible to degradation ¹. Modeling of folate reactions suggested that protein binding of folates helps to maintain folate homeostasis in

cases when folate status decreases ¹⁶⁶. For example, nuclear folate is more difficult to deplete than cytoplasmic folate. At the tissue level, high folate concentrations in the bone marrow of rodents has been observed, consistent with its high metabolic activity ¹⁶⁷. However, a higher kinetic turnover in bone marrow could explain why a folate deficient diet would result in a higher MF, if simply because it has a higher folate requirement and becomes more readily depleted under deficient conditions. The renewal rate of bone marrow is higher than that of colon producing ~1 trillion cells per day ^{168,169}. Folate requirements are higher for fast proliferating cells to support nucleotide synthesis ¹⁷⁰. We could hypothesize that FA supplementation did not cause an increase in MF in bone marrow due to its very high proliferation rate and high folate requirement. Bone marrow would require high intracellular folate concentrations in order to keep up with *de novo* nucleotide synthesis. This high kinetic turnover would also explain why FA deficiency would deplete the bone marrow and increase its MF, as the intracellular folate concentrations would be rapidly depleted in cells with a high rate of proliferation exposed to low folate conditions.

In contrast, while colon is considered a highly proliferative tissue, its rate of proliferation is slower than that of bone marrow. Around 10 million colon cells are regenerated a day with a full replacement of the colon epithelial surface every 3-5 days ¹⁷¹⁻¹⁷³, compared to the ~1 trillion cells produced in the bone marrow ¹⁶⁹. Also, it has been suggested that high concentrations of FA may saturate the DHFR enzymatic activity which can cause a functional folate deficiency by limiting the folate derivative cofactors available in the one-carbon metabolism ¹²⁴. Therefore, slower kinetic turnover and exposure to FA, rather than solely reduced folate derivatives could inhibit one-carbon metabolism and result

in an increase in MF in mice fed this FA diet. Moreover, naturally occurring folate-producing microbiota in the colon could explain the low MF observed in mice fed the FA deficient diet, as the folate produced by the microbiota is readily absorbed by the colon and could compensate, at least transiently, for the depletion caused by the diet^{131–135}.

As expected, the colon of mice treated with ENU exhibited a ~5-fold increase in MF regardless of FA diet when compared to those in the saline treatment. ENU is an alkylating agent that induces point mutations¹⁷⁴. We showed that the MF in the colon of mice treated with ENU followed the same direction as that observed in mice treated with saline, such that mice fed the FA supplemented diet had significantly a higher colon MF compared to mice fed the control diet. This differs from what was observed in the bone marrow, where MF across the FA diets did not show a significant difference meaning that FA intake did not have an effect on ENU-induced MF. Therefore, FA supplementation did not have a protective effect by decreasing the MF on either somatic tissue. In regard to the colon, FA supplementation exacerbated the ENU-induced MF possibly by inhibiting THF production because of the saturation of DHFR.

Identification of the FA-induced mutation spectra was accomplished by sequencing of mutant plaques. Mutation spectra varied between the two somatic tissues and among FA diets, suggesting a FA-intake dependent mutation spectrum. Base substitutions, including both transitions and transversions, comprised the majority of the types of mutations in both tissues. The proportions of transitions were higher in the bone marrow than in colon regardless of FA diet, while the opposite was observed for transversions.

Base-substitutions, as a result of spontaneous somatic mutations, are the most common mutations found in solid tumors¹³⁷. These mutations accumulate throughout life

and can arise from intrinsic or extrinsic factors, and the variability in cancer incidence among tissues is attributable in part to exposure to environmental risks factors^{5,109}. Cancer results from one or many mutations which confer a selective advantage to a cell to proliferate and escape the built-in and exogenous restraints of a normal cell^{5,136}. Unique combinations of mutation types generated from different mutational processes are called mutation signatures. The Catalog of Somatic Mutations in Cancer (COSMIC) is a large database that describes over 4 million coding mutations, from whole-genome and exome sequencing results and curation of publications^{138,139}. The database contains information for each type of signature, including the signature profile, the proposed etiology for the mutational processes underlying a particular mutation type, and the type of cancers in which the signature has been found^{138,140}. We compared the mutation spectra from our results to the characteristic mutation types of CRC and ALL, including base substitution subtypes, in order to identify a possible etiology behind the mutations. Many of signatures have been observed in ALL and CRC cancer types. The single base substitution (SBS) mutational signatures found in both cancers were: SBS [1, 2, 5, 13, 18, 40].

The G:C→A:T transition was the predominant mutation observed in both tissues in this study. This transition is characteristic of SBS [1, 2, 5, 40]^{139,140}. The proposed etiologies of this mutation type are age and the spontaneous deamination of 5-methylcytosine and deregulation of AID/APOBEC^{139–141}. SBS [1, 5, 40] have been associated with age, and age has been associated with the accumulation of somatic mutations^{108,175}.

SBS2 is associated with the G:C→A:T transition and it is associated with the deregulation of AID/APOBEC^{139,140}. The AID/APOBEC enzyme family is involved in the DNA cytosine deamination (C→U), a cellular mechanism with the purpose of inhibiting

viruses and retrotransposons by disrupting their genome^{145,146}. APOBEC can also target the host's genome to create genome diversity to help with antibody actions and evolution. However, it can also disrupt genome stability and lead to cancer if the enzyme's activities are uncontrolled¹⁴⁷. The APOBEC enzymes are considered one of the most significant carcinogens¹³⁸, which could explain the consistent proportions of G:C→A:T mutations in the *lacZ* gene of both tissues. Its role as a potent mutagen might mask some of the mutagenic effects of FA. We could propose that this etiology is the driver of this type of transition in both tissues and that it might have happened independently of FA intake, since there are no differences in the proportion of the transition across FA diets. However, it would be of great interest to investigate the relationship between FA and APOBEC since the type of mutation resulting from its deregulation involves purines and also misincorporation of uracil into the DNA.

In addition, APOBEC is also the etiology suggested for G:C→C:G transversion (SBS13). Transitions, G:C→A:T, are preferential over the G:C→C:G transversions because the transitions are a direct result from DNA polymerase repair. DNA polymerases insert an adenine opposite to the deaminated cytosine, causing a C>A transition. In contrast, the G:C→C:G transversion results from the action of REV1, a deoxycytidyl transferase, required to incorporate a cytosine opposite a deaminated cytosine^{145,148}. In our results, the G:C→C:G transversion was observed across all diets in the colon, but in the bone marrow it was only observed in the deficient diet. In the colon the G:C→C:G transversion could still be driven by a deregulation of APOBEC. However, we could suggest that the G:C→C:G transversion observed in mice fed FA deficiency is FA-diet induced in the bone marrow. FA deficiency enhances the number of cytosine deaminated

sites by misincorporation of uracil in the DNA such that G:C→C:G transversions could result by REV1.

Production of reactive oxygen species (ROS) was the proposed etiology for G:C→T:A transversion, characteristic of both ALL and CRC^{139,140}. FA deficiency has been associated with the production of ROS as a result of accumulation of homocysteine^{149,150}. From our results, the G:C→T:A transversion was the only transversion observed in both tissues across all diets. Although not significant, the proportions of this transversions were different across diets in both tissues. In colon, the proportion of the G:C→T:A transversion was highest in the group fed the FA deficient diet where in bone marrow the highest proportion was observed in those fed the FA supplemented diet. Because the G:C→T:A transversion was present across all diets and in both tissues, we could suggest that FA-intake did not have an effect on this type of mutation. However, future studies could find other somatic tissues where FA-intake may have an effect, given that FA has been associated with the etiology driving this transversion.

While all the signatures found in ALL were also found in CRC, there are a few signatures that were only found in CRC. Signatures that were only found in CRC (SBS [15, 20, 26]) are associated with DNA mismatch repair (MM). DNA repair mechanism proteins serve to detect and correct DNA damage and to initiate the cellular response to such damage. MM repair was determined to be the prevalent DNA repair in the colon, which relies on MM to repair base substitution mismatches during replication¹⁴². In contrast, the double-strand break repair (DSBR) is the canonical repair mechanism in leukemia¹⁴². DSBs occur naturally in B and T cells, which might explain the importance of the DBSR in the bone marrow. Identification of tissue-specific DNA repair mechanisms further

supports the potential of extrinsic mutagen factors, such as FA, to have a tissue-specific mutagenicity. Therefore, future research could focus on the FA-intake effect on specific DNA repair mechanism in specific tissues. This could further our understanding into how FA affects specific tissues by altering intrinsic mechanism such as DNA repair, especially since folate is involved in *de novo* nucleotide synthesis.

4.6 Conclusion

In this study, we used the MutaMouse to demonstrate that folate affects tissues independently. First, we observed contrasting MF results in response to FA diets between colon and bone marrow. In the colon, the highest MF was observed in mice fed the FA supplemented diet in mice treated with either saline or ENU. In contrast, the highest MF in the bone marrow was observed in mice fed the FA deficient diet and only in the saline treated mice. We hypothesized that the increase in MF in the colon may be a consequence of the saturation of DHFR by FA supplementation ultimately disrupting the formation of THF and one-carbon metabolism. This saturation is not experienced in the bone marrow since it is exposed primarily to 5-methylTHF circulating in the plasma. We also propose that gut microbiota produce folate, which is immediately available for uptake by colonocytes, which could potentially compensate, perhaps transiently, for the FA deficient diet, therefore minimizing the disruption to one-carbon metabolism and induction of mutations. However, this over compensation might be localized to the colon and therefore have no beneficial effect on other more distant tissues such as the bone marrow. We also show that FA adequacy kept MF at its lowest in both tissues. These results emphasize the importance of adequate folate intake may be optimal for normal and healthy development

and that it is when intakes are outside of an optimal range that aberrant development can occur.

Second, we determined that the FA-induced mutation spectrum differs by FA intake and between tissues. We identified the etiology that has been proposed for certain signatures found in ALL and CRC and determined if those type of mutations were observed in the FA-induced mutational spectra. Proposed etiologies underlying some types of mutations observed in our study include age, the spontaneous deamination of 5-methylcytosine, deregulation of AID/APOBEC, ROS production and deactivation of DNA repair mechanisms.

Ultimately, we determined a FA-related mutation spectrum in two somatic tissues from an *in vivo* study using the male MutaMouse. We show the potential of FA to induce single-base substitutions *in vivo*. In addition, we showed that these base substitutions are also found in ALL and CRC, two cancers in which FA intake has been associated with risk. Further *in vivo* studies will be necessary to evaluate and determine the mutation signature of FA and to determine its role and effect in intrinsic cell factors that may result in cellular transformation in CRC and ALL.

Chapter 5: Discussion

5.1 Context for the study

Folate deficiency has been associated with increased risk for gastro-intestinal cancers and has a causal role in the development of NTDs and megaloblastic anemia. FA supplementation has been associated with dual effects in CRC. We proposed that FA intake has an effect in somatic mutations of the male MutaMouse. Therefore, the purpose of this study was to investigate the effect of low, adequate, or high dietary FA intake in the colon of the male mice. Additionally, we sought to identify the folate- and tissue-specific mutation profiles of colon and bone marrow.

5.1.1 Mutagenic potential of folate

Folate has an important role in ensuring normal development, cell growth and maintenance of overall health. Therefore, disruption of any of the three folate-dependent biosynthetic pathways can have serious implications in a cell's survival, ultimately affecting organs and tissues as a whole. Impairments in *de novo* purine synthesis can result in a reduction of DNA replication and abnormal DNA repair^{13,14}. Insufficient dTMP synthesis may cause a nucleotide imbalance which results in the accumulation of uracil and its misincorporation into DNA, leading an increase in DNA double strand breaks due to the “futile cycle” of DNA repair^{15,16}. A decline in the production of AdoMet reduces the availability of methyl groups and cellular methylation capacity, thus altering methylation patterns across the genome. Hypomethylation of the genome can result in genome instability, disrupt gene expression, and alter risk for diseases such as cancer^{9,20}.

5.1.2 FA and CRC

Colon cancer is within the top three most common types of cancer in North America¹⁰⁰. It has been established that environmental and inherited factors play an important role in the risk of CRC¹⁰⁹. The colon is a highly proliferative tissue in which both folate deficiency and supplementation has been associated with a dual effect towards CRC. The protective or promoting effect folate on CRC risk likely depends on the folate status and the normal, versus cancerous, state of the tissue. Here we show that colon in the MutaMouse strain is susceptible to FA-induced DNA mutations, specifically that a FA supplemented diet resulted in a higher colon MF than those fed a control or deficient diet. Given that higher rates of mutation are associated with risk for cellular transformation, this suggests that, at least under the conditions this study, FA supplementation may promote the development of cancer. This was contrary to what we expected.

Folate deficiency has been associated with increased DNA instability and malignant transformation by disrupting the one-carbon metabolism. Folate deficiency can induce DBSs, nucleotide pool imbalance, impaired DNA repair and genome hypomethylation^{1,16,129,130}. These are candidate causal mechanisms of how folate deficiency can induce mutations. Folate deficiency mutagenicity potential has been determined *in vitro* where low folate concentrations induced hypomethylation in human-derived colon carcinoma cells and increase DNA strand breakage in human colonocytes^{59,114}. In rats, folate deficiency has been shown to induce DNA strand breaks and hypomethylation in the *Tp53* gene, a tumor suppressor gene and mutations in this gene are found in ~50% of CRC cases^{176,177}.

Moreover, our laboratory had previously demonstrated the genotoxicity of folate deficiency in mice where mice fed the FA deficient diet had significantly higher *Pig-a* MF in red blood cells when compared to those fed the FA supplemented diet ¹⁷⁸. More importantly, a previous study using tissues from the same mice used in this study, found a higher *lacZ* MF in bone marrow and micronuclei in red blood cells in mice fed the same FA deficient diet compared to those fed control and supplemented diets ⁸⁸.

Therefore, we expected mice fed the FA deficient diet to have a similarly higher MF in the colon than those in the control and supplemented diets. However, here we show that mice fed the FA deficient diet did not show an increase MF in the colon. The lack of effect of folate deficiency may potentially be explained by the presence of folate producing bacteria in the large intestinal tract ^{131–135}. Its local uptake by the colon may, at least temporarily, protect the folate status of the tissue and the effect of deficiency on folate-mediated nucleotide synthesis

Rather than having a null or beneficial effect, we observed that the FA supplemented diet increased the colon MF by 1.4-fold compared to mice fed the FA control and deficient diets among saline treated mice. Similarly, among the ENU treated mice, those fed the FA supplemented diet had a 1.2-fold higher MF than mice fed the FA deficient diet (Fig. 3.1 B). Some studies have suggested that high FA intake results in saturation of DHFR, thus decreasing the availability of intracellular folate cofactors, ultimately causing functional folate deficiency and disruption of one-carbon metabolism ¹²⁴. Other studies have suggested that high concentrations of FA leads to an increase in intracellular DHF concentrations, which has been shown to inhibit thymidylate synthase, AICAR transformylase and MTHFR, thus inhibiting the *de novo* synthesis of dTMP and purines,

and decrease methionine synthesis¹²⁵⁻¹²⁸. Therefore, high levels of FA could result in DNA instability and mutations, which could explain our results. Taken together these results suggest a tissue-specific action of FA.

Many studies have focused on physiological, cellular and methylation effects of FA intake in the colon. However, the underlying mechanism of mutation induction and the type of mutations induced by folate in the tissue remain unclear. Here we present, for the first time, the FA-induced mutation spectra in the colon of the male MutaMouse. While the majority of mutations in the colon were base substitutions regardless of diet, the mutation spectra varied, although not significantly, by FA-intake level suggesting there may be potentially unique mechanisms of action leading to mutagenesis depending on the level of intake. We also found that the mutation spectra observed in the colon were characteristic of CRC.

Mutation patterns for CRC, referred to as mutation signatures, were obtained from the COSMIC database. Each signature is a unique combination of mutation types generated ascribed to a particular mutational process^{139,140}. By comparing the FA-induced mutation types with those identified in CRC, we identified possible processes which may interact with FA to induce the mutation spectra seen in our results. We identified spontaneous deamination of 5-methyl-cytosine, activation of AID/APOBEC enzymes, and defective DNA mismatch repair (MM) as candidate etiologies for G:C→A:T transitions^{139,140}, which were the predominant mutation among the diets. The APOBEC enzymes are considered one of the most significant carcinogens (Alexandrov and Stratton, 2014) due to their inherent function to induce mutations by cytosine deamination. This enzymatic action can result in G:C→A:T transitions and G:C→C:G transversions, though transitions are predominant. In

the colon both of these types of mutations (G:C→A:T and G:C→C:G) were observed in all diets.

MM repair was another proposed etiology for G:C→T:A transversion, the most common transversion observed in all diet groups^{139,140}. MM repair defects have been shown to cause genome instability and to deregulate DNA damage-induced apoptosis¹⁷⁹. Defects in MM repair are known to cause microsatellite instability which has been associated with low folate concentrations; while folate supplementation has been shown to stabilize microsatellite sequences^{179,180}. MM repair is also required for error-free replication of expanded simple tandem repeats (ESTRs). Higher ESTR MF was observed in male mice weaned to a folic acid deficient diet^{181,182}. Taken together, these suggest that intracellular concentrations of folate could regulate the MM repair function. In addition, MM has also been shown to initiate apoptosis in response to folate deficiency-induced damage^{179,183}. Therefore, given previous findings we would expect that mutations associated with MM repair would have higher proportions in mice fed a FA deficient diet and that perhaps they would also be higher in those fed the FA supplemented diet since high concentrations of FA could induce functional folate deficiency in the colon, as mentioned previously. However, our results do not show a significant difference in the proportion mutation types which have been associated with MM repair (G:C→A:T and G:C→T:A).

G:C→T:A transversions can also result from damage caused by ROS^{139,140} and folate deficiency has been associated with the production of ROS^{149,150}. Although not significant, the proportion of G:C→T:A transversions in our results was higher in the group

fed the FA deficient diet compared to the control diet and FA control and supplemented diets.

Taken together we show the ability of FA supplementation to induce mutations *in vivo* in the colon. In addition, we showed that the types of mutations we observed in the colon are characteristic of CRC. Moreover, some of the etiologies proposed to induce such mutations have been associated with folate status, suggesting a connection between FA-intake, the mutagenic processes and CRC. Further research is required to determine if and how FA interacts with these intrinsic cellular mechanisms that cause these types of mutations. Additionally, since the FA-induced mutation spectra were similar across diets but the MF was significantly higher in the FA supplemented diet, we suggest that it is the rate of mutations rather than the type of mutation that is the factor which may influence the risk of CRC.

5.1.3 FA-induced mutation spectra and somatic tissue specificity

Variation in cancer incidence among tissues has been attributable to the tissue-specific mutagenicity of various intrinsic and extrinsic mutation-inducing factors (e.g., Rate of cell division; exposure to environmental factors) ^{108,109,142}. Folate is an exogenous factor that has been associated with some diseases and cancers, including CRC and ALL. Folate deficiency has been associated with an increase of DNA instability in human colonocytes and lymphocytes ^{59,60,66}. Folate deficiency has been shown to increase RBC micronucleus frequency, a measure of chromosomal damage, in mice ^{88,178}. FA supplementation has shown to decrease micronucleus frequency in the bone marrow of mice ⁷⁷ and to reduce the evolution of macroscopic neoplasms in the colorectum of rats ¹⁸⁴. However, in the colon, FA supplementation has also been shown to increase the number

neoplastic foci in cases where a neoplastic foci has been established ⁷⁹. Therefore, it is of interest to determine the effect of FA-intake in tissues affected by these cancers and to find if this effect of FA-intake is tissue-specific.

First, we compared the MF from the colon of the male MutaMouse with that obtained from the bone marrow of the same mice reported in a previous study ⁸⁸. We demonstrate that the FA-induced DNA mutations are tissue specific where in bone marrow FA deficiency results in higher MF and in colon FA supplementation results in higher MF. These contrasting results may be explained by inherent differences between the tissues such as cell proliferation, kinetic turnover of folate and the type of folate to which the tissues are exposed. Bone marrow has a higher kinetic turnover and cell renewal rate than colon ¹⁶⁷⁻¹⁶⁹, which could explain why it is more susceptible to a folate deficient diet. The bone marrow would have a higher folate requirement and under deficient conditions can potentially become depleted more readily than colon. In the same manner, this high folate requirement would also explain why FA supplementation did not result in high MF in bone marrow. Moreover, dietary folate, both naturally occurring folate and FA is absorbed in the intestinal epithelium. In our study, the vast majority of folate in the diet is in the form of FA, thus colon of our mice is theoretically directly exposed to FA where it first needs to be reduced to its bioactive form to enter the one-carbon metabolism ^{3,4}. On the other hand, bone marrow would be exposed to 5-methylTHF, the predominant form of folate in circulation, which can enter the one-carbon metabolism directly. Because FA has to be sequentially reduced by DHFR to DHF and then THF in the colon, we propose that the saturation of DHFR in the presence of high intracellular FA concentrations resulted in the observed higher MF in the colon in mice fed the FA supplemented diet.

We also compared the FA-induced mutation spectra from the bone marrow and colon. We found that the FA-induced mutation spectra were tissue-specific. Bone marrow had a higher proportion of transitions than colon regardless of FA diet, while the opposite was observed for transversions. While the proportion of indels in the bone marrow was higher in mice fed the FA supplemented diet and lower in mice fed the FA control and deficient diet when compared to those seen in the colon (Fig. 4.2).

We then compared the type of FA-induced mutations obtained from bone marrow and colon with mutations that have been found to be characteristic of ALL and CRC from in the COSMIC database. The single base substitution (SBS) mutational signatures found in both cancers were: SBS [1, 2, 5, 13, 18, 40].

The G:C→A:T transition was characteristic of SBS [1, 2, 5, 40]. Age, spontaneous deamination of 5-methyl-cytosine and deregulation of AID/APOBEC were the proposed etiologies for this type of mutation ^{139,140}. The proposed etiology of SBS [1, 5, 40] is age, which has been associated with accumulation of spontaneous somatic mutations in normal and adult stem cells ^{108,175}. While this transition was observed in both tissues in all diet groups, it was more prevalent in bone marrow.

AID/APOBEC dysregulation has also been proposed as the etiology for the G:C→A:T transition in both cancers (SBS2) ^{138,147}. As previously mentioned, APOBEC enzyme family involve in DNA editing by cytosine deamination (C→U) and it is considered a potent mutagen and significant carcinogen. Because G:C→A:T transition was observed at a similar proportion across all diets in colon and bone marrow we could propose that deregulation of APOBEC, causing its overactivity, is the main cause of this type of mutation.

In addition, APOBEC is also the etiology proposed to underlie the G:C→C:G transversion (SBS13) in both cancers. Cytosine deamination will, in most cases, result in a C→T transition (G:C→A:T) because most polymerases will insert an adenine opposite to an abasic site¹⁴⁵. However, low fidelity enzymes are also recruited at DNA damage sites. REV1 is a low fidelity enzyme which can take the place of DNA polymerase temporarily and will uniquely insert a cytosine opposite to an abasic site, which results in a C→G transversion (G:C→C:G)^{145,148}. In our results, the G:C→C:G transversion was observed across all diets in the colon. Therefore, in the colon, the capacity of FA-intake to induce the G:C→C:G transversion, could still be masked by intrinsic mutagenic factors such as deregulation of APOBEC. However, in the bone marrow, the G:C→C:G transversion was only observed in those fed the FA deficient diet. Folate deficiency has been shown to increase uracil in DNA of bone marrow cells^{16,185}. We could propose that FA deficiency increases the number of G:C→C:G transversions by increasing the number of sites with uracil in the DNA by misincorporation, thus increasing the sites where REV1 would be recruited and causing the transversion.

Production of ROS was the proposed etiology for G:C→T:A transversions observed in ALL and CRC and ROS has been associated with folate deficiency^{149,150}. From our results, the G:C→T:A transversion was the only transversion observed in both tissues and in all diet groups.

Whereas all the signatures found in ALL were also found in CRC, there are a few signatures that were only found in CRC. All of these signatures (SBS [15, 20, 26]) are associated with DNA mismatch repair (MM). DNA repair mechanisms serve to detect and correct DNA damage and to initiate the cellular response to such damage. MM repair is the

predominant DNA repair mechanism in the colon ¹⁴² and defects in MM are important in the initiation and progression of CRC ¹⁸⁶. MM deficiency is the etiology for G:C→T:A transversions (SBS20) and this transversion is the dominant type of mutation in the *Tp53* gene in CRC ¹⁸⁷. However, in our results we did not observe the effect of FA-intake on this type of mutation. Yet, we cannot completely discard the possibility of FA having some effect on this type of transition. Because mutations in the *Tp53* gene are found in ~50% of the cases of CRC and the G:C→T:A transversions is the most frequent type of mutation in the *Tp53* gene in CRC ^{176,187}, future studies could focus on the effect of FA-intake in the *Tp53* gene.

The mutation spectra of spontaneous *lacZ* gene mutations in a few tissues of the MutaMouse have been identified ^{141,188}. The predominant spontaneous base substitution in bone marrow was G:C→A:T transitions (~71%). The G:C→T:A transversion was the only transversion observed (~14%) ¹⁴¹. When compared to our results, we can propose that the transversions observed in bone marrow in mice fed the deficient diet were induced by an extrinsic factor, in this case FA-intake, because it was only in this diet where transversions other than G:C→T:A (the only transversion common to all diets and both tissues) were observed. Moreover, the presence of these transversions may have driven the increased MF seen in mice fed the FA deficient diet in the bone marrow.

Ultimately, we have shown that FA has the capacity to induced mutagenicity in tissues by increasing MF and that such mutagenicity is dependent on the intake level in different somatic tissues. Moreover, we have identified the FA-induced mutation spectra, which are also tissue-specific. Taken together, we propose that the high MF in bone marrow in mice fed the FA deficient diet is driven by transitions. In contrast, the high MF

observed in the colon of mice fed the FA supplemented diet is a result of enhancing the mutations inherent to the colon. Based on the mutation spectra, we have also identified possible mechanisms by which FA may impact mutagenesis including DNA repair mechanisms, such as MM, and intrinsic DNA mutation drivers, such as APOBEC.

5.2 Study strengths and limitations

One of the main strengths of this study was our ability to observe the effects of low, adequate and high FA intake *in vivo* in all tissues simultaneously. For the first time, we have identified unique FA-induced mutation spectra which informs the potential mechanism(s) of action of FA as it relates to mutagenicity, and how it differs in different tissue. These studies will guide further research in this area. For example, additional studies will need to be performed to confirm our findings in other mouse strains and tissues, and to determine whether they translate to cellular transformation and human cancer.

The use of the MutaMouse model and the *lacZ* gene to identify the FA-induced mutation spectra was both a weakness and a strength. While results observed in the *lacZ* gene are considered a very close representation of what would be observed across the genome, the *lacZ* gene is still an exogenous gene, and one that is not transcribed in the mouse⁸⁵. Therefore, any interactions that FA may have with enzymes and pathways involved in gene transcription would not have been observed in this study. However, the spontaneous mutant frequency in the *lacZ* gene in most somatic tissues of the MutaMouse is 5- to 10-fold higher than that observed in endogenous loci⁸⁵, making it an ideal model to test the mutagenic potential of comparably weak mutagens, such as FA.

The methods for detecting mutations used in this study could be further improved to increase the power of detection. Future studies should collect higher numbers of mutant

plaques for each sample and sequence the same number of mutant plaques for all samples for more consistency.

5.3 Future directions

The results of this study show FA has a tissue-specific effect on MF and the type of mutations it induces. This is a starting point for further research examining the mechanism(s) of action of FA, including its interaction with DNA repair mechanisms and endogenous cell enzymes that may increase mutagenicity. Also, future research could focus on identifying the FA-induced mutation spectra in other somatic tissues as well as the germline. Moreover, FA-induced mutation spectra should be identified in genes that are transcribed and typically mutated in cancers to determine the different effects, if any, of FA on transcribed and non-transcribed genomic sequences.

5.4 Conclusions

In conclusion, we have shown that FA intake has a mutagenic effect on somatic tissues that is tissue-specific and FA-intake level dependent. We determined the FA-induced mutation spectra for two somatic tissues, the colon and bone marrow, and we proposed possible mechanisms that could explain the contrasting results in MF between both tissues. Since FA intake has been associated with ALL and CRC, cancers from the tissues we studied, we proposed possible associations between FA and the etiologies for such cancers which will require further research. Our findings represent the beginning of a new focus in research involving the *in vivo* mutagenic properties of vitamins and other nutritional factors.

Ultimately, we have shown the carcinogenic potential of FA by inducing genomic mutations, which is a necessary preliminary step in carcinogenesis. We have also shown

that this potential is tissue-specific and depends on FA-intake at both the low and high intake levels, and further highlights the importance of achieving folate adequacy.

References

1. Stover, P. J. Physiology of Folate and Vitamin B 12 in Health and Disease. *Nutr. Rev.* **62**, S3–S12 (2004).
2. Shafizadeh, T. B. & Halsted, C. H. γ -Glutamyl Hydrolase, Not Glutamate Carboxypeptidase II, Hydrolyzes Dietary Folate in Rat Small Intestine. *J. Nutr.* **137**, 1149–1153 (2007).
3. Zhao, R., Matherly, L. H. & Goldman, I. D. Membrane transporters and folate homeostasis: intestinal absorption and transport into systemic compartments and tissues. *Expert Rev. Mol. Med.* **11**, (2009).
4. Stover, P. J. & Field, M. S. Trafficking of Intracellular Folates. *Adv. Nutr.* **2**, 325–331 (2011).
5. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489 (2015).
6. Beaudin, A. E. & Stover, P. J. Folate-mediated one-carbon metabolism and neural tube defects: Balancing genome synthesis and gene expression. *Birth Defects Res. Part C Embryo Today Rev.* **81**, 183–203 (2007).
7. An, S., Kumar, R., Sheets, E. D. & Benkovic, S. J. Reversible Compartmentalization of de Novo Purine Biosynthetic Complexes in Living Cells. *Science* **320**, 103–106 (2008).
8. Woeller, C. F., Anderson, D. D., Szebenyi, D. M. E. & Stover, P. J. Evidence for Small Ubiquitin-like Modifier-dependent Nuclear Import of the Thymidylate Biosynthesis Pathway. *J. Biol. Chem.* **282**, 17623–17631 (2007).

9. Duthie, S. J. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J. Inherit. Metab. Dis.* **34**, 101–109 (2011).
10. Matherly, L. H. & Goldman, D. Membrane transport of folates. *Vitam. Horm.* **66**, 405–457 (2003).
11. Selhub, J., Emmanouel, D., Stavropoulos, T. & Arnold, R. Renal folate absorption and the kidney folate binding protein. I. Urinary clearance studies. *Am. J. Physiol.-Ren. Physiol.* **252**, F750–F756 (1987).
12. Selhub, J., Nakamura, S. & Carone, F. A. Renal folate absorption and the kidney folate binding protein. II. Microinfusion studies. *Am. J. Physiol.-Ren. Physiol.* **252**, F757–F760 (1987).
13. Collins, A. R., Black, D. T. & Waldren, C. A. Aberrant DNA repair and enhanced mutagenesis following mutagen treatment of Chinese hamster Ade–C cells in a state of purine deprivation. *Mutat. Res. Repair Rep.* **193**, 145–155 (1988).
14. Kondo, M. *et al.* The Rate of Cell Growth Is Regulated by Purine Biosynthesis via ATP Production and G1 to S Phase Transition. *J. Biochem. (Tokyo)* **128**, 57–64 (2000).
15. Berger, S. H., Pittman, D. L. & Wyatt, M. D. Uracil in DNA: Consequences for carcinogenesis and chemotherapy. *Biochem. Pharmacol.* **76**, 697–706 (2008).
16. 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.* **94**, 3290–3295 (1997).
17. Bailey, L. B. *et al.* Biomarkers of Nutrition for Development—Folate Review. *J. Nutr.* **145**, 1636S–1680S (2015).

18. Basten, G. P. *et al.* Sensitivity of markers of DNA stability and DNA repair activity to folate supplementation in healthy volunteers. *Br. J. Cancer* **94**, 1942 (2006).
19. Linhart, H. G. *et al.* Folate Deficiency Induces Genomic Uracil Misincorporation and Hypomethylation But Does Not Increase DNA Point Mutations. *Gastroenterology* **136**, 227-235.e3 (2009).
20. Crider, K. S., Yang, T. P., Berry, R. J. & Bailey, L. B. Folate and DNA Methylation: A Review of Molecular Mechanisms and the Evidence for Folate's Role. *Adv. Nutr.* **3**, 21–38 (2012).
21. Duthie, S. J. Epigenetic modifications and human pathologies: cancer and CVD. *Proc. Nutr. Soc.* **70**, 47–56 (2011).
22. Jones, P. A. & Baylin, S. B. The Epigenomics of Cancer. *Cell* **128**, 683–692 (2007).
23. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
24. Gardiner-Garden, M. & Frommer, M. CpG Islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).
25. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41**, 178–186 (2009).
26. Miranda, T. B. & Jones, P. A. DNA methylation: The nuts and bolts of repression. *J. Cell. Physiol.* **213**, 384–390 (2007).
27. Hornstra, I. K. & Yang, T. P. High-resolution methylation analysis of the human hypoxanthine phosphoribosyltransferase gene 5' region on the active and inactive X

- chromosomes: correlation with binding sites for transcription factors. *Mol. Cell. Biol.* **14**, 1419–1430 (1994).
28. Payer, B. & Lee, J. T. X Chromosome Dosage Compensation: How Mammals Keep the Balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
 29. Guilleret, I. *et al.* Hypermethylation of the human telomerase catalytic subunit (hTERT) gene correlates with telomerase activity. *Int. J. Cancer* **101**, 335–341 (2002).
 30. Yadav, M. K., Manoli, N. M., Vimalraj, S. & Madhunapantula, S. V. Unmethylated promoter DNA correlates with p53 expression and apoptotic levels only in Vitamin B9 and B12 deficient megaloblastic anemia but not in non-megaloblastic anemia controls. *Int. J. Biol. Macromol.* **109**, 76–84 (2018).
 31. Klerk, M. *et al.* MTHFR 677C→T Polymorphism and Risk of Coronary Heart Disease: A Meta-analysis. *JAMA* **288**, 2023–2031 (2002).
 32. Friso, S. *et al.* A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. *Proc. Natl. Acad. Sci.* **99**, 5606–5611 (2002).
 33. James, S. J. *et al.* Abnormal folate metabolism and mutation in the methylenetetrahydrofolate reductase gene may be maternal risk factors for Down syndrome. *Am. J. Clin. Nutr.* **70**, 495–501 (1999).
 34. Botto, L. D. & Yang, Q. 5, 10-Methylenetetrahydrofolate Reductase Gene Variants and Congenital Anomalies: A HuGE Review. *Am. J. Epidemiol.* **151**, 862–877 (2000).

35. Chittiboyina, S., Chen, Z., Chiorean, E. G., Kamendulis, L. M. & Hocevar, B. A. The role of the folate pathway in pancreatic cancer risk. *PLoS ONE* **13**, (2018).
36. Canada, H. Dietary Reference Intakes. *aem* (2005). Available at: <https://www.canada.ca/en/health-canada/services/food-nutrition/healthy-eating/dietary-reference-intakes/tables/reference-values-vitamins-dietary-reference-intakes-tables-2005.html>. (Accessed: 1st June 2019)
37. Institute of Medicine (US) Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and its Panel on Folate, Other B Vitamins, and Choline. *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B6, Folate, Vitamin B12, Pantothenic Acid, Biotin, and Choline*. (National Academies Press (US), 1998).
38. Yates, A. A., Schlicker, S. A. & Suitor, C. W. Dietary Reference Intakes: The New Basis for Recommendations for Calcium and Related Nutrients, B Vitamins, and Choline. *J. Am. Diet. Assoc.* **98**, 699–706 (1998).
39. Bailey, L. B. *Folate in Health and Disease*. (CRC Press, 2017).
40. Halsted, C. H., Medici, V. & Esfandiari, F. Influence of alcohol on folate status and methionine metabolism in relation to alcoholic liver disease. *Folate Health Dis. Second Ed.* 429–448 (2009).
41. Hillman, R. S. & Steinberg, S. E. The Effects of Alcohol on Folate Metabolism. *Annu. Rev. Med.* **33**, 345–354 (1982).
42. Aslinia, F., Mazza, J. J. & Yale, S. H. Megaloblastic Anemia and Other Causes of Macrocytosis. *Clin. Med. Res.* **4**, 236–241 (2006).

43. Pitkin, R. M. Folate and neural tube defects. *Am. J. Clin. Nutr.* **85**, 285S-288S (2007).
44. Hibbard, B. M. The role of folic acid in pregnancy*. *BJOG Int. J. Obstet. Gynaecol.* **71**, 529–542 (1964).
45. Smithells, R. W., Sheppard, S. & Schorah, C. J. Vitamin deficiencies and neural tube defects. *Arch. Dis. Child.* **51**, 944–950 (1976).
46. Bower, C. & Stanley, F. J. Dietary folate as a risk factor for neural-tube defects: evidence from a case-control study in Western Australia. *Med. J. Aust.* **150**, 613–619 (1989).
47. MRC, V. S. R. G. Prevention of neural tube defects: Results of the Medical Research Council Vitamin Study. *The Lancet* **338**, 131–137 (1991).
48. Czeizel, A. E. & Dudás, I. Prevention of the First Occurrence of Neural-Tube Defects by Periconceptional Vitamin Supplementation. *N. Engl. J. Med.* **327**, 1832–1835 (1992).
49. De Wals, P. *et al.* Reduction in Neural-Tube Defects after Folic Acid Fortification in Canada. *N. Engl. J. Med.* **357**, 135–142 (2007).
50. Mills, J. L. & Signore, C. Neural tube defect rates before and after food fortification with folic acid. *Birt. Defects Res. A. Clin. Mol. Teratol.* **70**, 844–845 (2004).
51. Shakur, Y. A., Garriguet, D., Corey, P. & O'Connor, D. L. Folic acid fortification above mandated levels results in a low prevalence of folate inadequacy among Canadians. *Am. J. Clin. Nutr.* **92**, 818–825 (2010).
52. Friedberg, E. C., Walker, G. C., Siede, W. & Wood, R. D. *DNA Repair and Mutagenesis*. (American Society for Microbiology Press, 2005).

53. Lashner, B. A. Red blood cell folate is associated with the development of dysplasia and cancer in ulcerative colitis. *J. Cancer Res. Clin. Oncol.* **119**, 549–554 (1993).
54. James, S. J., Cross, D. R. & Miller, B. J. Alterations in nucleotide pools in rats fed diets deficient in choline, methionine and/or folic acid. *Carcinogenesis* **13**, 2471–2474 (1992).
55. Pogribny, I. P., James, S. J., Jernigan, S. & Pogribna, M. Genomic hypomethylation is specific for preneoplastic liver in folate/methyl deficient rats and does not occur in non-target tissues. *Mutat. Res. Mol. Mech. Mutagen.* **548**, 53–59 (2004).
56. Kruman, I. I. *et al.* Folic Acid Deficiency and Homocysteine Impair DNA Repair in Hippocampal Neurons and Sensitize Them to Amyloid Toxicity in Experimental Models of Alzheimer's Disease. *J. Neurosci.* **22**, 1752–1762 (2002).
57. Kim, Y.-I. Folate and carcinogenesis: evidence, mechanisms, and implications. *J. Nutr. Biochem.* **10**, 66–88 (1999).
58. Gibson, T. M. *et al.* Pre- and postfortification intake of folate and risk of colorectal cancer in a large prospective cohort study in the United States. *Am. J. Clin. Nutr.* **94**, 1053–1062 (2011).
59. Duthie, S. J., Narayanan, S., Blum, S., Pirie, L. & Brand, G. M. Folate Deficiency In Vitro Induces Uracil Misincorporation and DNA Hypomethylation and Inhibits DNA Excision Repair in Immortalized Normal Human Colon Epithelial Cells. *Nutr. Cancer* **37**, 245–251 (2000).
60. Duthie, S. J. *et al.* The Response of Human Colonocytes to Folate Deficiency in Vitro: Functional and Proteomic Analyses. *J. Proteome Res.* **7**, 3254–3266 (2008).

61. Snell, K., Natsumeda, Y., Eble, J. N., Glover, J. L. & Weber, G. Enzymic imbalance in serine metabolism in human colon carcinoma and rat sarcoma. *Br. J. Cancer* **57**, 87 (1988).
62. MacFarlane, A. J., Perry, C. A., McEntee, M. F., Lin, D. M. & Stover, P. J. Shmt1 Heterozygosity Impairs Folate-Dependent Thymidylate Synthesis Capacity and Modifies Risk of Apcmin-Mediated Intestinal Cancer Risk. *Cancer Res.* **71**, 2098–2107 (2011).
63. Huang, Y., Han, S., Li, Y., Mao, Y. & Xie, Y. Different roles of MTHFR C677T and A1298C polymorphisms in colorectal adenoma and colorectal cancer: a meta-analysis. *J. Hum. Genet.* **52**, 73–85 (2007).
64. Keku, T. *et al.* 5,10-Methylenetetrahydrofolate Reductase Codon 677 and 1298 Polymorphisms and Colon Cancer in African Americans and Whites. *Cancer Epidemiol. Prev. Biomark.* **11**, 1611–1621 (2002).
65. Roy Moulik, N., Kumar, A., Agrawal, S. & Mahdi, A. A. Folate deficiency in north Indian children undergoing maintenance chemotherapy for acute lymphoblastic leukemia—Implications and outcome. *Pediatr. Blood Cancer* **65**, n/a-n/a (2018).
66. Duthie, S. J. & Hawdon, A. DNA instability (strand breakage, uracil misincorporation, and defective repair) is increased by folic acid depletion in human lymphocytes in vitro. *FASEB J.* **12**, 1491–1497 (1998).
67. Skibola, C. F. *et al.* Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia. *Blood* **99**, 3786–3791 (2002).

68. Hishida, A. *et al.* Associations between polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and susceptibility to malignant lymphoma. *Haematologica* **88**, 159–166 (2003).
69. Skibola, C. F. *et al.* Polymorphisms in the methylenetetrahydrofolate reductase gene are associated with susceptibility to acute leukemia in adults. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12810–12815 (1999).
70. Dufficy, L. *et al.* G80A reduced folate carrier SNP influences the absorption and cellular translocation of dietary folate and its association with blood pressure in an elderly population. *Life Sci.* **79**, 957–966 (2006).
71. Jonge, R. de *et al.* Polymorphisms in folate-related genes and risk of pediatric acute lymphoblastic leukemia. *Blood* **113**, 2284–2289 (2009).
72. Mason, J. B. Folate, cancer risk, and the greek god, Proteus: a tale of two chameleons. *Nutr. Rev.* **67**, 206–212 (2009).
73. Oaks, B. M. *et al.* Folate intake, post-folic acid grain fortification, and pancreatic cancer risk in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Am. J. Clin. Nutr.* **91**, 449–455 (2010).
74. Giovannucci, E. *et al.* Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. *Ann. Intern. Med.* **129**, 517–524 (1998).
75. Lashner, B. A., Heidenreich, P. A., Su, G. L., Kane, S. V. & Hanauer, S. B. Effect of folate supplementation on the incidence of dysplasia and cancer in chronic ulcerative colitis. A case-control study. *Gastroenterology* **97**, 255–259 (1989).
76. Huang, R.-F. S., Yaong, H.-C., Chen, S.-C. & Lu, Y.-F. In vitro folate supplementation alleviates oxidative stress, mitochondria-associated death

- signalling and apoptosis induced by 7-ketocholesterol. *Br. J. Nutr.* **92**, 887–894 (2004).
77. Zhang, R., Wu, K., Zhan, C., Liu, X. & Gong, Z. Folic Acid Supplementation Reduces the Mutagenicity and Genotoxicity Caused by Benzo(a)pyrene. *J. Nutr. Sci. Vitaminol. (Tokyo)* **62**, 26–31 (2016).
78. Cole, B. F. *et al.* Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. *JAMA* **297**, 2351–2359 (2007).
79. Song, J., Medline, A., Mason, J. B., Gallinger, S. & Kim, Y. I. Effects of dietary folate on intestinal tumorigenesis in the *apcMin* mouse. *Cancer Res.* **60**, 5434–5440 (2000).
80. Stolzenberg-Solomon, R. Z. *et al.* Folate intake, alcohol use, and postmenopausal breast cancer risk in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Am. J. Clin. Nutr.* **83**, 895–904 (2006).
81. Kim, S. J. *et al.* Plasma folate, vitamin B-6, and vitamin B-12 and breast cancer risk in BRCA1- and BRCA2-mutation carriers: a prospective study. *Am. J. Clin. Nutr.* **104**, 671–677 (2016).
82. Gossen, J. A. *et al.* Efficient rescue of integrated shuttle vectors from transgenic mice: a model for studying mutations in vivo. *Proc. Natl. Acad. Sci.* **86**, 7971–7975 (1989).
83. Shwed, P. S., Crosthwait, J., Douglas, G. R. & Seligy, V. L. Characterisation of MutaTMMouse λ gt10-lacZ transgene: evidence for in vivo rearrangements. *Mutagenesis* **25**, 609–616 (2010).

84. Beal, M. A., Gagné, R., Williams, A., Marchetti, F. & Yauk, C. L. Characterizing Benzo[a]pyrene-induced lacZ mutation spectrum in transgenic mice using next-generation sequencing. *BMC Genomics* **16**, 812 (2015).
85. Lambert, I. B., Singer, T. M., Boucher, S. E. & Douglas, G. R. DETAILED REVIEW PAPER ON TRANSGENIC RODENT MUTATION ASSAYS. 20 (2008).
86. OECD. *OECD Guideline for the testing of chemicals: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays*. (2011).
87. Canadian Council on Animal Care. *Guide to the care and use of experimental animals. Vol. 1. Vol. 1*. (Canadian Council on Animal Care, 1993).
88. LeBlanc, D. P., Behan, N. A., O'Brien, J. M., Marchetti, F. & MacFarlane, A. J. Folate deficiency increases chromosomal damage and mutations in hematopoietic cells in the transgenic mutamouse model. *Environ. Mol. Mutagen.* **59**, 366–374 (2018).
89. Reeves, P. G., Nielsen, F. H. & Fahey, G. C. AIN-93 Purified Diets for Laboratory Rodents: Final Report of the American Institute of Nutrition Ad Hoc Writing Committee on the Reformulation of the AIN-76A Rodent Diet. *J. Nutr.* **123**, 1939–1951 (1993).
90. Klipstein, F. A. & Lipton, S. D. Intestinal Flora of Folate-Deficient Mice. *Am. J. Clin. Nutr.* **23**, 132–140 (1970).
91. Klipstein, F. A. & Samloff, I. M. Folate synthesis by intestinal bacteria. *Am. J. Clin. Nutr.* **19**, 237–246 (1966).

92. Trentin, G. A., Moody, J. & Heddle, J. A. Effect of maternal folate levels on somatic mutation frequency in the developing colon. *Mutat. Res. Mol. Mech. Mutagen.* **405**, 81–87 (1998).
93. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
94. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
95. Molloy, A. M. & Scott, J. M. Microbiological assay for serum, plasma, and red cell folate using cryopreserved, microtiter plate method. *Methods Enzymol.* **281**, 43–53 (1997).
96. Brevik, A. *et al.* Plasma concentration of folate as a biomarker for the intake of fruit and vegetables: the Hordaland Homocysteine Study. *Am. J. Clin. Nutr.* **81**, 434–439 (2005).
97. Gingerich, J. D., Soper, L., Lemieux, C. L., Marchetti, F. & Douglas, G. R. Transgenic Rodent Gene Mutation Assay in Somatic Tissues. in *Genotoxicity and DNA Repair: A Practical Approach* (eds. Sierra, L. M. & Gaivão, I.) 305–321 (Springer New York, 2014). doi:10.1007/978-1-4939-1068-7_18
98. Canadian Cancer Statistics. Canadian Cancer Statistics 2018. (2018). Available at: cancer.ca/Canadian-Cancer-Statistics-2018-EN.
99. World Health Organization. Cancer. *World Health Organization* Available at: <https://www.who.int/news-room/fact-sheets/detail/cancer>. (Accessed: 18th June 2019)

100. American Cancer Society. Cancer Facts & Figures 2019 | American Cancer Society. (2019). Available at: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html>. (Accessed: 18th June 2019)
101. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J. Gastrointest. Oncol.* **3**, 153 (2012).
102. Ricci-Vitiani, L., Fabrizio, E., Palio, E. & De Maria, R. Colon cancer stem cells. *J. Mol. Med.* **87**, 1097 (2009).
103. Munro, M. J., Wickremesekera, S. K., Peng, L., Tan, S. T. & Itinteang, T. Cancer stem cells in colorectal cancer: a review. *J. Clin. Pathol.* **71**, 110–116 (2018).
104. Cole, J. W. & McKalen, A. Studies on the morphogenesis of adenomatous polyps in the human colon. *Cancer* **16**, 998–1002 (1963).
105. Shih, I.-M. *et al.* Top-down morphogenesis of colorectal tumors. *Proc. Natl. Acad. Sci.* **98**, 2640–2645 (2001).
106. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
107. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
108. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
109. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).

110. Duthie, S. J. Folate and cancer: how DNA damage, repair and methylation impact on colon carcinogenesis. *J. Inherit. Metab. Dis.* **34**, 101–109 (2011).
111. Sanjoaquin, M. A., Allen, N., Couto, E., Roddam, A. W. & Key, T. J. Folate intake and colorectal cancer risk: A meta- analytical approach. *Int. J. Cancer* **113**, 825–828 (2005).
112. Cravo, M. L. *et al.* Folate Deficiency Enhances the Development of Colonic Neoplasia in Dimethylhydrazine-treated Rats. *Cancer Res.* **52**, 5002–5006 (1992).
113. Kim, Y. I. *et al.* Dietary folate protects against the development of macroscopic colonic neoplasia in a dose responsive manner in rats. *Gut* **39**, 732–740 (1996).
114. Wasson, G. R. *et al.* Global DNA and p53 Region-Specific Hypomethylation in Human Colonic Cells Is Induced by Folate Depletion and Reversed by Folate Supplementation. *J. Nutr.* **136**, 2748–2753 (2006).
115. Ma, J. *et al.* Methylenetetrahydrofolate Reductase Polymorphism, Dietary Interactions, and Risk of Colorectal Cancer. *Cancer Res.* **57**, 1098–1102 (1997).
116. Shane, B. *et al.* The 677C→T variant of MTHFR is the major genetic modifier of biomarkers of folate status in a young, healthy Irish population. *Am. J. Clin. Nutr.* **108**, 1334–1341 (2018).
117. Sohn, K.-J. *et al.* The methylenetetrahydrofolate reductase C677T mutation induces cell-specific changes in genomic DNA methylation and uracil misincorporation: A possible molecular basis for the site-specific cancer risk modification. *Int. J. Cancer J. Int. Cancer* **124**, 1999–2005 (2009).

118. Ma, J. *et al.* A Polymorphism of the Methionine Synthase Gene: Association with Plasma Folate, Vitamin B12, Homocyst(e)ine, and Colorectal Cancer Risk. *Cancer Epidemiol. Prev. Biomark.* **8**, 825–829 (1999).
119. Hirsch, S. *et al.* Colon cancer in Chile before and after the start of the flour fortification program with folic acid. *Eur. J. Gastroenterol. Hepatol.* **21**, 436–439 (2009).
120. Mason, J. B. *et al.* A temporal association between folic acid fortification and an increase in colorectal cancer rates may be illuminating important biological principles: a hypothesis. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **16**, 1325–1329 (2007).
121. Song, J. *et al.* Chemopreventive Effects of Dietary Folate on Intestinal Polyps in *Apc*^{+/-} *Msh2*^{-/-} Mice. *Cancer Res.* **60**, 3191–3199 (2000).
122. Kim, Y.-I. Folate and colorectal cancer: An evidence-based critical review. *Mol. Nutr. Food Res.* **51**, 267–292 (2007).
123. Olfert, E. D., Cross, B. M. & McWilliam, A. A. *Guide to the care and use of experimental animals.* (Canadian Council on Animal Care, 1993).
124. Bailey, S. W. & Ayling, J. E. The extremely slow and variable activity of dihydrofolate reductase in human liver and its implications for high folic acid intake. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 15424–15429 (2009).
125. Dolnick, B. J. & Cheng, Y. C. Human thymidylate synthetase. II. Derivatives of pteroylmono- and -polyglutamates as substrates and inhibitors. *J. Biol. Chem.* **253**, 3563–3567 (1978).

126. Allegra, C. J., Drake, J. C., Jolivet, J. & Chabner, B. A. Inhibition of phosphoribosylaminoimidazolecarboxamide transformylase by methotrexate and dihydrofolic acid polyglutamates. *Proc. Natl. Acad. Sci.* **82**, 4881–4885 (1985).
127. Matthews, R. G. & Baugh, C. M. Interactions of pig liver methylenetetrahydrofolate reductase with methylenetetrahydropteroylpolyglutamate substrates and with dihydropteroylpolyglutamate inhibitors. *Biochemistry* **19**, 2040–2045 (1980).
128. Smith, A. D., Kim, Y.-I. & Refsum, H. Is folic acid good for everyone? *Am. J. Clin. Nutr.* **87**, 517–533 (2008).
129. MacGregor, J. T., Schlegel, R., Wehr, C. M., Alperin, P. & Ames, B. N. Cytogenetic damage induced by folate deficiency in mice is enhanced by caffeine. *Proc. Natl. Acad. Sci.* **87**, 9962–9965 (1990).
130. James, S. J., Basnakian, A. G. & Miller, B. J. In Vitro Folate Deficiency Induces Deoxynucleotide Pool Imbalance, Apoptosis, and Mutagenesis in Chinese Hamster Ovary Cells. *Cancer Res.* **54**, 5075–5080 (1994).
131. Hekmatdoost, A. *et al.* The effect of dietary oils on cecal microflora in experimental colitis in mice. *Indian J. Gastroenterol. Off. J. Indian Soc. Gastroenterol.* **27**, 186–189 (2008).
132. SUGAHARA, H., ODAMAKI, T., HASHIKURA, N., ABE, F. & XIAO, J. Differences in folate production by bifidobacteria of different origins. *Biosci. Microbiota Food Health* **34**, 87–93 (2015).
133. Aufreiter, S. *et al.* Folate is absorbed across the colon of adults: evidence from cecal infusion of ¹³C-labeled [6S]-5-formyltetrahydrofolic acid. *Am. J. Clin. Nutr.* **90**, 116–123 (2009).

134. Rossi, M., Amaretti, A. & Raimondi, S. Folate production by probiotic bacteria. *Nutrients* **3**, 118–134 (2011).
135. Said, H. M. & Mohammed, Z. M. Intestinal absorption of water-soluble vitamins: an update. *Curr. Opin. Gastroenterol.* **22**, 140–146 (2006).
136. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
137. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).
138. Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).
139. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
140. Alexandrov, L. B. *et al.* The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv* 322859 (2018). doi:10.1101/322859
141. Douglas, G. R., Gingerich, J. D., Gossen, J. A. & Barlett, S. A. Sequence spectra of spontaneous lacZ gene mutations in transgenic mouse somatic and germline tissues. *Mutagenesis* **9**, 451–458 (1994).
142. Schaefer, M. H. & Serrano, L. Cell type-specific properties and environment shape tissue specificity of cancer genes. *Sci. Rep.* **6**, 20707 (2016).
143. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. *Mol. Cell* **10**, 1247–1253 (2002).
144. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646 (2002).

145. Harris, R. S. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res. BCR* **17**, 8 (2015).
146. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).
147. Knisbacher, B. A., Gerber, D. & Levanon, E. Y. DNA Editing by APOBECs: A Genomic Preserver and Transformer. *Trends Genet.* **32**, 16–28 (2016).
148. Masuda, K. *et al.* A Critical Role for REV1 in Regulating the Induction of C:G Transitions and A:T Mutations during Ig Gene Hypermutation. *J. Immunol.* **183**, 1846–1850 (2009).
149. Hsu, H.-C. *et al.* Folate Deficiency Triggers an Oxidative-Nitrosative Stress-Mediated Apoptotic Cell Death and Impedes Insulin Biosynthesis in RINm5F Pancreatic Islet β -Cells: Relevant to the Pathogenesis of Diabetes. *PLoS ONE* **8**, (2013).
150. Lai, K.-G. *et al.* Novel roles of folic acid as redox regulator: Modulation of reactive oxygen species sinker protein expression and maintenance of mitochondrial redox homeostasis on hepatocellular carcinoma. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **39**, 1010428317702649 (2017).
151. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45**, 136–144 (2013).
152. Friedenson, B. BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. *MedGenMed Medscape Gen. Med.* **7**, 60 (2005).

153. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49–53 (1991).
154. Barta, J. A. & McMahon, S. B. Lung-Enriched Mutations in the p53 Tumor Suppressor: A Paradigm for Tissue-Specific Gain of Oncogenic Function. *Mol. Cancer Res.* **17**, 3–9 (2019).
155. He, Z., Kosinska, W., Zhao, Z.-L., Wu, X.-R. & Guttenplan, J. B. Tissue-specific mutagenesis by N-butyl-N-(4-hydroxybutyl)nitrosamine as the basis for urothelial carcinogenesis. *Mutat. Res. Toxicol. Environ. Mutagen.* **742**, 92–95 (2012).
156. Ponomareva, O. N., Rose, J. A., Lasarev, M., Rasey, J. & Turker, M. S. Tissue-specific Deletion and Discontinuous Loss of Heterozygosity Are Signatures for the Mutagenic Effects of Ionizing Radiation in Solid Tissues. *Cancer Res.* **62**, 1518–1523 (2002).
157. Schmezer, P., Eckert, C. & Liegibel, U. M. Tissue-specific induction of mutations by streptozotocin in vivo. *Mutat. Res. Mol. Mech. Mutagen.* **307**, 495–499 (1994).
158. Garraway, L. A. & Lander, E. S. Lessons from the Cancer Genome. *Cell* **153**, 17–37 (2013).
159. Phillips, D. H. Mutational spectra and mutational signatures: Insights into cancer aetiology and mechanisms of DNA damage and repair. *DNA Repair* **71**, 6–11 (2018).
160. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836.e16 (2019).
161. Pui, C.-H. Acute Lymphoblastic Leukemia. in *Encyclopedia of Cancer* (ed. Schwab, M.) 23–26 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-16483-5_57

162. Terwilliger, T. & Abdul-Hay, M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J.* **7**, e577 (2017).
163. Chango, A. *et al.* A Polymorphism (80G->A) in the Reduced Folate Carrier Gene and Its Associations with Folate Status and Homocysteinemia. *Mol. Genet. Metab.* **70**, 310–315 (2000).
164. Laverdière, C., Chiasson, S., Costea, I., Moghrabi, A. & Krajcinovic, M. Polymorphism G80A in the reduced folate carrier gene and its relationship to methotrexate plasma levels and outcome of childhood acute lymphoblastic leukemia. *Blood* **100**, 3832–3834 (2002).
165. Scott, K. C. & Gregory, J. F. The fate of [3H]folic acid in folate-adequate rats. *J. Nutr. Biochem.* **7**, 261–269 (1996).
166. Nijhout, H. F., Reed, M. C., Budu, P. & Ulrich, C. M. A Mathematical Model of the Folate Cycle NEW INSIGHTS INTO FOLATE HOMEOSTASIS. *J. Biol. Chem.* **279**, 55008–55016 (2004).
167. Richardson, R. E., Healy, M. J. & Nixon, P. F. Folates of rat tissue: Bioassay of tissue folypolyglutamates and a relationship of liver folypolyglutamates to nutritional folate sufficiency. *BBA - Gen. Subj.* **585**, 128–133 (1979).
168. Edwards, J. L. & Klein, R. E. Cell Renewal in Adult Mouse Tissues. *Am. J. Pathol.* **38**, 437–453 (1961).
169. Ogawa, M. Differentiation and proliferation of hematopoietic stem cells. *Blood* **81**, 2844–2853 (1993).
170. Asante, I. *et al.* Alterations in folate-dependent one-carbon metabolism as colon cell transition from normal to cancerous. *J. Nutr. Biochem.* **69**, 1–9 (2019).

171. Barker, N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. *Nat. Rev. Mol. Cell Biol.* **15**, 19–33 (2014).
172. Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G. & Basler, K. GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. *Nature* **558**, 449 (2018).
173. Vermeulen, L. & Snippert, H. J. Stem cell dynamics in homeostasis and cancer of the intestine. *Nat. Rev. Cancer* **14**, 468–480 (2014).
174. Justice, M. J., Noveroske, J. K., Weber, J. S., Zheng, B. & Bradley, A. Mouse ENU Mutagenesis. *Hum. Mol. Genet.* **8**, 1955–1963 (1999).
175. Milholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget* **6**, 24627–24635 (2015).
176. Iacopetta, B. TP53 mutation in colorectal cancer. *Hum. Mutat.* **21**, 271–276 (2003).
177. Kim, Y. I. *et al.* Folate deficiency in rats induces DNA strand breaks and hypomethylation within the p53 tumor suppressor gene. *Am. J. Clin. Nutr.* **65**, 46–52 (1997).
178. MacFarlane, A. J. *et al.* Dietary folic acid protects against genotoxicity in the red blood cells of mice. *Mutat. Res.* **779**, 105–111 (2015).
179. Li, G.-M., Presnell, S. R. & Gu, L. Folate deficiency, mismatch repair-dependent apoptosis, and human disease. *J. Nutr. Biochem.* **14**, 568–575 (2003).
180. Choi, S.-W. & Mason, J. FOLATE AND COLORECTAL CARCINOGENESIS: IS DNA REPAIR THE MISSING LINK? *Am. J. Gastroenterol.* **93**, 2013–2016 (1998).

181. Swayne, B. G. *et al.* Investigating the effects of dietary folic acid on sperm count, DNA damage and mutation in Balb/c mice. *Mutat. Res. Mol. Mech. Mutagen.* **737**, 1–7 (2012).
182. Yauk, C. L. *et al.* Tandem repeat mutation, global DNA methylation, and regulation of DNA methyltransferases in cultured mouse embryonic fibroblast cells chronically exposed to chemicals with different modes of action. *Environ. Mol. Mutagen.* **49**, 26–35 (2008).
183. Gu, L., Wu, J., Qiu, L., Jennings, C. D. & Li, G.-M. Involvement of DNA mismatch repair in folate deficiency-induced apoptosis☆. *J. Nutr. Biochem.* **13**, 355–363 (2002).
184. Kim, Y. I. *et al.* Dietary folate protects against the development of macroscopic colonic neoplasia in a dose responsive manner in rats. *Gut* **39**, 732–740 (1996).
185. Wickramasinghe, S. N. & Fida, S. Bone marrow cells from vitamin B12- and folate-deficient patients misincorporate uracil into DNA. *Blood* **83**, 1656–1661 (1994).
186. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
187. Son, H., Kang, H., Kim, H. S. & Kim, S. Somatic mutation driven codon transition bias in human cancer. *Sci. Rep.* **7**, (2017).
188. Lynch, A. M., Gooderham, N. J., Davies, D. S. & Boobis, A. R. Genetic analysis of PHIP intestinal mutations in MutaTMMouse. *Mutagenesis* **13**, 601–605 (1998).