Roles for endogenous aldehydes in haematopoietic ageing and immunity

Oliver Beaven
St Edmund's Hall
This dissertation is submitted for the degree of Doctor of Philosophy
October 2025

Declaration

This dissertation is my own work, any instance where work has been included as an outcome of collaboration with others, has been explicitly stated in the text and Acknowledgements. This work has not been previously submitted, in part or whole, to the University of Oxford, or any other institution of higher education.
In accordance with the guidelines, this thesis does not exceed 50,000 words.

Oliver Beaven

Acknowledgements

I would first like to thank KJ for the opportunity to work with such an intellectually stirring group of people and projects. I am forever grateful for his insight, directness, and continued belief in me as a scientist when projects entered adversity. His continued willingness to discuss and explore scientific ideas have been foundational to helping me see beyond superficial interpretations of data and learn how to extract the rust from the gold.

Next, I thank my second supervisor, Felix Dingler. Without Felix, I honestly don't know where I would be. From day one Felix has stunned me with his willingness to help, his far-reaching knowledge and his ability to cut through the holes in any argument. For all my years in the lab, Felix has been there to guide and foster my curiosity and enthusiasm for ideas. Felix has been inspiring in his ceaseless embodiment of the scientific principles, while never parting from a profound kindness. I will always cherish the hours passed being hooked in a Felix conversation, even, and especially, those that begin at 10pm.

I thank all members of the lab for their insights, guidance and generous warmth. Everyone in the lab has always been there to help shape my science and working life in a significant way. I thank Christopher Millington, for guiding me through the adduct measurements (as well as climbing routes and spiritual chaos). I thank Ewa Gogola, for her endless supply of cutting insights in lab meetings, and for taking me through the steps of liver ploidy analysis. I thank Xinran Huang, for showing me the ways of the Wilkinson culture, for being the star recruit for big mouse take down days, and

opening my eyes to a fresh bounty of quality Chinese food. I thank Ashley Kamimae-Lanning for establishing the parental strains for the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice, her beautifully detailed protocols, and for the countless number of of delicious cakes that I couldn't do without. I thank Jill Brown for showing me the dark arts of the sister chromatid exchange assay, and for her aptitude for creating calm in tense moments. I thank the remaining members: Holly Russell, Kleitos Popadopolous, Samuel Jones, Madeline Foster-Smith, Caroline Scott; for making the lab the fun and stimulating place it was to work in. All will be missed dearly.

Next, I thank my collaborator, Matthias Gunther at DKFZ, for performing the variant allele analysis that forms the basis of all comments made about clonal haematopoiesis. I also thank Pramila Rijal for providing the haemagglutinin protein and her valuable insights into my immunity project. I thank Leanne Hodson, for offering to act as a second supervisor, and guiding me through a project on obesity resistance that was sadly only too short lived. I thank the WIMM Flow Cytometry facility and staff: Paul Sopp, Craig Waugh, Kevin Clark and Sally-Ann Clark; for all their expertise and help with FACS. I thank Kyle Clapton-Matheson and all of the JR-BMS staff for their marathon efforts in overseeing my mouse strains. I thank Philip Hublitz and Chris Babbs at the WIMM Genome Engineering Facility, for providing me with plasmids with which I could complete my gene editing project, and for their invaluable advice.

Finally, I will thank all my friends and family who have never left my side through the highs and the lows of my time in Oxford, making this an experience that I can look back on, rich in cherished memories and personal achievement.

Abstract

Formaldehyde is a fundamental driver of genomic instability. This has led to the evolution of a two-tiered defence system that engages metabolic detoxification ("tier-1" via *ALDH2* and *ADH5*) and DNA repair pathways ("tier-2" via Fanconi's anaemia repair (FA) and transcription coupled repair). Strikingly, ~540 million people of East Asian descent carry a variant in the *ALDH2* allele (*ALDH2*2*), which impairs tier-1 protection. When these carriers also incur mutations in *ADH5* or *FA* repair, they develop significant disease, characterised by leukaemia, bone marrow failure, developmental abnormality and shortening life span. Given this predisposition, the high prevalence of *ALDH2*2* is surprising, suggesting this allele has evaded selection by conferring an underappreciated selective advantage, or that its effects become significant only over time. In this thesis, I explore whether *ALDH2* deficiency may confer a selective advantage through enhancing humoral immunity. I compare the humoral immune response of *Aldh2*$^{-/-}$ mice to model antigens (HEL, OVA, BSA and HA), finding *ALDH2*-deficiency influenced humoral immunity in an antigen dependent manner, which may be significant in helping *ALDH2*2* carriers to resist certain infections. I also investigate how tier-1 deficiency drives ageing in the blood. An existing model of tier-1 deficiency, *Aldh2*$^{-/-}$*Adh5*$^{-/-}$, shows features of accelerated ageing, including cell-fate bias and somatic mosaicism, but its extreme perinatal lethality prevents its use to study ageing as a time-dependent process. To address this, I use *Vav1*^{iCre} to delete *ALDH2* specifically in the blood of *Adh5*$^{-/-}$ mice. This rescued the peri-natal lethality, and revealed features of haematopoietic ageing, though without clonal haematopoiesis. Together, these models demonstrate that endogenous formaldehyde can drive instructional changes on stem cell fate, providing insight into how endogenous formaldehyde may act as a fundamental driver of ageing. This work has significant implications for understanding mechanisms of ageing and disease in the 540 million individuals with compromised tier-1 protection.

Contents

List of Abbreviations

ADDS	Aldehyde Dehydrogenase Deficiency Syndrome
ADH5	Alcohol Dehydrogenase 5
ALDH2	Aldehyde Dehydrogenase 2
BMF	Bone Marrow Failure
CH	Clonal Haematopoiesis
DDR	DNA Damage Response
DSB	Double Stranded Break
ESCC	Oesophageal Squamous Cell Carcinoma
FA	Fanconi's Anaemia
GG-NER	Global Genome Nucleotide Excision Repair
GSH	Glutathione
HR	Homologous Recombination
HSC	Haematopoietic Stem Cell
LKS	Lineage- cKit+ Sca1+
LT-HSC	Long Term Haematopoietic Stem Cell
mNG	mNeonGreen
MPP	Multi-Potent Progenitor
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
ROS	Reactive Oxygen Species
SBS	Single Base Substitution
SNV	Single Nucleotide Variant
SSB	Single-Stranded Break
ST-HSC	Short Term Haematopoietic Stem Cell
TC-NER	Transcription Coupled Nucleotide Excision Repair
TLS	Translesion Synthesis
VAF	Variant Allele Frequency
WGS	Whole Genome Sequencing

Chapter 1

Introduction

1.1 DNA damage as a driver of ageing

1.1.1 DNA is an unstable molecule

Life hinges on the stable maintenance of genetic information, yet genetic information is carried in an inherently unstable molecule. The mammalian genome is estimated to undergo between 10,000 and 100,000 lesions a day – a product of spontaneous hydrolysis of purines, errors in DNA replication, and a constant barrage of exogenous and endogenous stressors (Lindahl, 1993). These stressors derive from elements encountered in everyday life, for example, UV exposure from sunlight drives formation of pyrimidine dimers and pyrimidine-pyrimidine 6-4 photoproducts between adjacent bases, forming bulky distortions in the DNA helix (Van Steeg & Kraemer, 1999). Reactive oxygen (ROS) and reactive nitrogen species produced from aerobic respiration, can cause oxidation of guanine molecules, forming 8-oxo-guanine, which can be mis-read as T bases during DNA replication (Kawanishi et al., 2006). Reactive aldehyde species, such as formaldehyde, can induce covalent crosslinks between bases and proteins, that block the progression of transcriptional and replication machinery (Hodskinson et al., 2020; Oka et al., 2024). These lesions can lead to generation of single stranded and double stranded breaks (SSBs and DSBs,

respectively). DNA damage results in changes to the underlying nucleotide sequence, introducing mutations (substitutions, insertions and deletions) and chromosomal aberrations, which underlie the risk of cancer and genetic disease (Schumacher et al., 2021).

Eukaryotes have evolved a plethora of enzymes that confer protection against these events. In the first instance, stalled replication forks can be bypassed by specialised translesion polymerases (TLS), which are error prone but enable tolerance of DNA damage (McCulloch & Kunkel, 2008). Alternatively, lesions can be directly repaired. One class of repair is that performed by excision repair enzymes, which recruit endonucleases to excise the lesion, and polymerases and ligases to fill in the generated gap. These pathways constitute mismatch repair, base excision repair and nucleotide excision repair (NER). During transcription, the stalling of RNA polymerase II upon encountering a lesion initiates a transcription-coupled NER (TC-NER) response, while lesions are also directly recognised by NER machinery independently of transcription, drive global genome NER (GG-NER) (Huang & Zhou, 2021).

DSBs are repaired by two mechanisms: non-homologous end joining (NHEJ) or homologous recombination (HR) (Iyama & Wilson, 2013). NHEJ directly ligates broken DNA ends together. This process is error prone, as compatible ends are generated by the removal of bases at the site of the break, however, it has its advantage in being operative during G1 of the cell cycle, and is rapidly activated (Mao et al., 2008). Homologous recombination only takes place during S/G2, but facilitates the accurate repair of the DNA break using the undamaged sister chromatid as a template across which the gap is restored, thus avoiding the loss of genetic information.

Interstrand crosslink repair by the Fanconi anaemia (FA) pathway, removes crosslinks between bases that block replication fork progression, and requires elements of excision, TLS and homologous recombination to complete its task. Mutations in proteins found in the FA repair pathway, result in severe developmental abnormalities, and rapid onset of bone marrow failure (BMF; on average by age 7) and significantly increased risk of cancer development, including 600-fold increased risk of acute myeloid leukaemia, and 5000-fold increased risk of myelodysplastic syndrome (Shimamura & Alter, 2010).

Activation of these repair pathways is controlled by a signalling cascade known as the DNA damage response (DDR). This is governed by the activity of three related kinases: ataxia-telangiectasia mutated (ATM), ATM and Rad3-related (ATR) and DNA-dependent protein kinase (DNA-PK) (Blackford & Jackson, 2017). These are responsible for direct recruitment of DNA repair proteins to sites of DNA damage, as well as activation of downstream effector pathways (such as p53, CHK2 and PPMD1), which coordinate wider cellular and transcriptional processes, such as autophagy and cell cycle response (Niedernhofer et al., 2018). The timing and strength of their activation determines whether the cell will undergo apoptosis, cell cycle arrest, or engage DNA repair (Blackford & Jackson, 2017). The combined action of these repair and signalling pathways protects the body from genomic instability and keeps the mutation rate of the mammalian genome to only ~20.7 substitutions and ~1.3 indels per year (Abascal et al., 2021).

1.1.2 Genomic instability drives processes observed in ageing

Inefficiencies and inaccuracies in DNA repair machinery means DNA damage accumulates in all tissues as we age (Abascal et al., 2021; Alexandrov et al., 2015). This is detected through markers that reflect incidence of DSBs, such as comet tail movement and phosphorylation of Ser139 on H2AX (γH2AX) (Beerman et al., 2014)) and accumulation of mutations (Abascal et al., 2021; Spencer Chapman et al., 2025). Capacity for DNA repair also declines (Beerman et al., 2014; Sedelnikova et al., 2008). As such, accumulating DNA damage is postulated as a key element of continual loss of somatic function that features in the ageing process (Figure 1.1; Kirkwood, 2005).

The contribution of DNA damage to driving the ageing process is emphasised by the occurrence of progeroid conditions arising from mutations in DNA repair enzymes. Patients with loss of function mutations in genes encoding TC-NER enzymes (*CSA, CSB, XPB, XPD* or *XPG*) suffer a profound condition known as Cockayne Syndrome (CS) (Karikkineth et al., 2017). These defects typically manifest in long-lived, non-replicative tissues that rely heavily on TC-NER; patients exhibit neurodegeneration, cataracts, osteoporosis, hearing loss, multiple-organ failure and a shortened life expectancy. Similar symptoms are also observed in patients with Werner's syndrome, who carry loss of function mutations in RecQ helicase, essential for unwinding DNA strands (Oshima et al., 2017). Accelerated features of ageing are also seen in genetic mouse models of DNA repair deficiency – genetic deletion of NHEJ, NER or telomere maintenance proteins in mice caused accelerated decline in the ability of haematopoietic stem cells (HSCs) to proliferate or transplant, analogous to the functional decline observed in aged HSCs (Nijnik et al., 2007; Rossi, Bryder, et al., 2007; Rossi et al., 2005). Deletion of ERCC1 in mice, ablating the essential

endonuclease common to GG-NER, TC-NER and FA pathways, drives premature ageing of the musculoskeletal, hepatic, neurological, cardiovascular and haematopoietic systems, and shortens lifespan to just 3 weeks (Niedernhofer et al., 2006). These patients and models highlight how endogenously produced DNA damage needs to be repaired, and that failure to repair these directly accelerates the ageing process.

While these are extreme scenarios, we are regularly exposed to carcinogenic agents that drive DNA damage, with deleterious consequences, for example, use of platinum and radiation-based chemotherapy, as well as lifestyle factors such as smoking and alcohol consumption. These agents induce DNA damage across tissues and accelerate the loss of diversity and acquisition of mutations in stem cell populations in somatic tissues, including in the blood (Mitchell et al., 2025), oesophageal epithelium (Martincorena et al., 2018; Yokoyama et al., 2019) and bronchial epithelium (Yoshida et al., 2020). This loss explains the decline of somatic function and regenerative capacity (López-Otín et al., 2013) and increases the risk of developing neoplasia through acquisition of secondary mutations (Knudson, 1971), or increased sensitivity to exogenous agents (Balmain, 2020). Indeed, children treated for primary cancer by cisplatin or radiation are at a significantly increased risk of developing severe chronic health conditions by the age of 35, particularly cardiac events or new cancer (Armstrong et al., 2014).

DNA damage-associated decline in function with ageing is mitigated in part by the DDR (reviewed by (Ou et al., 2018)). For example, hypermorphic mutations of p53 induce phenotypes of premature ageing in the blood, including atrophy of blood

populations and functional failure to engraft on transplantation (Dumble et al., 2007). Conversely, the accelerated incidence of BMF in FA patients is dependent on constitutive activation of p53 in haematopoietic stem and progenitor cells (HSPCs), which drives p21 induction and cell cycle arrest (Ceccaldi et al., 2012). When p53 was depleted, this arrest was no longer observed. In the ageing blood, HSCs are usually quiescent and accumulate DNA damage that is not actively repaired until they are induced to cycle (Beerman et al., 2014; Flach et al., 2014; Rossi et al., 2007). Upon cell cycle entry, DNA damage is recognised by stalling replication machinery, and a DNA damage response is induced. The scale of which increases dependent on the amount of DNA damage encountered (a function of time) resulting in transcriptional changes in HSCs, such as decreased ribosomal biogenesis, and can trigger apoptotic signalling, thus driving loss of function in the HSC pool over time (Flach et al., 2014; Walter et al., 2015). Persistent signalling of the DDR through p53, can also lead to permanent cell cycle arrest (i.e. senescence) (Rodier et al., 2009). The centrality of senescent cells to contributing to decline of tissue function with age is best demonstrated by experiments that genetically or pharmacologically clear senescent cells *in vivo*, in normally aged mice. For instance, targeted killing of p16 expressing cells (a key hallmark of senescence), which significantly increased life span and attenuated loss of tissue function in the heart and kidneys (Baker et al., 2016). Similar effects were observed following use of pharmacological agents, senolytics, which selectively clear senescent cells, which were able to restore HSC reconstitution efficiency in cells taken from aged mice (Chang et al., 2016).

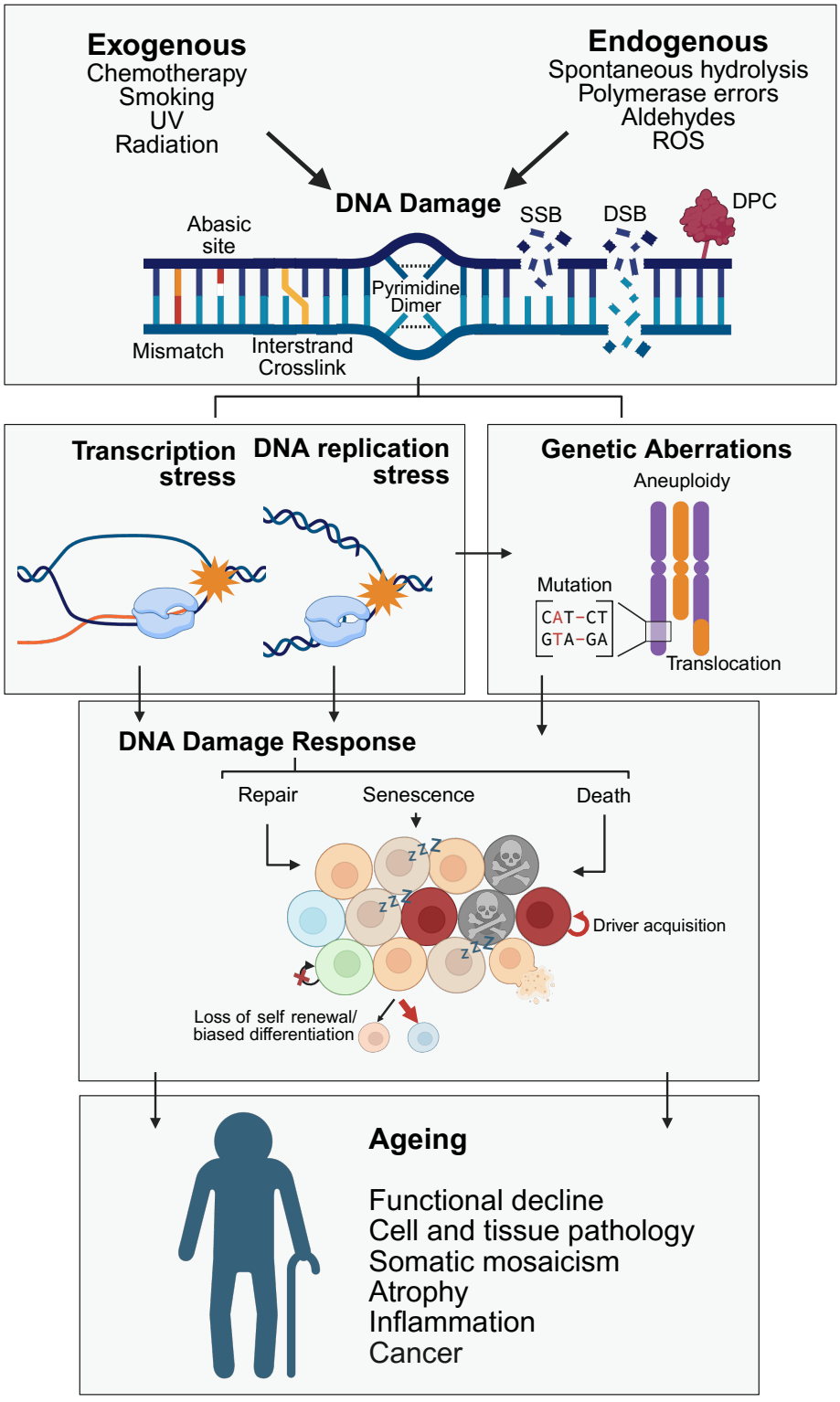


Figure 1.1 DNA damage as a driver of ageing

Model of DNA damage as a driver of ageing, adapted from Schumacher *et al.*, 2021. DNA molecules are damaged by endogenous and exogenous routes, changing the structure and sequence of DNA, resulting in transcription and replication stress, and genetic aberration. These can themselves, or through the DNA damage response, disrupt tissue processes and stem cell activity, which drives ageing as these occur continuously through life. ROS – Reactive Oxygen Species, UV – ultraviolet.

1.2 Aldehydes as an endogenous source of DNA damage

While the consequence of DNA damage is clear, the exact causative agents that generate the DNA damage observed in normal ageing is not. Study of single base substitution signatures (SBS) provides insight to the overall load and pattern of mutations present, which may point to the mutagenic agents driving DNA damage in that tissue. This is clear in cases with known mutational drivers, such as chemotherapeutic agents (Mitchell et al., 2025), as with common C>T mutations, resulting from spontaneous 5-methylcytosine deamination, driving the endogenous COSMIC SBS1 signature routinely observed in ageing (Alexandrov et al., 2015; Cagan et al., 2022). A candidate approach using CRISPR-Cas9 screening of iPSCs found that deletion of 8-oxoguanine glycosylase 1 (OGG1), which removes 8-oxo-dG adducts induced by ROS, significantly increased G>T (C>A) mutations, with similarity to SBS18, also shown to increase with age (Cagan et al., 2022; Zou et al., 2021). However, the aetiology of the most pervasive mutation signature, SBS5, which accumulates with age in all tissues across organisms, remains ambiguous (Abascal et al., 2021; Alexandrov et al., 2015; Cagan et al., 2022; Spencer Chapman et al., 2025).

Genetic models offer a convenient approach through which we can characterise the substrates of DNA repair enzymes, which contribute to the DNA damage and mutations observed with age. In a simple example, patients deficient in Xeroderma Pigmentosa proteins (involved in GG-NER) are hypersensitive to UV irradiation, which forms bulky 6-4 photoproducts and cyclobutene pyrimidine dimers that blocks DNA replication and transcription. As a result, these patients are at a 1000-fold increased

risk of developing skin cancer (Van Steeg & Kraemer, 1999). However, the agent of DNA damage requiring repair is not so clear for FA and CS. To make things harder, the major clinical manifestation of FA and CS are not observed in mice upon genetic deletion of FA and TC-NER repair factors. This is despite indications of functional decline, for example, LT-HSCs of *Fanca*$^{-/-}$ mice accumulate DNA damage *in vivo*, which results in their attrition with age (Walter et al., 2015) and perturbed foetal haematopoietic development in *Fancc*$^{-/-}$ (Kamimae-Lanning et al., 2013). This may suggest that the endogenous substrate necessitating TC-NER and FA is insufficiently present in mice. ROS is a widely postulated factor of DNA damage that accumulates in age, and FA mice do show increased sensitivity to ROS. Indeed, LT-HSCs in *Fanca*$^{-/-}$ mice present elevated 8-oxo-dG lesions and DSBs that could be partially reversed upon over-expression of superoxide dismutase 2 (SOD2; responsible for clearing ROS) (Walter et al., 2015). However, while genetic deletion of SOD1 in *Fancc*$^{-/-}$ mice induced reductions in red blood cell counts, bone marrow cellularity, and *ex vivo* colony forming capacity of bone marrow cells, it did not induce the profound BMF observed in human FA patients (Hadjur et al., 2001). Furthermore, *Ogg1*$^{-/-}$*Csb*$^{-/-}$ mice (deficient in 8-oxo-dG clearance and TC-NER) had more C>T mutations in 5-month-old livers, but did not exhibit clinical signs of CS (Trapp et al., 2007). Therefore, it is unlikely that ROS are the major substrates of endogenous DNA damage that necessitate these pathways.

1.2.1 Endogenous formaldehyde and acetaldehyde drive DNA damage in mice and humans

Formaldehyde and acetaldehyde are highly reactive chemical species that adduct DNA bases, as well as lysine and cysteine residues in proteins, which can evolve to

form covalent crosslinks between DNA, RNA and proteins (Fen et al., 1980; Hodskinson et al., 2020; Lu, Ye, et al., 2010; Tuma et al., 1987; Wang et al., 2000). Pioneering work by the Patel lab, revealed that endogenously derived aldehydes are substrates for the FA and TC-NER pathways *in vivo*. This was first shown through combined genetic deletion of mitochondrial aldehyde dehydrogenase 2 (*ALDH2*), which oxidises formaldehyde and acetaldehyde to formate and acetate, respectively, with loss of essential FA pathway protein, *FANCD2* (Figure 1.2). The subsequent *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ mice, had a significantly shortened lifespan (145 days vs >600 in controls) and display profound anaemia, BMF and/or acute leukemia, with significant numerical and functional decline of the HSC compartment, reminiscent of human FA patients (Garaycoechea et al., 2012; Langevin et al., 2011). These phenotypes coincide with significantly increased markers of DNA damage, such as γH2AX, increase in chromosomal translocations, red-blood cell micronuclei, sister-chromatid exchanges, and mutations in the blood compartment (Garaycoechea et al., 2012, 2018). This led to the description of a “two-tier” protection mechanism against aldehydes, the first tier operating through metabolic clearance, and the second through mechanisms of DNA repair. By ablating both, we highlight aldehydes as essential substrates of endogenous DNA damage.

Validation that endogenous aldehydes contribute to the DNA damage burden in humans, comes from a cohort of Japanese FA patients carrying the common *ALDH2* variant, rs671 [*ALDH2*2*, Glu504>Lys, the wild type demarked *ALDH2*1*] (Hira et al., 2013). The *ALDH2*2* variant encodes the ALDH2 protein with a lysine substitution at position 504, disrupting the formation of the functional homotetrameric structure of *ALDH2* in a homozygous dominant manner, making *ALDH2*2/*2* individuals unable to

detoxify acetaldehyde. *ALDH2*1/*2* individuals retain some detoxification capacity owing to residual *ALDH2*1* stabilising the tetrameric structure (Larson et al., 2005; Xiao et al., 1996; Zhou & Weiner, 2000). Inability to clear acetaldehyde leads to the characteristic “Asian flush” reaction in *ALDH2*2/*2* individuals following alcohol consumption (Harada et al., 1981). Hira *et al*. observed that FA patients with *ALDH2*2* alleles had significantly accelerated progression of BMF compared to carriers of the WT variant, with homozygotes showing faster progression than *ALDH2*1/*2* heterozygotes, mirroring observations of the *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ *mice*. This shows that endogenous aldehydes in humans are causing DNA damage requiring the FA repair pathway.

Later work focused on generation of genetic models that combined deletion of alcohol dehydrogenase 5 (ADH5), which clears formaldehyde by conversion of formaldehyde-glutathione conjugates to formate, with *FANCD2* deletion. The subsequent *Adh5*$^{-/-}$ *Fancd2*$^{-/-}$ mice, had a considerably more severe phenotype than *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ mice, shortening median lifespan to just 33 days, with additional induction of DNA damage and organ failure in the liver and kidneys (Pontel et al., 2015). This work led to the suggestion that ADH5 substrates are likely the dominant target of tier-1 protection. Combined genetic deletion of ADH5 with the essential TC-NER protein CSB (*Adh5*$^{-/-}$*Csb*$^{-/-}$), lead to mice developing severe progeroid manifestations reminiscent of human CS, including skeletal abnormalities, cachexia, profound neurodegeneration, and significant risk of renal failure (Mulderrig et al., 2021). The study of compound *Adh5*$^{-/-}$ and DNA repair deficient models extended our understanding of endogenous aldehydes as a substrate essential to multiple

mechanisms of DNA repair, across multiple tissues, thus serving as fundamental drivers of DNA damage in organisms.

1.2.2 Aldehyde clearance is essential to genomic stability even in the presence of DNA repair

Similar to how single knock-out models of FA and CS exhibit phenotypes mild compared to humans, single *Aldh2*$^{-/-}$ mice present no endogenous phenotype. *Adh5*$^{-/-}$ mice present increased risk of hepatocellular carcinoma, and increased number of Lineage^{-} cKit^{+} Sca1^{+} (LKS) cells with reduced functional activity upon inflammation stress or transplantation, but these endogenous phenotypes do not arise until late in life (Liu et al., 2004; Pontel et al., 2015; Wei et al., 2010). However, combined genetic deletion of *ALDH2* and *ADH5 in vivo*, reveals that these enzymes work in a redundant manner, resulting in severe perinatal lethality, developmental defects, anorexia, anaemia and increased risk of acute leukaemia, hepatocellular carcinoma and lymphoma in mice (Dingler et al., 2020; Oka et al., 2020). Analysis of DNA damage markers in the blood, including red blood cell micronucleation (markers of DSBs), sister chromatid exchanges (indicative of active HR-repair) and quantification of mutations, and more recently DPCs, highlights that the endogenous formaldehyde burden is sufficient to overwhelm the intact DNA repair machinery to drive disease (Dingler et al., 2020; Oka et al., 2024). Interestingly, analysis of the mutational signature within clonally expanded HSPCs of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice revealed a distinct mutational profile with high cosine similarity with SBS25, SBS40, SBS5 and SBS3, associated with multiple cancers (Alexandrov et al., 2020; Dingler et al., 2020). The similarity with SBS5 is particularly intriguing, as it suggests that aldehydes may well be contributing to the mutational processes that feature ubiquitously through life.

Human patients with combined *ALDH2* and *ADH5* deficiency were also described and mirrored the features of these mice, including aplastic anaemia, mental retardation, dwarfism, BMF and leukemic disposition early in life. This condition is now described as Aldehyde Dehydrogenase Deficiency Syndrome (ADDS) and emphasises the capacity for endogenous formaldehyde to compound in severe human disease even in the presence of intact DNA repair (Dingler *et al*., 2020; Oka *et al*., 2020).

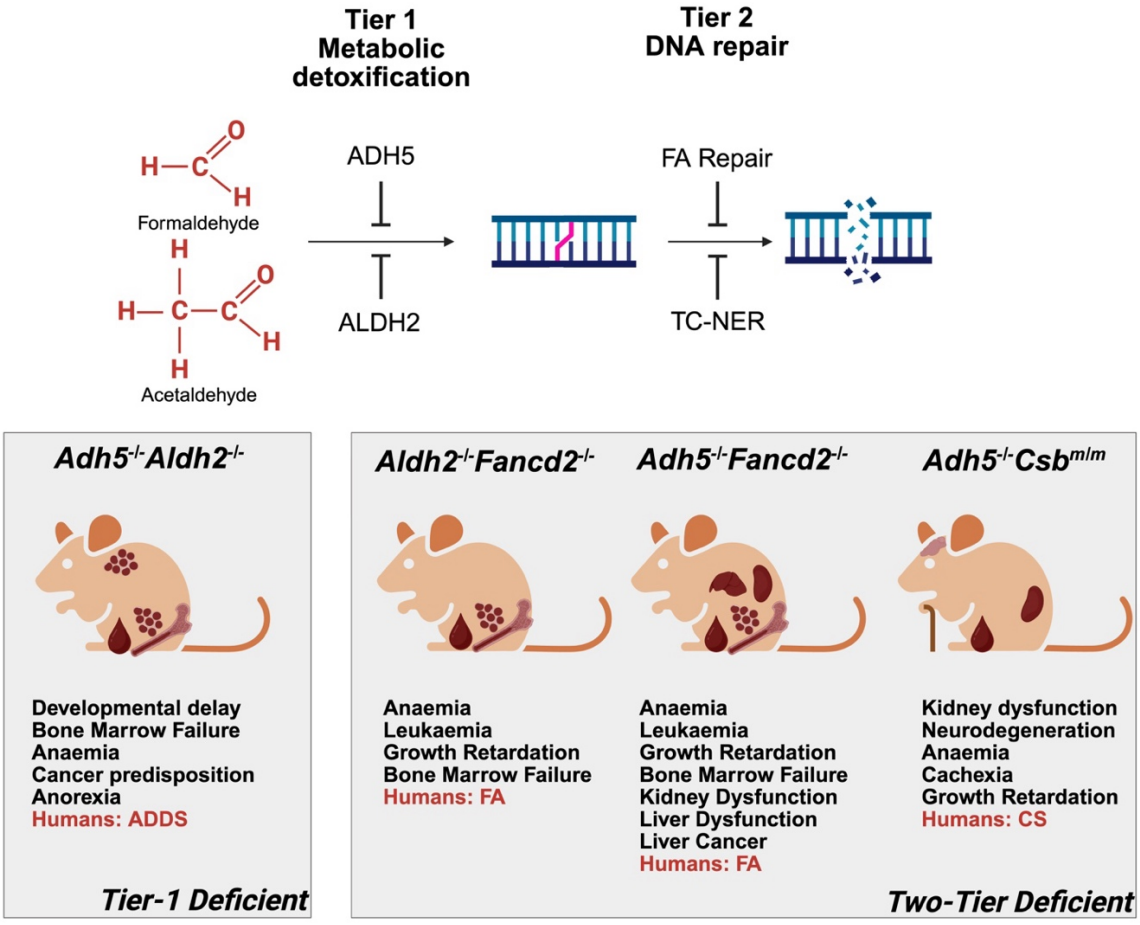


Figure 1.2 Two-tier protection and formaldehyde deficient models
(Top) Mechanism of two-tier protection against endogenous aldehydes. Formaldehyde and acetaldehyde are detoxified by *ADH5* and *ALDH2*, which serves as "tier-1" protection against accumulation of formaldehyde induced DNA damage. "Tier-2" operates through repair of lesions on DNA induced by aldehydes and is performed by both the FA (involving *FANCD2*) and TC-NER (involving *CSB*) pathways. (Bottom) Details of the severe phenotypes observed upon genetic ablation of tier-1 and tier-2 in tandem, or both arms of tier-1.

1.2.3 Pathways of endogenous aldehyde production

Formaldehyde is a common atmospheric pollutant, however, tracing of N2-hydroxymethyl-dG adducts (N^2-HOMe-dG, formed by formaldehyde adduction of guanine, reduced to N^2-Me-dG by $NaCNBH_3$, which is quantified by LC-MS), derived from inhaled ^{13}C-labelled formaldehyde relative to endogenous ^{12}C-formaldehyde, highlights that the vast majority of formaldehyde observed in the body is produced endogenously (Lu et al., 2010). The diet is an important source of aldehydes, for example, breakdown of pectin in the cell walls of fruits and vegetables by the microbiome, produces methanol (converted to formaldehyde by alcohol dehydrogenases), as well as alcohol consumption (ethanol being converted to acetaldehyde) (Dorokhov et al., 2015). However, aldehydes are also produced from a number of other metabolic processes. These include: decomposition of folate to formaldehyde during 1-carbon metabolism (Burgos-Barragan et al., 2017); conversion of glycine and alanine to formaldehyde and acetaldehyde, respectively, by myeloperoxidase (MPO) expressed in neutrophils (Hazen et al., 1998); demethylation of lysine on histones by JmjC-domain containing demethylases and LSD1 (Klose et al., 2006); metabolism of choline by dimethylglycine and sarcosine dehydrogenases (Porter et al., 1985); glycine metabolism by CYP2E1 (Clejan & Cederbaum, 1992; Hartman et al., 2017); and lipid-peroxidation of unsaturated fatty acids (Tamura et al., 1991). These routes of production have been extensively reviewed elsewhere (Dorokhov et al., 2015; Reingruber & Pontel, 2018).

In humans, the combined action of these pathways leads to an estimated serum concentration of between 10-100μM (Reingruber & Pontel, 2018). Interestingly, this is much lower in WT mice (~4μM), but increases upon genetic deletion of tier-1 enzymes

(Dingler et al., 2020), possibly explaining why mice are more tolerant of single genetic deletions of FA/TC-NER repair enzymes than humans. Furthermore, while these pathways of endogenous production exist, their relative contribution to the endogenous formaldehyde load is unknown. Similarly, the contribution of ALDH2 to formaldehyde catabolism *in vivo* only becomes apparent when combined with deletion of ADH5. It may be that there are other enzymes involved in clearing these aldehydes whose effects are under-appreciated due to the dominant effect of ADH5.

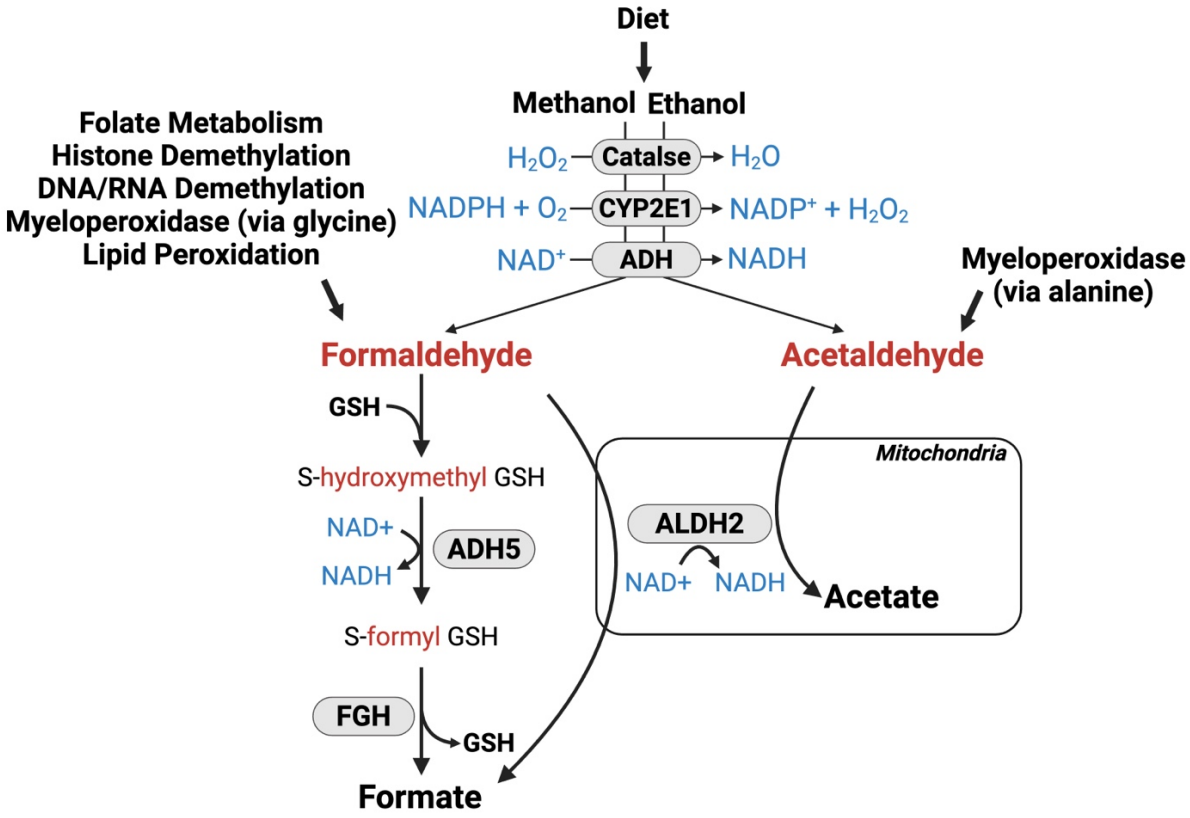


Figure 1.3 Mechanisms of aldehyde production and detoxification
Methanol and ethanol from the diet are converted to formaldehyde and acetaldehyde, respectively, by alcohol dehydrogenases (ADH), as well as catalase and cytochrome monooxygenase CYP2E1 (with ADHs likely representing the dominant route) (Dorokhov *et al*., 2015). Endogenous routes of formaldehyde production are detailed (Reingruber & Pontel, 2018). Formaldehyde reacts with free glutathione (GSH) in the cytoplasm, which produces S-hydroxymethyl-GSH, which ADH5 oxidises to S-formyl-GSH. This is subsequently converted to formate by S-formylglutathione hydrolase (FGH), regenerating free GSH. By contrast, mitochondrial ALDH2 acts on formaldehyde (or acetaldehyde) directly, producing formate (or acetate) (Reingruber & Pontel, 2018).

1.2.4 Questions on the mechanism of formaldehyde pathology

Most of the enzymes involved in the pathways of formaldehyde production are expressed in the liver, which may therefore be thought of as the centre of formaldehyde production (main site of ADH expression (Edenberg, 2007) and folate metabolism (Steinberg, 1984)). Nevertheless, the *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ and *Adh5*$^{-/-}$*Csb*$^{-/-}$ models demonstrate that formaldehyde drives pathology across tissues. This is further supported by quantification N^2-Me-dG adducts across the body of tier-1 deficient mice, finding high levels of adduction in liver, kidney and brains (Dingler *et al*., 2020; Mulderrig *et al*., 2021). This raises the questions of how much formaldehyde is produced intrinsically within a tissue and what determines the tolerance of formaldehyde in different tissues? As formaldehyde is highly detectable in the blood, it is likely that much of the formaldehyde encountered in tissues is extrinsically derived. Recent work by Holly Russell in the group showed that re-instating *ADH5* expression in the blood of *Adh5*$^{-/-}$*Csb*$^{-/-}$ mice alleviated features of neurodegeneration and kidney failure, with these tissues also having significantly fewer N^2-Me-dG adducts, indicative of reduced formaldehyde load, although still significantly higher than WT controls (H.Russell Thesis, Oxford). This highlights the significance of the blood in the systemic regulation of formaldehyde levels through the body and points to a potential avenue of therapeutic development. However, when *ADH5* expression was reinstated in the blood of *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ mice, kidney and liver function was not restored. Therefore, while a high formaldehyde load contributes to the progression of disease states, tolerance is governed by the intrinsic pathways that can be raised to clear the damage.

Aldh2$^{-/-}$*Adh5*$^{-/-}$ mice are distinct from two-tier deficient models as they have in-tact DNA repair. *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice have a milder haematopoietic defect than seen in *Aldh2*$^{-/-}$

Fancd2$^{-/-}$; *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice have 2-fold reduced LKS compartment against *Adh5*$^{-/-}$ controls, while *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ mice have a ~30-fold reduced at 8-12 weeks against *Aldh2*$^{-/-}$. Similarly, quantification of red-blood cell micronucleation sees *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ ~9.5-fold higher than WT, while *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ are ~5-fold higher, suggesting less DNA damage is incurred in the blood of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice. Nevertheless, most of these mice die shortly after birth. This suggests that, while overwhelming DNA repair, DNA damage may only feature as one element of the pathology. How formaldehyde drives damage in other tissues (kidney and liver) is not exactly known, as the peripheral tissue phenotypes have not been studied. We may expect features of tissue degeneration in overwhelming of TC-NER and FA repair pathways, as expected from the high adduct levels. Alternatively, however, formaldehyde is also capable of damaging wider cellular processes, e.g. through driving protein aggregation. This has already been suggested in case of *Adh5*$^{-/-}$ mice, where deletion of the integrated stress response (ISR) element CHOP, rescued the functional disadvantage of *Adh5*$^{-/-}$ HSCs upon transplantation (Yi et al., 2021). Although in this paper the effects were mild, *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice would be under a significantly greater formaldehyde burden, which may exaggerate these effects. Indeed, endogenous formaldehyde drives chronic expression of the anorexic hormone GDF-15 (growth/differentiation factor 15) in *Adh5*$^{-/-}$*Csb*$^{-/-}$ mice, which is responsible for the anorexic phenotype observed in these mice (Mulderrig *et al*., 2021). GDF-15 activation is downstream of p53-signalling, hence responsive to DNA damage, but can also be upregulated in response to physiological stressors, such as the unfolded protein response, nutrient deprivation, hypoxia and lipotoxicity (Lockhart et al., 2020; Wang et al., 2021). It is possible that these wider pathways drive expression of GDF-15 in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, which may account for their extreme anorexic phenotype.

1.2.5 The ALDH2 paradox in human populations

Given the necessity of functional ALDH2 in protection against FA progression and ADDS, deficiency in tier 1 protection carries the risk of developing further disorders in a manner reminiscent of the “two-hit hypothesis” of cancer development (Knudson, 1971). After losing ALDH2, there is increased propensity of acquiring second mutations, potentially in DNA repair genes, and increased dependence on functional mechanisms of DNA repair to protect against aldehyde induced damage. Indeed, profiling of oesophageal squamous cell carcinomas (ESCC) found that alcohol-consuming patients carrying the *ALDH2*2* allele, exhibit characteristic elevation of the mutational signature SBS16, which is frequently observed withinin driver genes of ESCC in alcoholics, such as *TP53* (Moody et al., 2021). More broadly, *ALDH2*2* carriers are at significantly increased risk of oesophageal, head and neck, stomach and lung cancers if alcohol is consumed (Im et al., 2022). Remarkably, despite these risks, the *ALDH2*2* allele is carried by 540 million people worldwide, originating from individuals of Han Chinese descent (Brooks et al., 2009). This raises an interesting evolutionary question as to how this variant is so common when its disadvantages are so clear. Even more so, approaches based on the measuring average haplotype lengths, suggest that the frequency of this allele has increased rapidly within the last 2000-3000 years (Cong et al., 2022; Field et al., 2016; Luo et al., 2023; Okada et al., 2018; Taliun et al., 2021). This raises the possibility that the *ALDH2*2* variant may convey an underappreciated positive effect that supports its positive selection human populations.

1.3 Endogenous formaldehyde as a driver of haematopoietic ageing

1.3.1 Haematopoiesis is dynamically maintained

Approximately 90% of daily cellular turnover in the human body takes place in blood, amounting to ~280 billion cells a day (Sender & Milo, 2021). This phenomenal proliferative output has its origins in a comparatively small number of HSCs, which can self-renew and retain a pluripotent state that allows them to differentiate into the cells of the blood system. HSCs are thought to exist as either long-term (LT) or short-term (ST) populations, classified based on their metabolic status and ability to reconstitute lineages upon transplantation (Yamamoto et al, 2018). *In vivo* lineage tracing in native haematopoiesis using genetic and fluorescent barcoding, revealed that HSCs contribute very infrequently to native haematopoiesis, contributing to downstream progeny in short-lived windows, which are asynchronously recruited such that only a limited number of HSCs contribute to blood production at any one time (Bernitz et al., 2016; Busch et al., 2015; Rodriguez-Fraticelli et al., 2018; Sun et al., 2014). HSCs not contributing to blood production are maintained in a quiescent state, believed to help prevent acquisition of somatic mutations that might facilitate neoplastic transformation, or amount epigenetic changes that reduce pluripotency (Beerman et al., 2014; Bernitz et al., 2016; Rossi et al., 2007).

Steady-state blood production is mostly carried out by multi-potent progenitors (MPP), which retain their proliferative capacity and differentiate into downstream progeny. MPPs can be subdivided based common surface markers (LKS $CD34^{+}$), but different expression of Flt3, CD150 and CD48, and distinct transcriptional profiles, which

reveals different lineage potential upon transplantation: MPP1 (also termed ST-HSCs, reflective of their limited ability to reconstitute lineages upon serial transplantation), MPP2 (megakaryocyte/erythroid biased), MPP3 (granulocyte/macrophage biased) and MPP4 (lymphoid biased) (Pietras et al., 2015). Single cell transplantation experiments and *in vivo* lineage tracing of HSC populations, revealed that the lineage biases that determine MPP fate are already present within the upstream HSC populations (Pei et al., 2020; Rossi et al., 2005; Sanjuan-Pla et al., 2013; Yamamoto et al., 2013), which stem from alterations in the epigenetic profile of LT-HSCs (Meng et al., 2023; Meng & Nerlov, 2025).

Lineage tracing of HSC/MPPs *in utero* has added complexity to the linear relationship thought to exist between HSCs and MPPs (Busch et al., 2015; Kobayashi et al., 2023; Patel et al., 2022; Pei et al., 2017). These studies suggest that MPPs that contribute to haematopoiesis arise in early development, coincident with the timing of HSC specification and emergence from the primitive endothelium, suggesting that these populations have distinct origins. This has since been supported by phylogenetic reconstruction of HSC/MPPs based on the tracing of shared somatic variants between existing HSC and MPP clones (Kapadia et al., 2025). Over time, the number of these long-lived MPPs (mostly MPP4s) decreases, while new MPPs are increasingly derived from upstream HSCs, which are biased towards myeloid differentiation (expanding MPP2/3) (Patel et al., 2022; Rodriguez-Fraticelli et al., 2018). Furthermore, lineage tracing and single cell transplantation studies have identified LT-HSC populations that already have entrained platelet bias, producing MPP2s that can directly differentiate into megakaryocytes, without going through the canonical route of megakaryocyte-precursor specification (Meng et al., 2023; Yamamoto et al., 2018). The proportion of

these platelet biased MPP2s also increases with age, further driving downstream myeloid bias. Therefore, where haematopoiesis was once understood to derive from distinct hierarchically arranged cell populations, we now understand it to exist in a state of flux between two clonal populations, each containing substantial heterogeneity (Figure 1.4). Understanding signals that drive the processes of fate determination and population flux through life is therefore important to understanding and mitigating the drivers of haematopoietic ageing.

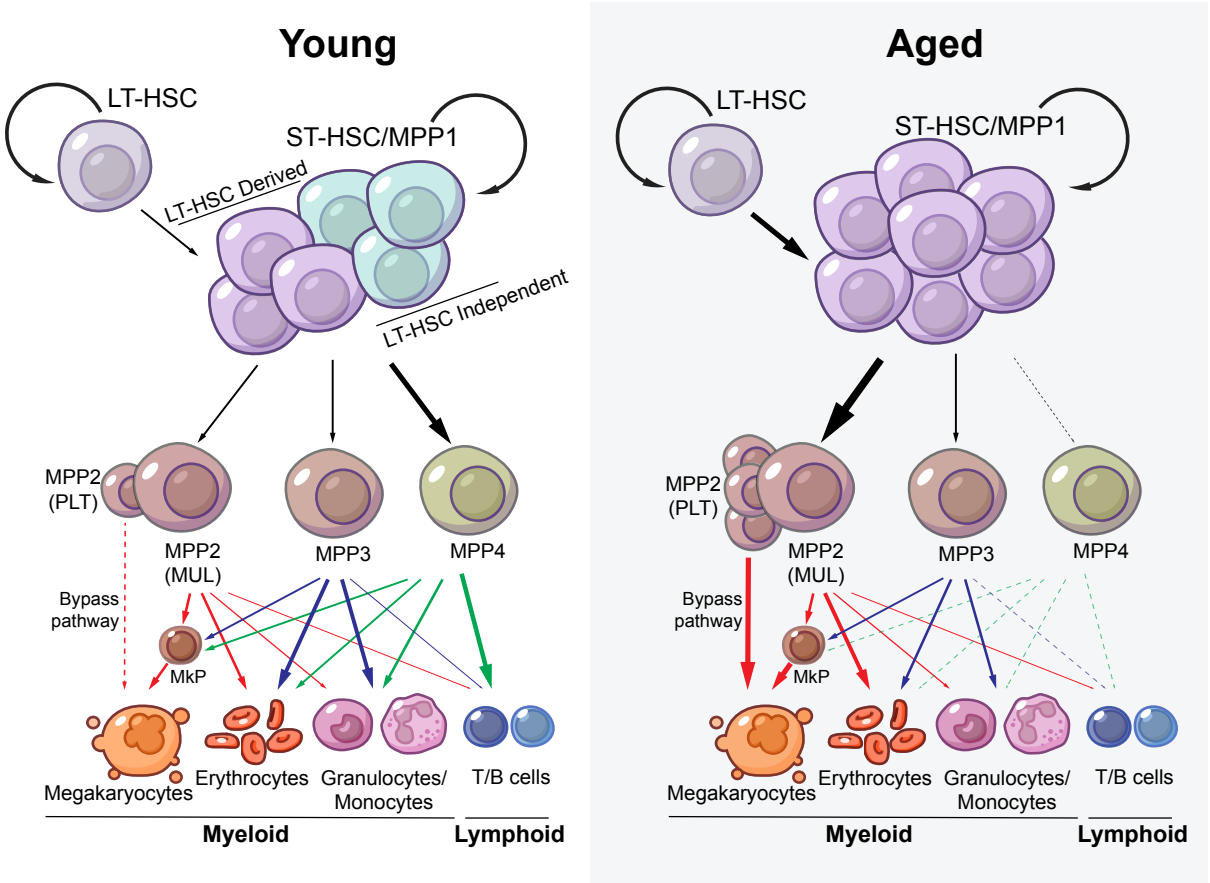


Figure 1.4 A modern model of haematopoiesis
Emerging model of haematopoiesis based on endogenous lineage tracing experiments. In young mice, the main blood producing population is composed of ST-HSCs/MPP1s derived from LT-HSCs and early progenitors concomitantly emerging from the AGM. These produce MPP2, MPP3 and MPP4, which have lineage biases, but contribute to production of all downstream populations. With age, the early-progenitors are depleted and replaced by progeny of LT-HSCs, which are increasingly myeloid bias, including enrichment of platelet biased MPP2s (PLT) that bypass the canonical route of megakaryocyte progenitor (MkP) production to produce megakaryocytes directly. MPP4s and the lymphoid compartment are no longer actively maintained as MPPs become biased for myeloid output.

1.3.2 Functional decline and loss of heterogeneity in haematopoietic ageing

The haematopoietic system loses functional capacity over time, evidenced by reduced transplantation potential of HSCs (Chambers et al., 2007; Dykstra et al., 2011; Rossi et al., 2005), cytopenia (Guralnik et al., 2004), and reduced capacity to respond to exogenous stressors (e.g. infection) (Mann et al., 2018), accompanied by an increased risk of blood cancer. Functional changes are mostly associated with the upstream HSC population, as downstream populations are shorter lived and require replenishing through life. However, ageing in the upstream populations does result in dysfunctional downstream progeny, which disrupts tissue homeostasis (Wong et al., 2023; Yousefzadeh et al., 2021) and immune protection (Elyahu et al., 2019; Montecino-Rodriguez et al., 2013). Indeed, selective depletion of myeloid biased HSCs that accumulate with age has recently been shown to rejuvenate the ageing immune system (Ross et al., 2024). Considering myeloid-biased HSCs are increasingly prevalent with age, it highlights how these age related changes are inherently dysfunctional.

This functional decline is a product of life-time accumulation of stressors, such as inflammation (Bogeska et al., 2022; Li et al., 2023), changes to the haematopoietic niche (Frisch et al., 2019; Ho et al., 2019; Mitchell et al., 2023; Pinho & Frenette, 2019), and, DNA damage (already discussed). These processes contribute to extensive epigenetic remodelling in HSCs, affecting genes related to cell cycle, response to stress, and lineage specification (Chambers et al., 2007; Mann et al., 2018; Meng et al., 2023; D. Sun et al., 2014) and is the subject of extensive study (reviewed by (Meng

& Nerlov (2025)). The existing cell context (e.g. cycling vs quiescence) affects how such signals are integrated, contributing to heterogeneity within the blood compartment. A recent meta-analysis on scRNAseq data comparing young and old HSCs, led to the description of an HSC “ageing score”, with heterogeneity in the HSC state increasing with age (Svendsen et al., 2021). This heterogeneity serves as the basis of competition between HSCs, leading to changes in stem cell architecture over time.

In humans, HSC/MPP populations deriving from clonal units can become disproportionally enriched with age, leading to an overall decline in the diversity of the HSC/MPP compartment, in a phenomenon termed Clonal Haematopoiesis (CH) (Jaiswal & Ebert, 2019). Classical understanding of CH came from the observation of recurrent mutations in a number of genes in human whole exome sequencing data (*DNMT3A, TET2, ASXL1, JAK2* and *TP53* being the most common) which increased with age, and are clinically associated with the development of haematological cancers (Genovese et al., 2014; Jaiswal et al., 2014; Jaiswal & Ebert, 2019; Xie et al., 2014). Work has since focused on trying to understand why these may confer a competitive advantage. Indeed, mechanisms through which driver events support expansion have been found to relate to resistance to factors that drive processes of ageing, such as *TP53* and *PPM1D* supporting resistance to DNA damage (Boettcher et al., 2019; Hsu et al., 2018); mutations in DNA demethylases *DNMT3A* and *TET2* support the maintenance of a pluripotent and self-renewing state, normally lost through ageing and inflammation (Jakobsen et al., 2024); mutations in *JAK2* support proliferation in response to growth factor signalling (Kralovics et al., 2005).

Adoption of unbiased whole genome sequencing (WGS) of human blood samples highlighted that known driver events in CH, only account for a small fraction of the CH observed in human populations (Zink et al., 2017). Importantly, incidence of CH without known driver mutations was still associated with increased risk of haematological malignancy and all-cause mortality. Sequencing of HSC/MPP colonies from human donors allowed detailed reconstruction of the phylogenetic relationship between HSC/MPP clones, which highlighted that after the age of 70, the blood is a composite of a large number of clonal populations (polyclonal), with most mutations located in non-exonic regions (Mitchell et al., 2022). This study and others, including longitudinal sampling of human patients (Fabre et al., 2022), and meta-analysis of patient data (Watson et al., 2020), also reveal that non-synonymous mutations in driver genes often arise early in life and often do not develop into clonal expansions until later in life, if at all. These studies have therefore matured our understanding of CH such that we now bare greater significance to the factors underpinning the functional heterogeneity within HSC/MPP populations, as crucially responsible for shaping the clonal architecture of the blood.

Precise determination of the relative importance of the stimuli behind this heterogeneity in native haematopoiesis is hard to show. These are pleiotropic and act in concert to different degrees throughout life. Exogenous challenge models, such as irradiation to induce DNA damage (Yu et al., 2016), polyinosinic:polycytidylic acid and LPS to model inflammatory stress (Bogeska et al., 2022; Mann et al., 2018), and transplantation to test functional capacity (Yamamoto, Wilkinson, Ooehara, et al., 2018) are useful for understanding how acute stress can alter haematopoiesis, but do not represent the native situation. Genetic models of accelerated ageing through

dysregulated inflammation (García-García et al., 2021) and ablated DNA repair enzymes (Rossi, Bryder, et al., 2007) are also not indicative of physiological ageing. A further challenge is that CH is not observed in mice, suggesting that the factors that drive functional decline are distinct from those driving CH (Kapadia *et al*., 2025). Phylogenetic analysis of CH in ageing mice revealed some important distinctions between the mouse and human situation that might explain this: (1) the pace of HSC expansion is greater than HSC/MPP loss through death or differentiation (~6 weeks vs ~18 weeks) in mice, facilitating exponential growth of the HSC pool and increasing growth advantage required for a clone to be detected (2) mouse HSC/MPPs accumulate mutations at a much slower rate than other somatic tissues, and far slower than observed in humans. This directly contrast to estimates in humans, where division rates are much lower (once every 2-20 months) (Lee-Six et al., 2019), and approximately equal to the rate of loss (Körber et al., 2025). This would make human clonal populations more sensitive to perturbations that increase the rate of loss. The lower mutation rate in mice (despite higher proliferation rate) suggests protection against somatic mutations, i.e. DNA damage, is a potential mechanism by which mice safeguard against clonal acquisition. Indeed, in humans, chemotherapy and smoking are both factors that increase the incidence of CH (Bolton et al., 2020; Mitchell et al., 2025; Zink et al., 2017), and CH is observed in FA patients who are highly sensitive to endogenous DNA damage (Sebert et al., 2023); in both cases, clones are enriched where the DDR is downregulated. However, in an endogenous setting, we lack mechanistic understanding of how significant endogenous DNA damage is in promoting this driverless selection.

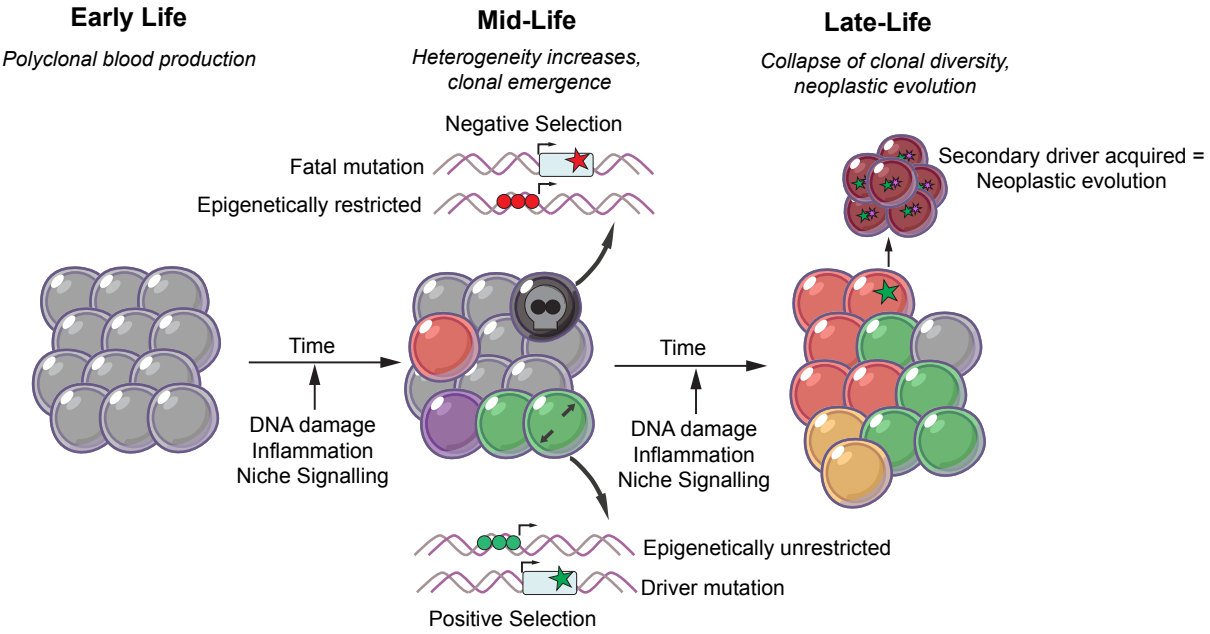


Figure 1.5 Evolution of clonal haematopoiesis
(Left) Early in life, blood production is highly diverse and polyclonal, with limited functional heterogeneity between clones. (Middle) As we age, life-time exposure to DNA damage, inflammation, and ongoing changes in the bone marrow niche microenvironment, introduce mutations and epigenetic changes to HSC/MPPs, which may confer an advantage, or disadvantage, depending on the cellular context. (Right) These forces accumulate over time, leading to expansion of clones able to tolerate changes, or which have acquired a mutation that supports proliferation, and loss of cells with reduced fitness. Clones that have a proliferative advantage are typically at greater risk of developing into a proliferative neoplasm when additional driver mutations supporting proliferation are acquired.

1.3.3 Aldehydes as drivers of haematopoietic ageing

The severe haematopoietic phenotypes of the *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ and *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ mice highlight the influence of aldehyde induced DNA damage on the blood and recent work has advanced our understanding of the mechanistic factors responsible for the phenotypes observed.

As discussed, *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ mice develop a severely anaemic phenotype, with early onset of BMF, significant mutation accumulation, and functional decay.

Interestingly, these effects were substantially recovered by deletion of p53 (Garaycoechea et al., 2018). Subsequent scRNAseq profiling of HSPCs from *Aldh2*$^{-/-}$ *Fancd2*$^{-/-}$ mice revealed that these had a prematurely increased transcriptional ageing score, with many of the genes upregulated being direct targets of p53 (Wang et al., 2023). Interestingly, many of these p53-dependent genes are also over-expressed when this modelling was applied to other published scRNAseq datasets of normally aged blood, reinforcing the contribution of DNA damage to epigenetic remodelling in normal ageing.

Aldh2$^{-/-}$*Adh5*$^{-/-}$ blood also shows signs of haematopoietic ageing, including more mutations, transcriptional upregulation of the DDR, increased transcriptional ageing score, myeloid biased haematopoiesis, and reduced transplantation capacity, although not as severe as in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ and *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ models (Dingler *et al*., 2020; Oka *et al*., 2020; Wang *et al*., 2023). This suggests that endogenous formaldehyde alone is sufficient to accelerate the ageing process, thus providing a physiologically relevant model of ageing, based on amplification of an endogenous metabolite, rather than permanent alteration of transcriptional networks (e.g. p53 or NF-κB) or DNA repair enzymes.

Recently, Felix Dingler from the group, performed WGS of granulocytes sorted from *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, and worked with Matthias Gunther (DKFZ, Heidelberg), to assess the distribution of variant allele frequencies (VAF) to infer clonal selection, as performed in Korber *et al*., 2025 (Figure 1.6a). This method involves plotting the inverse VAF against the cumulative number of single nucleotide variants (SNVs) at that frequency. When haematopoiesis occurs without clonal emergence, a distinct

upward curve shape forms, with low frequency variants having accumulated more mutations as they emerge later in life. When clonality is observed, a distinctive shoulder forms at a higher VAF, where this variant is enriched in the downstream population. The height of the shoulder is indicative of the timing of clonal emergence, as younger clones will not have accumulated as many mutations, and older clones will have more. In *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, distinctive shoulders are observed at high VAF, with a low height (Figure 1.6b). This is significant as it tells us that metabolic formaldehyde can induce clonal haematopoiesis in mice and that this emerges early in life.

However, using the *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mouse a model of haematopoietic ageing is compromised by the severe peri-natal lethality and multi-morbid effects observed in these mice. Furthermore, it is not clear if the progressive formaldehyde burden experienced in life is able to shape the clonal architecture, or if the clonality observed is a unique feature of intense selection that occurs in mice during the weaning window. Therefore, to understand if chronic formaldehyde is a driver of haematopoietic ageing and clonal haematopoiesis, which may carry significance to the 540 million individuals carrying the *ALDH2*2* variant allele, we need a less extreme model of tier-1 deficiency.

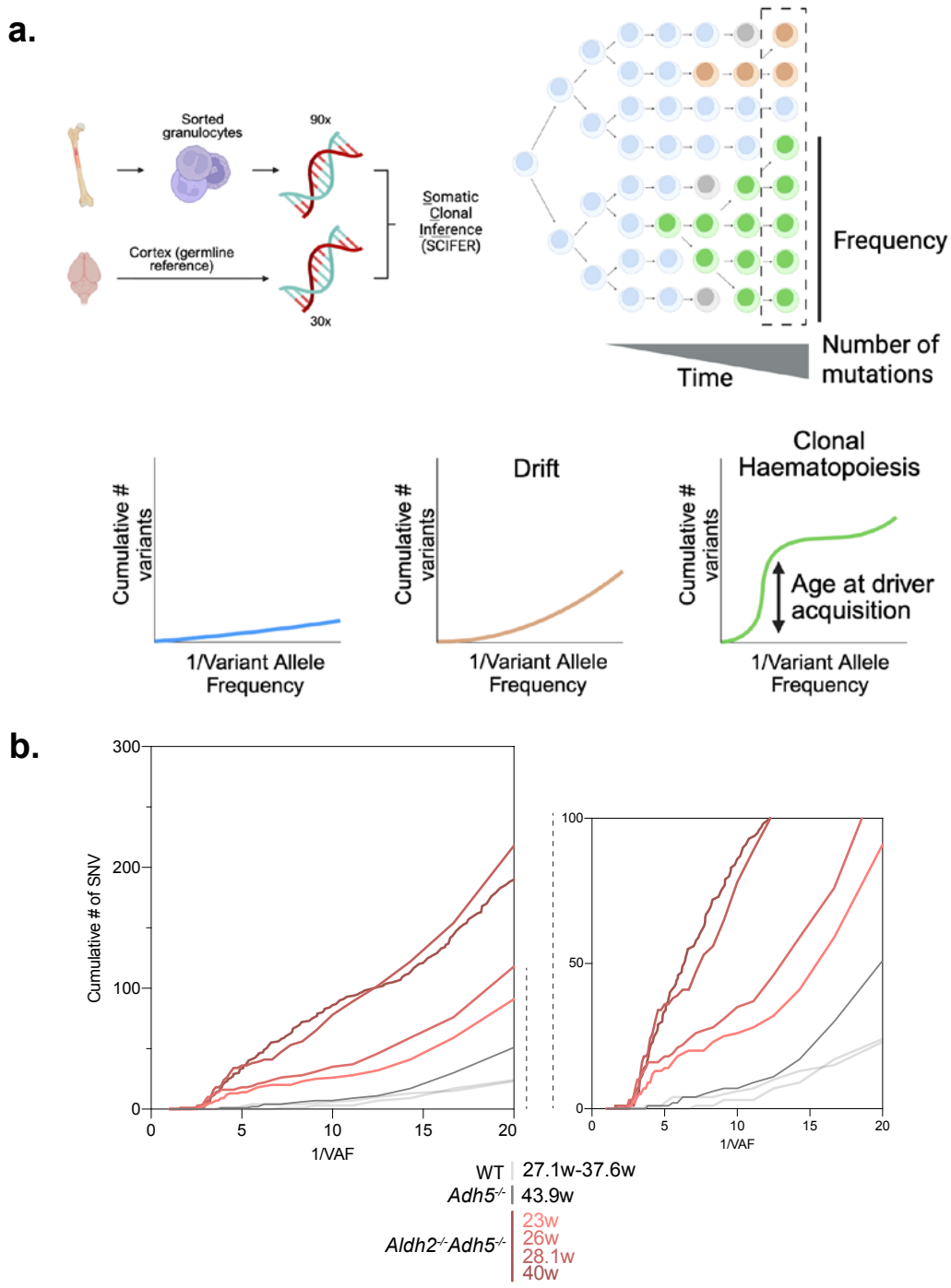


Figure 1.6 Inference of clonality in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice by WGS of granulocytes
(a) Schematic illustrating the accumulation of variants in a granulocyte population and their spread in populations, and how these shape the VAF/SNV curve. Low frequency variant alleles expanding through drift will gradually increase in frequency over time, reflected in an upwards inflected curve. Clonal expansions will be detected by more SNVs detected at that VAF, with the height reflecting the age of driver acquisition (Korber *et al*., 2025). (b) Cumulative number of SNVs and their frequency as detected by bulk (90x) WGS of bone marrow granulocytes from *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ and WT and *Adh5*$^{-/-}$ mice at their indicated ages. Graph is repeated (right) with a different cumulative # of SNV scale to emphasise the presence of shoulders at high VAF in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice.

1.4 Thesis aims

Mammalian evolution of two independent routes of clearing aldehyde mediated damage, and the potentially devastating consequences of their ablation, highlights how protection against aldehydes can be a driving force of natural selection. However, mechanistic insights into how metabolically produced formaldehyde drives features of ageing, and a plausible explanation for the surprising prevalence of the *ALDH2*2* East Asia, is lacking. Furthermore, given the number of alternative routes through which formaldehyde can be produced and responded to, we have little understanding of the relative importance in these pathways in contributing to the formaldehyde load or protection.

Therefore, my thesis has three aims:

1. **How do cells recognise and respond to formaldehyde? (Chapter 3)** I sought to develop a reporter cell line based on the expression of GDF-15 (responsive to formaldehyde and integrating multiple cellular stress response pathways) that can be developed for use in a CRISPR-Cas9 screen for genes involved in defence against and sensing of formaldehyde.
2. **Does ALDH2 deficiency confer a selective advantage? (Chapter 4)** I explore the hypothesis of enhanced immune protection mediated by ALDH2 deletion using an immune challenge study in *Aldh2*$^{-/-}$ mice.
3. **Does blood formaldehyde clearance protect against ageing of the blood and peripheral tissues? (Chapter 5)** I functionally characterise a blood-

specific version of the *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ model, increasing the formaldehyde load in the blood but circumventing the extreme multimorbid effects observed in the constitutive, enabling the study of formaldehyde induced ageing in haematopoiesis, and the significance of this on peripheral tissues.

Chapter 2

Materials and Methods

2.1 Mouse Strains

2.1.1 *ALDH2* Strains (Chapter 4)

ALDH2-deficient mice were generated from embryonic stem cells from EUCOMM (*Aldh2*$^{tm1a(EUCOMM)Wtsi}$; MGI ID: 4431566, C57BL/6N; hereafter *Aldh2*$^{-}$) as previously described (Langevin et al., 2011; Skarnes et al., 2011). Heterozygous *Aldh2*$^{+/-}$ mice were crossed with *Aldh2*$^{+/-}$ or *Aldh2*$^{-/-}$ mice to produce the experimental *Aldh2*$^{-/-}$ strain. *Aldh2*$^{+/-}$ mice do not show haploinsufficiency, hence are grouped with *Aldh2*$^{+/+}$.

Aldh2$^{tm1c(EUCOMM)Wtsi}$ (hereafter *Aldh2*^{c}) mice were generated from crossing Aldh2$^{tm1a(EUCOMM)Wtsi}$ with mice expressing constitutive FLP recombinase, inducing germline removal of FRT-lacZ-loxP-Neo-FRT cassette and restoring the functional allele, now flanked by LoxP that can be deleted upon recombination with Cre. Mice were outbred to remove the FLP allele. Generation detailed in Skarnes *et al*., 2011.

Aldh2$^{tm1d(EUCOMM)Wtsi}$ mice (hereafter *Aldh2*$^{-}$) were produced from breeding *Aldh2*^{c} mice, with mice constitutively expressing Cre-recombinase, facilitating germline deletion of removing exon 4 (critical exon). The *Aldh2*$^{-}$ allele was then maintained by interbreeding, and the Cre-recombinase was removed in this process.

2.1.2 *Adh5*$^{-/-}$ Strains

Adh5^{tm1Stam} (MGI 3033711, C57BL/6N; hereafter Adh5^{-}) mice, previously described (Liu *et al*., 2004), were gifted from Dr L. Liu at UCSF and maintained through inbreeding with periodic backcrossing.

2.1.3 *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ Strain (Chapter 5)

Aldh2^{c} and *Aldh2*$^{-}$ alleles were crossed with *Adh5*^{tm1Stam} (described above) carrying mice and maintained through interbreeding. *Vav1*$^{iCre+}$ (*Commd10*$^{Tg(Vav1\text{-}icre)A2Kio}$; MGI ID: 2449949, C57BL/6N) was first bred into *Aldh2*$^{-/-}$*Adh5*$^{+/-}$, and the *Vav1*$^{iCre+}$ allele was maintained in this colony through interbreeding, mice carrying both *Vav1*$^{iCre+}$ and *Aldh2*^{c} were never used for breeding. *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice were generated by the following parental crosses: Father *Aldh2*$^{c/c}$*Adh5*$^{-/-}$ x Mother *Aldh2*$^{+/-}$*Adh5*$^{+/-}$ *Vav1*$^{iCre+}$; Father *Aldh2*$^{c/c}$*Adh5*$^{-/-}$ x Mother *Aldh2*$^{-/-}$*Adh5*$^{+/-}$ *Vav1*$^{iCre+}$; Father *Aldh2*$^{c/c}$*Adh5*$^{-/-}$ x Mother *Aldh2*$^{+/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$; Father *Aldh2*$^{c/+}$*Adh5*$^{-/-}$ x Mother *Aldh2*$^{-/-}$*Adh5*$^{+/-}$*Vav1*$^{iCre+}$. Neither ALDH2, nor ADH5 present haploinsufficiency in mice, including in combination (N.Oberbeck's Thesis, Cambridge; Dingler *et al*., 2020), therefore the following genotypes are experimentally grouped: WT = *Aldh2*$^{+/-}$*Adh5*$^{+/-}$ *Vav1*$^{iCre+}$, *Aldh2*$^{+/-}$*Adh5*$^{+/-}$, *Aldh2*$^{c/-}$*Adh5*$^{+/-}$, *Aldh2*$^{c/c}$*Adh5*$^{+/-}$; *Adh5*$^{-/-}$ = *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$, *Aldh2*$^{+/-}$*Adh5*$^{-/-}$, *Aldh2*$^{c/c}$*Adh5*$^{-/-}$, *Aldh2*$^{c/+}$*Adh5*$^{-/-}$, *Aldh2*$^{c/+}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$.

2.1.4 Mouse husbandry

Animal experiments were performed under the approval of the MRC Weatherall Institute of Molecular Medicine animal welfare and ethical review body, and the UK

Home Office under the Animal (Scientific Procedures) Act 1867, under Project Licenses PFC007716E (expired 20th September 2023) and PP2564045 (post 20th September 2023). Mice were housed at 19-23 °C in individually ventilated cages with environmental enrichment, a light-dark cycle between 7:00 am – 7:00 pm and with SAFE® A03 diet provided *ad libetum*.

2.4.7 Timed Matings

Aldh2$^{+/-}$*Adh5*$^{-/-}$ x *Aldh2*$^{-/-}$*Adh5*$^{+/-}$ and *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ x *Aldh2*$^{+/-}$*Adh5*$^{+/-}$ crosses were set up by pooling males and females into the same cage overnight. Mice were separated the following morning, at which point females were weighed and checked for vaginal plugs. This time point was defined as E0.5. Embryo stage was defined by days post conception.

2.1.5 Mouse genotyping

Genotyping of mice strains was performed by Transnetyx (Cordova, TN) using ear-biopsies performed at 2-3 weeks of age, performing genotyping by real-time PCR with specific probes designed against the genes of interest.

Embryos were genotyped using DNA extracted from the tip of the foetal tail. DNA extraction was performed using Thermo Scientific Phire Tissue Direct PCR Master Mix #F-170S Kit, following the Dilution and Storage Protocol (DNA was stored at -20 °C until further use). The PCR was performed using this kit with the following thermocycling parameters after an initial denaturation at 98 °C for 5 mins:

Denaturation 98 °C for 5 s, annealing at 60.1 °C for 5 s, extension at 72 °C for 5 s. This cycle was performed 40 times, with a final extension at 72 °C for 1 min. Samples were stored at 4 °C until run on a 1% agarose/TBE gel for 30 mins at 95 V.

Genotyping of ALDH2 exon 4 in peripheral blood and *ex vivo* cultures was performed using KOD polymerase (Sigma-Aldrich, 71086-3) in a mutli-plex PCR reaction. The PCR reaction mix was made up as follows (for 1x25 µl reaction): 2.5 µl KOD 10X master mix, 1.5 µg $MgSO_4$, 2.5 µl dNTPs, 0.75 µl primer mix (at 10 µM); 0.5 µl KOD polymerase; 1 µl gDNA template; 16.25 µl H_2O. The PCR was performed with the following thermocycling parameters after an initial denaturation at 95 °C for 2 mins: Denaturation 95 °C for 20 s, annealing 62 °C for 10 s, extension at 70 °C for 10 s. This cycle was performed 35 times, with a final extension at 70 °C for 2 min. Samples were stored at 4 °C until run on a 1% agarose/TBE gel for 30 mins at 95 V.

2.1.6 ALDH2 western blot

10-20 mg frozen tissue was homogenized in 400µl RIPA buffer (1X TBS 1% v/v IPGEPAL-CO-630, 0.1% v/v Sodium Deoxycholate, 0.1% v/v SDS, 5 mM $MgCl_2$) supplemented with Benzonase (1:1000, Millipore) and Protease/Phosphatase Inhibitor (1x ThermoScientific™, A32965), adding 2x7 mm metal balls to the tube and shaking in a tissue lyzer at 30 Hz for 4 minutes at 4 °C. Solution was incubated on ice for 30 minutes, before the solution was cleared by centrifuging at 21,300 x g for 25 minutes at 4 °C, collecting the supernatant. Protein concentration was determined using a Pierce™ BCA Protein Assay Kit (ThermoFisher, 23225) to manufacturer's instructions, protein standards between 0-20µg/ml were used, and absorbance at 562

nm was measured on a CLARIOstar plus microplate reader. Samples were diluted to achieve a final concentration of 1.5 µg/µl, and made up to 20 µl, and 5µl Invitrogen™ NuPAGE™ LDS Sample Buffer (Invitrogen™, NP0007) with added β-mercaptoethanol. Samples were reduced for 5 minutes at 85 °C, then spun down and 20µl was loaded onto Bolt™ 4-12% Bis-Tris Plus Gels in 1x NuPAGE™ MOPS SDS Running Buffer (Invitrogen,NP0001), and run at 140V for 60 minutes. Samples were then immediately transferred onto a 0.45 µm PVDD Transfer Membrane (ThermoFisher, 88518) for 4 hours at 35 V at 4 °C. Membranes were washed in TBS, and blocked for 1 hour with 5% milk powder in TBS with 0.05% Tween-20 (TBST) (Sigma-Alderich, P1379). Membranes were then stained at with rabbit anti-ALDH2 antibody (developed in-house by Nina Oberbeck) at 1:2000 and mouse anti-vinculin (Millipore, 05-386) at 1:10,000. Staining was performed overnight at 4 °C on a rolling platform. Membranes were then washed and stained with secondary antibodies – 1:10,000 goat-anti-Mouse IRDYE® 680RD (LICORbio™, 926-68070), 1:10,000 anti-Rabbit IRDYE® 800CW (LICORbio™, 926-32211), for 2 hours at room temperature in the dark. Membranes were then washed in TBS, and blots were visualised on a ChemiDoc™ imaging system (Bio-Rad).

2.2 *In vitro*

2.2.1 Cell-Line Culture

Cell lines were maintained in Dulbecco's modified Eagle's medium with 10 % Foetal Bovine Serum (FBS) (v/v), 2 mM GlutaMAX™ and 50 µM β-mercaptoethanol. 32D cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 Medium

supplemented with 10 % FBS, 2 mM GlutaMAX™ and 50 μM β-mercaptoethanol. β-mercaptoethanol was removed from media when cells were treated for cytotoxicity.

2.2.2 Formaldehyde Cytotoxicity Assay

Cell lines were seeded in a 96 well plate in 100 μl media, at 15,000 cells/cm^2. Formaldehyde was diluted in culture media to 3x test concentration, and 50 μl of this was added to each well, making the final volume 150 μl at 1x concentration formaldehyde. Quintuplicates were used for each concentration indicated. Cells were then cultured in a humidified CO_2 incubator at 37 °C for 24 h. Viability was determined by measuring reduction of resazurin to resorufin using a CellTiter-Blue® (CTB) Cell Viability Assay (Promega). CTB was applied to each well at a final concentration of 1:20 and incubated until saturation was apparent (1-4 hours). Fluorescence was then measured at 579 nm excitation and 584 nm emission using a CLARIOstar Plus Plate Reader. Wells containing no cells were used as a baseline for fluorescence.

2.2.3 Cell Line GDF-15 Expression

Cell lines were seeded in a 6 well plate in 2ml culture media, at ~15 000 cells/cm^2. Formaldehyde and cisplatin were prepared at 3x final concentration, and 1 ml was added to each well. Cells were cultured for 24 h, after which the supernatant was removed, and cells were lysed directly in 350 μl guanidinium thiocyanate lysis buffer (RLT, Qiagen).

2.2.4 mRNA extraction and RT-qPCR for GDF-15 expression

Total RNA was extracted from cell lines and mouse tissues using QIAGEN RNeasy mini prep kit (QIAGEN), according to the manufacturer protocol. RNA was extracted in 30 µl elution buffer. RT-qPCR was performed using the NEB Luna® Universal Probe One-Step qPCR kit, containing HotStart Taq polymerase. 5' nuclease probes and primers for quantification of mRNA levels were purchased from IDT (PrimeTime™ RT-qPCR sequences and probe dyes are described in **Table 1**).

2.2.5 GDF-15-mNG transcriptional reporter integration

A GDF-15-mNG donor plasmid was cotransfected with a pX458(mRuby)-gRNA_GDF1-5 plasmid (both designed by Philip Hublitz of WIMM Genome Engineering Service) containing a gRNA targeting exon 2 of the GDF-15 locus (guide sequence: CCACTGCATATGAGCAGTCCTGG) into cell lines as described above. The efficiency of indel induction was determined by deconvolution of Sanger sequencing traces using TIDE (Brinkman et al., 2014). The GDF-15-mNG donor plasmid was linearised overnight prior to transfection with the restriction enzyme PsiI-v2 (NEB, R0744S) in NEB CutSmart Buffer and the remaining digested vector was purified with the Zymo Clean and Concentrator kit (Zymo Research, D4014). Cells were transfected with pX458(mRuby)- gRNA_GDF-15 and GDF-15-mNG donor plasmids were via electroporation with the Neon Transfection system (ThermoFischer Scientific). This was performed in a 10 µl reaction with $5x10^5$ cells, 1 µg of each plasmid, with 2 pulses at 1150 V, 40 ms duration (HepG2); 2 pulses at 1000V, 35ms (HeLa). Cells were then seeded into a 6 well plate. Control cells were prepared in the same way but not electroporated. 72 hours after transfection, cells were single cell

sorted into 96 well plates based on expression of mRuby and mNeonGreen, performed on a BD FACSARIA™ Fusion. Positively integrated clones were verified via PCR, using primers located outside of the homology arm and within the mNG sequence. Reporter activity was also verified by flow cytometry of clones, separated by replica plating and treated with cisplatin for 24h. Pharmacological enhancement of integration by homologous recombination was performed with SCR7 (inhibiting DNA ligase IV), NU7441 (inhibiting DNA-PK), Alt-R HDR enhancer v2 (IDT), treatment regimens of these drugs are quoted in the text.

2.2.6 GDF-15 exogenous reporter plasmid production

An empty pCMV6 vector was generated from an existing pCMV6-MT1 (Origene, NM_013602) vector by excision of the CMV6 promoter and MT1 locus by restriction enzymes BsrGI-HF (NEB, R3575S) and SacII (NEB, R0157S) to manufacturer's instructions. The cleaved vector was isolated using gel electrophoresis and QIAquick PCR & Gel Cleeanup Kit (Qiagen, 28506). The human GDF-15 promoter locus was cloned from HEK293 gDNA, and mNeonGreen sequence was cloned from the Donor template plasmid, by PCR. Primers were designed to produce sticky ends that could be used to ligate the vector and template DNA fragments together, also extended to introduce nuclear localisation and Gly-Ser linker sequences, where relevant. Gibson assembly of fragments was performed using NEBuilder® Hifi DNA Assembly Master Mix (NEB, E2621S) to manufacturer instructions. Plasmids were transfected into Escherichia coli in the presence of Kanamycin. Bacterial colonies with successful integration were verified by PCR and these colonies were expanded. Plasmids were purified using Qiagen Plasmid Plus Midi Kit (Qiagen, 12943), and the sequence of the

final plasmid was verified by NanoPore sequencing (Plasmidsaurus, Louisville, Kentucky). Plasmids were linearised with PsiI-v2 before electroporation (as above). Stable clones were selected by supplementation of growth media with 750 µg/ml G418 (Sigma-Aldrich, A1720) for 1-2 weeks. Surviving clones were assessed for mNG reporter activity by flow cytometry and RT-qPCR.

2.3 *Ex vivo*

2.3.1 HSC extraction from adult bone marrow

HSCs were extracted from femur, tubia, pelvis and spine of adult mice, using a protocol adapted from Wilkinson *et al*. (2020). Bones were crushed using a pestle and mortar and washed with cold PBS. Bone marrow was filtered through a 70 µm filter and made up to 50 ml with cold PBS. Cell number was then quantified with using Solution 13 (Chemometec, cat. no. 910-3013), at a 1:20 concentration, before measuring on a NucleoCounter NC-3000 (ChemoMetec, cat. no. 991-3001). 5-10 million cells were collected from these for single colour controls. Bone marrow cells were also taken at this stage for quantification/other experiments. Remaining cells were pelleted and resuspended in 350µl PBS, after which cKit+ cells were stained with anti-cKit APC conjugated antibody (0.2 mg/ml, BioLegend 105812), at 0.2 µl per $1x10^7$, for 15 minutes at room temperature. Cells were then washed and resuspended in 350µl cold PBS with anti-APC microbeads (Miltenyi, cat. no. 130-090-855) at a concentration of 0.2 µl stock per $1x10^7$, for 15 minutes at 4 °C. Cells were then washed are resuspended in 10 ml PBS. cKit enrichment was performed using a MACS LS column (Miltenyi, cat. no. 130-042-401), attached to a MidiMACS separator magnet,

pre-washed with 3 ml PBS. The cell suspension was passed onto the column through a 50 µm cell strainer. The number of cKit enriched cells was quantified as before and washed. cKit+ cells were either directly used for culture at density of 100,000 cells per well in a 96 well Corning™ CellBind tissue culture plate (5 wells per mouse) or used for sorting. Cells being sorted were incubated with the biotin-conjugated antibodies of the *Ex vivo* LT-HSC sorting panel (Table 1) at 300 µl master mix per $1x10^7$ in the dark for 15 minutes at room temperature. Cells were then washed and resuspended in the remaining antibodies for 45 minutes in the dark at room temperature. Cells were then washed and transferred to FACS tubes in 500 µl PBS with propidium iodide at 1:1000 and sorted immediately. LT-HSCs were sorted on BD FACSARIA Fusion sorter into a 96 well Corning™ CellBind tissue culture plate containing 200 µl PVA-HSC culture media (below).

Table 1 *Ex vivo* LT-HSC sorting Panel

Antigen	Conjugate	Supplier	Code	Dilution
Ly6G/Ly6C (Gr-1)	Biotin	BioLegend	108403	1:1000
Ter119	Biotin	BioLegend	116203	1:1000
CD4	Biotin	BioLegend	100507	1:1000
CD8a	Biotin	BioLegend	100703	1:1000
CD45R/CD220	Biotin	BioLegend	103203	1:1000
CD127 (IL-7Ra)	Biotin	BioLegend	135005	1:1000
CD34	FITC	ThermoFisher	11-0341-82	1:100
cKit	APC/Cy7	BioLegend	105812	1:400
Sca1	PE	BioLegend	108108	1:400

Lineage Markers	APC/Cy7	BioLegend	405208	1:400
CD150	PE/Cy7	BioLegend	115914	1:400
CD48	BV421	BioLegend	115914	1:400

2.3.2 HSC extraction from foetal liver and foetal bone marrow

The foetal liver (E16.5 and E18.5) and limb bones (E18.5) were collected from foetuses and stored in a boju jar containing cold PBS. Tissues were homogenised and collected into a 50ml tube using the wide edge of a 1 ml syringe plunger over a 70µm filter, washed through to a final volume of 10ml with cold PBS. The cell count was recorded using solution 13 (described above). Foetal bone marrow cells were stained with foetal stem and progenitor staining panel (Table 2) and sorted 25 cells/well based on LKS+. Homogenised foetal liver cells were seeded at a density of $2x10^5$ cells per well in a 96 well Cell-bind plate. Foetal cultures were grown in 200µl PVA-HSC culture medium and cultures were maintained the same as adults (below).

Table 2 Foetal stem and progenitor staining panel

Antibody	Conjugate	Supplier	Cat Number	Dilution
CD3e	FITC	eBioscience	35-0031-U500	1:100
CD4	FITC	BD	130308	1:100
CD8a	FITC	BD	553031	1:100
CD11c	FITC	eBioscience	11-0114-85	1:100
CD45R/B220	FITC	BD	553088	1:100
Gr-1	FITC	eBioscience	553127	1:100
Ter119	FITC	BD	116206	1:100

cKit	APC eF780	eBioscience	47-1171-82	1:1600
Sca1	PE-Cy7	eBioscience	25-5981-82	1:200
CD150	BV785	BioLegend	115937	1:200
CD41	BV421	BioLegend	133911	1:200
CD48	BV421	BioLegend	103427	1:200

2.3.2 *Ex vivo* PVA-HSC culture

HSCs were cultured *ex vivo* using culture conditions described in Wilkinson *et al*., 2020. The complete media contained: Ham's F-12 Nutrient Mix (Gibco), HEPES (final concentration is 10 mM, Gibco), Penicillin-streptomycin-glutamine (final concentration is 1X, Gibco), Insulin-transferrin-selenium-ethanolamine (final concentration is 1X, Gibco), Recombinant animal-free murine thrombopoietin (TPO, final concentration is 100 ng/ml, Preprotech), Recombinant animal free murine stem cell factor (SCF, final concentration is 10 ng/ml, Preprotech), Polyvinyl alcohol (PVA, 87-90% hydrolysed, used at 1 mg/ml). Cells maintained in 96-well plates were cultured in 200µl media, with the first media change performed after 7 days, then every 2-3 days.

2.3.3 *Ex vivo* proliferation analysis

Proliferation of foetal liver and adult HSC cultures was performed by flow cytometry as described in Wilkinson *et al*., 2020. 50µl (except for the day 7 measurement, in which the full well is taken) culture was collected from each well and transferred to a staining plate, mixed with the biotinylated Lineage antibodies from *Ex vivo* haematopoietic stem and progenitor staining panel (Table 3), and incubated at room

temperature for 15 minutes. After this, 2.5µl of cells from each well are separated and combined to provide the control samples. Cells are washed with 200µl PBS and resuspended in 50µl fresh antibody cocktail from the remainder of the panel. For unstained controls, only PBS was added, or FMOs, an incomplete antibody cocktail was used. Staining was performed for 15 minutes at room temperature in the dark. After this, the cells are washed in 200µl PBS and resuspended in 100µl PBS containing propidium iodide at 1:1000 concentration. Cells were then analysed on LSR Fortessa X20 using HTS.

Table 3 *Ex vivo* Haematopoietic stem and progenitor proliferation panel.

Panel A original panel, Panel B is alternative panel not using APC (see Chapter 5), Panel C is used to assess competitive proliferation

Antigen	Conjugate	Supplier	Code	Dilution	Panel
Ly6G/Ly6C	Biotin	BioLegend	108403	1:1000	Panel A/B/C
Ter119	Biotin	BioLegend	116203	1:1000	Panel A/B/C
CD4	Biotin	BioLegend	100507	1:1000	Panel A/B/C
CD8a	Biotin	BioLegend	100703	1:1000	Panel A/B/C
CD45R/CD220	Biotin	BioLegend	103203	1:1000	Panel A/B/C
CD127	Biotin	BioLegend	135005	1:1000	Panel A/B/C
CD201	APC	ThermoFisher	17-2012-82	1:400	Panel A/C
	PE	Invitrogen	12-2012-82	1:200	Panel B
cKit	BV421	BioLegend	105828	1:400	Panel A/C
	BV785	BioLegend	105841	1:200	Panel B
Sca1	PE	BioLegend	108108	1:400	Panel A/C
	FITC	BioLegend	108106	1:400	Panel B
Biotin	APC/Cy7-Streptavidin	BioLegend	405208	1:400	Panel A/B/C
CD150	PE/Cy7-CD150	BioLegend	115914	1:400	Panel A/B/C

CD45.1	BUV395	BD	565212	1:200	Panel C
CD45.2	eFluor450	ThermoFisher	48-0454-82	1:200	Panel C

2.4 Flow Cytometry

2.4.1 Haematopoietic immunophenotyping

Haematopoietic tissues were harvested as performed above (spleen and thymus were also crushed into single cell suspensions as performed with foetal livers). Peripheral blood was acquired by exsanguination and 100µl was collected into an K_3EDTA MiniCollect tubes (Greiner Bio-One, 15748169). Before centrifugation, cell counts were performed using solution 13 at a 1:20 dilution. With the exception of peripheral blood where the entire volume was used, up to $1x10^7$ cells were used for staining. Haemolysis was performed for 10 minutes at 4 °C on bone marrow, spleen and peripheral blood samples using 10ml haemolytic buffer (155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA). Cells were washed and resuspended in associated pre-mixed staining buffers in FACS Buffer (1% BSA in PBS) at 100µl/$1x10^7$ cells.

HSC and progenitor staining was performed on bone marrow cells using haematopoietic stem and progenitor panel (Table 4). CD16/32 was added at in half of the staining volume for 15 minutes at 4 °C, before the remaining cocktail was added and cells were stained for 45 minutes at 4 °C in the dark. Cells were washed in 3ml FACS Buffer and the pellet was resuspended in Step-BV421(1:100 in PBS/BSA) and stained for a further 15 minutes in the dark at 4 °C. Cells were washed with a further

3ml FACS Buffer, and the pellet was resuspended in 400µl FACS Buffer on ice and immediately proceeded to analysis by flow cytometry. The maturation and mature lineage staining panel (Table 5) was applied to the bone marrow and spleen followed the same protocol as above, but the staining cocktail was applied in full and samples were resuspended in a final volume of 200µl FACS Buffer. The maturation and mature lineage panel was applied to the spleen and peripheral blood suspensions in the same manner as the lineage maturation panel. The thymic development panel (Table 6) was performed with a 15-minute incubation of the main antibody panel and resuspended in 300µl FACS Buffer. Foetal haematopoietic tissues were analysed with the foetal haematopoietic stem and progenitor staining panel (Table 3) and stained in 50µl cocktail per $1x10^7$ cells for 45 minutes, before resuspension in 400µl PBS. Immediately before samples were analysed by flow cytometry, 0.5 mg/ml 7-AAD was added to the cell suspension and the cells were passed through a 70µm cell strainer.

Table 4 Haematopoietic Stem and Progenitor Staining Panel

Antigen	Conjugate	Supplier	Code	Dilution
CD3e	FITC	eBioscience	35-0031-U500	1:440
CD4	FITC	BD	130308	1:440
CD8a	FITC	BD	553031	1:440
CD11b	FITC	BD	553310	1:440
CD11c	FITC	BD	11-0114-85	1:440
B220/CD45R	FITC	BD	553088	1:440
FceR1a	FITC	eBioscience	11-5898-85	1:440
Gr-1 (Ly-6G)	FITC	eBioscience	553127	1:440
Ter119	FITC	BD	116206	1:440
CD41	FITC	BD	553848	1:100
CD48	Biotin/Strep-BV421	BioLegend	103410 / 405225	1:100
CD127 (IL7Ra)	BV605	BD	135025	1:200
CD150	BV786	Biolegend	115937	1:200
CD135 (Flt3)	PE	eBioscience	12-1351-82	1:200
Sca1	PECy7	eBioscience	25-5981-82	1:200
CD16/32 (FcgRII/III)	APC	eBioscience	17-0161-82	1:100
CD34	AF700	eBioscience	56-0341-82	1:10
CD117 (cKit)	APCeF780	eBioscience	47-1171-82	1:1600

Table 5 Maturation and Mature Lineages Staining Panel

Antibody	Conjugate	Supplier	Code	Dilution
CD4	BV421	BD	740024	1:200
CD3e	BV605	BD	563004	1:200
Mac/CD11b	BV711	BD	563168	1:200
CD8a	BV785	BD	563332	1:200
Gr1/Ly-6G	FITC	eBioscience	553127	1:200
CD71	PE	eBioscience	12-0711-82	1:200
Ter119	PE-Cy7	BD	557853	1:200
B220/CD45R	APC	BD	553092	1:200
IgM	APCef780	eBioscience	47-5790-82	1:200

Table 6 Thymic Development Staining Panel

Antibody	Conjugate	Supplier	Cat Number	Dilution
CD4	BV421	BD	740024	1:200
CD44	BV605	BD	563058	1:200
B220/CD45R	FITC	BD	553088	1:200
CD11b	FITC	BD	553310	1:200
CD3e	FITC	eBioscience	35-0031-U500	1:200
Gr-1 (Ly-6G)	FITC	eBioscience	553127	1:200
Ter119	FITC	BD	116206	1:200
CD25	PE-Cy7	eBioscience	25-0251-82	1:200
CD8	APC	BD	553035	1:200
CD45	AF700	BD	560510	1:200

2.4.2 Micronucleus assay

20µl blood was collected from mice via tail vein sampling, or exsanguination, and added to 110µl Heparin (500 USP/ml in PBS, Sigma H9399) on ice (all subsequent steps are performed at 4 °C unless otherwise stated). 120µl of blood suspension was added to 1.35ml methanol at -70 °C, and samples were fixed at -70 °C for a minimum of 12 hours before analysis. 500µl fixed blood was washed with 3ml bicarbonate buffer (0.9% NaCl, 5.3 mM $NaHCO_3$). Samples were washed and resuspended in 75µl bicarbonate buffer. For each sample, 20µl was mixed with 80µl staining mix (staining mix composed of 72µl bicarbonate buffer, 1µl FITC-CD71 (ThermoFisher, R17217.1.4), 7µl RNase (Sigma, R4642)) and stained in the dark for 45 minutes. The samples were washed with 700µl bicarbonate buffer and resuspended in 250µl bicarbonate buffer. Propidium Iodide (Sigma P4864) was added to each sample to a final concentration of 5µg/ml, 10 minutes before analysis on BD Fortessa X20 with HTS.

2.4.3 Nuclear ploidy analysis

0.25g liver was homogenised through a 40µm cell strainer into a 50 ml falcon using cold LA buffer (250mM sucrose, 5mM $MgCl_2$, 10mM Tris-HCl pH 7.4). Samples were centrifuged for 10 minutes at 600 x g, 4 °C. The supernatant was discarded and pellet fully resuspended in 7ml cold LA buffer and spun again. The supernatant was discarded and fully resuspended in 9 volumes cold LB buffer (2M sucrose, 1mM $MgCl_2$, 10mM Tris-HSC pH 7.4). Cells were the centrifuged at 16,000 x g for 30 minutes at 4 °C, forming two distinct layers, corresponding to cell debris (upper) and nuclei (lower). The upper layer was removed, and nuclei were fully resuspended in

500µl LA buffer, then transferred to a 15ml falcon tube. Cells were fixed in 5ml 70% ethanol (v/v in water), applied in dropwise manner, and stored at -20 °C until analysed. Fixed nuclei were spun for 5 minutes at 300 x g, supernatant discarded, and nuclei resuspended in 1ml PBS. Nuclei were quantified using solution 13 on NucleoCounter (Chemometec), and nuclear concentrations were equalised for each sample. Propidium iodide was added to final concentration of 40µg/ml, and RNase A was added to a final concentration of 100µg/ml. Nuclei were then incubated for 20 minutes on ice in the dark and analysed by flow cytometry on Attune NxT Flow Cytometer (ThermoScientific).

2.5 Sister Chromatid Exchange Assay

25mg BrdU slow-release pellet (Innovative Research of America) was surgically implanted subcutaneously into 8-10 week-old mice. After 24h, mice were given 100µl colchicine (0.5% w/v in water, Sigma) via intraperitoneal injection, and mice were sacrificed after 30 minutes by cervical dislocation. The femurs were harvested into ice-cold PBS. Bone marrow cells were flushed from the femurs with cold PBS through a 70-µm cell strainer and spun at 300 x g for 10 minutes at 4 °C. The cell pellets were resuspended in 10 ml pre-warmed hypotonic solution (75mM KCl) and incubated at 37 °C in a water bath for 15 minutes. After incubation, 1ml fixative (3:1 methanol:acetic acid) was added dropwise to hypotonic suspension and the suspension was centrifuged for 10 minutes at 250 x g. Supernatant was removed leaving 50µl volume, and a further 3ml fixative was slowly added in a dropwise manner. Fixative was then added to a final volume of 10 ml and cells were incubated at room temperature for 30

minutes before being stored at -20 °C until further use. Before transfer to slides, cells were centrifuged at 300 x g for 5 minutes and the fixative was replaced. The samples were centrifuged again and resuspended in a small volume of fixative. Using a glass pipette, 1-2 drops of cells were place onto clean Superfrost slides (Avantor 631-0912), followed by 1-2 drops fixative and allowed to dry. Slides were washed in PBS for 5 minutes at room temperature, then stained for 10 minutes with 5µg/ml Hoechst 33258 (H3568, Molecular Probes) in PBS. Slides were rinsed in Sorensen's buffer (67mM solution pH 7: 100ml H_2O, 50ml 133mM KH_2PO_4, 50ml 133mM Na_2HPO_4), then placed horizontally and covered with 1cm fresh Sorensen's buffer. Slides were irradiated 5cm away from 15 W 365nm UV light source at 50C for 20 mintues. The slides were removed and placed in 2xSSC buffer (0.3M NaCl, 30mM $Na_3C_6H_5O_7$, pH 7) at 60 °C for 1 hour. Slides were then rinsed in Gurr's buffer and stained in 10% Giemsa in Gurr's buffer for 10 minutes at room temperature. Slides were washed with water and once dry, mounted in Entellan mountant. Metaphase events were imaged by light microscopy on an Olympus BX51 at 40X magnification with oil-emersion, images were analysed using FIJI.

2.6 Immunology methods

2.6.1 Acetaldehyde modification of BSA and HEL

44mg of Bovine Serum Albumin (BSA; Sigma-Aldrich, A7030) and Hen egg lysozyme (HEL; Roche, 10837059001). dissolved in 10ml PBS and filtered through a 22µm syringe filter. 5ml of this preparation was mixed with 500µl 550mM acetaldehyde (final concentration at 500µM), and the remaining 5ml HEL in PBS was mixed with 500µl

PBS (1X Gibco™ 10010031). These were transferred to 1.5ml twist capped tubes and mixed in ThermoMixer (Eppendorf™ 5382000031) at 37 °C Celsius for 12-16 hours. After which, the antigen preparations were transferred to -70 °C Celsius until further use.

2.6.2 Antigen purification by size exclusion chromatography (SEC)

Incubated antigens were thawed at room temperature, before being centrifuged at 10,000 x g for 5 minutes. The supernatant from individual tubes was collected and pooled into a 5ml tube. Size exclusion chromatography (SEC) was performed on a Cytiva HiLoad 16/600 Superdex 75 pg size exclusion chromatography column (Cytiva, 28989333), connected to an AKTA avant Chromatography System. The column was washed with 1.5 column volume (CV) PBS, before 2.5 ml modified antigen was loaded through the 5 ml manual injection tube. The fractions were eluted by passing 2 CV PBS through the column, collecting 1.8ml fractions of eluted antigen into a 2ml V-bottom 96 deep well plate (Thermo Scientific™ 95040452) and stored at 4 °C. The column was then washed with a further 1 CV PBS and a deep cleaning procedure was performed according to the manufacturer's instructions, before a fresh antigen was loaded.

Identification of the monomeric fraction was performed by loading 10µl of each eluted fraction, mixed with 4X NuPAGE™ LDS Sample Buffer (ThermoFisher™ NP0007), onto a Bolt™ Bis-Tris Plus Mini Protein Gels, 4-12% (Invitrogen™ NW04127BOX). Fractions were assessed only from the posterior half and tail of the absorbance curve generated by the AKTA. The gel was the run for 1 hour at 140V, then stained with

Coomassae Blue. Fractions that contained a single band corresponding to the size of the monomeric antigen, were pooled together and sterile filtered using a syringe filter. The protein concentration was quantified using a NanoDrop™ One (ThermoFisher 701-508112), and the protein was then made up to 500 µg/ml with sterile PBS. The concentrated protein was then diluted into 500 µl aliquots and stored at -70 °C until further use.

2.6.3 Immunisations

C57Bl6/c mice aged 8-12 weeks were immunised with 50µg of modified or unmodified monomeric HEL, BSA in 50µl PBS via subcutaneous (s.c.) injection. Mice received a booster immunisation of the same dose of antigen via s.c. injection on weeks 2, 4 and 6. 25µg HA in 50µl PBS was administered by intramuscular (i.m.) injection on weeks 0, 3, 6, 9 and 18. Addavax (InvivoGen, vac-adx-10) immunisations were performed in 1:1 ratio Addavax:PBS, prepared fresh and used immediately prior to immunisation. Addavax was only used in priming reaction, with 50µg/50µl BSA or 25µg/50µl HA used for subsequent boosters of Addavax primed mice. Mice were exsanguinated on the final week of their protocol and serum was collected and stored at -70 °C. Serum was stored at -70 °C.

2.6.4 ELISA

ELISA was performed using 5µg/ml unmodified antigen absorbed into 384-well MaxiSorp Plates (ThermoScientific™ 464718) in 20µl freshly prepared coating buffer (8.4g $NaHCO_3$, 9.6g Na_2CO_3, in 1L Mill-Q-H_2O). Plates were sealed using a Polyolefin

seal (Starlab, E2796-9792) and spun at 2000 x g for 1 minutes, then incubated overnight at 4 °C. The following day, plates were equilibrated to room temperature, coating buffer was removed, and the plates are washed 3 times with freshly prepared wash buffer (1X PBS, 0.05% Tween-20 (Merck, P7949)) using a 50™ TS Microplate Washer (BioTek, 40-301). The plate washer was set with the default wash settings with some modifications: prime was performed before every round of washing, washing was performed with 30 seconds shaking, dispense step was performed at the start. The wells were blocked with 70 µl freshly prepared 1% BSA in PBS (HEL/HA) or 2% Milk in PBS (BSA) and shaken at 1000 rpm for 1 hour. Blocking solution was then removed and 30µl serum samples was added. Serum samples were prepared first as a 1:40 dilution (5µl serum into 195µl blocking solution), followed by a 2.5-fold dilution, then 4-fold serial dilution series was performed, such that the final dilution was 1:409,600. A standard curve was prepared for each plate using pooled serum from high responding mice, or a mouse monoclonal antibody diluted 1:1000 then in a 3-fold serial dilution (BSA) or 10-fold (HEL) series (HEL – N/A-CP150; BSA – Proteintech 66201-1-1g). Serum incubation was performed at room temperature for 1 hour while shaking. Serum was then removed, plates were washed 5 times, and IgG detection antibody (anti-mouse IgG-HRP conjugate) was applied at a concentration of 1:40 000 blocking solution, shaking for 45 minutes at room temperature. Detection antibody was then removed, and the plate is washed 7 times. Signal was produced using 50µl per well TMB-ELISA, incubating in the dark for 30 minutes. The reaction was then stopped using 25µl per well 2M HCl. Signal was measured using a ClarioStar Plate reader, measuring absorbance at 450nm.

Absorbance readings were normalised against background signal, and a standard curve was generated from the positive control curve. Antibody titres at the highest dilution at which the absorbance was greater than or equal to the threshold set by the positive control (where the absorbance becomes exponential).

2.7 Animal experiments and processing

2.7.1 Methanol treatment

Methanol treatment of mice was performed by intraperitoneal (IP) injection with 2 g/kg methanol for the time points indicated before mice were sacrificed by cervical dislocation and organs were dissected and stored at -70 °C.

2.7.2 Blood Sampling

Serial blood sampling from live mice was performed by tail vein blood sampling, collecting 10% blood volume (body weight (kg) x 58.5 (ml) x 0.1) at biweekly or monthly intervals. Blood was collected into K_3EDTA MiniCollect tubes (Greiner Bio-one) for blood counting, or Microvette® 100 Serum CAT or 500 CAT-gel tubes (SARSTEDT) for serum collection. Serum was isolated by centrifugation at 10,000 x g for 5 minutes, and supernatant was stored at -70 °C.

2.7.3 Serum biochemistry

All serum biochemical analysis was performed by Core Biochemical Assay Laboratory (CBAL) (Cambridge). Urea, creatinine, albumin, bilirubin, aspartate aminotransferase and alanine aminotransferase were performed using a Siemens Dimension EXL200 analyser. GDF-15 was measured using DuoSet ELISA (R&D system), developed in-house.

2.8 N^2-Me-dG quantification

10-15mg tissue was homogenised for 4 minutes at 30 Hz (Qiagen/Retsch tissue lyser) in 700µl Qiagen Purgene Cell Lysis Solution (Qiagen, 158113) with 4µl Proteinase K (20mg/ml; Qiagen, 158146), then incubated at 37 °C for 30 minutes at 900rpm in a ThermoMixer. 4µl RNAse A (Qiagen, 19101) was added and incubated for a further 60 minutes. Lysates were transferred to a tube containing 265µl cold Protein Precipitate Solution (Qiagen, 158126) and cooled on ice for 5 minutes. Samples were centrifuged at 21,300 x g for 5 minutes at 4 °C and the supernatant was transferred to a tube containing 600µl cold i-PrOH (Merck, 67-63-0). Samples were inverted and held on ice for 5 minutes, before centrifugation at 21,300 x g for 2 minutes. The supernatant was discarded, and tubes samples were placed open upside down for 5 minutes, allowing the remaining i-PrOH to evaporate. The pellet was resuspended in 600µl 70% ethanol (v/v in water) and centrifuged at 21,300 x g for 2 minutes. The supernatant was discarded, and remaining ethanol was evaporated as before. Samples were fully resuspended in 500µl $NaCNBD_3$ (50mM, pH 5.2 in 200mM NaOAc) and incubated in a ThermoMixer at 37 °C for 48 hours at 900 rpm.

After incubation, 900µl i-PrOH was added to each sample and pelleted by centrifuged at 21,300 x g for 5 minutes. The supernatant was discarded, and tubes were allowed to dry fully, and pellets resuspended in 900µl 70% ethanol. Samples were centrifuged at 21,300 x g for 5 minutes, the supernatant removed and redissolved in 80µl ultra-pure water overnight at room temperature. DNA concentration was quantified by nanodrop. 10µg DNA used for digest and made up to volume of 76µl with ultra-pure water, before pre-mixed digestion cocktail was added: 2 U shrimp alkaline phosphatase, 0.5µl (New England Biolabs), 3.8 U snake venom phosphodiesterase I, 1µl (Sigma, P3243) and 10 U DNase I, 0.5µl (Roche), 10x Digestion Buffer, 10µl. 5µl ^{15}N-N^2-MedG and ^{15}N-dA internal standards were added to each digest. A standard curve was generated using 5µg genomic DNA extracted from WT 32D cells spiked with ^{15}N-dA (dilution series: 272nmol-8.5nmol), ^{15}N-N^2-MedG (dilution series: 100 fmol-0.24fmol) over 6 points, as well as a sample without added standard. Digestion was performed for a minimum 18 hours at 37 °C, before the samples were filtered through a 1000 MWCO Vivacon® 500 column (Sartorius, VN01H32) and centrifuged at 16,000 x g at 12 °C for 20 minutes. Flow through was transferred to mass spectrometry vials (Waters, 186000385c) and analysed on TSQ Altis™ Triple Quadrupole Mass Spectrometer with DIONEX UltiMate 3000 RS Autosampler.

2.9 Granulocyte sorting and WGS

~10^7 bone marrow cells were collected and haemolysed, before staining with Granulocyte Sorting Staining Panel (Table 7) in 1% BSA in PBS for 15 minutes in the dark at 4 °C. 0.5 mg/ml 7-AAD was added to identify viable cells and $2x10^5$-$1x10^6$ granulocytes were sorted into PBS on FACSAria Fusion (BD). Cells were pelleted and

the supernatant was removed; the cell pellet was sorted at -70 °C. Brain cortex from the same animal was dissected to be used as a germline reference. gDNA was prepared using Zymo Quick-DNA Miniprep kit (Zymo, D3024). Sequencing libraries were prepared by Novogene and sequenced on platform Illumina NovaSeq6000 by Novogene. WGS data processing and variant calling was performed by Matthias Gunther (DKFZ, Heidelberg), as previously described Korber *et al*. (2025).

Table 7 Granulocyte Sorting Staining Panel

Antigen	Conjugate	Supplier	Code	Dilution
Gr1.Ly-6C	FITC	BD	553127	1:200
B220/CD45R	APC	BD	553092	1:200
CD3e	BV605	BD	563004	1:200
Ter119	PE-Cy7	BD	557853	1:200

2.10 List of Primers

Table 8 List of primers used through study

Mouse Genotyping		
ALDH2	AAACTTTGCACACACTGTCCC CCCAGATCCAACTGTTAGGAATAC GCTTCACTGAGTCTCTGGCATCTC ACTGTCGAGCCTTTAGCCAG	
ADH5	TCTCTTGATTCCCACTTTGTG AATGGACGAGTGGAGATTTC AGTCTGGAGTCACTCATTTTG	
GDF-15 endogenous promoter integration verification		
Primer Set 1	5' Integration	CTGCACCTGCGTATCTCTCG ACCCGATATGAGGGACCAGA
	3' Integration	CACCTACACCTTTGCCAAGCC GGCACTTGGTTGCAGATAGAC
Primer Set 2	5' Integration	GTTGGGCATGATGGCATGCACTTGTAGTC CGCTGCGGCGTCAGCTCAGC
	3' Integration	GACTTCCTCTGCCCTCTCCGGATC GGACGAGCTGTACAAGGCTAGCTGAG
GDF-15 Cas9 cutting verification primers		
AGGTCAGGGAGAGGAGGATC GCTTGGCGTGTTAGGGGAA		
GDF-15 cloning exogenous promoter		
Promoter 1	GDF-15	TTGTATCTATATCATAATATGTACATTCCTCCGCCCTCTGGGTTCAA TCCGGAGCCACCGCCTCCTTGGAAACTTGCGCGGCTCGCC
	mNG	AAGGAGGCGGTGGCTCCGGAGGCGGTGGCTCCGTCGACATGGTGAGCAAGGG TGTTCAGGAAACAGCTATGACCGCGGCTCAGCTAGCCTTGTACAGCTCGTCCATGC
Promoter 2	GDF-15	TAATATGTACATTTAGTCTAGAACTCTTGACGTCA CGCCCTTGCTCACCATGAGTTCTTGCCCGGGCATG
	mNG	CCGGGCAAGAACTCATGGTGAGCAAGGGCGAGGAG AGCTATGACCGCGGCCTACACCTTGCGCTTCTTCTTGGGCTTGTACAGCTCGTCCAT
No Promoter	mNG	TGTTCAGGAAACAGCTATGACCGCGGCTCAGCTAGCCTTGTACAGCTCGTCCATGC GTTGTATCTATATCATAATATGTACATTTGTCGACATGGTGAGCAAGG
qPCR Primers		
Human GDF-15(6FAM/Zen Iowa Black)	GAAGATTCGAACACCGACCTC CCCGAGAGATACGCAGGT Probe CGGATCCCAGCCGCACTTCT	
Human TBP (5-SUN/Zen Iowa Black)	GCTGTTTAACTTCGCTTCCG CAGCAACTTCCTCAATTCCTTG Probe TGATCTTTGCAGTGACCCAGCATCA	
Mouse GDF-15(6FAM/Zen Iowa Black)	ACTCGAACTCAGAACCAAGTC AGACCCTGACTCAGCGA Probe ACCCCAATCTCACCTCTGGACTGA	
Mouse TBP (5-SUN/Zen Iowa Black)	TGTATCTACCGTGAATCTTGGC CCAGAACTGAAAATCAACGCAG Probe ACTTGACCTAAAGACCATTGCACTTCGT	
mNG (6-FAM/Zen Iowa Black)	GGGTAAGTCTTCTTGCTCCTG GACTGTGAACTACCGCTACAC Probe ACGAGGGAAGCCACATCAAAGGAG	

Chapter 3

Reporting Formaldehyde Stress in Cell Lines

3.1 Introduction

Current understanding of biological protection against formaldehyde is limited to a small number of metabolic and DNA repair enzymes. Discovery of genes involved in detoxification and protection is made challenging due to the significant redundancy observed between them *in vivo*. RNAseq experiments of two-tier and tier-1 deficient mice highlight significant transcriptional changes in response to formaldehyde stress, emphasising a gene regulatory network that is sensitive to formaldehyde that we poorly understand (Dingler et al., 2020; Mulderrig et al., 2021). CRISPR/Cas9 knock-out screens are a useful tool for rapidly identifying genes involved in these responses (Bock et al., 2022). However, traditional synthetic lethality screens are limited in their sensitivity to detect genes that induce subtle changes in gene regulatory networks. To get around this limitation, transcriptional reporters can be used to detect changes in expression of a gene of interest, in response to genetic perturbation (Bock *et al*., 2022).

GDF-15 is an anorectic hormone induced in response to a broad range of stressors, including DNA damage, unfolded protein stress, nutrient deprivation, lipotoxicity, hypoxia, and is induced by common pharmaceuticals (e.g. NSAIDS and metformin) (Wang *et al*., 2021). These pleiotropic means of GDF-15 induction are facilitated by its complex promoter, mapped in Figure 3.1. Notably, both chronic and acute formaldehyde exposure can induce GDF-15 expression and secretion, as observed in *Adh5*$^{-/-}$*Csb*$^{-/-}$ deficient mice, and in *Adh5*$^{-/-}$ mice treated with methanol (Mulderrig *et al*., 2021). By integrating stress signals from across the cell, GDF-15 expression levels may serve as a transcriptional biomarker for the overall burden of formaldehyde-induced stress. This property could therefore be exploited in reporter-based CRISPR/Cas9 screen, to identify genes involved in (1) formaldehyde detoxification and protection (increased GDF-15 on deletion) (2) formaldehyde production and stress signalling (decreased GDF-15 on deletion).

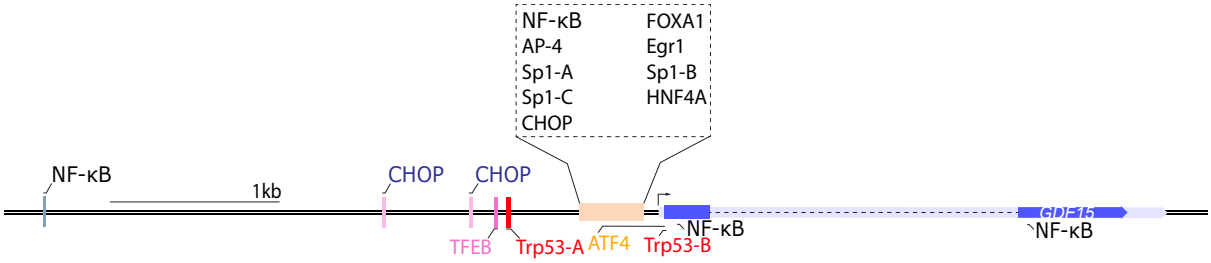


Figure 3.1 Map of human GDF-15 locus
Known GDF-15 promoter binding motifs indicated (SP1, NF-kB (Böttner et al., 1999), SP1, EGR1 (Baek et al., 2001, 2004), p53 (Osada et al., 2007), TFEB (Kim et al., 2021), ATF4 (Miyake et al., 2021), HNF4A (Delaforest et al., 2019), CHOP (Osman et al., 2023).

In this chapter, I look to identify a cell line that induces GDF-15 transcription upon formaldehyde treatment and attempted to develop a transcriptional reporter of GDF-15 expression, by either targeting the endogenous locus, or using exogenous promoter constructs.

3.2 Results

3.2.1 GDF-15 expression in cell lines

To have a cell line that would act as a suitable candidate for a genome wide screen, it needed to express GDF-15, which is significantly elevated in response to formaldehyde, and be edited in culture with high efficiency. I sought to identify a suitable cell line by examining the basal, and formaldehyde induced, GDF-15 expression by RT-qPCR using TATA-box binding protein (TBP) as a reference (Nygard et al., 2007). To serve as a positive control for DNA damage induced GDF-15 expression, I also treated cell lines with cisplatin. These treatments were performed for 24h. This was initially performed on HepG2, HEK293, Jurkat, HK-2, 32D and RPE1 cell lines, for which *Adh5*$^{-/-}$ cell lines had been generated in the lab, and have been gene edited with CRISPR-Cas9 (Malong et al., 2025; Olivieri et al., 2020; Shang et al., 2018; Veach & Wilson, 2018). *Adh5*$^{-/-}$ cells would be the preferred choice as this removes one arm of redundancy in the protection against aldehydes, which might increase the impact of a second mutation in a CRISPR/Cas9 screen. Across all cell lines, GDF-15 expression was only induced in response to either formaldehyde or cisplatin in HepG2 cells (Figure 3.2). *ADH5*-deficiency increased sensitivity of all cell lines to formaldehyde, and in HepG2, this also corresponded to an increase in GDF-15 expression.

The highest formaldehyde concentration tested in HepG2 (150μM) was not substantially beyond the EC_{50} of formaldehyde sensitivity (135μM), although the measurements of formaldehyde toxicity in HepG2 *Adh5*$^{-/-}$ cells had considerable

variability. As this cell line had the greatest expression of GDF-15 and is amenable to gene editing, I selected it to generate a transcriptional reporter of the GDF-15 locus.

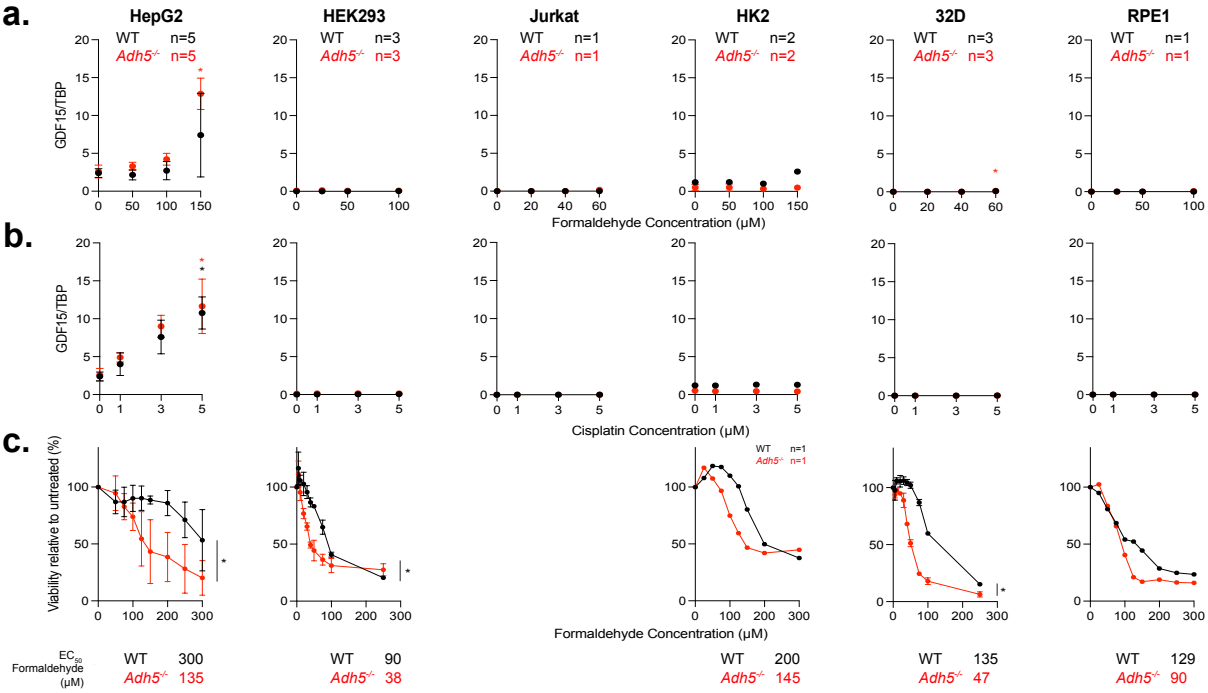


Figure 3.2 GDF-15 induction and formaldehyde sensitivity in cell lines
RT-qPCR measurement of GDF-15 mRNA expression relative to TBP in different cell lines in response to 24h formaldehyde (a) or cisplatin (b) treatment. Determination of formaldehyde toxicity in 24h treatment by Cell Titre Blue assay, the estimated EC_{50} is given below (c). Points represent mean of experimental replicates ± SD; n = number of experimental replicates; statistical comparisons were only performed when n ≥ 3 (a-b) * = p<0.05 (Dunn's Test, comparing to 0μM within genotype) (c) * = p<0.05 (Paired Wilcoxon Rank Test comparing WT vs *Adh5*$^{-/-}$).

3.2.2 Endogenous Reporter Construction

A transcriptional reporter construct was engineered by Philip Hublitz in the WIMM Genome Engineering core facility, designed to integrate an mNeonGreen (mNG) sequence into exon 2 of the GDF-15 locus (referred to as Donor template; Figure 3.3a), with homologous recombination into the target locus stimulated by means of the Ruby variant of the pX458 vector expressing Cas9 and a gRNA targeting exon 2 of the GDF-15 locus, hereafter pX458R-GDF-15. The donor template contains a GSG linker and T2A sequence that allows translational ribosome skipping between mNG and GDF-15 translation products and removes the endogenous PAM containing template sequence to prevent re-editing by Cas9. I verified effective Cas9 cleavage of the GDF-15 locus using HepG2 cells transfected with pX458-GDF-15, which I sorted based on mRuby expression. Efficiency was calculated using TIDE analysis of Sanger sequencing products of the GDF-15 locus in HepG2, reporting cutting efficiency of 61.7%.

To produce the reporter cell line, I co-transfected HepG2 cells with donor template and pX458-GDF-15 by electroporation (Figure 3.3b). To increase the efficiency of integration (Yao et al., 2018) and reduce persistence of non-integrated DNA, I linearised the donor template using PsiI-v2. After 48h-72h, cells were single cell sorted based on expression of both mNG and mRuby signal. I first attempted to identify clones with successful functional reporter integration using flow cytometry. To do this, I single cell sorted mRuby+mNG HepG2 cells and allowed colonies to grow for 2 weeks. I then split colonies to replica plates and after 1 further week, I treated one plate with 5μM cisplatin for 24h. I analysed mNG expression by flow cytometry, and did not identify any clones that expressed mNG either basally, or after GDF-15

induction with cisplatin (Figure 3.3c). I then assessed whether there was any sign of genomic integration by PCR, repeating the electroporation on new cells and bulk sorting mNG and mRuby double positive cells, from which I extracted gDNA. To provide a positive control sequence, I produced a modified version of the donor template plasmid that extended the homologous sequence at both the 5' and 3' ends of the plasmid. I could not detect any integration into the homologous locus within any primer pair, including when testing across a range of annealing temperatures, or when using DMSO in the PCR mix (Figure 3.3d).

One method of improving the efficiency of integration of the donor template, is the use of pharmacological inhibitors of NHEJ, including SCR7 (inhibitor of DNA ligase IV) and NU7741 (inhibitor of DNA-PK) (Chu et al., 2015; Maruyama et al., 2015; Robert et al., 2015), or promoters of HDR (commercially produced cocktail Alt-R HDR enhancer v2) (Kath et al., 2022). Therefore, I attempted to achieve reporter integration by culturing HepG2 cells in the presence of SCR7 before and after transfection, then assessing reporter integration by treating the cells in bulk with cisplatin and observing mNG signal via flow cytometry (Figure 3.3e). Again, I observed no induction of mNG in any group (Figure 3.3f). I next attempted to improve integration using varying combinations of NHEJ inhibitors and HDR enhancers through the same treatment protocol. Here, I assessed integration by PCR, using a range of annealing temperatures (Figure 3.3g). At some concentrations/annealing temperatures, integration was apparently observable, but there was considerable non-specific primer binding in the untransfected control samples for both 5' and 3'. A clean PCR method using just the 5' and 3' primers targeting the endogenous locus may help resolve this non-specific

binding, while Sanger sequencing may help clarify the identity of the non-specific bands.

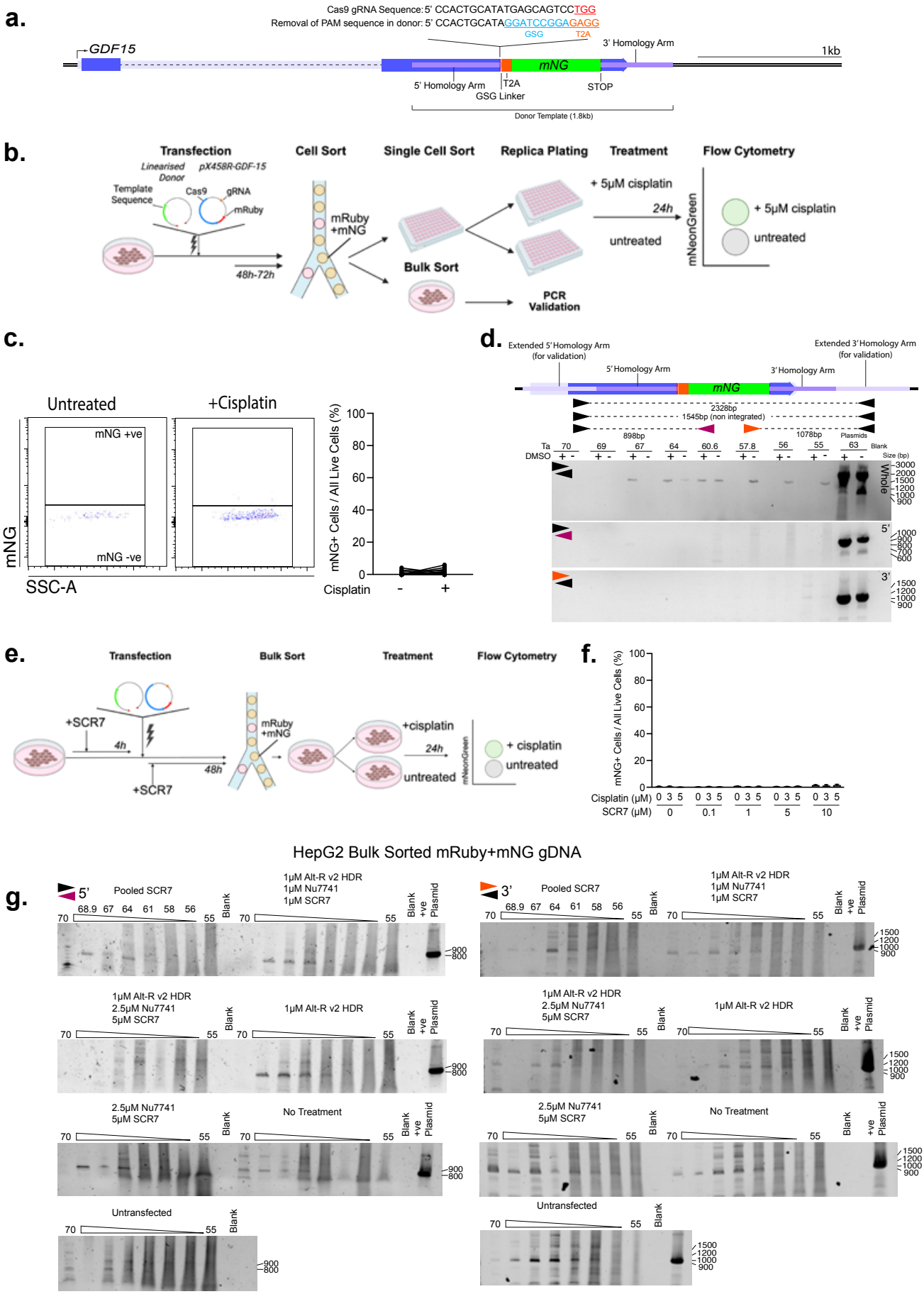

a.
Cas9 gRNA Sequence: 5' CCACTGCATATGAGCAGTCCTGG
Removal of PAM sequence in donor: 5' CCACTGCATAGGATCCGGAGAGG
GSG
T2A
GDF15
3' Homology Arm
1kb
mNG
5' Homology Arm
T2A
GSG Linker
STOP
Donor Template (1.8kb)
b.
Transfection
Linearised Donor
pX458R-GDF-15
Template Sequence
Cas9
gRNA
mRuby
48h-72h
Cell Sort
mRuby +mNG
Single Cell Sort
Bulk Sort
Replica Plating
Treatment
+ 5μM cisplatin
24h
untreated
Flow Cytometry
mNeonGreen
+ 5μM cisplatin
untreated
PCR Validation
c.
Untreated
+Cisplatin
mNG +ve
mNG -ve
mNG
SSC-A
mNG+ Cells / All Live Cells (%)
Cisplatin
d.
Extended 5' Homology Arm (for validation)
5' Homology Arm
mNG
3' Homology Arm
Extended 3' Homology Arm (for validation)
2328bp
1545bp (non integrated)
898bp
1078bp
Plasmids
Ta
DMSO
Size (bp)
Whole
5'
3'
e.
Transfection
Bulk Sort
Treatment
Flow Cytometry
+SCR7
4h
48h
+SCR7
mRuby +mNG
+cisplatin
24h
untreated
mNeonGreen
+ cisplatin
untreated
f.
mNG+ Cells / All Live Cells (%)
Cisplatin (μM)
SCR7 (μM)
HepG2 Bulk Sorted mRuby+mNG gDNA
g.
5'
Pooled SCR7
1μM Alt-R v2 HDR
1μM Nu7741
1μM SCR7
Blank
+ve
Plasmid
1μM Alt-R v2 HDR
2.5μM Nu7741
5μM SCR7
2.5μM Nu7741
5μM SCR7
No Treatment
Untransfected
3'

Figure 3.3 Endogenous reporter production
(a) Diagram of endogenous GDF-15 locus with the reporter template integrated. The Cas9 gRNA sequence used to generate a targeted break is removed following homologous integration of the reporter element. (b) Schematic of reporter integration and validation approaches. (c) (left) representative flow cytometry gating of replica clones treated with either 5μM cisplatin or untreated (right) quantification of mNG positive cells, clones are connected with a line (73 clones screened). (d) (Top) Schematic of PCR validation of template integration, extended homology arms detail the additional sequence added to the positive control plasmid, connected arrows depict expected PCR product from the primer pair used (Bottom) Representative gels of PCR to assess integration of donor template into bulk sorted HepG2 cells. (e) Schematic of SCR7 screening protocol. (f) Quantification of functionally integrated mNG at different SCR7 concentration, points represent 3 technical replicates, 1 experiment. (g) PCR gels of bulk sorted HepG2 gDNA treated with varying drug concentrations.

3.2.3 Exogenous Reporter Construction

Having encountered continual difficulty in making an endogenous reporter line, I attempted an exogenous reporter approach. Here, I cloned a promoter fragment of the GDF-15 promoter (containing both p53 binding sites), and fused these to an mNG sequence (Promoter 1, Figure 3.4a), and inserted these into a pCMV6 plasmid, which contains a Kanamycin/Neomycin resistance cassette driven by an SV40 promoter. I also designed a control plasmid that did not contain a promoter element (No Promoter, Figure 3.4a), and another construct that contained the promoter construct identical to that described in Osada *et al*, 2007 (Promoter 2, Figure 3.4a). This promoter sequences was previously able to report GDF-15 expression through driving luciferase expression (Osada et al, 2007). Importantly, the SV40 promoter and GDF-15 promoter are oriented in opposite directions, hence mNG expression should be driven specifically by GDF-15 promoter elements. Again, I used a PsiI-v2 restriction enzyme to linearise the plasmid, and electroporated these into HepG2 cells. I selected for cells that had stably integrated the plasmid into the host genome by addition of G418

(protection conferred by the neomycin resistance cassette) to the growth media for 2 weeks. Interestingly, I found no HepG2 cells survived this process.

As HepG2 cells failed to integrate the plasmids, I searched for an alternative cell line that may be more amenable to editing. I identified Huh7 and HeLa cell lines as potential candidates, both reported to transcribe GDF-15, which is significantly elevated upon stimulation of the integrated stress response by tunicamycin treatment (Patel et al., 2019) and have both been used in to make endogenous reporter lines and in CRISPR/Cas9 screens (Roberts et al., 2023; Zimmermann et al., 2018). I evaluated GDF-15 transcription and toxicity in both HeLa and Huh7 in response to formaldehyde and cisplatin as above and explored whether ADH5-inihibtion increased sensitivity to formaldehyde using the pharmacological inhibitor N6022 (Green et al., 2012; Wit et al., 2023). Both cell lines expressed GDF-15 in response to cisplatin, but only HeLa expressed GDF-15 in response to high dose formaldehyde. Huh7 did increased GDF-15 induction at high doses of formaldehyde following N6022 treatment, while N6022 treatment of HeLa cells yielded no further GDF-15 increase (Figure 3.4b). It is of note, however, that N6022 did not increase toxicity of formaldehyde, suggesting that ADH5 was not sufficiently inhibited, and higher concentrations of N6022 may be tolerated to support further induction (Figure 3.4c). Toxicity of formaldehyde to HeLa cells after N6022 treatment gave variable results, although the absence of a difference in GDF-15 induction following treatment suggests ADH5 was not sufficiently inhibited in these either. Further experiments are needed to establish the dose of N6022 at which ADH5 is inhibited in these cell lines, or producing an *Adh5*$^{-/-}$ cell line.

I electroporated both HeLa and Huh7 cell lines with pCMV6 plasmids containing either no-promoter control, promoter 1 or promoter 2 (promoter 2 only in HeLa) (Figure 3.4d). I was able to produce stable cell lines with all 3 plasmids in both HeLa and Huh7. I then tested induction of GDF-15 expression by flow cytometric analysis of cells, treated with formaldehyde or cisplatin for 24h (Figure 3.4e). Again, I found no induction of mNG in response to treatment. Constitutive expression of mNG was observed for all 3 plasmids in HeLa, but not in Huh7 (Figure 3.4f). I verified that GDF-15 was still induced in these new cell lines by RT-qPCR of GDF-15. I also performed RT-qPCR of the mNG transcripts, to see if there was transcribed protein that had not been translated. If GDF-15 is induced in the cell, the promoter elements driving expression should have been driving mNG (Figure 3.4g). I found strong GDF-15 induction in response to cisplatin but not formaldehyde in these lines, but this did not correspond to induction of mNG transcription.

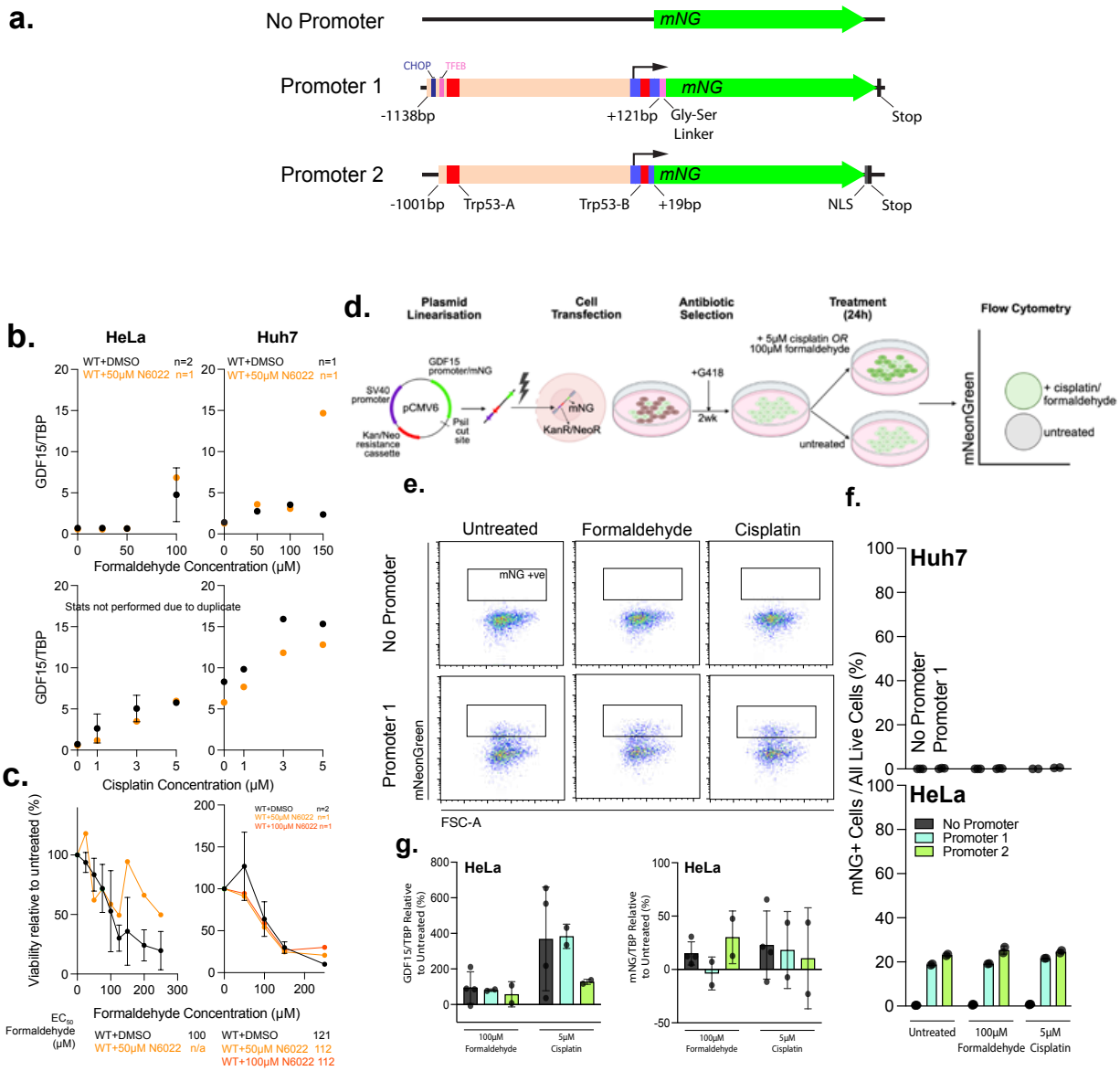


Figure 3.4 Exogenous reporter production
(a) Schematics of the promoter constructs designed and cloned into pCMV6 vector backbone. (b) GDF-15 induction and (c() toxicity assessment in HeLa and Huh7. Points represent mean of experimental replicates ± SD; n = number of experimental replicates. (b) Schematic of stable reporter cell line generation and functional validation using cisplatin or formaldehyde. e) Representative FACS gating of flow cytometry assessment of exogenous reporter fluorescence. (f) Quantification of mNG expressing cells in stably produced Huh7 and HeLa cell lines. (g) RT-qPCR of GDF-15 and mNG expression in HeLa, relative to untreated cells of the same reporter line. (f-g) One experimental replicate for each, performed in duplicate.

I made one final attempt in developing an endogenous construct using the HeLa cell line, transfecting cells without drugs and bulk sorting mRuby+mNG cells. To validate integration, I used an additional set of primers to validate 5'/3' integration (Set 2, Figure 3.5). I observed background bands corresponding to the correct size in 5'-primer set 1 and 3'-primer set 2, but taking the 3'-primer set 1 and 5'-primer set 2, there did not appear any signal of successful integration in HeLa cells. Assessment of functional integration by flow cytometry was not performed and would be important for future work.

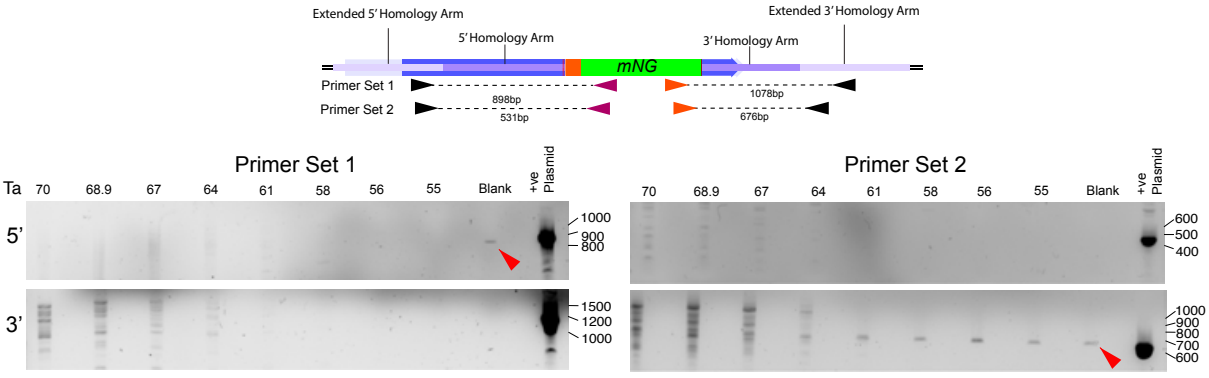


Figure 3.5 Endogenous reporter production in HeLa cells
(Top) Map of the regions of amplification by the primer sets to validate 5' and 3' integration (Bottom) PCR gel images of endogenous GDF-15 locus of HeLa cells after bulk sorting, red lines mark contaminating bands, numbers to the right represent size (bp). Numbers above represent annealing temperature used (Ta) in °C. Positive control and blank were run with annealing temperature 63 °C.

This last experiment represents the last I conducted on this project. Having encountered substantial difficulty in generating a transcriptional reporter of the GDF-15 locus using an exogenous or endogenous approach, I decided not to pursue further work.

3.3 Discussion

In this chapter, I identified that formaldehyde treatment could induce GDF-15 expression in multiple cell lines but encountered substantial challenge in producing a transcriptional reporter for GDF-15.

3.3.1 Variability in GDF-15 induction

Of all the cell lines that were tested in this study, only HepG2, Huh7 and HeLa cells could induce in response to formaldehyde or cisplatin GDF-15. DNA-damage induced p53 signalling is a key driver of GDF-15 expression, with two p53 binding elements within the promoter region of GDF-15 (Osada *et al*, 2007), and p53 deficiency completely perturbing GDF-15 induction in mice in response to cisplatin (Mulderrig *et al*., 2021). It appears p53 status of selected cell lines typically corelated with GDF-15 expression in this study: HepG2 and HeLa have functional p53 (Scheffner et al., 1990; Vollmer et al., 1999) and Huh7 contains a mutated form with high stability (Bressac et al., 1990), on the other hand, Jurkat cells have attenuated p53 activity (Gioia et al., 2018) and HEK293 express proteins E1A/E1B from adenovirus-5, which directly antagonises p53 activity (Savelyeva & Dobbelstein, 2011). p53 alone does not provide a full explanation for the different expression levels, RPE1 express WT p53 but have no observable induction (Haapaniemi et al., 2018), and genetic and epigenetic differences in regulatory proteins will differ between cells. Indeed, HK-2 cells express E6/E7 proteins from human papilloma virus, of which E6 degrades p53 (Pal & Kundu, 2020), still had very small induction was observed in HK-2 cells. This highlights the important functions of different promoter elements, binding transcription factors such

as ATF4 and CHOP, activated upon integrated stress response signalling (Li et al., 2018; Miyake et al., 2021). A key limitation of this experiment was that only the 24h time point was considered. *In vivo* GDF-15 expression in response to stress is acute: GDF-15 is expressed within 2 hours of acute kidney injury (Liu et al., 2020). Indeed, I measured GDF-15 mRNA in *Adh5*$^{-/-}$ mouse tissues in response to acute methanol challenge, which identified GDF-15 expression peaking in the kidneys within 6h (males) and 12h (females) (Figure 3.6a). It is likely that this would be more so the case *in vitro*, hence the lack of induction in cell lines following treatment may be more a reflection of missing the window of expression, rather failure of induction. Furthermore, this may have necessitated higher concentrations of formaldehyde to have persistent effects of induction over 24h. However, I did look at 6h time point for HeLa cells, and here I did not see strong induction, but this only has one experimental replicate (Figure 3.6b).

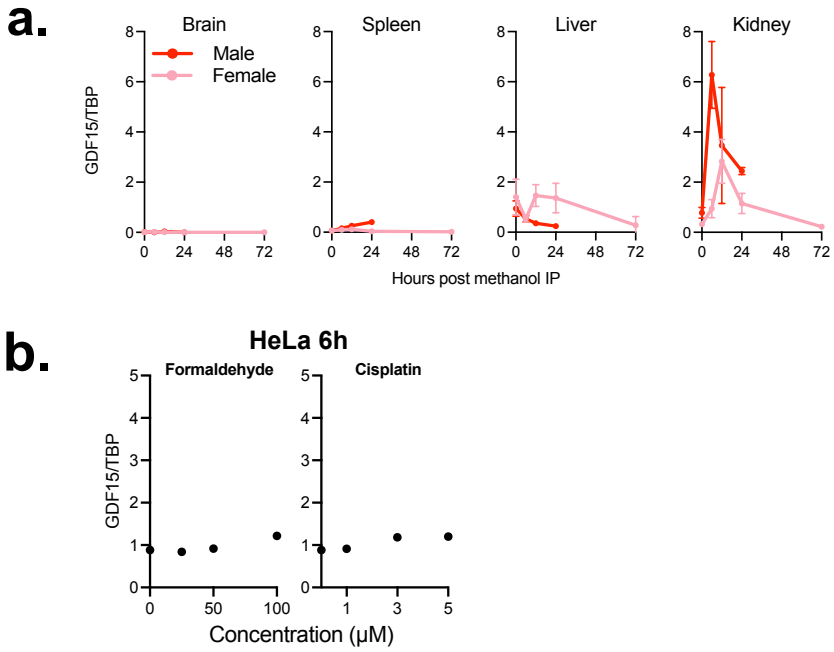


Figure 3.6 Dynamics of GDF-15 Induction in mice and HeLa cells
(a) GDF-15 transcription induction in 8–18-week-old *Adh5*$^{-/-}$ mice by 2g/kg methanol IP. Bars represent mean ± SD, 2-3 mice per time point (b) GDF-15 expression in HeLa cells following 6h incubation with formaldehyde/cisplatin. Points represent mean of technical duplicates ± SD, 1 experiment performed.

There are further limitations to be aware of. For one, only one *Adh5*$^{-/-}$ clone was assessed for each cell line. Single cell clones acquire different characteristics through mutation acquisition or epigenetic changes, which can alter their response to drug treatment (Hanlon et al., 2019; Mao et al., 2008; Maruyama et al., 2015). Therefore, effects on induction may be a clone specific artefact and not due to activity of ADH5 deletion (although formaldehyde sensitivity in *Adh5*$^{-/-}$ cells is expected and consistent across cell lines). Further work will be needed to validate these findings in independent clones of the cell line selected, before a screen is to be performed. For another, several cell lines were only tested in one experimental replicate (HK-2, RPE1, Huh7, HeLa + N6022).

3.3.2 Reporter Construction

HepG2 was initially selected as the cell line of choice because its strong induction and relatively low sensitivity to formaldehyde would mean relative increases in induction can be readily detected. HepG2 cells are also amenable to editing, observing cutting efficiency of 61.7% at the GDF-15 locus. However, these cells were apparently resistant to integrating plasmid DNA into the genome. Developing good PCR based approaches of validation were complicated by the absence of a non-plasmid based positive control template, which requires different optimisation conditions that may not translate to a successful PCR of genomic template. However, to mitigate for this, I also tested whether any cells had integrated a functional reporter directly using flow cytometry, and none were observed. Therefore, even if successful integration was achieved in a small subset of cells, there was not an identifiable signal that corresponded to functional activity.

Integration of donor templates through HDR is inefficient due to the rapid incidence of NHEJ in response to DNA breaks (Mao et al., 2008; Maruyama et al., 2015). A range of concentrations of NHEJ inhibitors was used, including inhibition before cutting was carried out, but none of these conditions resulted in improved efficiency of integration (although a high-quality PCR is still pending). It may be that NHEJ could still operate through non-canonical NHEJ, which utilises DNA ligase I and III, rather than IV (inhibited by SCR7), and does not depend on DNA-PK-Ku70 holoenzyme formation (DNA-PK inhibited by NU7741) (Boboila et al., 2010; Panier & Boulton, 2014). Furthermore, NHEJ is active throughout the cell cycle, rather than in S/G2 (Panier & Boulton, 2014), and synchronisation of cells in the cell cycle has been reported to improve rates of HDR mediated integration in human induced pluripotent stem cells (Yang et al., 2016). This may be a way to increase targeting of the GDF-15 locus. Integration efficiency may also relate to the sequence context of GDF-15 locus, but GDF-15 locus is in an active region of the genome (HepG2 constitutively express GDF-15 in basal conditions), hence HDR should be favoured (Aymard et al., 2014).

The exogenous reporter approaches had a similar fate, with HeLa and Huh7 cell lines able to successfully integrate plasmid (determined by selection of cells stably expressing antibiotic resistance), but at least in the HeLa line, this did not correspond to functional GDF-15 reporter. It remains to be shown whether this was the same in Huh7. The GDF-15 promoter 2 design was identical to that used in Osada *et al*., 2007, which was linked to a luciferase reporter, so why did it work previously? The first difference lies in that the vector backbone used by Osada *et al*., was a pGL3-Basic plasmid (Addgene plasmid #212936). In this construct, the multiple cloning site where

the GDF-15 promoter was inserted, is flanked in the 5' region by an RNA polymerase II termination signal. The pCMV6 vector I used contains no such element, which may explain constitutive promoter activity observed in HeLa. Another difference is the cell line used. Osada *et al*., used Saos2 cells, which don't express p53 themselves, but were modified to overexpress p53. Therefore, activation of the promoter observed by Osada *et al*., is a product of over-expression in unstimulated conditions, not a reflection of the regulatory environment of the cell. It would be interesting to see what happens if the reporter-mNG constructs detailed in this chapter were cloned into the pGL3-Basic background and placed into HeLa cells, and stimulated with formaldehyde or cisplatin. If mNG is induced, this would support the promoter elements within the range detailed being sufficient to drive GDF-15 induction in response to DNA damage.

While generating endogenous and exogenous reporter methods were challenging, a third, non-reporter-based method may be worth considering. Flow-FISH involves use of fluorescent in-situ hybridisation probes specific to RNA of a gene of interest, allowing changes in gene expression associated with specific guide-RNAs to be detected by flow cytometry. This has been applied in CRISPRi-screens to identify regulatory elements that modulate expression of GATA2 and HDAC6 (Fulco et al., 2019), and has been successfully applied to HepG2 cells (Reilly et al., 2021). Such an approach would allow one to circumvent additional rounds of cell line modification needed to produce a reporter, and get around the apparent difficulties of generating a GDF-15 reporter itself, to provide a more direct method of detecting expression.

3.3.2 Conclusion

The purpose of designing a reporter cell line was to use GDF-15 induction as a sensitive readout of formaldehyde stress in cells *in vitro*, using conditions that would not induce substantial toxicity. This could reveal novel enzymes involved in tier-1 and tier-2 protection against formaldehyde stress, and genes involved in signalling to the GDF-15 locus. Unfortunately, none of the approaches used here yielded a functional reporter cells line. In the process I discovered several limitations in the strategies I used, which inform the choice of avenues for future work, but I did not pursue these further.

Chapter 4

The influence of *ALDH2* deficiency on humoral immunity

4.1 Introduction

4.1.1 An immunological hypothesis for the *ALDH2*2* paradox

The *ALDH2*2* allele decreases the catalytic activity of *ALDH2* protein (Figure 4.1a-b), causing carriers who consume alcohol to accumulate very high levels of acetaldehyde, resulting in the famous “Asian Flush” response (Crabb et al., 1989; Harada et al., 1981; Xiao et al., 1996; Yoshida et al., 1984). As a result, alcohol consumption in *ALDH2*2* carriers significantly increased risk of cancer development, most commonly oesophageal cancer (Im et al., 2022; Yokoyama et al., 1996, 2003). Beyond the adverse reaction to alcohol, *ALDH2*2* is a major risk allele in the congenital disease FA and ADDS, highlighting the significant burden of aldehydes on the mechanisms of DNA repair in these carriers (Hira *et al*., 2013; Dingler *et al*., 2020; Oka *et al*., 2020). It is therefore curious that this allele has persisted in human populations, now affecting ~540 million people worldwide (Figure 4.1c; Brooks et al., 2009; H. Li et al., 2009).

Even more so, estimations based on archaeological data and averaging of haplotype lengths in current populations, suggest *ALDH2*2* first arose underwent expansion within the last 2000-3000 (Cong et al., 2022; Field et al., 2016; Luo et al., 2023; Okada et al., 2018; Taliun et al., 2021). This recent expansion suggests this variant has been positively selected.

Interestingly, incidence of *ALDH2*2* overlaps with another common variant in the enzyme aldehyde dehydrogenase 1b (ADH1B), *ADH1B*2* [rs1229984, Arg47>His] (Edenberg, 2007). This polymorphism is even more widespread among East Asians (affecting ~89%), believed to predate *ALDH2**2, having arose ~7,000 years ago and recently expanded from ~4,000 years ago (Cong et al., 2022; Eng et al., 2007; Koganebuchi et al., 2017; Li et al., 2007; Luo et al., 2023; Okada et al., 2018; Taliun et al., 2021). *ADH1B* is responsible for the conversion of ethanol to acetaldehyde, and *ADH1B*2* encodes a non-synonymous mutation that accelerates this rate of conversion (Edenberg, 2007). *ADH1B*2* does not directly induce a “flushing” response to alcohol and only induces milder increases in blood acetaldehyde after consumption, but it is negatively associated with drinking behaviour independently of *ALDH2*2*, and can have synergistic effects in cancer promotion (Chen et al., 1999; Peng et al., 2014; Peng & Yin, 2009). It is striking that the *ADH1B*2* and *ALDH2*2* mutations originated independently yet converge on acetaldehyde metabolism (Figure 4.1d), possibly suggesting acetaldehyde is an agent that underpins the natural selection of these alleles (Oota et al., 2004).

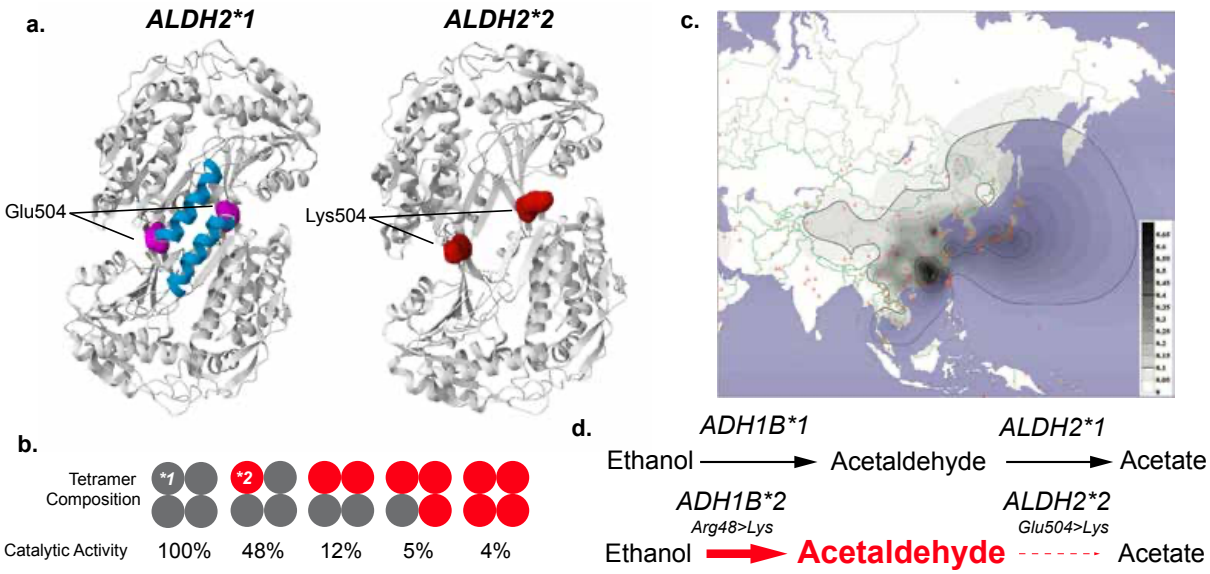


Figure 4.1 – The *ALDH2*2* paradox
(a) Homodimer structures of *ALDH2*1* (PDB1O05) and *ALDH2*2* (PBD1ZUM) variants described by (Larson et al., 2005). Residue 487 in each monomer are highlighted. Lysine substitution prevents the formation of the central alpha helix structures (highlighted blue) in the oligomerisation domain, decreasing tetramer stability, as well as reducing affinity to NAD+ (essential cofactor for acetaldehyde oxidisation). **(b)** Reduced catalytic activity of final homotetrameric structure of ALDH2 upon increasing ratio of *ALDH2*2* subunits in the final tetramer (Xiao et al., 1996; Zhou & Weiner, 2000). **(c)** Distribution of *ALDH2*2* variant in East Asia, taken from (Figure 2 from (Li et al., 2009); open triangles represent population sample locations, grey scale represents allele frequency). **(d)** Convergent activities of the common *ADH1B*2* and *ALDH2*2* on acetaldehyde metabolism.

Potential protective roles of *ALDH2*2* have been observed in GWAS studies. A study assessing the Japanese BioBank, found the *ALDH2*2* variant was significantly associated with 47 characteristics, including a protective relationship against cardiovascular disease, ischemic stroke and liver cirrhosis (Sakaue et al., 2021). However, these protective effects may be largely attributable to the dominant action of *ALDH2*2* in promoting alcohol abstinence (which was not investigated in this study). Indeed, these cases where *ALDH2*2* is associated with obesity and cancer, are both dependent on alcohol consumption (Im et al., 2022; Spracklen et al., 2020; Wang et al., 2016). Given the protective effect of alcohol abstinence, this may be thought to drive selection, however, this argument has limitations. *ALDH2*2* heterozygotes

continue to consume alcohol in abundance, so the notion of behavioural change is incomplete (Crabb et al., 1989; Higuchi et al., 1996; Xiao et al., 1996). Second, cancer incidence associated with *ALDH2*2* is only increased later in life, well after the reproductive window begins, making it unlikely to significantly affect reproductive fitness (Lee et al., 2022; Yokoyama et al., 2019).

One of the most important drivers of positive selection in human populations in resistance to infectious disease (Fumagalli et al., 2011; Karlsson et al., 2014). There is evidence to suggest that aldehydes normally cleared by *ALDH2*, can positively influence actions of the innate and adaptive immune system (see Wang et al., 2024 for an extensive review of the innate and adaptive immune system).

First, in innate immunity. Innate immunity involves a fast acting, non-specific response to clear invading pathogens. Neutrophils are essential in this response, which release an oxidative burst that directly kills engulfed pathogens (Segal, 2005). Experiments with human neutrophils cultured *ex vivo*, found myeloperoxidase (MPO) can produce aldehydes through the oxidation of free amino acids, using H_2O_2 and Cl^- as cofactors, including the production of acetaldehyde from free alanine (Anderson et al., 1997; Hazen et al., 1998). Loss of MPO results in immunodeficiency, as MPO deficient neutrophils lose their ability to kill bacteria (Lehrer et al., 1969.; Segal, 2005). Based on the known cytotoxic effects of aldehydes to bacteria (O'Brien et al., 2005), it is possible that *ALDH2* deficiency may support direct pathogen killing by neutrophils by increasing the aldehyde concentration, facilitating more rapid pathogen killing.

Second, adaptive immunity. The adaptive immune response is instigated by T (cell-mediated) and B (antibody-mediated) cells, which have specialised surface receptors to detect antigens directly, or antigens processed and presented by the innate immune system (epitopes), allowing them to mount specific responses to the pathogens. In work by Allison & Fearon (2000), direct modification of antigens by glycolaldehyde, which can be produced through L-serine oxidation by MPO, was reported to significantly increase the immunogenicity of model antigens ovalbumin (OVA), Hen Egg Lysozyme (HEL) and chicken gamma globulin. Additionally, antigen presentation by macrophages and dendritic cells, and subsequent T-cell activation was accelerated when presenting modified antigens *in vitro*. These experiments suggest that increased antigenicity (ability of antigen to activate immune response by binding to immune cells receptors) is conferred by aldehyde modification that may occur *in vivo* by products of the innate immune system. Therefore, by increasing the aldehyde concentration *in vivo* in *ALDH2*2* individuals, the chance of these modifications may increase, supporting adaptive immunity.

4.1.2 Evidence for *ALDH2* deficiency affecting humoral immunity

To explore whether ALDH2 deficiency may confer a humoral immune advantage, fellow group member Felix Dingler (F.Dingler), immunised *ALDH2* proficient ($Aldh2^{+}$ (see methods)) and deficient ($Aldh2^{-/-}$) mice with OVA. Remarkably, the anti-OVA IgG-titres against $Aldh2^{-/-}$ mice were significantly increased against WT (Figure 4.2a). Based on the convergent effects of the common *ADH1B*2* and *ALDH2*2* variants on acetaldehyde, and the ability for acetaldehyde to be produced from MPO metabolism, F.Dingler explored whether the differential effect observed may be result of direct

modification of OVA acetaldehyde, similar to the proposal by Allison & Fearon. Immunisation with acetaldehyde modified OVA suggested modification may indeed drive an increase in the mean anti-IgG titre (although there was considerable viability) (Figure 4.2b). This preliminary data gives credence to the argument that ALDH2 deficiency may confer an immune advantage, and that direct modification of antigens by acetaldehyde may be a mechanistic route.

In this chapter, I ask whether these observations with OVA were generalisable to other antigens, which would further support an argument that improving antigenicity by direct modification of antigens occurs in *Aldh2*$^{-/-}$ mice.

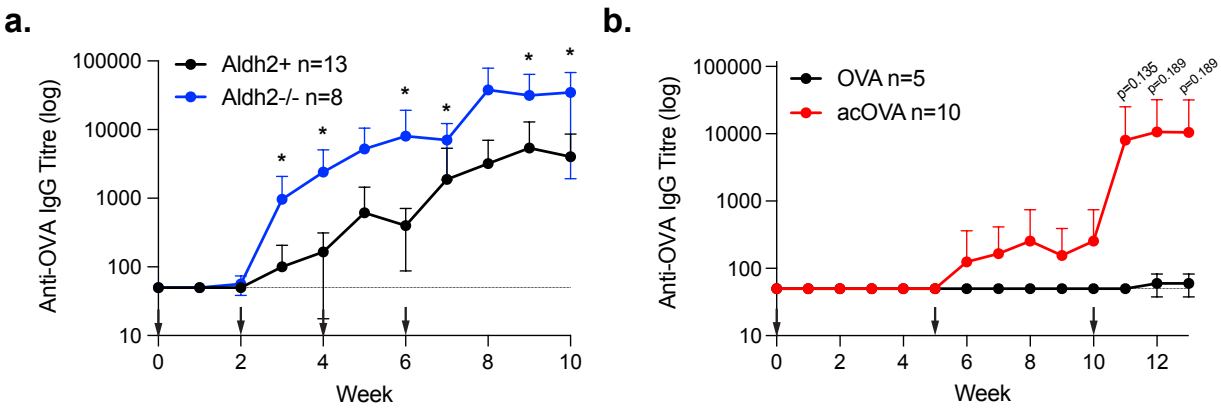


Figure 4.2 – *Aldh2*$^{-/-}$ mice have greater humoral response to OVA
(a) Mean Anti-OVA IgG titres of serum from *Aldh2*$^{+}$ and *Aldh2*$^{-/-}$ mice, immunised with 50µg/50µl at weeks 0, 2, 4, 6 (arrows). Points represent mean ±SD, * = p<0.05 t-test. (**b)** Mean anti-OVA IgG titres of serum from *Aldh2*$^{+}$ mice, immunised with 50µg/50µl OVA or acetaldehyde modified (ac)OVA, at weeks 0, 5 and 10 (arrows). Points represent mean ±SD, significance where p<0.05, Mann-Whitney U test. Dotted line at y=40 represents limit of detection by ELISA.

4.2 Results

4.2.1 Direct acetaldehyde modification does not increase antigenicity of HEL or BSA

I first set about addressing whether acetaldehyde modification induces adjuvant like properties, as was suggested in the OVA condition. I performed this using HEL and BSA, which both contain surface lysine residues that enable modification, and have lysine residues present in previously described epitopes, supporting potential these proteins to be stably modified by acetaldehyde (Figure 4.3). I treated HEL and BSA with the same protocol used to modify OVA, incubating with 500mM acetaldehyde for 12-16 hours at 37 °C. To prevent avidity effects of protein aggregation (Benne et al., 2016), I isolated the monomeric fraction of these proteins by size exclusion chromatography, identifying the monomeric fraction by SDS-PAGE stained with Coomassie blue. I repeated this process with antigens treated with PBS, to serve as the monomeric control antigens used throughout this chapter (Figure 4.4).

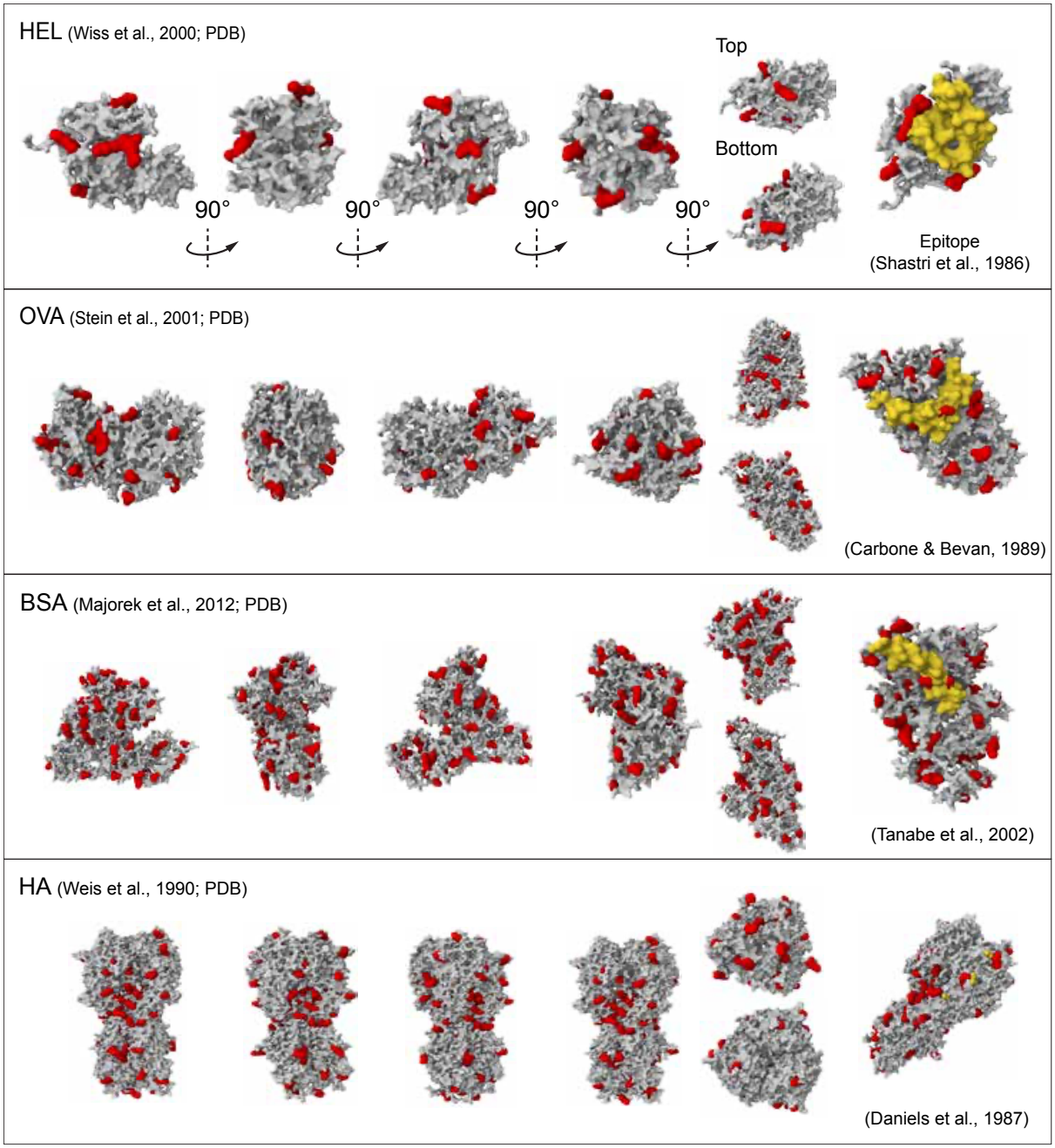


Figure 4.3 – PDB Structures of antigens used in this study
PDB structures of proteins used for immunisation in this chapter. Molecular surface representations are depicted with lysine residues highlighted in red. Residues that contribute to characterised epitopes are highlighted in yellow.

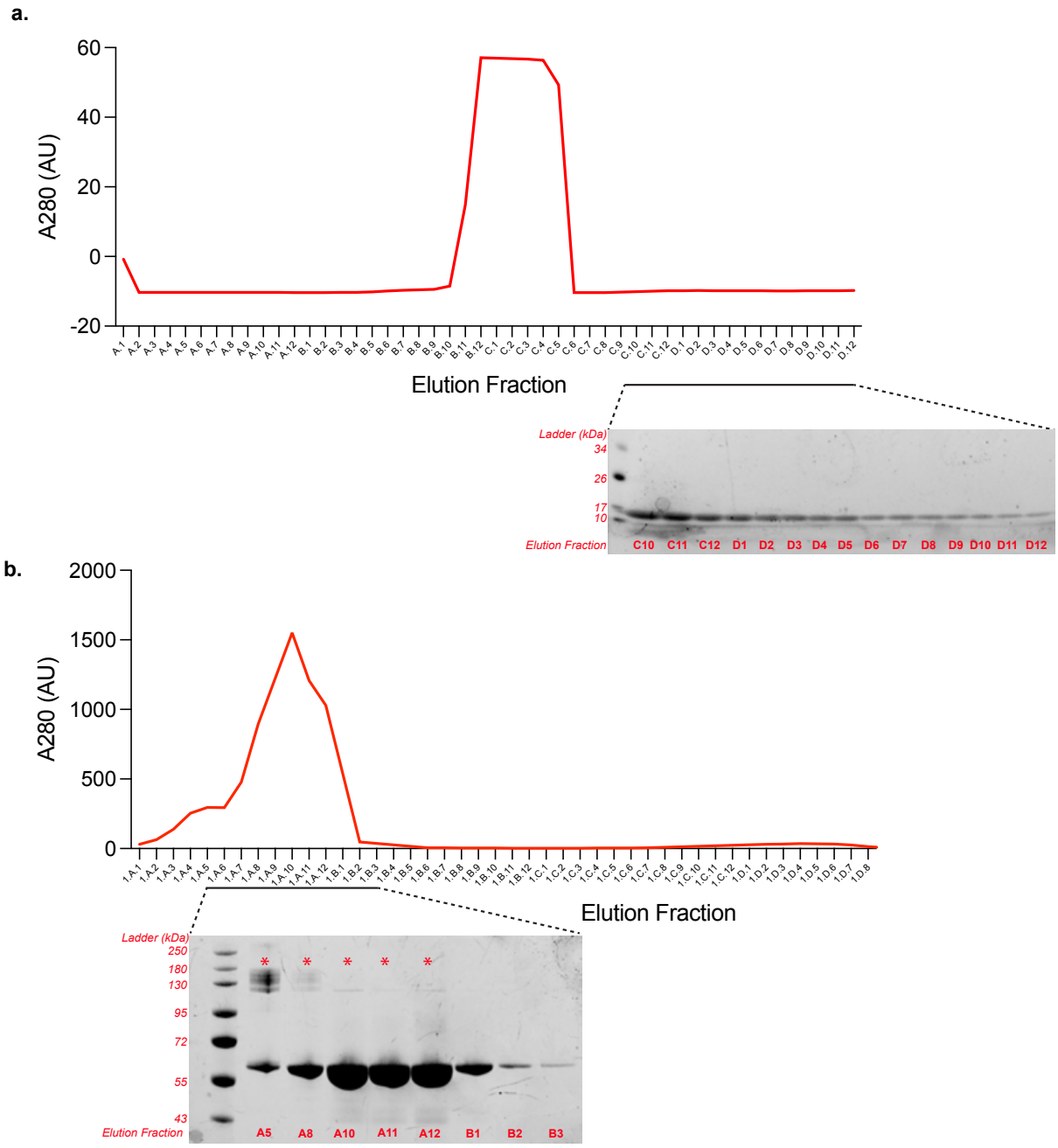


Figure 4.4 – Purification of HEL and BSA by SEC
(Top) SEC trace of with (bottom) SDS-PAGE gel stained with Coomassie blue of eluted protein fractions indicated from trace, for HEL **(a)** and BSA **(b)**. Multimeric aggregates marked by *.

Based on work on OVA, a compressed immunisation schedule produced a more robust antibody response and decreased variability between groups (Figures 4.1a versus 4.1b), hence I opted for a similar schedule (Figure 4.5a). I first immunised *Aldh2*$^{+}$ mice with 5µg/50µl HEL or acHEL, based on the dose used by Allison & Fearon, 2000, but this did not induce a detectable immune response in mice (data not shown). Therefore, I increased the dose to 50µg/50µl HEL, where an immune response was detected (Figure 4.5b). This became the working dose for all subsequent immunisations unless otherwise stated. Acetaldehyde modification did not induce a significant difference to the IgG titre of either HEL or BSA at any time point (Figure 4.5b-c; Mann-Whitney U test, $p>0.05$). However, it is worth highlighting the tendency for both unmodified HEL and BSA to confer a greater immune response, over the modified form, which may suggest a negative effect of acetaldehyde modification on these antigens, counter to OVA.

To rule out the possibility of acetaldehyde modification producing new epitopes that are not recognised on the unmodified antigen by ELISA, I repeated the ELISA of BSA and acBSA immunised mice, using both unmodified and modified BSA as the test substrate (Figure 4.5d). I found no difference in anti-BSA IgG titres between the different test antigens, suggesting that any acetaldehyde specific epitopes that may have been produced were not working in an immunodominant manner. I also performed this using week 12 serum samples from the original OVA/acOVA experiment, again finding no difference relating to the test substrate, and reassuringly corroborating the pattern of results observed previously (Figure 4.5e).

These results suggest that direct modification of antigens by acetaldehyde does not increase antigenicity in mice in a manner generalisable across antigens.

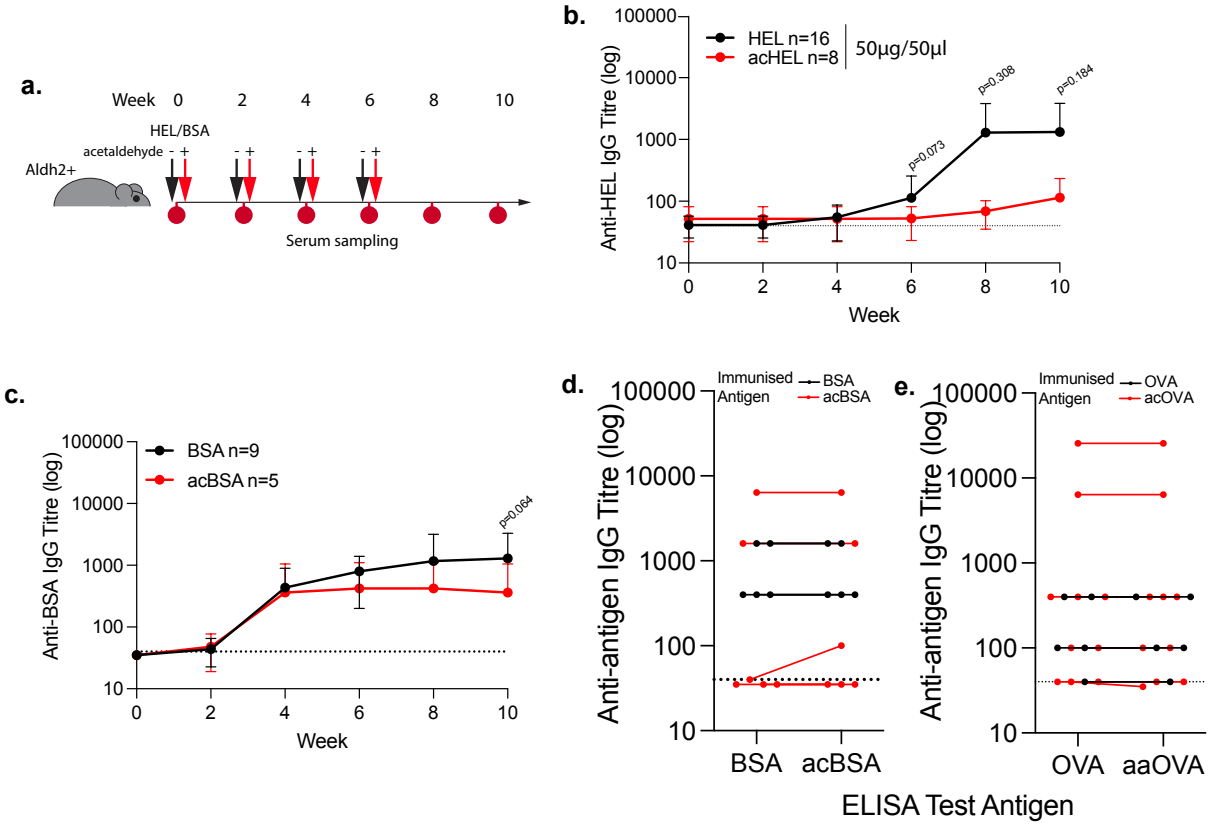


Figure 4.5 – Acetaldehyde modification does not increase immunogenicity of HEL or BSA
a. Immunisation protocol for experiment, mice were immunised with either 50µg/50µg modified (ac) or unmodified antigen on weeks 0, 2, 4, and 6. **b.** Anti-HEL IgG titre for mice immunised with HEL or acHEL. **c.** Anti-BSA IgG titre for mice immunised with BSA or acBSA. Points are mean ±SD. Significant values where $p<0.05$, Mann-Whitney U. **d.** Titre for individual mice from week 6 BSA experiment in this figure to BSA and acBSA, as measured by ELISA. **e**. Titre for individual mice from week 14 of OVA experiment in figure 4.2a, to OVA and acOVA, as measured by ELISA. Lines connect measurement in samples from the same mouse. Dotted line at y=40 represents limit of detection by ELISA.

4.2.2 *Aldh2*$^{-/-}$ mice do not exhibit a greater immune response to HEL, BSA or HA

Next, I sought to determine whether *Aldh2*$^{-/-}$ mice had a greater humoral immune response to immunisation against HEL or BSA, as observed in OVA condition. I began by confirming the phenotype observed in OVA, repeating the ELISA described in Figure 4.3d-e, again reporting a significant increase in IgG titre in *Aldh2*$^{-/-}$ against *Aldh2*$^{+}$, without indication of an acetaldehyde specific immunodominant epitope (Figure 4.6a). I immunised *Aldh2*$^{+}$ and *Aldh2*$^{-/-}$ mice with 50µg/50µl HEL or BSA as before (Figure 4.6b) and determined the anti-HEL/BSA IgG titre by ELISA (Figure 4.6c, 4.6d). A humoral response was detected against both conditions, but this did not yield a significant difference between genotypes.

To examine an additional antigen relevant to vaccine development, I also tested using recombinant HA (variant H3N2 A/Aichi/2/1968, kindly gifted from Pramila Rijal, Oxford), routinely used for influenza vaccinations (Daniels et al., 1987; Slifka et al., 1998; Wang et al., 2018). For these experiments I administered mice with 25µg/50µl HA via intramuscular injection on weeks 0, 3, 6, and 9 (Figure 4.6e). I was also interested in whether there was a difference in immunological memory between *Aldh2*$^{+}$ and *Aldh2*$^{-/-}$ mice, so I extended the protocol to provide a 4th booster at week 18 (leaving time for the active plasma cell count and antibody concentration to decline). Again, however, I did not observe a significant difference between genotypes (Figure 4.6f). Anti-HA IgG titres remained high in the time between weeks 9-18, indicative of persistence of long-lived plasma cells, suggesting that this regime was not a sufficient test of immune memory (Slifka *et al*., 1998).

Together, these results suggest that the increased humoral response to OVA in *Aldh2*$^{-/-}$ must operate through a mechanism that is not generalisable to other antigens.

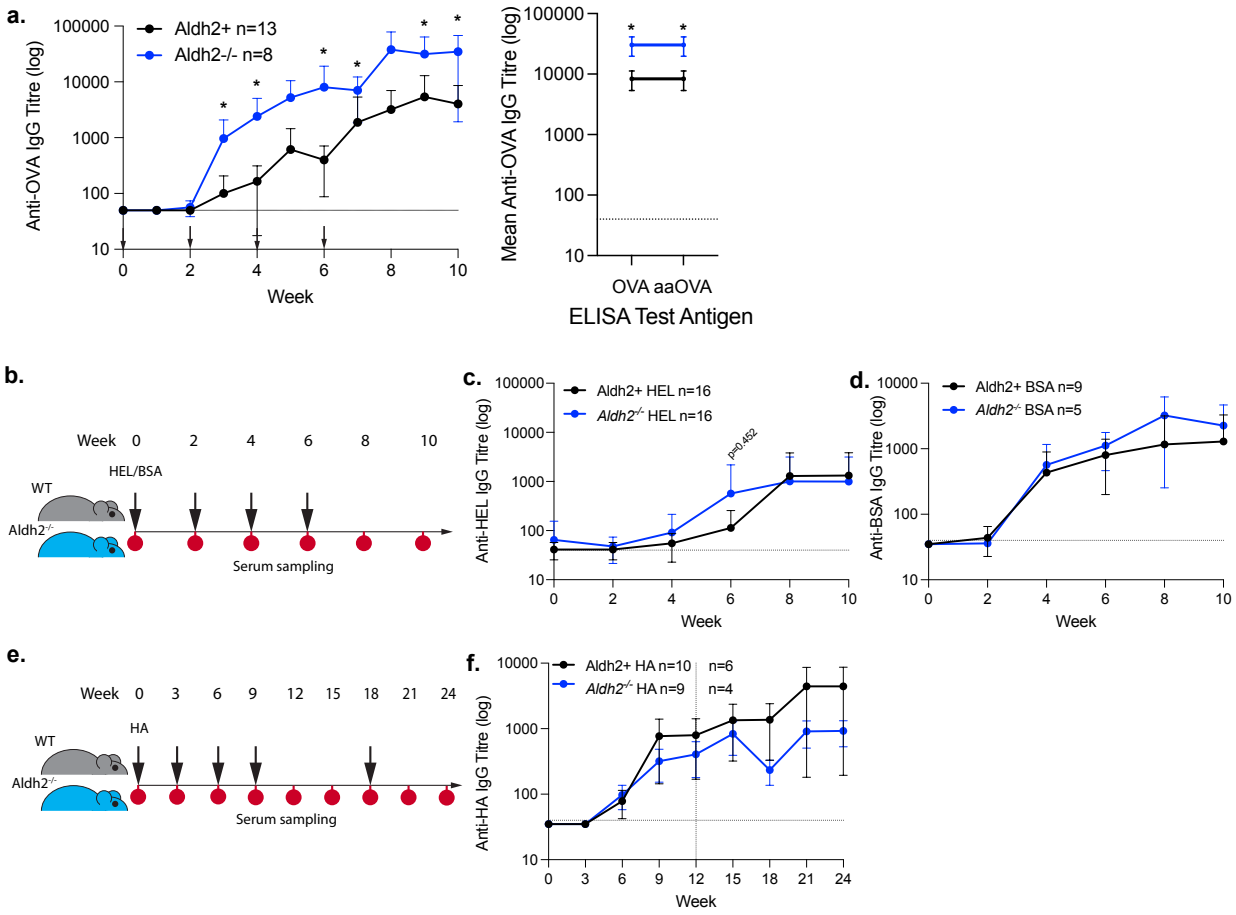


Figure 4.5 – *Aldh2*$^{-/-}$ mice do not have elevated humoral response to HEL, BSA or HA

(a) (left) figure 5.2a repeated (right) ELISA against OVA or acOVA using week 10 serum collected from the same experiment. Points represent mean ±SD, * p<0.05, Mann-Whitney U test. (**b**) Immunisation protocol for mice were immunised with 50µg/50µg unmodified antigen. (**c**) Anti-HEL IgG titre. **(d)** Anti-BSA IgG titre. U. **(e)** Immunisation protocol mice immunised with 25µg/50µg unmodified HA. **(f)** Anti-HA titre, a cohort of the experimental mice were not used beyond week 12, indicated by vertical line change in n numbers. Points are mean ±SD. Significant values where p<0.05, Mann-Whitney U test. Dotted line at y=40 represents limit of detection by ELISA.

4.2.3 Adjuvant stimulation does not induce a differential effect in *Aldh2*$^{-/-}$ mice

MPO activity is not restricted to generation of acetaldehyde (Hazen et al., 1998; Zhang et al., 2002), nor is *ALDH2* restricted to detoxification of acetaldehyde (O'Brien et al., 2005). Indeed, *ALDH2* does have affinity to glycolaldehyde, established by Allison & Fearon to increase antigenicity, and can be produced by serine oxidation. I hypothesised that by hyperactivate the innate immune system, I may increase the total aldehyde burden produced by MPO, thereby increasing endogenous tagging and adaptive immune activation, which may be differentially larger upon loss of *ALDH2*.

To test this hypothesis, I immunised *Aldh2*$^{+}$ and *Aldh2*$^{-/-}$ mice either BSA or HA, combined with the MF-59-like, oil-in-water based emulsion, Addavax. Addavax was selected as it has been characterised to promote neutrophil recruitment to the injection site, and subsequent antigen presentation and humoral increase (Calabro et al., 2011; O'Hagan et al., 2012; Zhang et al., 2025), verified to increase the humoral response to the same HA variant used in this study (Wang et al., 2018). In the BSA group, Addavax treatment failed to significantly increase the antibody titre against non-Addavax treated controls at any week (Mann-Whitney U, $p>0.05$), and also failed to induce a differential effect between genotypes at any time point (Figure 4.5a). This might suggest a sub-optimal priming strategy with Addavax for delivery of BSA, which may relate to its route of administration (subcutaneous), which has been reported to decrease responses (Bian et al., 2022). For the HA group, Addavax treatment significantly increased anti-IgG titres against non-Addavax treated groups at all time points (Mann-Whitney U, $p<0.05$), and, interestingly, in WT mice, there was a

significantly faster induction of the immune response, observed at week 6, which then disappeared by week 8 (Figure 4.5b).

These experiments suggest that *Aldh2*$^{-/-}$ may contribute to the differential effects of the innate immune system, but that these effects are not generalisable to different antigens, and can be negative.

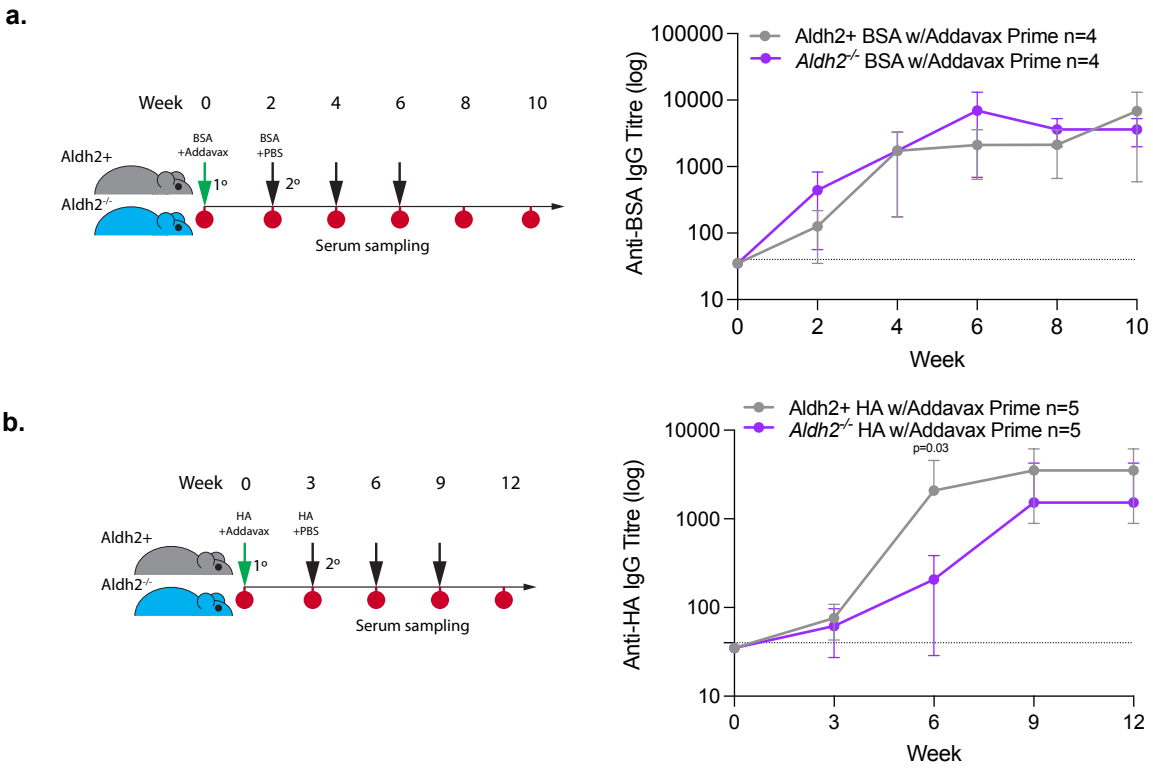


Figure 4.6 – Priming with Addavax causes differential immune response in *Aldh2*$^{-/-}$ against HA but not BSA
a. (Left) Immunisation protocol for mice primed with 10µg/50µl BSA in a 1:1 ratio with Addavax:PBS, subsequent boosters were performed with 50µg/50µl BSA in PBS (Right) anti-BSA IgG titres. **(b)** (Left) Immunisation protocol for mice primed with 10µg/50µl HA in a 1:1 ratio with Addavax:PBS, subsequent boosters were performed with 25µg/50µl HA in PBS (Right) anti-HA IgG titres. Points are mean ±SD. Significant values where $p<0.05$, Mann-Whitney U test. Dotted line at y=40 represents limit of detection by ELISA.

4.3 Discussion

Work in this chapter assessed whether the apparent increase in adaptive immune response against OVA observed in *Aldh2*$^{-/-}$ mice, was generalisable to other antigens, and operated through a mechanism where antigens are directly modified by acetaldehyde. I found that any effect due to *Aldh2*$^{-/-}$ or acetaldehyde modification, on the immunogenicity of HEL, BSA or HA, tended to be negative, although these were subject to considerable variability and in almost all cases appeared statistically insignificant, suggesting the study was underpowered to detect such differences. It is interesting to note that there was a significantly faster induction of a humoral responses to HA in response to Addavax treatment in *Aldh2*$^{+}$ mice, which may suggest *in vivo* modification dependent on ALDH2 conferring a negative effect on the humoral response, but does not confirm this as the mechanism the action. When considering the humoral response to HEL, OVA, BSA and HA, it appears there is some interaction with *ALDH2*, which can be positive or negative, depending on the antigen, which may indeed translate to antigens encountered in human history to facilitate selection.

4.3.1 Antigen specificity in aldehyde mediated immunity

The fact that the response of *Aldh2*$^{-/-}$ mice was different between antigens, is not an unusual finding, and highlights the highly specific nature of immune responses. Epitopes produced after antigen processing have specific properties, facilitating highly specific binding interactions with T-cell receptors (TCRs) and surface antibodies on B-cells (Allen et al., 1987; Majorek et al., 2012; Pishesha et al., 2022). The influence of modifications on these interactions, is dependent on the epitope sequence. This has

been described in detail with relation to formaldehyde modifications on different epitopes of diphtheria toxin, which can both increase and decrease binding affinity to specific antibodies in an epitope specific manner (Tommaso et al., 1994; Metz et al., 2003). If we consider the direct modification hypothesis as the mechanism through which *Aldh2*$^{-/-}$ mice alter the antigenicity of epitopes, this may explain the differential effects in OVA (and negative effects on HA in Addavax stimulation). Indeed, the delayed onset of a strong response in *Aldh2*$^{-/-}$ mice primed with HA+Addavax, may reflect a decrease in receptor affinity conferred by endogenous modification, which necessitates affinity maturation in germinal centres, meanwhile high affinity interactions could support direct plasma cell production and rapid immunity in *Aldh2*$^{+}$ mice (Kwak et al., 2019; Paus et al., 2006).

The present work suggests that direct modification of antigens by acetaldehyde did not have a strong effect on antigenicity, which would argue against the convergent effects of *ADH1B*2* and *ALDH2*2* on acetaldehyde, interacting with the immune response. However, there are some considerations to be made. Acetaldehyde is believed to stably adduct lysine residues on proteins (Hoffmann et al., 1993; Tuma et al., 1987) and acetaldehyde moieties on proteins have been described to confer immunogenicity. This is observed from presence of auto-antibodies from alcoholic patients with a higher affinity to acetaldehyde-modified proteins than unmodified (I Niemela et al., 1987; Worrall et al., 1991). Similar observations have been made when mice injected with acetaldehyde-modified KLH, assessed for their antibody titre against acetaldehyde-modified human plasma and haemoglobin (Israel et al., 1986). In these studies, antibodies were no further reactive to the unmodified protein. I did not observe this when comparing the titre of antibodies reactive to unmodified and

modified forms. A key difference between the present work and these past studies, is that I did not stabilise acetaldehyde modifications on antigens used for immunisation and ELISA using $NaBCNH_3$. Based on the apparent stability previously reported (Tuma *et al*., 1987; Hoffman *et al*., 1993), and the fact that modified antigens were stored at -20°C, I assumed there would not be significant reversal. However, Isreal *et al*. also probed whether $NaBCNH_3$ treatment affected binding affinity, and found there was a significant increase when adducts were stabilised. Indeed, Tuma *et al*., 1987 observed more stable adducts when $NaBCNH_3$ was used. Furthermore, Niemela *et al* and Worrall *et al* both also observed higher binding affinity from the serum antibodies of the non-alcoholic patients to acetaldehyde modified haemoglobin, which would argue in favour of a generalised impact of acetaldehyde in increasing antibody affinity, which would not have been observed in my study. Instability of acetaldehyde adducts may have resulted in incomplete adduction of the injected protein, preventing recognition of acetaldehyde specific moieties.

These results contrast to the observations of glycolaldehyde adduction increasing antigenicity observed by Allison & Fearon, but this may be expected based on the chemical differences between acetaldehyde and glycolaldehyde. For one, glycolaldehyde adduction forms adducts believed to be more stable, producing ketoamine structures that can resist chemical destabilisation (Acharya & Manning, 1983). Furthermore, where glycolaldehyde and acetaldehyde both form Schiff bases upon reaction with amine residues, glycolaldehyde undergoes Amadori rearrangement, generating new aldehyde groups that can go onto crosslink with other amino acids, while acetaldehyde is less likely to undergo further spontaneous reaction (Acharya & Manning, 1983; Glombt & Monnier, 1995; Kamps et al., 2019). Therefore,

the chemical nature of the modification mediated by the aldehyde will significantly affect its influence on antigenicity.

While not observed here with acetaldehyde adducts, other aldehyde modifications that are abundant owing to lack of detoxification by ALDH2, may be significant and contribute to the variable responses of *Aldh2*$^{-/-}$ mice. Formaldehyde modification of proteins has been reported to increase the immunogenicity of antigens, as observed in vaccination against toxins of *Haemophilus ducreyi* (Lagergård et al., 2007) or *Bacillus anthracis* (Little et al., 2007). As well as more readily forming adducts on amino acid residues (Kamps et al., 2019), formaldehyde adducts are well reported to produce protein-protein crosslinks (Kast & Klockenbusch, 2010), which could increase immunogenicity by promoting protein aggregation (Benne *et al*., 2016). Besides formaldehyde, larger, more complex aldehydes, such as 4-hydroxynonenal (4-HNE) adduction and malondialdehyde (MDA), are capable of producing distinct epitopes on proteins that are specific to the modification, rather than the native protein (Thiele et al., 1998). These are both substrates of *ALDH2*, and can be produced from lipid peroxidation (Vasillou et al., 2004), another significant product of MPO activity (Zhang et al., 2002). Indeed, 4-HNE and MDA adducts are often elevated in alcoholic patients and have been associated with auto-immune disorders, facilitating production of self-active antibodies (Khatoon et al., 2012; Tuma, 2002). These modifications would be undetected by ELISA against the native antigen, so it would be interesting to look for alternative adducts that might support protection.

The measure of IgG binding activation by ELISA, is an efficient means through which we can integrate the activities of innate and adaptive immune system into a final

antibody response. However, this only captures a fraction of activity of the adaptive system. First, not all IgG antibodies are the same, and our ELISA does not differentiate between the abundance of neutralising antibodies, which would be more effective in killing pathogens early in the immune response (Kwak et al., 2019; Paus et al., 2006; Roost et al., 1995). Furthermore, this study did not describe the effect on conferring immunological memory. Antigen affinity to the B-cell receptors, and TCRs on T-follicular helper cells present in germinal centres, are both essential determinants of B-cell differentiation, which may indeed be influenced by changing affinity relationships due to aldehyde modification (Benson et al., 2007; Crotty, 2014; Shih et al., 2002). Finally, while focus was on antibody production, the other roles of T-cells were not considered, including those involved in driving cell-mediated immunity and innate immune regulation, which are also directed in part by affinity-based reactions with the TCR (Crotty, 2014; Tubo et al., 2013). Intriguingly, improved cell mediated protection has been suggested in a cohort of Japenese participants following COVID-19 vaccination. Here, *ALDH2*2* carriers had a weaker IgG response, but enhanced T-cell response (Bogahawaththa et al., 2024; Matsumoto et al., 2022). These studies are confounded by a very small sample size, but nevertheless highlight the necessity of further study.

4.3.2 The direct killing hypothesis and limitations of the immunity hypothesis

Besides supporting an adaptive immune response, there is also the hypothesis where aldehydes contribute to the direct killing of pathogens carried out by the innate immune system, as laid out by Darwin & Stanley, 2022. These authors applied this hypothesis to a study in *Aldh2*$^{-/-}$ mice, (Berry et al., 2023), to ascertain whether ALDH2 deficiency

conferred protection to *Mycobacterium tuberculosis* and *Francisella tularensis*. Their study did not provide strong evidence for this. They found that *in vitro*, the aldehydes formaldehyde, 4-HNE and MDA (all catabolised by ALDH2) only induced toxicity at high concentrations (the lowest concentration of formaldehyde tested was 100μM), and only increased when co-administered with nitric oxide. However, the effect of nitric oxide alone was not reported, hence synergy cannot be concluded. Endogenous serum concentrations of formaldehyde in *Aldh2*$^{-/-}$ mice, is only ~9μM (Dingler *et al*., 2020), while the actual amount of aldehyde production by MPO metabolism *in vivo*, is unknown. Furthermore, the reduction in colony forming units of bacteria in *Aldh2*$^{-/-}$ mice was very mild, and there was no comment on the overall survival of the mice, hence there is little to support *Adlh2*$^{-/-}$ conferring meaningful resistance to infection.

It is worth considering that in producing an aldehyde burst significant enough to directly kill pathogens, this may also contribute to significant cell death in an *ALDH2* deficient host. Similarly, significant endogenous modifications of host proteins induced during such a burst may prompt auto-immune responses against (Theofilopoulos et al., 2017) (although association between *ALDH2*2* and auto-immune disorders has not been reported in GWAS studies of East-Asian, including studies of Grave's disease (Nakabayashi et al., 2011), or lupus (Han et al., 2009; Yin et al., 2021)). Therefore, while immune protection supports the clearance invading pathogens, it may come at a significant trade off to the host. Indeed, a recent publication by Sun et al., 2024, described loss of ALDH1B1 in mice, a mitochondrial dehydrogenase with 75% homology to *ALDH2*, and with significant capacity to detoxify formaldehyde and acetaldehyde (Stagos et al., 2010), to worsen symptoms of acute influenza A viral infection, arguing against a protective effect of aldehydes in the immune system.

Future work would benefit from extended study of *Aldh2*$^{-/-}$ mice to infection challenge to appreciate whether the hypothesised mechanisms of disease protection indeed carry significant trade-offs in host tolerance. It would also be interesting to use Biobanked samples from human *ALDH2*2* carriers to directly assess whether they have mechanistic differences in their ability to respond to different infections, and in their disease tolerance, thus understanding whether *ALDH2*2* does indeed promote resistance independently of alcohol consumption, or if this is too expensive.

Finally, an intriguing association between *ALDH2*2* and reduced incidence of tuberculosis has been described in a cohort of Korean men (Park et al., 2014). Unfortunately, the authors did not investigate whether this was separable from the strongest effect of *ALDH2*2* in reducing alcohol consumption, known to worsen tuberculosis (as well as other infectious diseases) (Barr et al., 2016; Imtiaz et al., 2017). Such associations have also not been reported in other GWAS studies of tuberculosis risk in China (where *ALDH2*2* is common) (Chang et al., 2024; Zheng et al., 2018). Nevertheless, while I have assumed that effects on drinking behaviour will not affect the reproductive population by conferring risk to the young individual, the ability of alcohol abstinence to protect against infectious diseases such as tuberculosis, and others (e.g. HBV, particularly prevalent in East Asia (Lin & Cheng, 2002)), may support the success of the parental population, thus improving reproductive success, and may indeed still be significant.

4.3.3 Study limitations

Substantial variation in the immune response was observed across treatment groups, especially in treatments with adjuvants, and long-term HA administration, and in the work with acOVA, which reflects a need for larger N-numbers to increase statistical power. Furthermore, titre determination by ELISA was performed based on 4-fold serial dilutions (see methods), which may have been too steep a gradient to provide the resolution required to observe more subtle effects between treatments, and contributed to the variability (see figure Figure 4.5d-e). This becomes more apparent upon estimations of high antibody titres, where resolution is decreased. Finally, given the absence of *Aldh2*$^{-/-}$ effects against HEL, BSA and HA, the immunisation schedule with OVA should be reproduced, accounting for the effects of facility variation (OVA experiments performed by F.Dingler in Cambridge, I performed my experiments in Oxford) and genetic backcrossing of the *Aldh2*^{tm1a} strain (important to remove acquisition of confounding genetic variants that may introduce problems) (Keane et al., 2011).

4.3.4 Conclusion

The work herein suggests that ALDH2-deficiency can influence the adaptive immune response in an antigen specific manner, in the absence of exogenous alcohol consumption, giving strength to the argument that *ALDH2*2* individuals may present an immune advantage in certain contexts. The phenomenal complexity of the immune system makes it difficult to dissect where exactly this may arise, however, this study argues against the role of acetaldehyde modifications that may be encountered *in vivo* in *ALDH2*2* carriers (and indeed, *ADH1B*2* carriers), as contributing. To support the direct modification hypothesis, future work may focus on characterising how aldehyde

modification affects the affinity of known epitopes, and whether these higher affinity interactions are observed in antibodies derived from *Aldh2*$^{-/-}$ mice. Other work may take advantage of the growing abundance of population data, such as in the UKBiobank, Japanese BioBank and Kadoorie biobank, to more specifically explore the relationship between *ALDH2*2* incidence and disease protection.

Chapter 5

The role of systemic formaldehyde on haematopoietic ageing and tissue homeostasis

5.1 Introduction

DNA damage accumulates in the blood with age (Beerman *et al*., 2014), but understanding the functional consequences of this in driving haematopoietic ageing is complicated to study, as it is hard to decouple DNA damage from other pleiotropic contributors of the ageing process. Much of our understanding of DNA damage as a driver of haematopoietic decline comes from analysis of genetic models where DNA repair is deficient (Rossi *et al*., 2007), or damage is induced by chemical or radiation treatment (Pilzecker et al., 2017; Yu et al., 2016), which are not representative of normal ageing. Furthermore, HSCs in mice are known to accumulate mutations at a slower rate than in humans, suggesting mice may be protected from DNA damage, supporting the maintenance of a healthy, polyclonal blood population into the end of life (Kapadia *et al*., 2025).

Analysis of two-tier deficient mice (*Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ and *Adh5*$^{-/-}$*Fancd2*$^{-/-}$) have demonstrated how endogenously produced formaldehyde is a significant driver of

endogenous DNA damage in ageing, and can drive characteristic features of haematopoietic ageing, such as anaemia, myeloid-bias, transcriptional ageing score, and functional decline (Langevin *et al*., 2011; Garaycoechea *et al*., 2012; Pontel *et al*., 2015; Garaycoechea *et al*., 2018; Wang *et al*., 2020). However, as these models lack functional DNA repair, they are not applicable to study physiological ageing.

The *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ model also demonstrates hallmarks of aged haematopoiesis, with clear indication that this is rooted in DNA damage (increased micronuclei, sister chromatid exchanges and mutation accumulation) (Dingler *et al*., 2020; Oka *et al*., 2020). These functional changes coincide with the onset of clonal haematopoiesis, suggesting that formaldehyde is an important agent of selection in the haematopoietic compartment. This mouse therefore has potential as a model of physiological ageing driven by an endogenous metabolite in the presence of functional effector signalling and DNA repair (tier-2) (Dingler *et al*., 2020; Oka *et al*., 2020). However, this model is limited as most mice die within the peri-natal window, and survivors develop stochastic multi-morbidities (e.g. lymphoma and hepatocellular carcinoma present in the same mice (Dingler *et al*., 2020; N.Oberbeck's Thesis, Cambridge)). The early origins of clonal emergence, suggests that extreme selection happens early in the life of these mice, and the resulting mice are mosaics adapted to tolerate a high formaldehyde burden. Therefore, it is hard to study formaldehyde-driven haematopoietic ageing in these mice as a gradual, time-dependent process.

Extending from this, aged blood is associated with the decline of other tissues, such as secreting proteins that drive inflammation and senescence (Yousefzadeh *et al*., 2021; Wong *et al*, 2023). Haematopoietic capacity for formaldehyde detoxification also

has significant consequence on ageing of peripheral tissues, as shown by H.Russell, when reinstating formaldehyde detoxification specifically in the blood of *Adh5*$^{-/-}$*Csb*$^{-/-}$ mice, suppressed formaldehyde driven disorders in the brain and kidney through reducing the formaldehyde load in these tissues (H.Russell's Thesis, Oxford). In the inverse situation, we may expect formaldehyde produced, or not cleared, from the blood to increase the load in other tissues, thus driving pathology. However, whether the effect of the blood on other tissues is related its role in regulating the systemic formaldehyde load, or due to its own intrinsic ageing, is unclear. Resolving this distinction is key to understanding disease progression of formaldehyde-related disorders and guiding therapeutic strategies.

In this chapter, I use an *Vav1*^{iCre} inducible system to genetically remove *ALDH2* specifically in the blood of globally *ADH5*-deficient mice. This is designed to uncouple the wider pathological phenotypes observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, from the phenotypes of ageing found in the blood, allowing the impact of formaldehyde in driving ageing specifically in the blood, to be studied. I also use this model to ask what role the blood plays in controlling the formaldehyde load in wider tissues, and whether the blood contributes significantly to their functional decline.

5.2 Results

5.2.1 Peripheral *ALDH2* is sufficient to suppress the severe phenotype of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice

To restrict the formaldehyde detoxification deficiency of the constitutive *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mouse to the blood, I utilised the tm1c-allele of *Aldh2* (*Aldh2*^{c}) (Skarnes *et al*., 2011), where the critical exon 4 of *Aldh2* is flanked by loxP sites in intronic regions, which results in a functional tm1c-allele that can be converted to a null allele (*Aldh2*$^{-}$) by Cre-mediated recombination (Figure 5.1a). I crossed males carrying *Aldh2*^{c} with females hemizygous for Vav1-iCre (homozygosity of Vav1-iCre is embryonic lethal (Joseph et al., 2013)), therefore all Vav1-iCre carrying mice are hemizygous (*Vav1*$^{iCre+}$). Deletion of *Aldh2*^{c} upon Vav1-iCre expression is complete (Figure 5.1b-d) and I observed no instance of emergence of non-deleted clones in ageing blood populations (Figure 5.1e).

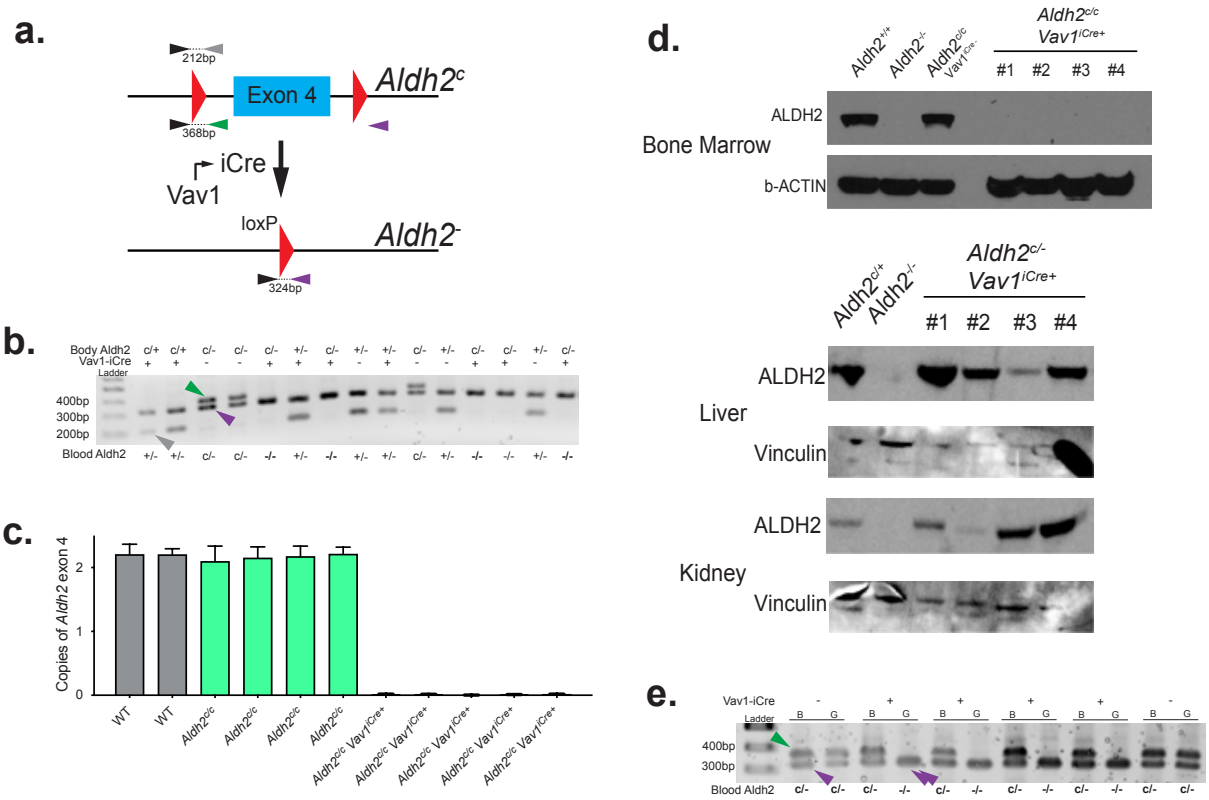
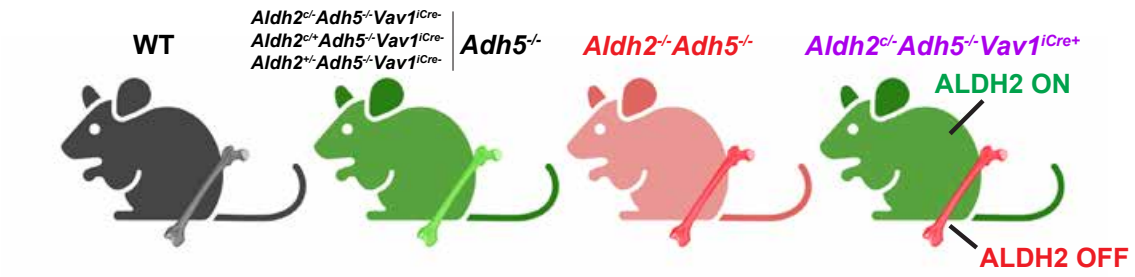


Figure 5.1 Construction and validation of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ model
(a) In the Aldh2(tm1c) (*Aldh2*^{c}) allele, the critical exon 4 is flanked by 2 loxP sites that undergo recombination in the presence of Cre, removing Exon 4 and one loxP site, preventing recombination, producing Aldh2(tm1d) (*Aldh2*$^{-}$). Cre expression is driven by the published Vav1-iCre allele, which is active in haematopoietic lineages from ~E11, producing *Aldh2*$^{-}$ specifically in the haematopoietic system. Presence of recombined alleles can be detected by multiplex-PCR, with *Aldh2*^{c} detected by the

black+green primer pair, giving a 368bp product, and *Aldh2*$^{-}$ black+purple primer pair giving a 324bp product. WT Aldh2 can be detected by black+grey primer pair, giving a 21bp product (WT allele does not contain loxP site). (b) Representative multiplex PCR gel for genotyping Aldh2 allele on gDNA from peripheral blood samples of adult mice. Arrows highlight bands corresponding to alleles present. (c) Quantification of efficiency of *Aldh2*^{c} deletion, performed by Juan Garayacoechea. (d) Western blot for ALDH2 protein presence in whole bone marrow (performed by Fred Langevin), liver and kidney. Loading controls are b-actin (bone marrow) or vinculin (liver, kidney). (e) Representative multiplex PCR gel for genotyping Aldh2 allele against gDNA from brain (B) sorted granulocytes (G) from 1 year old mice.

These mice are bred in a tier-1 proficient background with at least one copy of either *ALDH2* or *ADH5*. Haploinsufficiency has not been reported for *Aldh2* or *Adh5* in the context of tier-2 proficiency, with a single copy of *Aldh2*$^{+}$ or *Adh5*$^{+}$ sufficient to compensate for *Adh5*$^{-/-}$ or *Aldh2*$^{-/-}$ respectively (N.Oberbeck's Thesis, Chapter 6 (6.3.2); Dingler *et al*., 2020). This resulted in the mouse model, *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$, which is tier-1 proficient in the body (*Aldh2*$^{+/-}$*Adh5*$^{-/-}$, functionally *Adh5*$^{-/-}$) but deficient in the haematopoietic tissues (*Aldh2*$^{-/-}$*Adh5*$^{-/-}$). *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice were completely protected from the profound perinatal lethality observed in *Aldh2*$^{-/-}$ *Adh5*$^{-/-}$ animals (Figure 5.2a-c), and there have been no naturally occurring deaths or instances of morbidity so far from mice kept up to 1 year, in stark contrast to *Aldh2*$^{-/-}$ *Adh5*$^{-/-}$ mice (Dingler *et al*., 2020).

a.



b.

Genotype of offspring at 2-3 weeks

Aldh2c/c Adh5-/- x *Aldh2+/- Adh5+/- Vav1iCre+/-*

	Aldh2c/+ Adh5+/-	*Aldh2c/- Adh5+/-*	*Aldh2c/+ Adh5-/-*	*Aldh2c/- Adh5-/-*	
Exp.	12.50%	12.50%	12.50%	12.50%	
Obs. Vav1iCre-	15%, n = 12	16.25%, n = 13	3.75%, n = 3	23.8%, n = 19	
Obs. Vav1iCre+	11.25%, n = 9	3.75%, n = 3	6.25%, n = 5	20%, n = 16	p=0.151

Aldh2c/+ Adh5-/- x *Aldh2-/- Adh5+/- Vav1iCre+/-*

	Aldh2c/-Adh5+/-	*Aldh2+/- Adh5+/-*	*Aldh2+/- Adh5-/-*	*Aldh2c/- Adh5-/-*	
Exp.	12.50%	12.50%	12.50%	12.50%	
Obs. Vav1iCre-	8.8%, n = 3	20.6%, n = 7	5.9%, n = 2	14.7%, n = 5	
Obs. Vav1iCre+	8.8%, n = 3	11.8%, n = 4	14.7%, n = 5	14.7%, n = 5	p=0.5557

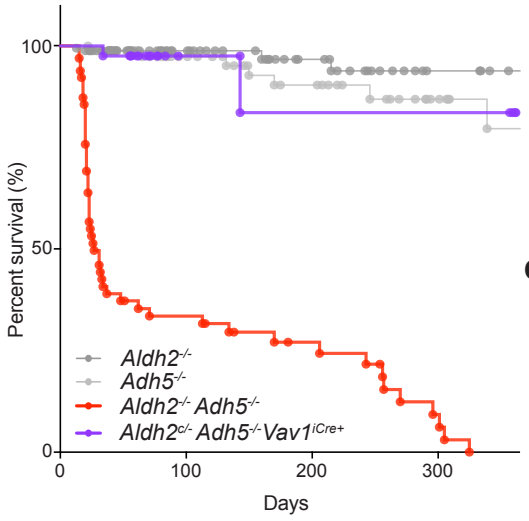
Aldh2c/c Adh5-/- x *Aldh2-/- Adh5+/- Vav1iCre+/-*

	Aldh2c/- Adh5+/-	*Aldh2c/- Adh5-/-*	
Exp.	25%	25%	
Obs. Vav1iCre-	19.64%, n = 11	26.79%, n = 15	
Obs. Vav1iCre+	17.86%, n = 10	35.71%, n = 20	p=0.5839

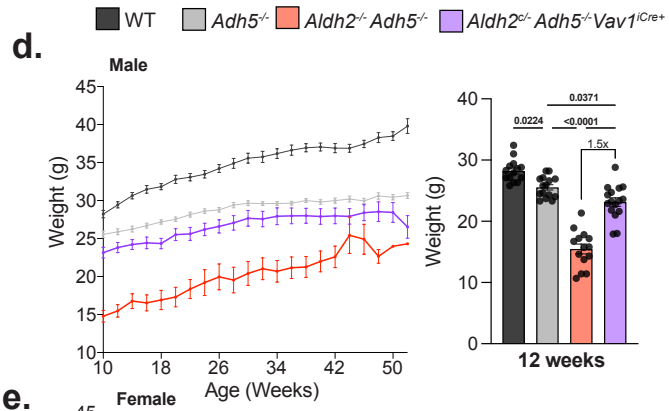
Aldh2c/c Adh5-/- x *Aldh2+/- Adh5-/- Vav1iCre+/-*

	Aldh2c/+ Adh5-/-	*Aldh2c/- Adh5-/-*	
Exp.	25%	25%	
Obs. Vav1iCre-	15.9%, n = 7	15.9%, n = 7	
Obs. Vav1iCre+	34.1%, n = 14	34.1%, n = 14	p>0.9999

c.



d.



e.

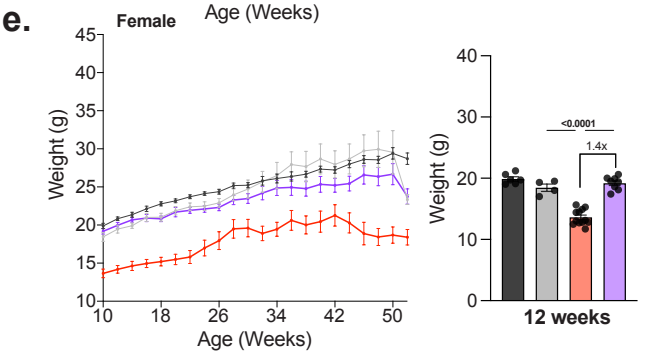


Figure 5.2 *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* have suppressed the perinatal lethality and severe growth defect of *Aldh2$^{-/-}$Adh5$^{-/-}$* mice
(a) Explanation of genotyping classification of mice used in this study – Adh5$^{-/-}$ mice are grouped based on presence of at least one functional *ALDH2* allele globally, while *ADH5* deficient globally. In *Aldh2$^{-/-}$Adh5$^{-/-}$* mice are deficient in *ADH5* and *ALDH2* globally. In *Aldh2$^{c/-}$ Adh5$^{-/-}$ Vav1^{iCre+}*, *ALDH2* deletion is restricted to the blood and peripheral tissues contain one copy of functional *ALDH2*. (b) Observed and expected frequencies of all pups from breeding combinations (displayed above), genotyped at 3 weeks age. *Aldh2$^{c/-}$ Adh5$^{-/-}$ Vav1^{iCre+}* genotype is highlighted in purple, P values calculated from Fisher's test with 95% confidence interval. (c) Kaplan-Meier survival plot demonstrating survival of *Aldh2$^{c/-}$ Adh5$^{-/-}$ Vav1^{iCre+}* mice in this study (n = 40) overlaid with survival data for *Aldh2$^{-/-}$*, *Adh5$^{-/-}$*, and *Aldh2$^{-/-}$ Adh5$^{-/-}$* mice (n = 166, 89, 67 respectively), published in Dingler *et al*., 2020. Each point represents death of individual mouse, experimental mice are included and taken down for experiments as censored events. (d) Total body mass of male mice as mean ± SEM over time (left, n at start = 14, 15, 17, 16) and at 12 weeks age (right, n = 14, 15, 14, 16). (e) Total body mass of female mice as mean ± SEM over time (left, n at start = 6, 8, 19, 8) and at 12 weeks age (right, n = 6, 4, 12, 8). p-values determined by ANOVA with Tukey's post hoc test (d-e, right) where significance $P<0.05$.

Male *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ mice are significantly smaller than WT mice (consistent with the understood *Adh5*$^{-/-}$ phenotype), male *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ were significantly smaller than male *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ mice, and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ was significantly smaller than all genotypes (Figure 5.2d-e). The difference between *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ and *Aldh2*$^{c/-}$ *Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice was sexually dimorphic at 12 weeks (not observed at this age in females, although a difference appears to emerge with age; consistent with previous reports of sexual dimorphism in other *Adh5*$^{-/-}$ models (Mulderrig *et al*., 2021; C.Millington, manuscript in preparation)), but there was no significant difference between male and female *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice (sex difference within each genotype determined by t-test, $p>0.005$, in *Aldh2*$^{-/-}$ *Adh5*$^{-/-}$ $p=0.06$).

These results mean the expression of ALDH2 in the peripheral tissues is sufficient to confer protection to the body from the systemic formaldehyde load, while the weight phenotype suggests formaldehyde stress is still conferred by increased endogenous production.

5.2.2 *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice have a high formaldehyde burden in peripheral tissues

To understand how *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice are protected from the profound peri-natal lethality, I quantified the formaldehyde burden to peripheral tissues. I did this through measuring the level of reduced N^2-HOMe-dG (N^2-Me-dG) adducts on DNA, by LC-MS, a biomarker of formaldehyde exposure (Lu *et al*., 2010), in different mouse tissues (Figure 5.3). The mean burden of adducts in the bone marrow of male mice was significantly higher in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice than in *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ mice,

consistent with the higher formaldehyde burden in the absence of tier 1 in haematopoietic tissues (also observed in Dingler *et al*., 2020). The number of bone marrow samples in female *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$, or male and female *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, was too low to perform statistical analysis. However, in the spleen, there was a clear increase in adduct level of these groups compared to *Adh5*$^{-/-}$, while being genetically equivalent to each other. Therefore, the haematopoietic formaldehyde burden across these mice is increased.

Strikingly, despite the peripheral organs all retaining *ALDH2*, the adduct levels in kidney and liver of male *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice was similar to that in *Aldh2*$^{-/-}$ *Adh5*$^{-/-}$ animals, while that in the brain was intermediate between the *Adh5*$^{-/-}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$. Furthermore, there was sexual dimorphism in the adduct level in the kidney and liver of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice that was not observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$, also consistent with the weight phenotype, suggesting that males are under a greater formaldehyde burden than females. As these sexually dimorphic effects in the brain, liver and kidney are not also observed in the spleen (haematopoietic tissue), it suggests that the sexual dimorphism stems from *ALDH2* regulation intrinsic to that tissue. Overall, more adducts were observed in the peripheral tissues than in the haematopoietic tissues, which may reflect different tissue dynamics (more cycling in the bone marrow/spleen) reducing the rate at which adducts accumulate.

These results suggest that tier-1 protection in peripheral tissues by *ALDH2*, does not operate by directly decreasing the formaldehyde load on DNA in peripheral tissues, and emphasises the blood as an essential route of systemic detoxification.

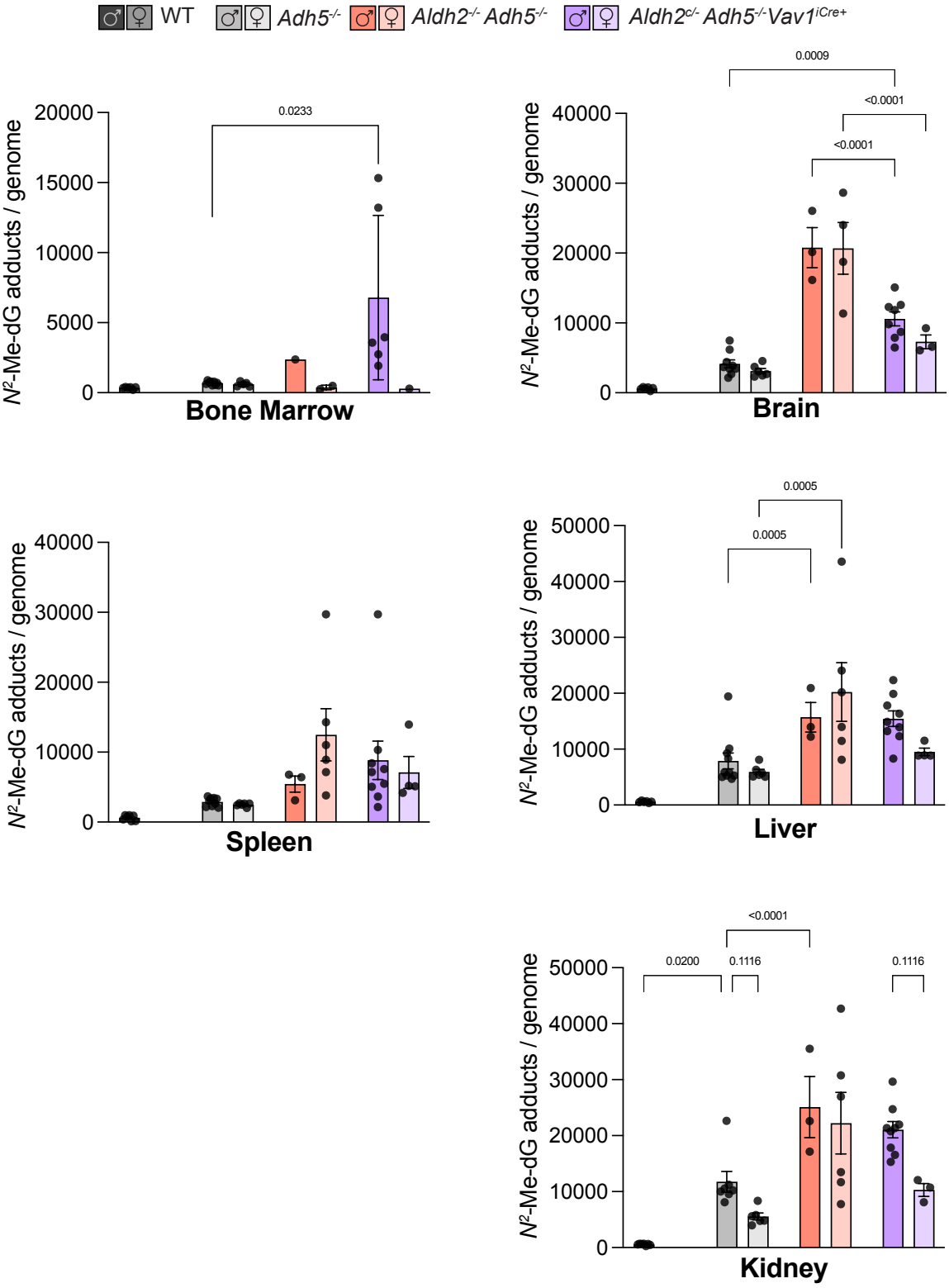


Figure 5.3 Quantification of N^2-Me-dG adducts in tissues at 8-12 weeks
Quantification of adducts in bone marrow, brain, spleen, liver and kidney in 8-12 week-old mice. Data represent mean±SEM, male n = 7, 7, 3, 9; female n = 0, 7, 6, 5. Data points on graph represent individual mice, analysis performed by two-way ANOVA with Tukey's post hoc test, significance where P<0.05 (one-way ANOVA with Tukey's post hoc performed on male WT, *Adh5*$^{-/-}$, *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ bone marrow samples only).

5.2.3 High formaldehyde burden in tissues does not correlate with tissue damage

Intrigued by the observation that *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice shared similarly high formaldehyde burden in peripheral organs, but diverged markedly in their survival at 8-12 weeks, I asked whether this correlated with markers of genotoxic and organ stress in these tissues.

Nuclear polyploidisation in the liver occurs in response to genotoxic stress, and is observed in two-tier deficient *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ mice, correlating with a high adduct burden in *Adh5*$^{-/-}$ mice (Pontel *et al*., 2015). Interestingly, there was no indication of increased nuclear polyploidisation in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ or *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice over controls, despite their significant formaldehyde burden (Figure 5.4).

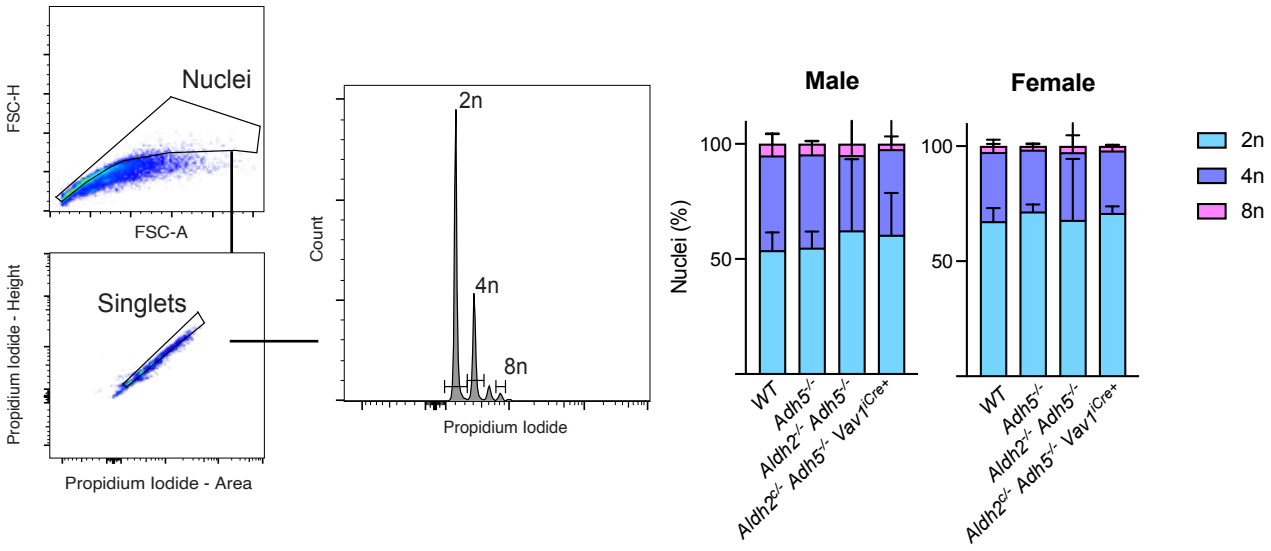


Figure 5.4 Liver ploidy analysis
Representative flow cytometry gating of purified nuclei extracted from livers (6-12 weeks old) and quantification for chromatin content by PI staining (left) quantification results (right). 8n was the largest detected. Bars represent mean±SD, males n = 6, 2, 5, 8; females n = 4, 4, 8, 5. p values measured by ANOVA with Tukey's post-hoc test, significance where P < 0.05 (*Aldh2*$^{+/-}$*Adh5*$^{-/-}$ were excluded from statistical comparison in males due to low n).

Turning to biochemical markers, circulating GDF-15 is a generalised marker of stress (Lockhart et al., 2020; Wang et al., 2021) and is inducible by formaldehyde stress, contributing to the anorexic phenotype observed in *Adh5*$^{-/-}$*Csb*$^{-/-}$ mice (Mulderrig *et al*., 2021). Circulating GDF-15 was significantly higher in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice (consistent with their failure to gain weight), while *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice were indistinguishable from WT, suggesting *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice are not experiencing the same chronic stress observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$. Consistent with this, *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice were indistinguishable from WT across markers of kidney function (assessed by presence of waste products, creatinine and urea, filtered by kidney), liver damage (asparate aminotransferase (AST) and alanine transaminase (ALT)), or liver function (Albumin synthesis, bilirubin) (Figure 5.5). By contrast, *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ males showed signs of kidney dysfunction with significantly elevated urea (although only in males), as well as liver stress with significantly elevated ALT,

elevated AST and decreased albumin synthesis, although the variance was high across these. At 1 year, these markers in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ males significantly increased, indicative of liver and kidney dysfunction. By contrast, at 1 year, *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ mice still had lower levels of circulating GDF-15 and no sign of increased kidney or liver damage, but did have significantly elevated GDF-15 compared to *Aldh2*$^{+/-}$*Adh5*$^{-/-}$mice, indicative of elevated chronic stress to tissues (Figure 5.6). Ageing data for females followed a similar trend but is confounded by low number of mice in each group.

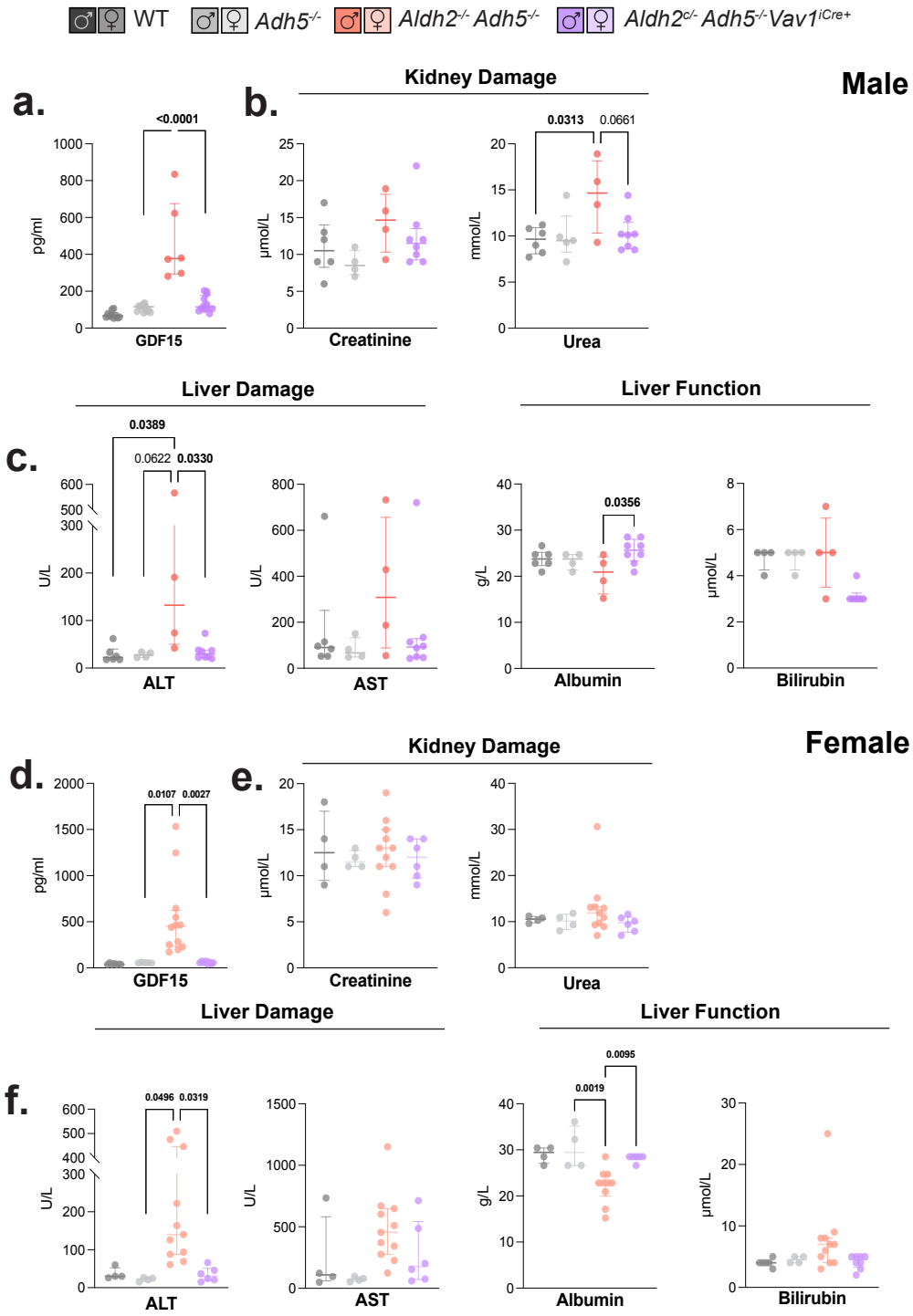


Figure 5.5 Serum measurement of GDF-15 and tissue damage markers at 8-12 weeks

(a-c) Serum concentration from terminal bleeds of 8-12-week males (except *Aldh2*$^{-/-}$ *Adh5*$^{-/-}$ which are terminal samples from 4-27 weeks) of circulating GDF-15 (a), kidney function markers, creatinine and urea (b), liver damage markers, ALT and AST, and function markers, albumin and bilirubin (c). (d-f) Serum concentrations in 8-12-week females (except *Aldh2*$^{-/-}$ *Adh5*$^{-/-}$, which are all terminal samples from 5-26 weeks, 1 mice 33 weeks) of GDF-15 (d), kidney function (e) and liver damage and function (f). Bars represent media ± inter-quartile range, males n = 10, 11, 6, 12; females n = 4, 11, 6. Data points indicate individual mice, p values measured by ANOVA with Tukey's post-hoc test, significance marked where $P < 0.05$.

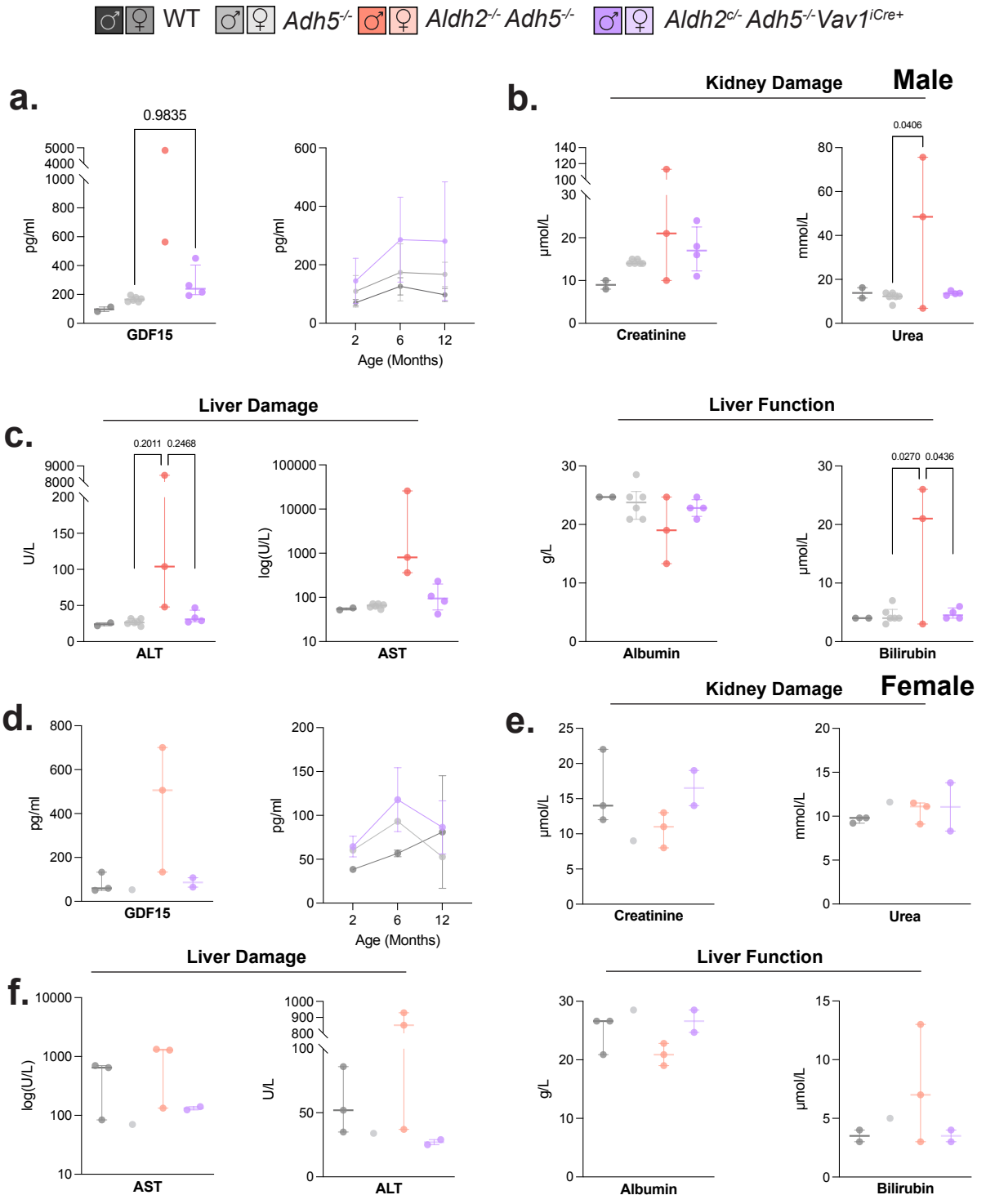


Figure 5.6 Serum measurement of GDF-15 and tissue damage markers at 1 year
Serum concentration from terminal bleeds of 1-year-old males (a-c) and females (d-f) of circulating GDF-15 (a males, d females), kidney function markers, creatinine and urea (b males, d females), liver damage and function markers (c males, f females). Bars represent media ± inter-quartile range, males n = 2, 6, 3, 4; females n = 3, 1, 3, 2. Data points indicate individual mice, p values measured by ANOVA with Tukey's post-hoc test where ns > 0.05 (not performed on females).

The absence of correlation between adduct burden in kidney and liver in both *Aldh2*$^{c/-}$ *Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, and liver nuclear polyploidisation or significant stress, contrast with phenotypes of liver and kidney organ failure observed in *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ and *Adh5*$^{-/-}$*Csb*$^{-/-}$ mice (Pontel *et al*., 2015; Mulderrig *et al*., 2021). This suggests that the mechanism by which formaldehyde is driving tissue pathology in presence of intact DNA repair machinery, is not through over-whelming tier-2 protection.

5.2.4 *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice exhibit normal haematopoiesis at 8-12 weeks

Aldh2$^{-/-}$*Adh5*$^{-/-}$ mice develop anaemia, with a decline in all stem and progenitor populations and myeloid bias, and exhibit a thymic defect (loss of cellularity and maturation) in young animals. As *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+/-}$ mice are genetically equivalent in their haematopoietic tissues and experience a high formaldehyde burden, I asked whether this correlated to the same perturbations of haematopoiesis.

Figure 5.7 details the immunophenotypic markers used to identify different haematopoietic populations across this study and their relationship to each other in the haematopoietic hierarchy. These markers are consistent with Dingler *et al*. (2020), allowing phenotypic comparison with *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, but I also included examination of MPP sub-populations (Pietra *et* al., 2015; Rodriquez-Fraticelli *et al.*, 2018).

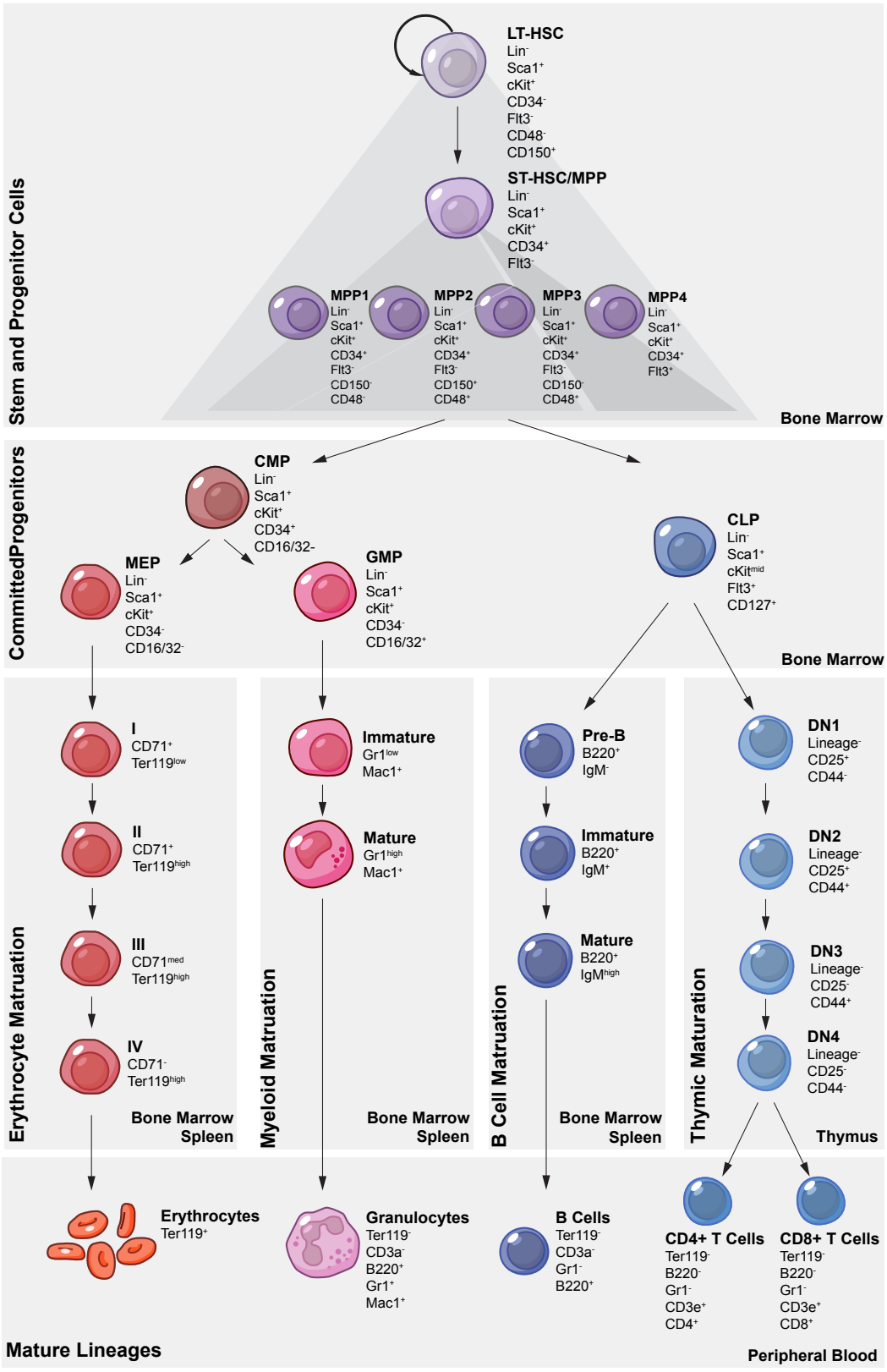


Figure 5.7 Haematopoietic tree with immunohistochemical markers used to identify cell populations within the different blood compartments, adapted from Juan Garaycoechea's Thesis, Cambridge.

I did not observe any difference in the abundance of immunophenotypic haematopoietic stem and progenitor (Figure 5.8), or maturing lineages in the bone marrow (Figure 5.9a-c), thymus (Figure 5.9g), or mature populations in the peripheral blood (Figure 5.9d-f, 5.9h), in male *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+/-}$ mice against control mice. This was also the case in females (Supplementary Figures 5.1, 5.2). *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ mice did present a reduced red blood cell count, and increased mean corpuscular volume, suggestive of macrocytosis and cell death (Figure 5.10). A small increase in MCV was also observed in females. Overall, there appears to be a suppression of the quantitative decline in HSC numbers observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$, which may reflect cell-extrinsic protection conferred by *ALDH2* in peripheral tissues.

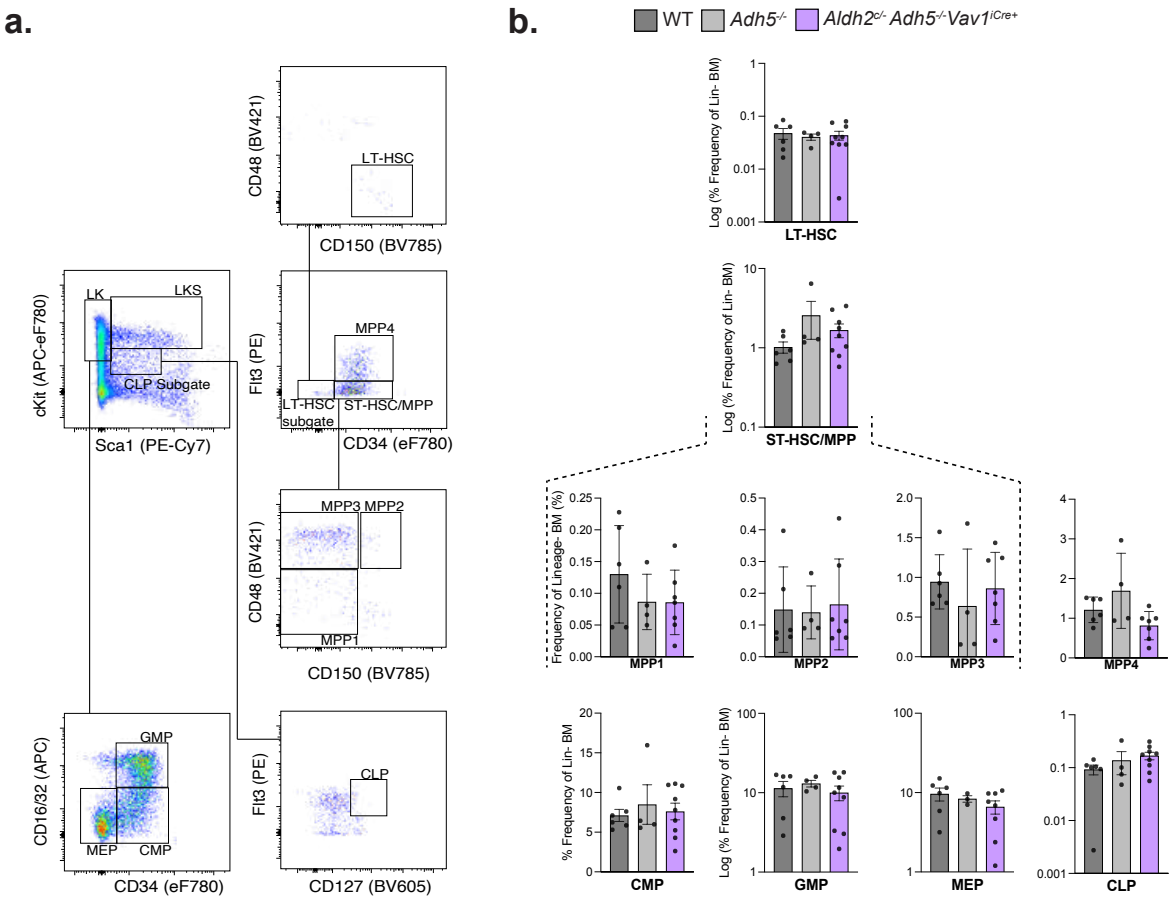


Figure 5.8 Haematopoietic stem cell, multipotent progenitor and committed progenitor quantification in bone marrow in 8-12 week-old males.
(a) Flow cytometry gating strategy used to identify LT-HSCs, ST-HSCs, MPPs, LKS, CLPs, GMPs, CMPs and MEPs from Lineage$^-$ whole bone marrow (b) Quantification of respective populations in 8-12-week male old mice, as a proportion of total number of Lineage$^-$ cells sampled. Bars represent mean ± SEM, data represent individual mice, n = 6, 4, 9, p values determined by Dunn's test, significance marked where $P < 0.05$.

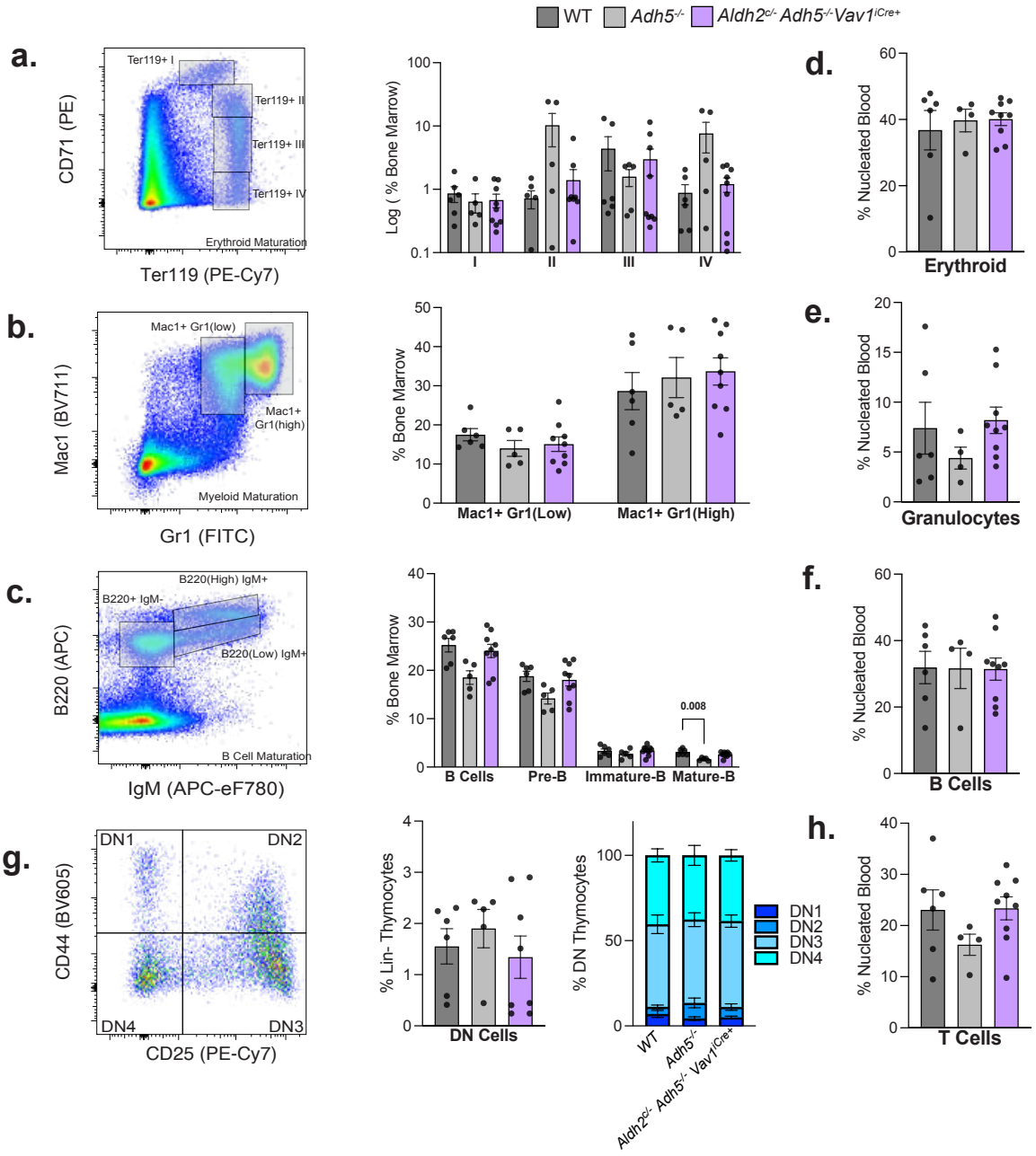


Figure 5.9 Maturing blood populations in 8-12 week-old males
(a-c) Representative flow cytometry gating (left) and quantification of maturing lineages (right) in the whole bone marrow, for erythroid (a) myeloid (b) and B-cell (c) development. (d-f) Quantification of terminal blood populations in peripheral blood erythroid (d) myeloid (e) and B-cell (f). (g) Representative flow cytometry gating (left) and quantification of developing Lineage$^-$ CD45$^+$CD4$^-$CD8$^-$ (DN) thymocytes (middle) and quantification of populations in each maturation stage, defined by CD25 and CD44 staining (right). (h) Quantification of T cell populations (Ter119$^-$ B220$^-$ Gr1$^-$ CD3e$^+$) in peripheral blood. All data representing mean ± SD, points represent individual mice, n = 6, 5, 9, p values determined by Dunn's test, significance marked where $P < 0.05$.

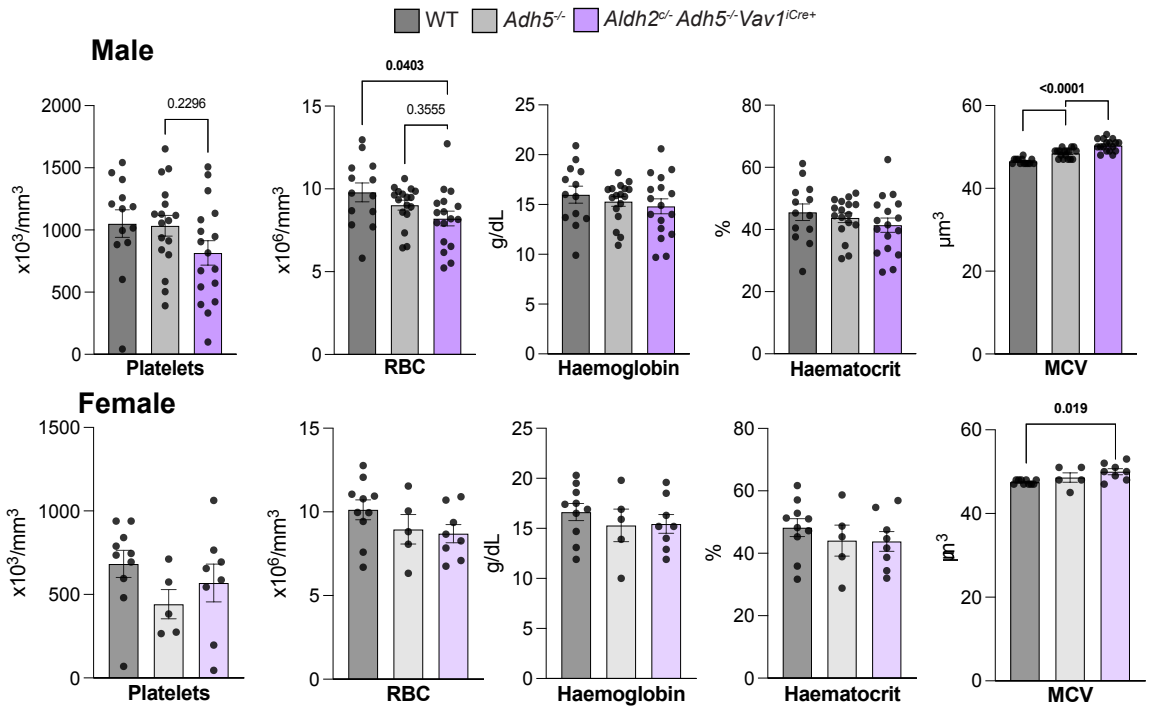


Figure 5.10 Complete blood count and parameters in 8-12 week-old mice
Measured parameters of peripheral blood in at 8-12-week mice (Males: n = 13, 17, 17; Females: n=8,5,7 respectively). Data are mean ± SEM, each dot represents individual mice, p values measured by ANOVA with Tukey's post-hoc test, significance marked where $P < 0.05$

5.2.5 *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice present elevated genomic instability in the blood

I then determined whether the recovery of haematopoiesis was related to protection against genotoxic stress induced by formaldehyde. I assessed this by quantifying the presence of micronuclei in red blood cells, a product of double stranded breaks incurred during red blood cell development (O'Connell et al., 2015). *Aldh2$^{c/-}$Adh5$^{-/-}$ Vav1^{iCre+}* were at an intermediate level between *Aldh2$^{-/-}$Adh5$^{-/-}$* and controls. Again, there was significant sexual dimorphism in the *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice, not present in *Aldh2$^{-/-}$Adh5$^{-/-}$* (Fig 5.11a). I next quantified the number of sister chromatid exchange events (SCEs) present in whole bone marrow, an indicator of active DNA repair by homologous recombination, in *Aldh2$^{+/-}$Adh5$^{-/-}$* controls and *Aldh2$^{c/-}$Adh5$^{-/-}$*

Vav1^{iCre+} mice. Here, I observed a 1.7-fold increase in the number of SCE events in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ over *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ (Figure 5.11b), which is lower than the >2-fold increase seen in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ (Dingler *et al.*, 2020).

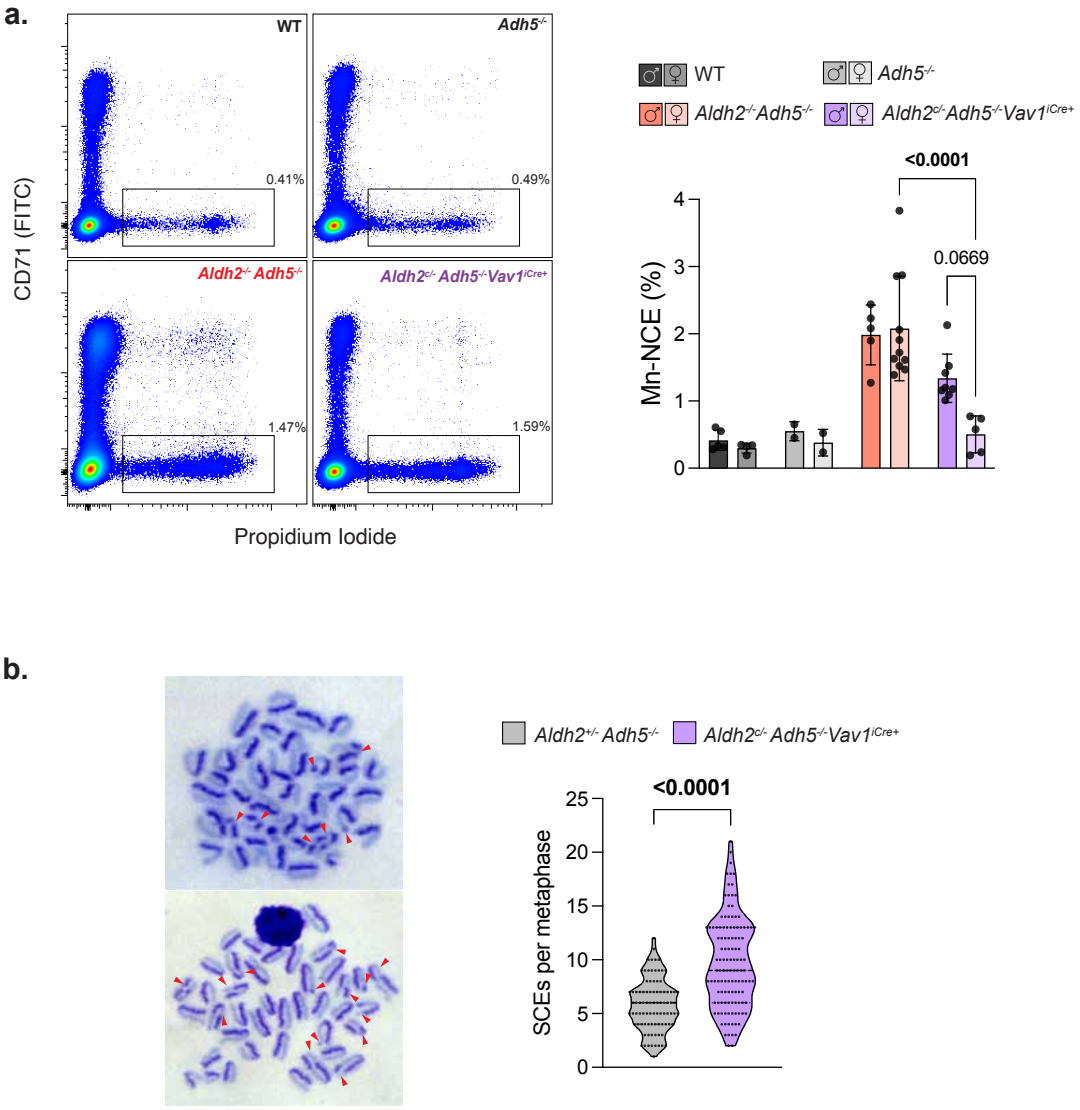


Figure 5.12 Markers of genomic instability in 8-12 week-old mice
(a) Micronucleus assay. Representative flow cytometric gating strategy of micronucleus assay (left) and quantification (right) of samples from peripheral blood of 8-12-week old mice. Red blood cells are differentiated from reticulocytes by absence of CD71 expression, micronuclei determined by presence of PI staining. Male n = 5, 2, 5, 8 female n = 4, 2, 11, 5. Bars represent mean ± SD, p values measured by two-way ANOVA with Tukey's post-hoc test, significance marked where P < 0.05 (female *Aldh2*$^{+/-}$*Adh5*$^{+/-}$ and *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ were excluded from statistical comparison due to low n). (b) Sister chromatid exchange assay**.** (Left) Representative images of low SCE (top) and high SCE (bottom) metaphase spreads, with SCE events highlighted in by red arrowheads. (Right) Quantification of SCE events from 3 individual mice of each genotype, data points represent number of SCEs in an individual metaphase spread. n = 89, 121. Comparison of genotype effect by t-test.

These results highlight that there are active processes of DNA damage and repair in the blood of the *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice at 8-12 weeks, but less than is observed in *Aldh2$^{-/-}$Adh5$^{-/-}$* mice, suggesting that the level of DNA damage experienced in the blood of *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice was below a threshold needed to drive the haematopoietic phenotype observed in *Aldh2$^{-/-}$Adh5$^{-/-}$*.

5.2.6 HSCs from *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice retain their proliferative capacity *ex vivo*

While the genotoxic stress determined by Mn and SCE assays, allow us to appreciate elevated DNA damage in the blood compartment, they do not tell us if this has functional consequence to the immunophenotypic HSC population. Accumulating DNA damage significantly reduces the ability of HSCs to respond to proliferative challenge, with replication machinery now detecting DNA damage and upregulating checkpoint regulators (Flach *et al*., 2014; Walter *et al*., 2015). I therefore asked whether elevated endogenous formaldehyde stress decreased the ability of LT-HSCs to respond to proliferative challenge. To do this, I used an *ex vivo* PVA-based culture method, that uses proliferative cytokines TPO and SCF, to drive proliferation of sorted HSCs (Wilkinson et al., 2019).

First, I measured the proliferation of LT-HSCs that had been sorted at a density of 50 cells per well over a 1-month period. WT, *Aldh2$^{-/-}$* and *Adh5$^{-/-}$* grew at the same rate over 28 days, consistent with their absence of *in vivo* haematopoietic phenotype (Figure 5.14a-c). Conversely, *Aldh2$^{-/-}$Adh5$^{-/-}$* cultures could not expand over the 28-day period. Interestingly, *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}*cultures, were indistinguishable in their proliferative capacity to the controls. This difference is important as *Aldh2$^{-/-}$Adh5$^{-/-}$* and *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* cultures are genetically equivalent *ex vivo*, indicating that formaldehyde produced or encountered *ex vivo*

cannot account for the proliferative defect induced by tier-1 deficiency, hence the difference must arise in differences in their epigenetic states *in vivo*. To rule out confounding effects of incomplete deletion of the conditional allele, I confirmed the absence of contaminating non-deleted *Aldh2*^{c} containing cells by PCR (Figure 5.14d).

In this experiment, I noticed an auto-fluorescent signal in the APC channel, normally used to stain CD201 present in LT-HSCs (Figure 5.15a). I, therefore, designed an alternative flow cytometry antibody panel that did not use the APC fluorophore, to quantify proliferation of the LT-HSC population of all genotypes from cKit+ cultures. This revealed that the proliferative defect in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ emerged as a failure of LT-HSC proliferation *ex vivo* (Figure 5.15b-c). However, in one of two *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ cultures, expansion began to emerge by d28, which may highlight the heterogeneity of HSC state *in vivo* and suggests the action of clonal selection in the *ex vivo* culture.

I was also interested to see if HSCs from the different genotypes were different in their competitive capacity, as has been observed upon transplantation (Pontel *et al.,* 2015; Dingler *et al*., 2020). I set up a competitive proliferation assay of CD45.2^{+} WT, *Aldh2*$^{-/-}$, *Adh5*$^{-/-}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ cells against CD45.1^{+} WT sorted LT-HSCs *ex vivo*, and found that while *Aldh2*$^{-/-}$ did not have a proliferative disadvantage, *Adh5*$^{-/-}$ cells were subcompetitive from d14, and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ cells are immediately disadvantaged from d7 (Figure 5.14e). This data reflects the sub-competitive nature of *Adh5*$^{-/-}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ HSCs observed upon transplantation (Pontel *et al*., 2015; Dingler *et al*., 2020), indicating selection is operating *in vivo*, accelerated by proliferative induction.

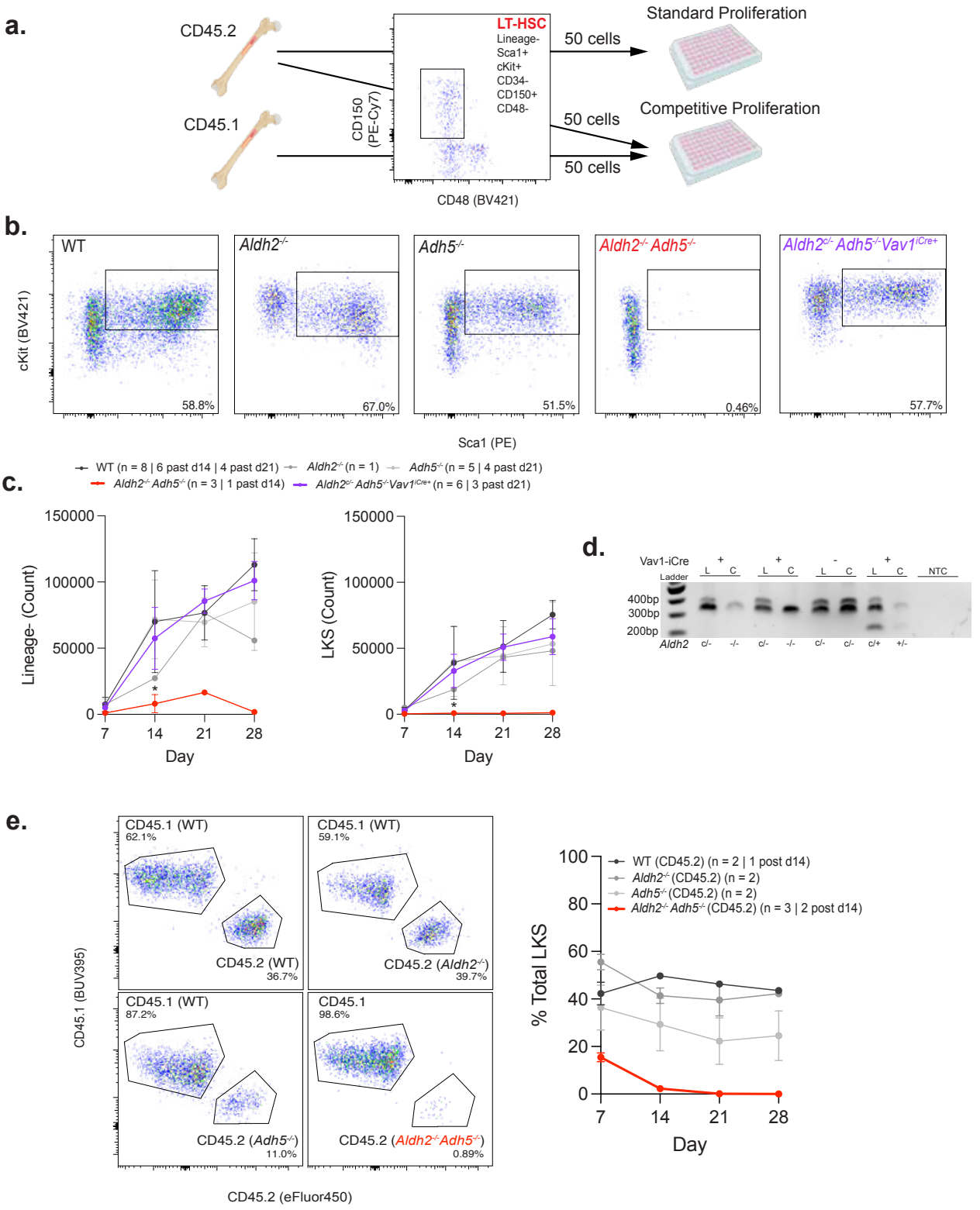


Figure 5.14 *Ex vivo* proliferative potential of sorted LT-HSCs
(a) Schematic of sorting procedure of HSCs used in this experiment, 50 LT-HSCs (LKS-CD34-CD150+CD48-) were sorted from cKit+ enriched whole bone marrow, from CD45.1 WT mice, or CD45.2 WT, *Aldh2*$^{-/-}$, *Adh5*$^{-/-}$, *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ or *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice, all mice were 8-12 weeks old. Competitive proliferation assays were formed with 50 CD45.1 cells and 50 CD45.2 cells in each well, standard assays only with 50 CD45.2 cells, assayed every 7 days. (b) Representative flow cytometry gating

of LKS cells in Lineage- population in standard proliferation assay, numbers representing percentage of parent population. (c) Quantification of mean number of Lineage- cells, and LKS+ cells in each culture, initial n = 8, 1, 5, 3, 6. (d) Representative multiplex-PCR on gDNA of d28 *ex vivo* cultures to verify absence of contaminating *Aldh2*^{c} allele in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$Vav1^{iCre+} cultures, internal reference for *Aldh2*^{c} provided by liver PCR, bands are paired according to mice, L – Liver, C – culture. (e) Representative flow cytometry gating of the proportion of CD45.1 and CD45.2 LKS+ cells (left) and quantification of mean proportion of CD45.2 LKS+, n = 2, 2, 2, 3. All cultures were formed with 3-5 technical replicates; data points represent mean of biological replicates ± SD. *p<0.05. Note that not all biological replicates are completed to d28, indicated in key.

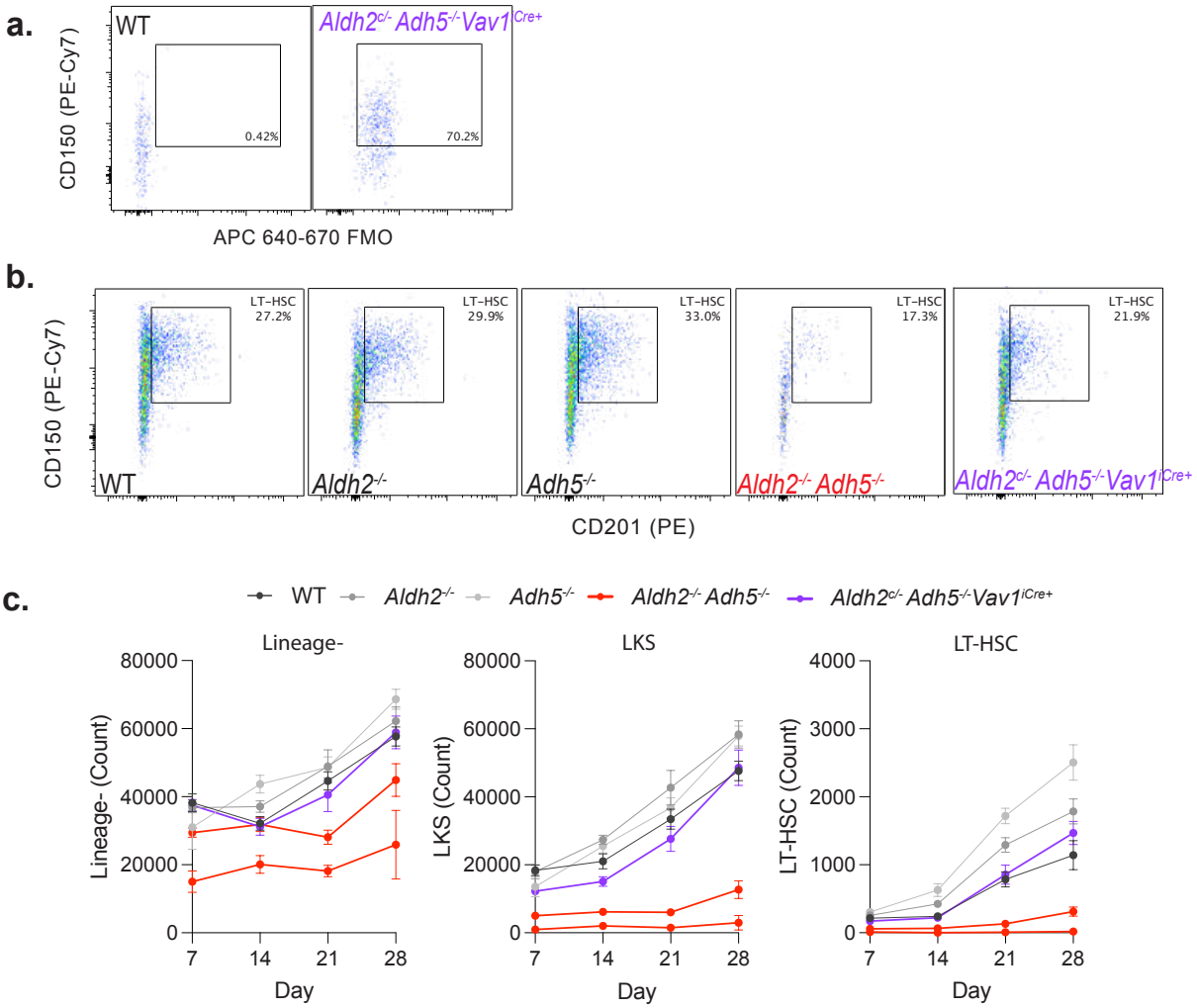


Figure 5.15 – APC autofluorescence and cKit+ enriched *ex vivo* cultures
(a) Representative gating of CD201:APC FMOs for *ex vivo* WT and *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ cultures demonstrating background autofluorescence in APC flow channel. (b) Representative flow cytometry gating of LT-HSCs from LKS population of cKit+ cultures using CD201:PE channel (Panel B, table 3). (c) Proliferation of cKit+ enriched cultures (100,000 cells per well) over 28 days, each line represents culture from an individual mouse, points represent mean ± SEM across 5 technical replicates.

These *ex vivo* experiments highlight that *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ LT-HSCs exist in different epigenetic state *in vivo*, which may relate to the reduced DNA damage burden encountered through life, as conferred by peripheral *ALDH2*. It also highlights the endogenous formaldehyde burden may be a key driver of selection by increasing replicative stress, which may influence the structure of aged haematopoietic populations.

5.2.7 1 year old *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice show signs of haematopoietic ageing

As formaldehyde induces genotoxic stress in the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ blood, and functional decline upon accumulation DNA damage is a hallmark of the haematopoietic ageing process, I looked for signs of ageing in the haematopoietic compartment of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice at 1 year old.

Aged *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice presented clearer signs of anaemia, with significantly reduced RBC count, haemoglobin, haematocrit, and increased MCV (macrocytosis), relative to *Adh5*$^{-/-}$ controls (Figure 5.16a). This worsening red blood cell phenotype is further consistent with increased micronuclei incidence (measured again at 6 months) in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ males, such that they were indistinguishable from *Aldh2*$^{-/-}$*Adh5*$^{-/-}$, reflecting the accumulation of DNA damage in the blood with age (Figure 5.16b). This phenotype was continuously sexually dimorphic, with female *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ having significantly fewer micronuclei than males at 6 months, and do not present markers of anaemia at 1 year (Figure 5.16). Intriguingly, the platelet counts for both male and female *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$

mice decreased, which is in contrast to *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, where this is increased (Dingler *et al*., 2020).

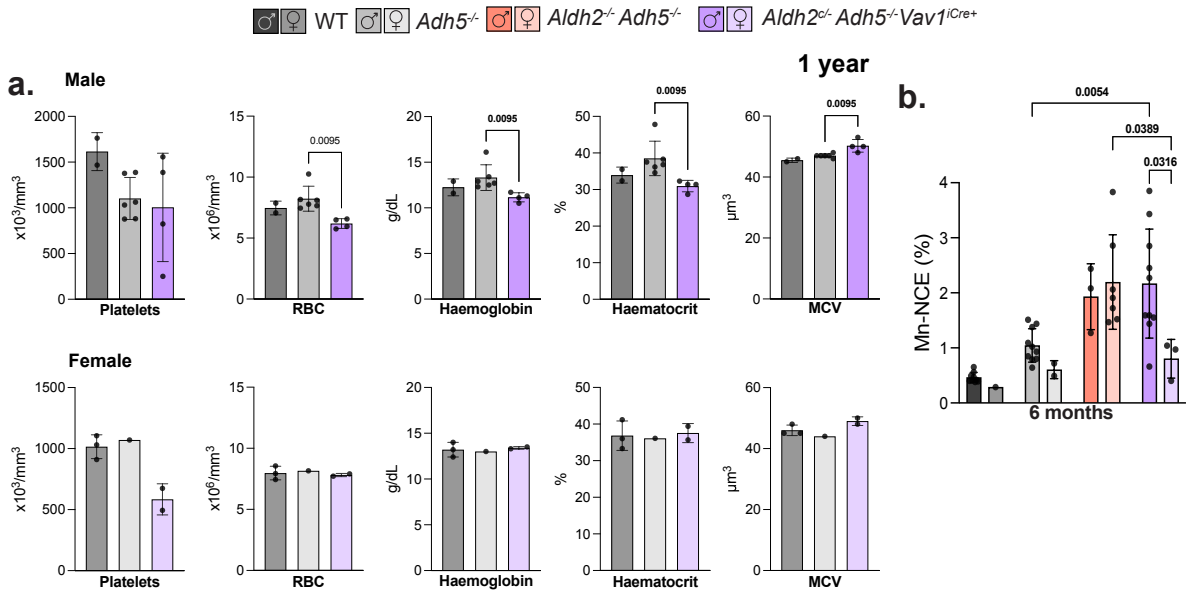


Figure 5.16 Peripheral blood count and parameters and micronucleation in aged animals
(a) Peripheral blood parameters of male (top - n = 2, 6, 4) and female (bottom – n = 2, 1, 2) mice at 1 year old. Bars represent mean ± SD, p values measured by Mann-Whitney U test, significance marked where P < 0.05 (*Aldh2*$^{+/-}$*Adh5*$^{+/-}$ males were not included in comparisons, nor were comparisons performed in females due to low number of biological replicates). (b) Quantification of micronucleated RBCs in 6 month mice (males n = 9, 10, 3, 10; females n = 1, 2, 7, 3). Bars represent mean ± SD, points represent individual mice, p values measured by ANOVA with Tukey's post-hoc test where ns > 0.05 (data with n < 3 values were not included).

At 1 year old, the haematopoietic compartment of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ male mice present some interesting changes (Figure 5.17, 5.18). There was an increase in the LT-HSC population in both *Adh5*$^{-/-}$ and *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ against the WT (this has been previously reported for *Adh5*$^{-/-}$ (Pontel *et al*., 2015)). The ST-HSC/MPP population retained its size but there was a striking increase in the MPP1 and MPP2 populations of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ compared to *Adh5*$^{-/-}$, while MPP3s significantly contracted. There was also a significant decline in the primarily lymphoid biased MPP4 population, coinciding with a significant loss of downstream CLPs in the bone marrow, and B cell count in the peripheral blood. All other populations were indistinguishable

from *Adh5*$^{-/-}$. Despite their lymphoid origins, the thymus and T-cell populations were maintained, likely reflecting the long-lived nature of T-cells and limited replacement from bone marrow progenitors (Soerens et al., 2023; Sun et al., 2014). *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ female mice also showed reduction in the MPP4 population, but not in other blood populations, potentially reflective of their lesser formaldehyde burden (Supplementary Figure 5.3, 5.4).

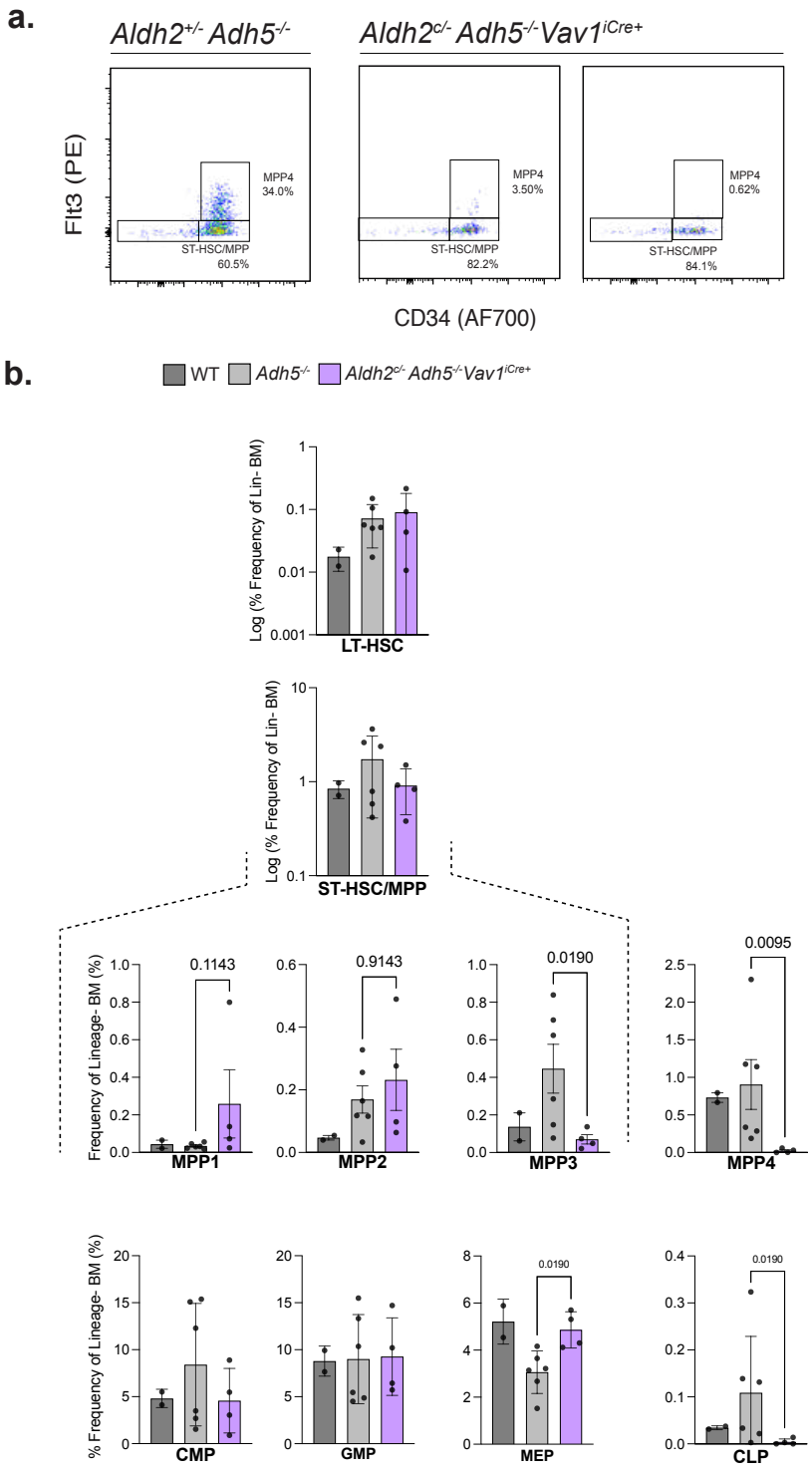


Figure 5.17 Haematopoietic stem cell, multipotent progenitor and committed progenitor quantification in bone marrow of 1 year old males
(a) Representative flow cytometry gating of live LKS+ bone marrow, highlighting gates for MPP4 and ST-HSC/MPP populations for *Aldh2*$^{+/-}$*Adh5*$^{-/-}$ and *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ mice, numbers representing percentage of parent population. (b) Quantification of HSC and committed progenitor populations from whole bone marrow (as performed in Figure 5.10). n = 2, 6, 4. Bars represent mean ± SD, p values measured by Mann-Whitney U test, significance marked where P < 0.05 (excluding comparisons to *Aldh2*$^{+/-}$*Adh5*$^{+/-}$).

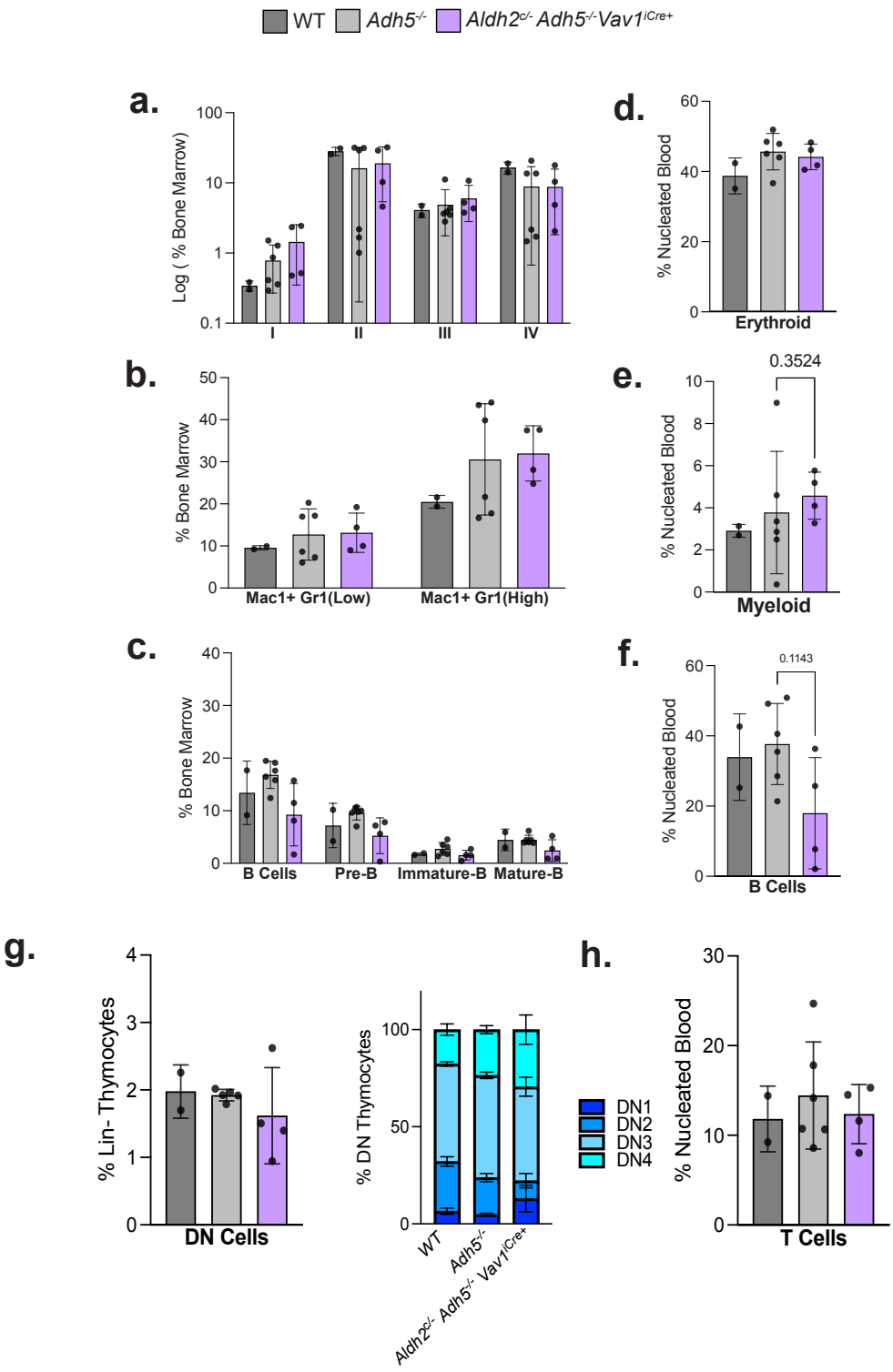


Figure 5.18 Maturing blood populations in 1-year aged males
(a-c) quantification of maturing lineages in the bone marrow for erythroid (a) myeloid (b) and B-cell (c) development. (d-f) Quantification of terminal blood populations in peripheral blood erythroid (d) myeloid (e) and B-cell (f). (g) Quantification of developing DN thymocytes (left) and quantification of populations in each maturation stage (right). (h) Quantification of T cell populations in peripheral blood. n = 2, 6, 4. Bars represent mean ± SD, p values measured by Mann-Whitney U test, significance marked where $P < 0.05$ (excluding *Aldh2*$^{+/-}$*Adh5*$^{+/-}$).

5.2.8 1 year old *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice do not have clonal blood

With the contraction of the MPP3 compartment producing blood and signs of DNA damage within the haematopoietic system of *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice, it is likely that pressures of selection are operating within the blood of the *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mouse to support HSC/MPPs that can tolerate the intermediate formaldehyde load. Next was to determine whether this corresponded to changes in the clonal architecture of the blood, using the same method applied in *Aldh2$^{-/-}$Adh5$^{-/-}$* mice (chapter introduction).

I performed WGS at 90X sequencing depth on granulocytes sorted from the bone marrow of two *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* males at 1 year old, using gDNA extracted from the cortex of the same mice, sequenced at 30X depth, to act as a germline reference. Identification of variant alleles, and derivation of variant allele frequencies and number of single nucleotide variants was performed by M.Gunther, and these were plotted against each other (Fig 5.19). In stark contrast to the *Aldh2$^{-/-}$Adh5$^{-/-}$* mice, the SNV/VAF curves for the two *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice, suggests clonality did not emerge. Instead, the blood fits models of neutral drift model of haematopoiesis, lacking the distinctive “shoulder” at high VAF would emerge from clonal populations. However, there was a high cumulative number of SNVs reported at low VAF compared to WT and *Adh5$^{-/-}$* controls, which suggests an increased mutation rate within the HSC/MPP pool, similar to *Aldh2$^{-/-}$Adh5$^{-/-}$*. This suggests that DNA damage is accumulating in the HSC/MPP pool, which may be driving accelerated ageing, but that CH in *Aldh2$^{-/-}$Adh5$^{-/-}$* mice is driven by an additional pressure present early in life.

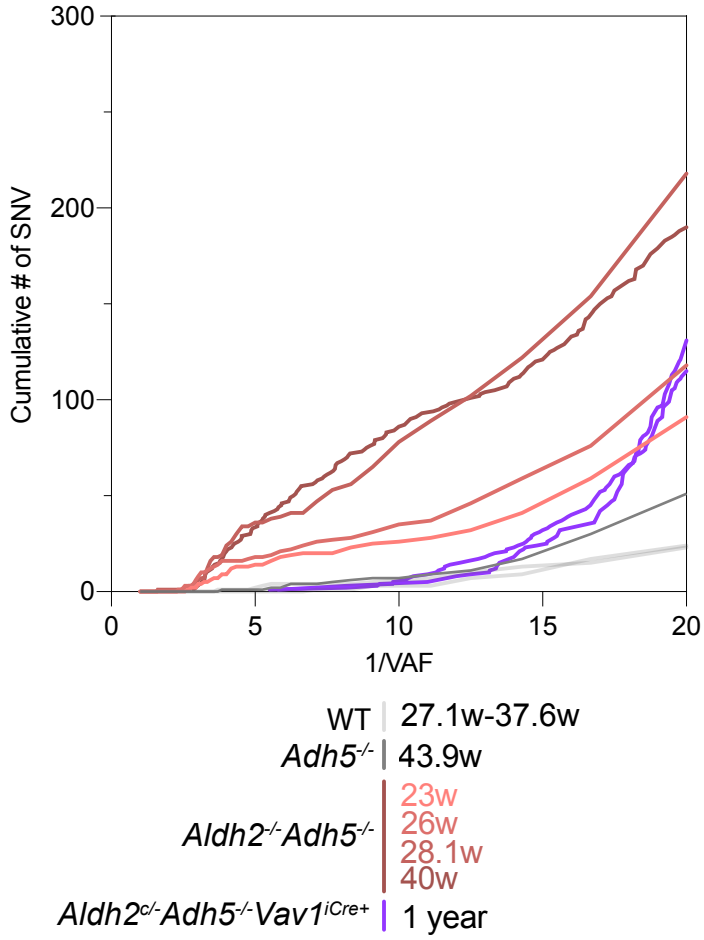


Figure 5.19 Clonal haematopoiesis is not observed in *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice at 1 year age
Cumulative number of SNVs and their frequency as detected by bulk (90x) WGS of bone marrow granulocytes from two male *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice

5.2.9 Formaldehyde exposure in development drives HSC selection in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice

HSC/MPPs are expected to have accumulated ~50 SNVs by the time of birth (Kapadia *et al*., 2025). Given that the cumulative number of SNVs at the point of clonal emergence in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice is <50 (Figure 5.19), this suggest that formaldehyde is shaping the clonal population early in life. I therefore investigated the phenotype of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ HSCs during development.

During foetal haematopoiesis, cells emerging from the primitive endothelium, including the dorsal aorta, emerge and migrate to the foetal liver, where they proliferate between E12-E16. From E17.5, they begin to migrate to the bone marrow, where they begin to expand from birth (Christensen et al., 2004; Ema & Nakauchi, 2000). I first quantified the number of reduced N^2-Me-dG adducts in E16.5 and E18.5 foetal livers, to understand whether there was a significant increase in the formaldehyde burden in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ foetuses (Figure 5.20). At E16.5, there was a subtle increase in the adduct level in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, which was much more substantial by E18.5, also observable to a lesser degree in liver of *Adh5*$^{-/-}$ foetuses. The overall burden is still very low compared to the level observed at 8-12 weeks (Figure 5.3), nevertheless, there is clearly a large engagement of metabolic clearance of formaldehyde in development.

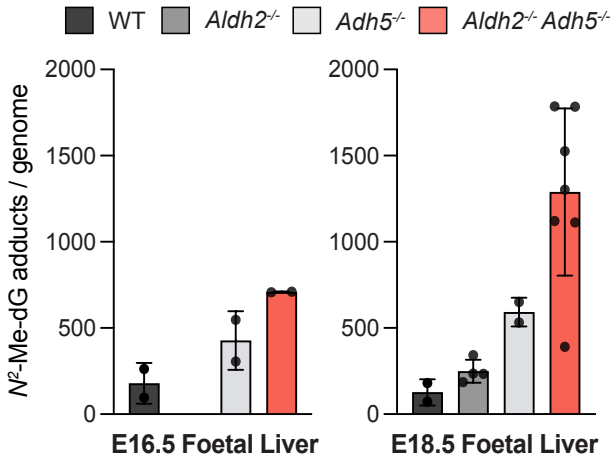


Figure 5.20 N^2-Me-dG adduct quantification in foetal liver at E16.5 and E18.5
E16.5 n = 2, 0, 2, 2; E18.5 n = 2, 4, 2, 7. Bars represent mean ± SD

I then quantified the immunophenotypic HSC stem and progenitor populations in the foetal liver and bone marrow, at E18.5 (Figures 5.21a, 5.21c), and observed no striking difference in frequency of these cell types was observed between genotypes.

To test whether the increased formaldehyde load in development restricted the proliferative capacity of HSC, I placed whole foetal liver samples into *ex vivo* PVA-cultures (described above) and measured proliferation over 1 month (Figure 5.21b). Interestingly, I found *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ HSCs failed to grow, while all other genotypes were equivalent, in keeping with the adult situation (Figure 5.14c). I was also interested in the *ex vivo* proliferative function HSCs in the foetal bone marrow. As the immunophenotypic HSC population is much rarer, I sorted LKS cells into wells at 25 cell/well, and quantified growth by measuring the cultured area of each well after 14 days (Figure 5.21d). Here, I found WT cells were advantaged against all other genotypes (consistent with the competitive advantage observed in the adults), and a complete failure of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ HSPCs to establish colonies. However, it is important to note that the number of mice used in these experiments is very low (in many cases n=1), therefore in need of repetition.

These preliminary findings are consistent with those of the adults, suggesting that formaldehyde incurred *in vivo* in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ is significant to restrict proliferative potential, and drive the selection that culminates in clonal dominance early in young *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice.

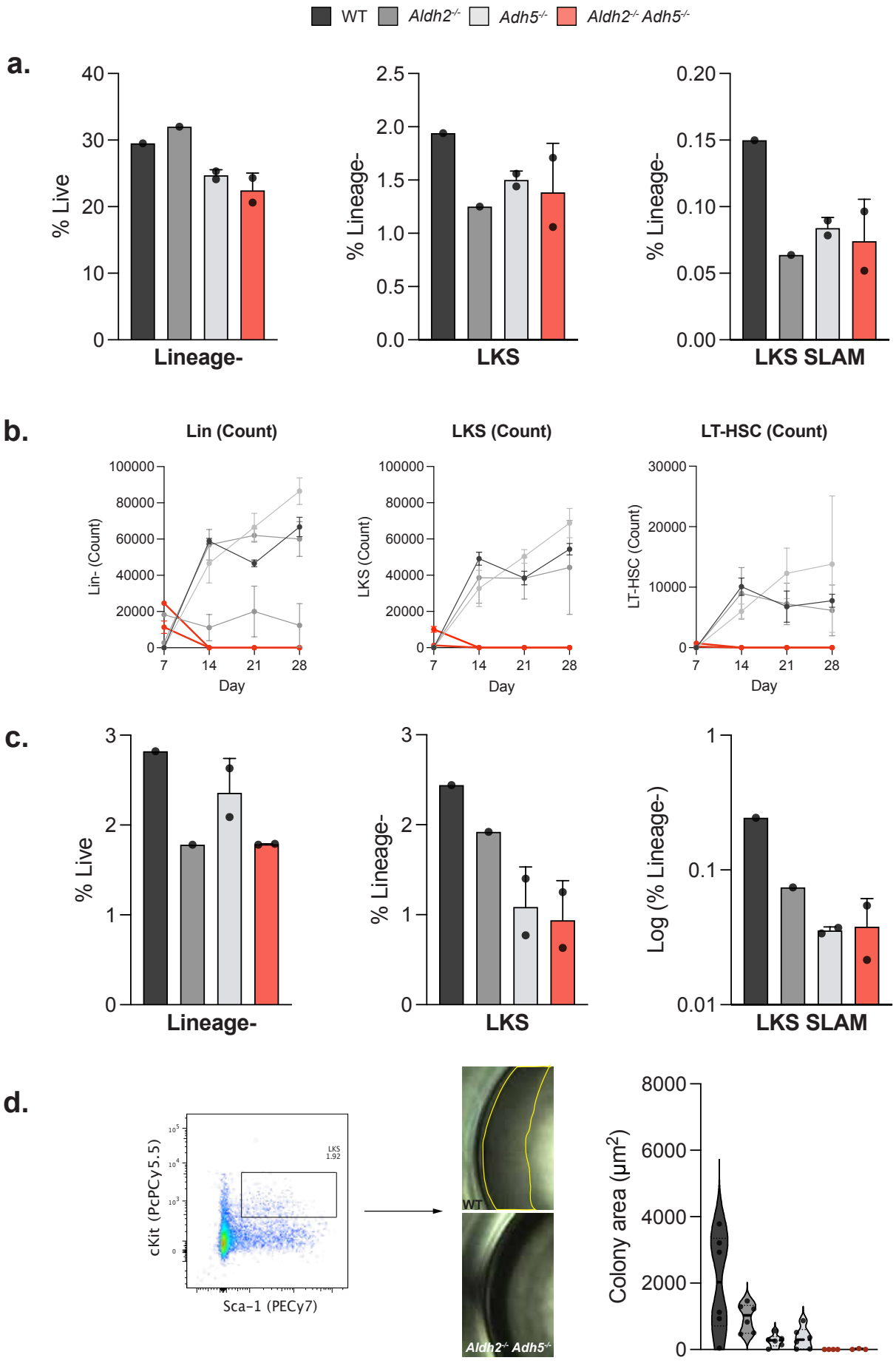

WT
Aldh2-/-
Adh5-/-
Aldh2-/- Adh5-/-
a.
% Live
Lineage-
% Lineage-
LKS
% Lineage-
LKS SLAM
b.
Lin (Count)
LKS (Count)
LT-HSC (Count)
Lin- (Count)
LKS (Count)
LT-HSC (Count)
Day
c.
% Live
Lineage-
% Lineage-
LKS
Log (% Lineage-)
LKS SLAM
d.
cKit (PcPCy5.5)
Sca-1 (PECy7)
LKS
1.92
WT
Aldh2-/- Adh5-/-
Colony area (µm2)

Figure 5.21 *In vivo* quantification and *Ex vivo* proliferative potential of E18.5 haematopoietic stem cells from foetal haematopoietic tissues
(a) Quantification of HSPC (LKS) and HSC (LKS-SLAM) populations in foetal liver analysed by flow cytometry (lines are mean ± SD). (b) *Ex vivo* proliferation of cultured whole foetal liver ($2x10^5$ seeded per well). Lines represent individual foetal cultures, points are mean ± SEM of five technical replicates. (c) Quantification of HSPC (LKS) and HSC (LKS-SLAM) populations in foetal bone marrow analysed by flow cytometry (lines are mean ± SD). (d) Representative gating for sorting 25 LKS foetal bone marrow cells per well (left) example of imaging cultures after 14 days by widefield light microscopy (middle) quantification of colony growth as area (μm^2) for all colonies sorted, each point represents an individual colony, lines indicate median ± interquartile range (right). n = 1, 1, 2, 2

5.3 Discussion

By restricting *ALDH2* deficiency to the blood of *Adh5*$^{-/-}$ mice, I was able to develop a model where the severe formaldehyde toxicity observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice was suppressed, but the formaldehyde level was sufficient to induce DNA damage in the blood and accelerate the ageing process. This model therefore allows us to perform closer analysis of the processes that underpin the pathology of chronic formaldehyde exposure, extending to mechanisms of tissue damage and blood homeostasis, and suggest roles of formaldehyde in driving clonal haematopoiesis.

5.3.1 Chronic formaldehyde exposure drives haematopoietic ageing

Loss of *ex vivo* proliferative capacity in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ LT-HSCs *in utero*, together with early signs of haematopoietic ageing (myeloid-bias, higher transcriptional ageing score), indicates that their phenotype arises from high systemic formaldehyde exposure already incurred during development, leading to long-term epigenetic reprogramming and adaptation. This is consistent with Oka et al. (2020), who reported HSC defects and myeloid bias in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice already at 3 weeks, and Dingler

et al. (2020), who studied mice >4 weeks, confirming the phenotype emerges early rather than progressively with age. In contrast, *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice only display ageing features (e.g. myeloid bias) later in life, making them a more faithful model of formaldehyde-driven haematopoietic ageing. Thus, while *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ reflects developmental adaptation to high formaldehyde, *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ provides a clearer window into how chronic formaldehyde exposure accelerates haematopoietic ageing.

Chronic formaldehyde appears to drive macroscopic changes in blood populations in manner typical of natural ageing (Rossi et al., 2005). There is a collapse of the MPP4 population, and downstream CLP and B-cell lineages, normally derived from MPP4 cells (Patel *et al*., 2022). This loss of lymphoid progenitors is believed to derive from loss of long-lived progenitors (predominantly MPP4s) from early in development, which are only marginally replaced by MPPs derived from LT-HSCs in adulthood, with these adopting an increasingly myeloid-bias program (Meng et al., 2023; Rodriguez-Fraticelli et al., 2018). It is also interesting to observe a loss of the MPP3 compartment, despite the expansion of the MPP1 population (upstream of MPP3), when relative to *Adh5*$^{-/-}$. This MPP3 decline is not typically observed in ageing models (Singh et al., 2025), but reduced rate of renewal has been described in a genetic model where tolerance to DNA damage was lost through disabling the ability to undergo TLS, thus increasing the burden of DNA damage in the haematopoietic compartment, potentially analogous to the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ model (Pilzecker et al., 2017). In the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice, formaldehyde exposure may expediate the processes of MPP3/MPP4 cell loss by driving apoptosis or differentiation in these populations, while

facilitating epigenetic changes in upstream populations to disable replacement as output is biased to MPP2s, and future work should look into how exactly this occurs.

Identifying the mechanistic drivers of these changes is complex. DDR signalling is an obvious candidate, as markers of DNA damage (micronuclei, SCEs, mutations) are increased in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, and DNA repair genes are upregulated in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ HSCs **(**Dingler *et al*., 2020). The p53 score of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ was described to be relatively low (Wang *et al*., 2023), but chronic low-level activation of p53 during a lifetime would support epigenetic changes, such as those inducing senescence (Chambers et al., 2007; Rodier et al., 2009). We may also consider non-DNA damage dependent mechanisms may be in play, such as mitochondrial stress (Sun *et al*., 2025), suggested to drive myeloid-bias and have potential for direct induction by formaldehyde. This may also be a product of inflammation, which can be further induced by DNA damage (Harding et al., 2017) (although induction of inflammatory cytokines and inflammatory genes was not observed in *Aldh2*$^{-/-}$*Fancd2*$^{-/-}$ HSCs (Wang *et al*., 2023)). Additionally, formaldehyde may promote senescence of cells in the niche environment, which supports myeloid fate bias (Frisch *et al*., 2019). As there is basis for formaldehyde to influence each of the pleiotropic contributors to ageing, future work would benefit from combined scRNA-seq and ATAC-seq experiments, to understand whether formaldehyde drives haematopoietic ageing through a generalised process, or more specifically a product of DNA damage signalling, thus revealing targets for further study.

The *ex vivo* culture experiments point towards cell state changes in LT-HSCs as a route of this heamatopoietic remodelling. This is seen as *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and

Aldh2$^{-/-}$*Adh5*$^{-/-}$ LT-HSCs, have significantly different capacities to proliferate *ex vivo*, despite being genetically equivalent *ex vivo*. Accumulation of DNA damage within quiescent HSCs has been shown to become deleterious upon forced exit from quiescence (Beerman *et al*., 2014; Walter *et al*., 2015), and the haematopoietic compartment of both *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice show signs of significant genomic instability. Therefore, we may have expected *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ HSCs to also fail to proliferate *ex vivo* when exiting from quiescence, which may serve as a model through which attrition of HSCs operates in aged mice. This hypothesis has limitations, however; LT-HSCs typically accumulate less DNA damage than downstream populations (Beerman et al., 2014; Mohrin et al., 2010), and the inferred markers of DNA damage (micronuclei and SCEs) do not survey DNA damage specifically in LT-HSCs. Indeed, adduct measurements in the haematopoietic tissues may further suggest accumulation of formaldehyde on DNA that could form interstrand crosslinks, but these are no higher in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice at the same age (when looking at the spleen). Therefore, there is little evidence to support *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ LT-HSCs in having significantly higher DNA damage accumulated at the same age than *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice. Another key limitation of this hypothesis is that limited proliferation can still be observed in downstream populations (Lin- populations expand in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ cultures until d21), and in cKit-enriched cultures. If driving apoptotic signalling, it is unlikely that DNA damage is persisting beyond the first division that would facilitate this continued proliferation. Finally, the fact that foetal liver HSCs are already in a state of proliferation (Morrison et al., 1995), fail to expand, suggests that changes in the cell fate are an epigenetic instruction that may be entrained by the DDR, but are not overwhelming it.

It is interesting to observe the sub-competitive nature of *Adh5*$^{-/-}$ LT-HSCs *ex vivo*, which may suggest damage signalling *ex vivo* operates to drive selection. Indeed, recent work by fellow group member, Samual Jones, used a Cre-inducible system to delete *ERCC1* (hence all capacity for NER) *ex vivo* in WT and *Adh5*$^{-/-}$ LT-HSCs, which induced complete cell death specifically in *Adh5*$^{-/-}$ cultures. This suggests that formaldehyde is present in *ex vivo* cultures at a level sufficient to necessitate DNA repair. In a competitive situation, this may just be enough to confer a disadvantage through activation of the DDR (e.g. p53 activation (Bondar & Medzhitov, 2010)). Based on the fact that *Adh5*$^{-/-}$ are still maintained in competition at 28 days, this suggests that these differences are a product of responses early in the culture. Considering this will be a period of intense epigenetic change as cells adapt to culture, it may be a point to consider the potential of JmjC-domain containing histone demethylases, which produce formaldehyde following the oxidative removal of methyl marks, and are known to regulate stemness in haematopoietic cells (Kerenyi et al., 2013; Klose et al., 2006; Li et al., 2018). Therefore, in an *Adh5*$^{-/-}$ HSC, formaldehyde produced by these proteins may be sufficient to induce a p53 response that confer a disadvantage in competitive situations, thus serving as a candidate for an endogenous formaldehyde source with functional consequence in haematopoiesis.

5.3.2 Formaldehyde drives clonal haematopoiesis under certain conditions

Despite accelerated ageing in the haematopoietic phenotype, there was no sign of clonal emergence from the SNV/VAF curves at 1 year in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice, in marked contrast to *Aldh2*$^{-/-}$*Adh5*$^{-/-}$, which appear to develop CH early in life. The fact

that an *ex vivo* defect can be observed in HSCs from *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ foetuses, provides key insight into how formaldehyde can shape the clonal architecture of the blood.

During haematopoietic development, HSC/MPPs in the foetal liver begin migrating to the bone marrow from ~E17.5, where they undergo significant expansion from the time of birth (Hall et al., 2022). This puts HSC/MPPs under two pressures – first, they need to be able to migrate and adapt to the bone marrow microenvironment, second, they need to resist replication stress. A high burden of systemic formaldehyde in the foetal liver niche of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ foetuses, will induce changes in the cell state (e.g. through the DDR inducing epigenetic reprogramming) that precede this window, restricting their ability to proliferate. Hence HSC/MPPs that survive will be selected through their ability to tolerate the formaldehyde burden. This is similar to the mechanism of "pruning selection" described in cases of clonal emergence in humans following autologous transplantation (Spencer Chapman et al., 2024). Here, the authors describe a model of selection, where clonal diversity was reduced upon autologous HSC transplantation, with expansion of clones that present before transplantation which were better adapted for survival and expansion in new bone marrow niche. In this model, there is frequent presence of clones with driver mutations that support positive selection by "growth". To further support this as a mechanism of action in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, extended analysis of the WGS data can be performed (Korber *et al*., 2025), including estimation of the date of clonal emergence, which would reveal whether these clones do have early origins, as well as mapping the variant alleles to genes to assess their potential significance in acting themselves as driver events.

Cell extrinsic detoxification of formaldehyde by *ALDH2* protects HSC/MPPs in the foetal liver of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice, hence these early pressures won't be present, explaining the absence of an early clonal phenotype. It remains to be shown whether *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ may develop a clonal blood phenotype upon further ageing, as may have been expected from the presentation of other signs of haematopoietic ageing (myeloid bias, loss of MPP populations and accumulation of mutations). Currently, we do not have a gauge of the functional potential of the quiescent HSC pool in aged *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice, and this may be expected to contract based on the observations of accumulating DNA damage and fate restriction. If the lifetime exposure to formaldehyde stress accelerates the loss of functional HSC/MPPs upstream of active blood production, we may observe a sudden collapse of this heterogeneity when surveying granulocytes at a later time point, as is observed in human ageing populations (Mitchell *et al*., 2022, Korber *et al*., 2025; a hypothetical mechanism of clonal acquisition in WT, *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice is summarised in Figure 5.22). How long this would take is uncertain, at least outliving the lifespan of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, and the number of LT-HSCs does not appear to be contracting in the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ (although this is not a reflection of the functional size) by 1 year.

Waiting for clonality to emerge in a sterile environment is not reflective of life. Exposure to pathogens and inflammatory agents (such as pollution (Hill et al., 2023)) are a daily occurrence, which can drive HSC proliferation through inflammatory signalling (Baldridge et al., 2011). The ability of HSCs to tolerate these stressors changes with age (Mann *et al*., 2018). If formaldehyde is accelerating the ageing of HSCs, we may expect the HSC/MPP population to be more extensively remodelled in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$

Vav1^{iCre+} mice in response to inflammatory challenge (e.g. pI:pC or LPS treatment), than in an *Adh5*$^{-/-}$. If this is the case, it would suggest a mechanism through which formaldehyde could shape the functional compartment in natural populations, in that it influences pruning of the HSC/MPP pool as we encounter exogenous stimuli through life.

Using granulocytes to measure the incidence of clonal blood is a significant advance in the ability to rapidly identify the onset of clonal haematopoiesis and expansion of variants known to confer risk to neoplastic transformation, which is significant for screening patients to understand their risk of developing cancer (Körber et al., 2025; Steensma et al., 2015), and has potential for applying to understand how clonality arises through native haematopoiesis following experimental perturbation. However, there are key limitations that restrict interpretations of these results. First, we are only able to detect clones at a frequency of above 5% VAF, which is a significantly lower resolution compared more advanced colony sequencing methods, being able to detect clones from 1% enrichment (Lee-Six et al., 2019). This threshold could be reduced by performing WGS at higher depth (Korber *et al*., 2025). Second, by only surveying granulocytes, which have a lifespan of ~3 weeks (Sun *et al*., 2014), we are only surveying a portion of haematopoietic output, and not directly the upstream HSC/MPP population, further underestimating clonal changes, as not all HSC/MPPs will be producing granulocytes at any one time. Third, as this data is drawn from WGS of a bulk population, we cannot deconvolve the direct relationship between variants and their progeny, so we can't confidently classify which variant in the "shoulder" population may be acting as a driver, or whether this population is made up of multiple clones with similar VAF. Finally, the data presented here is still awaiting mathematical

modelling to confirm whether the clones observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice fit that of a model of clonal evolution, or whether they still fit a model of neutral drift. This work is still active at the time of writing.

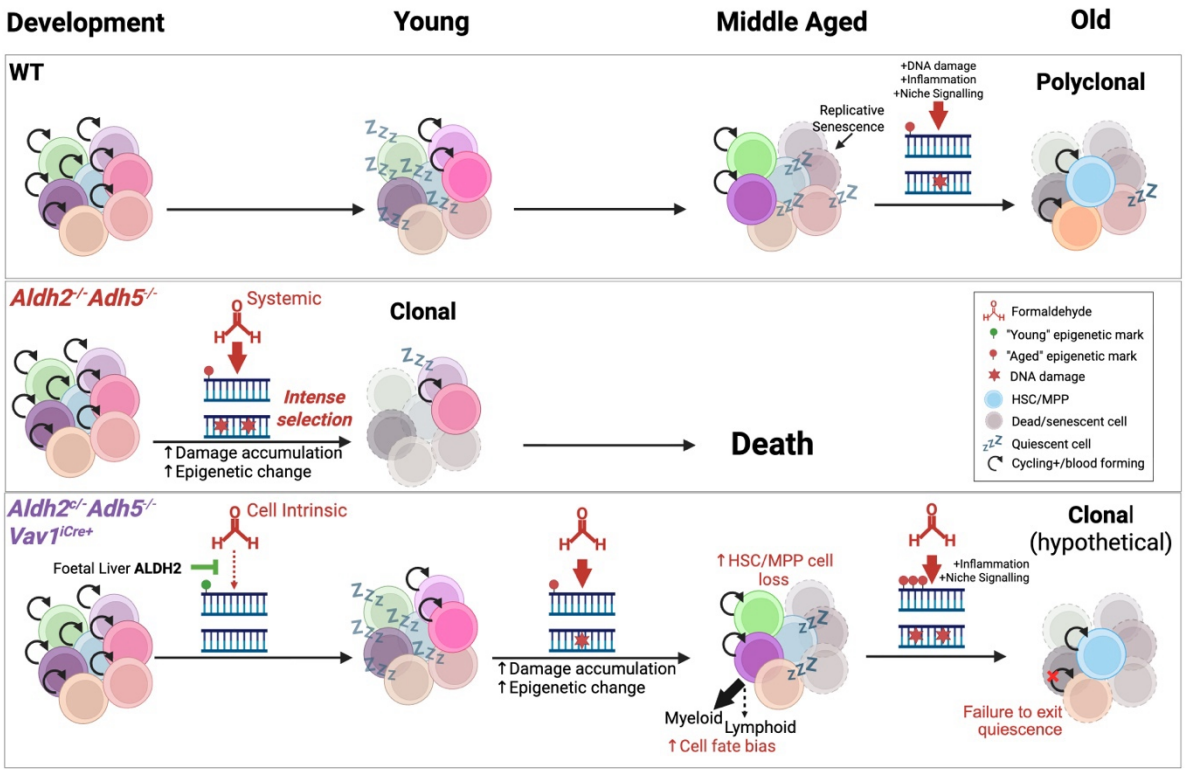


Figure 5.22 Model of clonal emergence in *Aldh2$^{-/-}$Adh5$^{-/-}$* and *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre}* mice

(Top) Clonal emergence in WT mice. During development, the HSC/MPP pool undergoes rapid expansion, with HSCs adopting a quiescent fate upon entry into the bone marrow, with only a limited number of HSC/MPP clones producing blood at any one time. Over time, these populations change as active HSC/MPPs undergo replicative senescence, and new HSC/MPPs are recruited. The reseve of functional HSC/MPPs decreases as signals such as DNA damage, inflammation and niche signalling, influence cell potential. However, in mice, there is a large enough reserve pool that this does not become limiting by the time of death. (Middle) Clonality in *Aldh2$^{-/-}$Adh5$^{-/-}$* mice. In these mice, the high formaldehyde burden *in utero* drives DNA damage and epigenetic change, which drives intense selection of HSC/MPPs such that the young mouse is a mosaic of only a significantly reduced pool of clones. Further clonal evolution is not observed as these mice die early in life. (Bottom) Clonality in *Aldh2$^{c/-}$Adh5$^{-/-}$Vav1^{iCre+}* mice. HSC/MPPs in development are protected by cell extrinsic ALDH2, hence pass through development without undergoing formaldehyde mediated reprogramming. However, chronic formaldehyde accumulation during life drives accumulation of DNA damage and epigenetic reprogramming that may drive functional decline of HSCs and cell attrition. As this accumulates, this reduces the size of the reserve HSC pool capable of replenishing the HSC/MPPs that are currently producing blood as they enter replicative senescence, leading to a collapse in clonal diversity in old mice.

5.3.3 Cell intrinsic *ALDH2* confers protection to peripheral tissues

Peripheral *ALDH2* was able to confer protection against the severe phenotypes observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, notably the severe perinatal lethality and multi-morbidities, as well as the reduced body weight and markers of tissue pathology. The fact that only a marginal increase in circulating GDF-15 was observed in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice with age, suggests that chronic formaldehyde stress and ageing blood did not directly contribute to worsening tissue pathology, as may have been expected from accelerated ageing or progression of chronic disease (Adela & Banerjee, 2015; Ho et al., 2013; Mulderrig et al., 2021; Tanaka et al., 2018; Yousefzadeh et al., 2021).

Interestingly, the signs of alleviated tissue pathology in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice arise despite equivalent levels of N^2-Me-dG adducts as seen in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice. N^2-Me-dG adducts serve as a biomarker of formaldehyde on DNA in tissues (Lu *et al*., 2010). Formaldehyde bound to guanine bases can be readily converted to DNA-DNA and DNA-protein crosslinks, promoting replication/transcription stalling and double stranded break formation (Wang et al., 2022; C.Millington, manuscript in preparation) hence more adducts are believed to correlate with greater DNA damage. This is further supported by correlation between adduct levels in tissues (e.g. kidney) and pathological phenotypes observed upon tier-2 deficiency (Pontel *et al*., 2015; Dingler *et al*., 2020; Mulderrig *et al*., 2021), and greater DNA-protein crosslink formation in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice (Oka et al., 2024). Given that the adduct levels are largely equivalent, this suggests that directly overwhelming tier-2 may not be the route through which tissue pathology is driven in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice. This is further supported by the absence of nuclear polyploidisation in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, which is induced by DDR proteins and observed in models of oxidative stress and DNA repair

deficiency (Chipchase et al., 2003; Gentric et al., 2015; Lau & Poon, 2023) and is observed in *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ mice (Pontel *et al*., 2015).

To date, the significance of ALDH2 compartmentalisation to the mitochondria, the essential difference between peripheral tissues of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ and *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ mice, has not been investigated. However, mitochondrial dysfunction is a well-recognised driver of diseases across tissues, including the kidney (Che et al., 2014), brain (Lin & Beal, 2006) and liver (Mansouri et al., 2018). Mitochondria are a major source of endogenous formaldehyde, with several pathways of production found within: spontaneous decomposition of tetrahydrofolate, 5,10-methyl-tetrahydrofolate and 10-formyl-tetrahydrofolate in the mitochondria can release formaldehyde (Burgos-Barragan *et al*., 2018); choline metabolism (Porter et al., 1985); lipid peroxidation (Bindoli, 1988); CYP2E1 localisation and metabolism of glycine (Clejan & Cederbaum, 1992; Knockaert et al., 2011). As a hub of formaldehyde production, the mitochondrial genome and proteins might be under an additional burden of formaldehyde, sufficient to exceed an apoptotic threshold, when loss of *ALDH2* is no longer compensated by *ADH5* (Kazak et al., 2012; Nguyen et al., 2023). In this way, the *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ phenotype may be additive not only through increased DNA damage, but also mitochondrial pathology. A proof-of-concept system to study this may be by combining tier-1 deficiency with pol-γA deficiency (compromising proof reading during mtDNA replication and is associated with premature ageing in mice (Trifunovic et al., 2004)). Formaldehyde induced increases in the mtDNA mutation rate may exacerbate mitochondrial dysfunction and drive premature ageing, even in the presence of tier-2 protection.

An alternative explanation may relate to the intense bottlenecking of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice during the perinatal window. Somatic mosaicism is a feature not just related to the blood, but is observed across tissues, including the oesophagus (Martincorena *et al*., 2018), breast tissue (Nishimura et al., 2023), bronchial epithelium (Gómez-López et al., 2025), liver (Ng et al., 2021) and male germ line (Neville et al., 2024), with expanding clones frequently presenting mutations in DNA damage proteins (notably TP53). In the blood, mutations that arise during development are now understood to contribute to the clonal architecture of adult tissues (Kapadia et al., 2025; Liu et al., 2025), as we hypothesise in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$. Given the high adduct burden in the foetal liver of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, profound mosaicism may arise in these tissues very early in life that selects for clones which can tolerate high formaldehyde but also decreases regenerative capacity, thus accelerating ageing. These clones may have deficiencies in DDR pathways, limiting DNA damage signalling in the liver that may also inhibit polyploidisation, while promoting cancer predisposition observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$. Future work to substantiate this mechanism may involve direct assessment of clonal architecture in the liver (and other tissues) of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice to identify clonal expansions, or epigenetic adaptations, that enrich for tolerance to formaldehyde.

It is important to note, that the biochemical markers of tissue damage, elevated specifically in the *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, may be explained by tissue-tissue connections that cannot be resolved with this model, rather than formaldehyde damage specific to that tissue. For one, renal dysfunction is also a common product of chronic liver disease, resulting from changing portal hypertension and renal vasoconstriction and drop in the kidney's ability to filter waste products (Arroyo et al., 2007). In another instance, hypoalbuminaemia (failure of the liver to produce albumin) is used routinely as an indicator of malnourishment, besides functional failure, which is consistent with

the anorexic phenotype observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice (Seltzer et al., 1979; Soeters et al., 2019). Histological characterisation of tubular damage and liver fibrosis would be a valuable addition to this work to help establish whether the route of the stress markers observed is indeed cell intrinsic.

It is interesting to observe that the largest difference in formaldehyde exposure (quantified by N^2-Me-dG adducts) between *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, was observed in the brain. This is interesting in the context of ADDS patients; a consistent phenotype amongst these patients is a significant mental retardation (Dingler *et al*., 2020; Oka *et al*., 2020). *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice had a 5-fold increase in the adduct level in the brain compared to *Adh5*$^{-/-}$ mice. In the absence of TC-NER, *Adh5*$^{-/-}$*Csb*$^{-/-}$, this formaldehyde burden is sufficient to induce profound neurodegeneration, in resemblance with human Cockayne syndrome (Mulderrig *et al*., 2021). Moreover, using a Cre-Switch model that reinstates ADH5 expression in the blood system, the adduct level is significantly lowered and partially rescues this neurodegenerative phenotype, despite the remaining absence of *Csb*$^{-/-}$, hence the phenotype correlates with the formaldehyde burden measured by the adduct level (H. Russell's Thesis, Oxford; L. Mulderrig's Thesis, Cambridge). *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice have never been assessed for their neurological phenotype, so it would be interesting to see if the clear increase in formaldehyde burden correlates with any signs of neurodegeneration, and if blood rescue restores function. If there is no phenotype, this raises the question of what is the critical amount of formaldehyde necessary to overload tier-2 protection, and are disorders derived from stress to other organelles?

5.3.4 Limitations

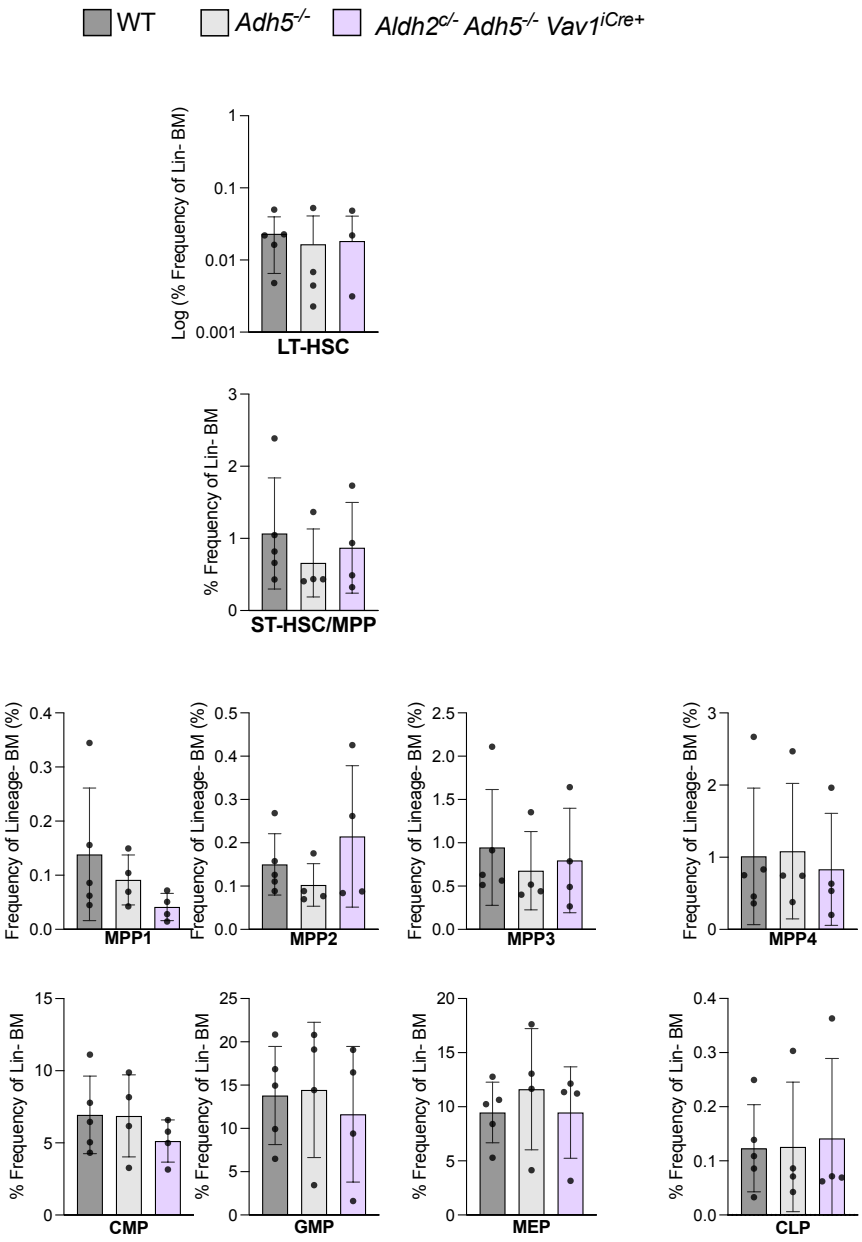
A recurring limitation of this study is the low number of biological replicates throughout, particularly regarding the *in utero* phenotypic characterisation of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ haematopoietic phenotype and need more replicates to increase the statistical power. In addition, there are complications associated with the use of the reduced N^2-Me-dG adduct in understanding of formaldehyde exposure across all tissues. The rate of cell division is a factor in the rate of adduct accumulation, such that more turnover reduces the adduct level, despite being in environments of equivalent formaldehyde. This study would benefit from direct serum formaldehyde quantification using GC-MS methods as used in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice (Dingler *et al*., 2020), to understand whether the systemic load of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ and *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ is indeed equivalent, or if the blood of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ is still under still under a greater load, which may account for their differences.

5.3.5 Conclusion

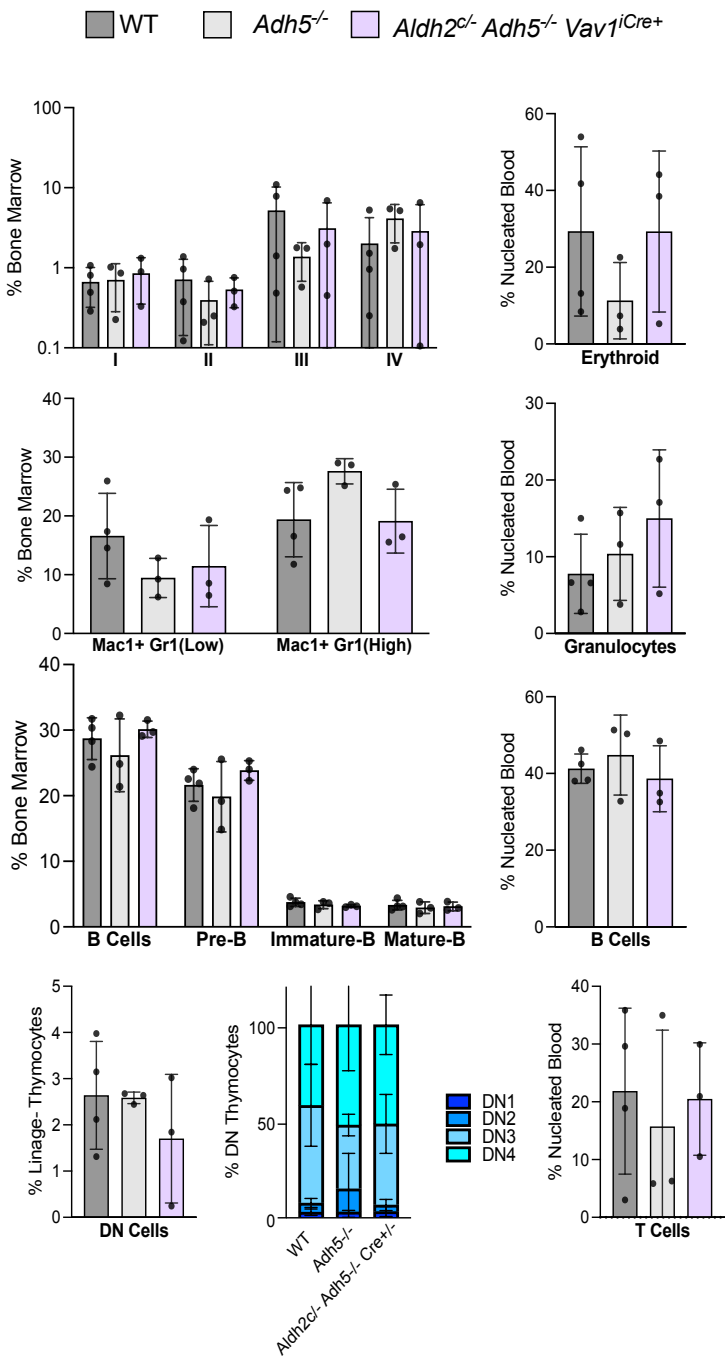
To conclude, peripheral clearance of formaldehyde by ALDH2 is sufficient to suppress the significant pathology observed in the *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice. It also extends understanding of the origins of *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ pathology to essential changes experienced *in utero*, which may help explain their rapid onset of peri-natal lethality. By contrast, the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mouse develops an ageing haematopoietic phenotype directly as a function of formaldehyde exposure through time, thus serving as a more suitable model to study how formaldehyde may drive the ageing process. Interestingly, despite the similar formaldehyde burden in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ and *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ tissues, as measured by N^2-Me-dG adduct quantification, there was a

striking difference in health of the haematopoietic, kidney and liver tissues. I have discussed how this may indicate mechanisms of adaptation in the *Aldh2*$^{c/-}$*Adh5*$^{-/-}$ *Vav1*$^{iCre+}$ tissues to high formaldehyde, potentially underappreciated roles of mitochondria, and a route by which formaldehyde drives mosaicism in haematopoietic tissues in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice that accelerates ageing of these mice. Future work should focus on extending the phenotypic analysis performed here to develop a more detailed understanding of the functional changes that take place in the different organs, and will benefit from multi-OMIC profiling to provide a mechanistic understanding of how exactly formaldehyde drives these changes. Furthermore, while the changes in HSC/MPP output of *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice did not translate to changes in clonal architecture, signs of accelerated ageing were observed (downstream MPP compartments, epigenetic reprogramming, and mutation accumulation), which may translate to clonality when encountering exogenous stress, while in a native setting, emergence of clonality may only be a matter of time.

5.4 Supplementary Figures

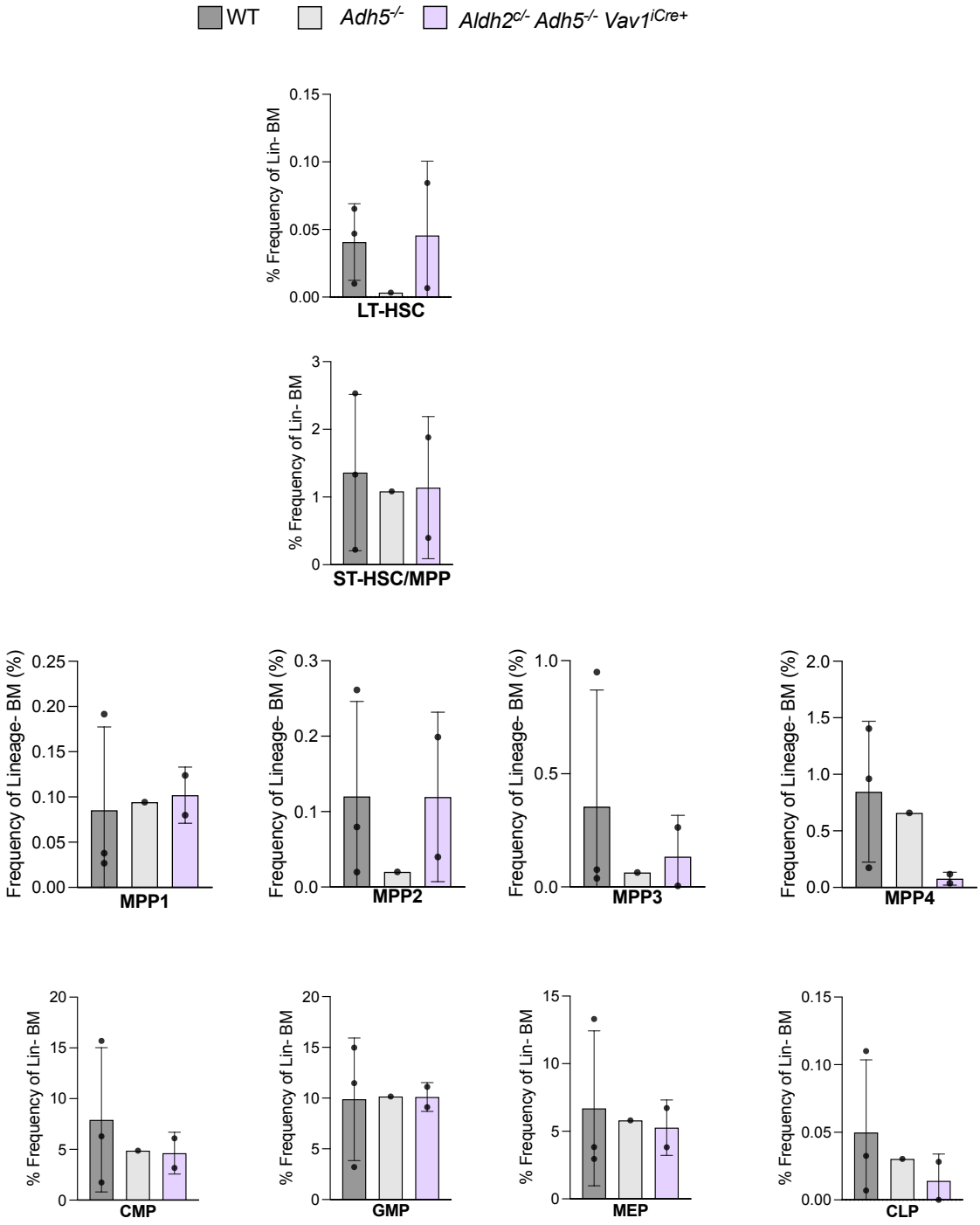


Supplementary Figure 5.1. – Haematopoietic stem cell, multipotent progenitor and committed progenitor quantification in bone marrow in 8-12-week females
Quantification of respective populations in 8-12-week male old mice, as a proportion of total number of Lineage$^-$ cells sampled. Bars represent mean ± SD, data represent individual mice, n = 5, 4, 3, p values determined by Dunn's test, ns > 0.05.

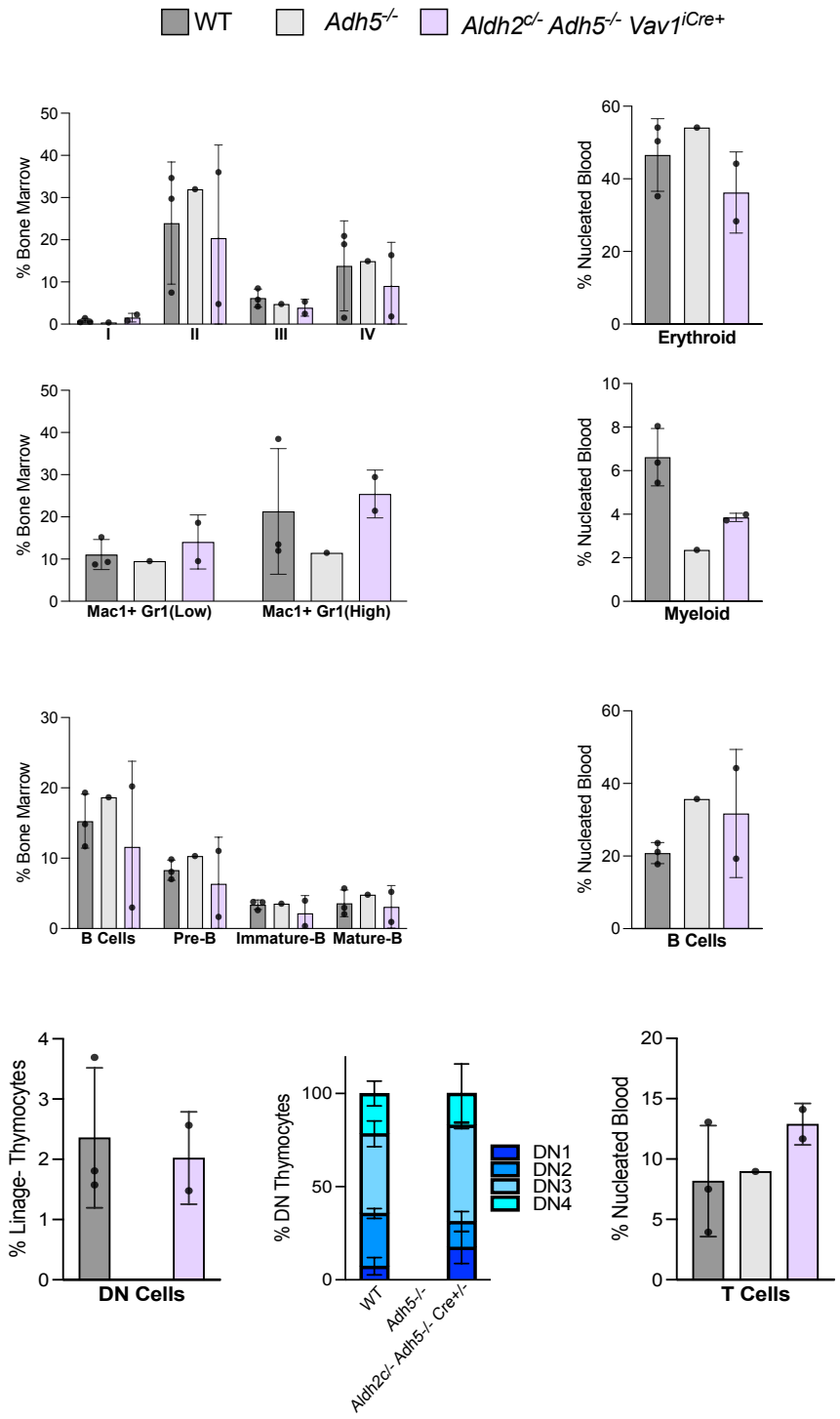


Supplementary Figure 5.2 Maturing blood populations in the bone marrow in 8-12 week-old females

Quantification of maturing lineages (left) in the whole bone marrow and subsequent peripheral blood populations (right), for erythroid, myeloid, B-cell, thymic development. All data representing mean ± SD, points represent individual mice, n = 5, 4, 3, p values determined by Dunn's test, ns > 0.05.



Supplementary Figure 5.3 Haematopoietic stem cell, multipotent progenitor and committed progenitor quantification in bone marrow of 1-year old females, n = 3, 1, 2, bars represent mean ± SD



Supplementary Figure 5.4 Bone marrow maturation in 1-year aged females
(Left) Quantification of maturing lineages for erythroid, myeloid, and B-cell and thymic development. (Right) Quantification of terminal blood populations in peripheral blood. n = 3, 0, 2. Bars represent mean ± SD

Chapter 6

Concluding remarks

Formaldehyde is an endogenous driver of DNA damage, hence a driver of ageing. By studying enzymes with combined deletions of metabolic enzymes of detoxification (*ALDH2*, *ADH5*) and DNA repair (*FANCD2*, *CSB*), we understand this one carbon molecule as a significant source of genomic instability. In this way, formaldehyde can act as a driving force of natural selection at the human population level (in evolution of redundant mechanisms of two-tier protection) and cellular level (driver of CH).

The severity of the *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ phenotype, highlights the significance of the redundant tier-1 mechanism that support formaldehyde clearance. Meanwhile, scRNAseq analysis has highlighted significant transcriptional remodelling in contexts of high endogenous formaldehyde, suggesting signalling responses that are responsive to formaldehyde that may coordinate an adaptive response. However, the relative significance of these pathways in mitigating the toxic effects of formaldehyde in unknown. My thesis therefore had its **first aim: How do cells recognise and respond to formaldehyde (Chapter 3)**. I attempted to develop a fluorescent reporter of GDF-15 transcription to provide a sensitive readout of formaldehyde stress, which could be used in a genome-wide CRISPR-Cas9 screen. In this screen, increased GDF-15 induction would suggest genes involved in protection against formaldehyde,

while decreased induction would suggest genes involved in damage signalling or formaldehyde sources. I attempted to produce this reporter using both endogenous and exogenous routes but was unsuccessful in both avenues. There were technical limitations in these experiments that would need to be addressed if the work would be revisited, which have been detailed. However, given reports of Flow-FISH in successfully monitoring transcriptional output in CRISPR-Cas9 screens, which gets around the need to produce a genetically modified cell line that could potentially remove elements necessary for transcription, I would recommend this route in future work.

While the significance of tier-1 and two-tier deficiency in humans is evident, it remains curious how the loss of function *ALDH2*2* variant has risen to such prominence among East Asians. Therefore, my thesis had its **second aim: Does ALDH2 deficiency confer a selective advantage (Chapter 4)**. I dismissed the argument of alcohol abstinence based on the late onset of associated diseases that alcohol may induce, which would be unlikely to influence selection in the reproductive population. I instead focused on potential for immune protection, routing from work by Douglas & Fearon, who described direct modification of antigens by glycolaldehyde (which can be produced from MPO expressed in neutrophils) to increase the humoral response. In an ALDH2-deficient background, absence of aldehyde clearance by ALDH2 may increase the degree of detoxification of these aldehydes in the immune system, thus conferring increased immune response. Investigating this hypothesis, F.Dingler observed *Aldh2*$^{-/-}$ mice indeed had a greater response to OVA. I looked to see if this was a generalisable feature in other antigens, HEL, BSA or HA, but did not observe the same increases, although in HA, a decelerated humoral response was observed.

The fact that *Aldh2*$^{-/-}$ mice do show differences in the immune response in OVA and HA, suggests that differential regulation of the adaptive immune response may indeed be a mechanism by which *ALDH2*-deficiency confers an advantage, although is highly dependent on the antigen properties. This study focuses on a general output of the immune system, binding IgG antibodies, and future work should focus on understanding the direct mechanism through which this may operate, and whether this confers protection to infectious disease in *in vivo* challenge models, or human patients carrying the *ALDH2*2* allele. However, in this chapter I also discuss limitations of the aldehyde hypothesis, in that the mechanisms hypothesised to increase immune protection, could have a significant trade-off in disease tolerance.

One of the tissues most susceptible to formaldehyde induced damage is the blood. Chronic formaldehyde exposure drives features of haematopoietic ageing in *Aldh2*$^{-/-}$ *Fancd2*$^{-/-}$, *Adh5*$^{-/-}$*Fancd2*$^{-/-}$ and *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ models, observed by loss of repopulation capacity, anaemia, myeloid-bias, and clonal haematopoiesis. However, we don't know exactly the mechanism by which formaldehyde may drive ageing in the blood in a normal ageing scenario. Furthermore, as ageing in the blood is known to accelerate the functional decline of peripheral tissues, it is important to understand whether this is through its roles in mediating formaldehyde detoxification, or through functional decline in supporting tissue homeostasis. However, *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ experience severe peri-natal lethality and multi-morbidity. This makes it difficult to studying changes to the blood as a function of time (i.e. age related). I, therefore, characterised a mouse model, where ALDH2 was deleted specifically in the blood of *Adh5*$^{-/-}$ mice using a *Vav1*$^{iCre+}$ system. This relieved the extreme peri-natal defect and multi-morbid phenotype, prompting my thesis' **third aim: does blood formaldehyde**

clearance protect against ageing in the blood and peripheral tissues? I found peripheral clearance of formaldehyde by ALDH2 was sufficient to protect the haematopoietic system and peripheral tissues from the chronic formaldehyde burden observed in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$. Interestingly, there was still a significant formaldehyde load in all tissues. In the blood, this high formaldehyde burden drove DNA damage and accelerated acquisition of myeloid biased haematopoiesis and mild signs of anaemia, indicative of accelerated ageing. In the peripheral tissues, however, there was no clear indication of formaldehyde inducing genomic instability (assessed through liver polyploidisation) or tissue damage in young or 1 year old mice. This may highlight the importance of non-DNA damage channels of pathology in *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mice, such as mitochondrial stress, as a route of formaldehyde mediated disease. Protection of the haematopoietic and peripheral tissues from toxic effects of formaldehyde, suggests significant epigenetic changes are induced that facilitate protection, hence future work would benefit from multi-omic profiling of these tissues to understand what pathways are upregulated that facilitate formaldehyde tolerance, and how do they affect cellular functions, such as lineage bias in haematopoietic stem cells.

Finally, this work puts important light on how formaldehyde may act as a driver of clonal haematopoiesis. In driving instructional changes to cell fates, formaldehyde may act as a driver of selection, e.g. by promoting cells better adapted to tolerate high formaldehyde loads. In the constitutive *Aldh2*$^{-/-}$*Adh5*$^{-/-}$ mouse, the perinatal window is a period of rapid expansion that may amplify the forces of selection, resulting in clonal haematopoiesis, and likely mosaicism of other tissues. In a chronic situation (as seen in *Aldh2*$^{c/-}$*Adh5*$^{-/-}$*Vav1*$^{iCre+}$ mice) the same underlying epigenetic changes may incur, but the stimulus for selection is not observed. Proliferative stimuli, such as those driven

by acute infection, are a feature of life, but not of a laboratory mouse in a sterile environment. I predict that in underlying the ageing process in the blood (and likely other tissues), formaldehyde mediated instruction will accelerate the pruning of stem cell populations when exposed to extrinsic stressors. In doing so, formaldehyde becomes a fundamental agent of somatic decay through life.

References

Abascal, F., Harvey, L. M. R., Mitchell, E., Lawson, A. R. J., Lensing, S. V., Ellis, P., Russell, A. J. C., Alcantara, R. E., Baez-Ortega, A., Wang, Y., Kwa, E. J., Lee-Six, H., Cagan, A., Coorens, T. H. H., Chapman, M. S., Olafsson, S., Leonard, S., Jones, D., Machado, H. E., … Martincorena, I. (2021). Somatic mutation landscapes at single-molecule resolution. *Nature*, *593*(7859), 405–410. https://doi.org/10.1038/s41586-021-03477-4

Acharya, A. S., & Manning, J. M. (1983). Reaction of glycolaldehyde with proteins: latent crosslinking potential of alpha-hydroxyaldehydes. *Proceedings of the National Academy of Sciences*, *80*(12), 3590–3594. https://doi.org/10.1073/pnas.80.12.3590

Adela, R., & Banerjee, S. K. (2015). GDF-15 as a target and biomarker for diabetes and cardiovascular diseases: A translational prospective. *Journal of Diabetes Research*, *2015*. https://doi.org/10.1155/2015/490842

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nature Genetics*, *47*(12), 1402–1407. https://doi.org/10.1038/ng.3441

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Ashiqul Islam, S. M., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Getz, G., … Yu, W. (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3

Allen, P. M., Matsueda, G. R., Evans, R. J., Dunbar, J. J. B., Marshall, G. R., & Unanue, E. R. (1987). Identification of the T-cell and Ia contact residues of a T-cell antigenic epitope. *Nature*, *327*(6124), 713–715. https://doi.org/10.1038/327713a0

Allison, M. E. D., & Fearon, D. T. (2000). Enhanced immunogenicity of aldehyde-bearing antigens: A possible link between innate and adaptive immunity. *European Journal of Immunology*, *30*(10), 2881–2887. https://doi.org/10.1002/1521-4141(200010)30:10<2881::AID-IMMU2881>3.0.CO;2-9

Anderson, M. M., Hazen, S. L., Hsu, F. F., & Heinecke, J. W. (1997). Human neutrophils employ the myeloperoxidase-hydrogen peroxide-chloride system to convert hydroxy-amino acids into glycolaldehyde, 2-hydroxypropanal, and acrolein: A mechanism for the generation of highly reactive α-hydroxy and α,β-unsaturated aldehydes by phagocytes at sites of inflammation. *Journal of Clinical Investigation*, *99*(3), 424–432. https://doi.org/10.1172/JCI119176

Armstrong, G. T., Kawashima, T., Leisenring, W., Stratton, K., Stovall, M., Hudson, M. M., Sklar, C. A., Robison, L. L., & Oeffinger, K. C. (2014). Aging and risk of severe, disabling, life-threatening, and fatal events in the childhood cancer survivor study. *Journal of Clinical Oncology*, *32*(12), 1218–1227. https://doi.org/10.1200/JCO.2013.51.1055

Arroyo, V., Terra, C., & Ginès, P. (2007). Advances in the pathogenesis and treatment of type-1 and type-2 hepatorenal syndrome. In *Journal of Hepatology* (Vol. 46, Issue 5, pp. 935–946). https://doi.org/10.1016/j.jhep.2007.02.001

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., & Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks.

Nature Structural and Molecular Biology, *21*(4), 366–374. https://doi.org/10.1038/nsmb.2796

Baek, S. J., Horowitz, J. M., & Eling, T. E. (2001). Molecular cloning and characterization of human nonsteroidal anti-inflammatory drug-activated gene promoter. Basal transcription is mediated by Sp1 and Sp3. *Journal of Biological Chemistry*, *276*(36), 33384–33392. https://doi.org/10.1074/jbc.M101814200

Baek, S. J., Kim, J. S., Nixon, J. B., DiAugustine, R. P., & Eling, T. E. (2004). Expression of NAG-1, a Transforming Growth Factor-β Superfamily Member, by Troglitazone Requires the Early Growth Response Gene EGR-1. *Journal of Biological Chemistry*, *279*(8), 6883–6892. https://doi.org/10.1074/jbc.M305295200

Baker, D. J., Childs, B. G., Durik, M., Wijers, M. E., Sieben, C. J., Zhong, J., A. Saltness, R., Jeganathan, K. B., Verzosa, G. C., Pezeshki, A., Khazaie, K., Miller, J. D., & Van Deursen, J. M. (2016). Naturally occurring p16 Ink4a-positive cells shorten healthy lifespan. *Nature*, *530*(7589), 184–189. https://doi.org/10.1038/nature16932

Baldridge, M. T., King, K. Y., & Goodell, M. A. (2011). Inflammatory signals regulate hematopoietic stem cells. In *Trends in Immunology* (Vol. 32, Issue 2, pp. 57–65). https://doi.org/10.1016/j.it.2010.12.003

Balmain, A. (2020). The critical roles of somatic mutations and environmental tumor-promoting agents in cancer risk. In *Nature Genetics* (Vol. 52, Issue 11, pp. 1139–1143). Nature Research. https://doi.org/10.1038/s41588-020-00727-5

Barr, T., Helms, C., Grant, K., & Messaoudi, I. (2016). Opposing effects of alcohol on the immune system. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *65*, 242–251. https://doi.org/10.1016/j.pnpbp.2015.09.001

Beerman, I., Seita, J., Inlay, M. A., Weissman, I. L., & Rossi, D. J. (2014). Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell Stem Cell*, *15*(1), 37–50. https://doi.org/10.1016/j.stem.2014.04.016

Benne, N., van Duijn, J., Kuiper, J., Jiskoot, W., & Slütter, B. (2016). Orchestrating immune responses: How size, shape and rigidity affect the immunogenicity of particulate vaccines. In *Journal of Controlled Release* (Vol. 234, pp. 124–134). Elsevier B.V. https://doi.org/10.1016/j.jconrel.2016.05.033

Benson, M. J., Erickson, L. D., Gleeson, M. W., & Noelle, R. J. (2007). Affinity of antigen encounter and other early B-cell signals determine B-cell fate. In *Current Opinion in Immunology* (Vol. 19, Issue 3, pp. 275–280). https://doi.org/10.1016/j.coi.2007.04.009

Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H., & Moore, K. (2016). Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell*, *167*(5), 1296-1309.e10. https://doi.org/10.1016/j.cell.2016.10.022

Berry, S. B., Espich, S., Thuong, N. T. T., Chang, X., Dorajoo, R., Khor, C.-C., Heng, C.-K., Yuan, J.-M., Fox, D., Anaya-Sanchez, A., Tenney, L., Chang, C. J., Kotov, D. I., Vance, R. E., Dunstan, S. J., Darwin, K. H., & Stanley, S. A. (2023). Disruption of Aldehyde Dehydrogenase 2 protects against bacterial infection. *BioRxiv : The Preprint Server for Biology*. https://doi.org/10.1101/2023.08.24.554661

Bian, L., Zheng, Y., Guo, X., Li, D., Zhou, J., Jing, L., Chen, Y., Lu, J., Zhang, K., Jiang, C., Zhang, Y., & Kong, W. (2022). Intramuscular Inoculation of AS02-Adjuvanted Respiratory Syncytial Virus (RSV) F Subunit Vaccine Shows Better Efficiency and Safety Than Subcutaneous Inoculation in BALB/c Mice. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.938598

Bindoli, A. (1988). LIPID PEROXIDATION IN MITOCHONDRIA. In *Free Radical Bkdogy & Medicine* (Vol. 5).

Blackford, A. N., & Jackson, S. P. (2017). ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. In *Molecular Cell* (Vol. 66, Issue 6, pp. 801–817). Cell Press. https://doi.org/10.1016/j.molcel.2017.05.015

Boboila, C., Yan, C., Wesemann, D. R., Jankovic, M., Wang, J. H., Manis, J., Nussenzweig, A., Nussenzweig, M., & Alt, F. W. (2010). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase. *Journal of Experimental Medicine*, *207*(2), 417–427. https://doi.org/10.1084/jem.20092449

Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., Lu, T., Maroc, L., Norman, T. M., Song, B., Stanley, G., Chen, S., Garnett, M., Li, W., Moffat, J., Qi, L. S., Shapiro, R. S., Shendure, J., Weissman, J. S., & Zhuang, X. (2022). High-content CRISPR screening. In *Nature Reviews Methods Primers* (Vol. 2, Issue 1). Springer Nature. https://doi.org/10.1038/s43586-021-00093-4

Boettcher, S., Miller, P. G., Sharma, R., McConkey, M., Leventhal, M., Krivtsov, A. V., Giacomelli, A. O., Wong, W., Kim, J., Chao, S., Kurppa, K. J., Yang, X., Milenkowic, K., Piccioni, F., Root, D. E., Rücker, F. G., Flamand, Y., Neuberg, D., Lindsley, R. C., … Ebert, B. L. (2019). A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science*, *365*(6453), 599–604. https://doi.org/10.1126/science.aax3649

Bogahawaththa, S., Hara, M., Furukawa, T., Iwasaka, C., Sawada, T., Yamada, G., Tokiya, M., Kitagawa, K., Miyake, Y., Kido, M. A., Hirota, Y., & Matsumoto, A. (2024). Asian Flush Gene Variant Enhances Cellular Immunogenicity of COVID-19 Vaccine: Prospective Observation in the Japanese General Population. *Vaccines*, *12*(9), 1015. https://doi.org/10.3390/vaccines12091015

Bogeska, R., Mikecin, A.-M., Kaschutnig, P., Fawaz, M., Büchler-Schäff, M., Le, D., Ganuza, M., Vollmer, A., Paffenholz, S. V., Asada, N., Rodriguez-Correa, E., Frauhammer, F., Buettner, F., Ball, M., Knoch, J., Stäble, S., Walter, D., Petri, A., Carreño-Gonzalez, M. J., … Milsom, M. D. (2022). Inflammatory exposure drives long-lived impairment of hematopoietic stem cell self-renewal activity and accelerated aging. *Cell Stem Cell*, *29*(8), 1273-1284.e8. https://doi.org/10.1016/j.stem.2022.06.012

Bolton, K. L., Ptashkin, R. N., Gao, T., Braunstein, L., Devlin, S. M., Kelly, D., Patel, M., Berthon, A., Syed, A., Yabe, M., Coombs, C. C., Caltabellotta, N. M., Walsh, M., Offit, K., Stadler, Z., Mandelker, D., Schulman, J., Patel, A., Philip, J., … Papaemmanuil, E. (2020). Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nature Genetics*, *52*(11), 1219–1226. https://doi.org/10.1038/s41588-020-00710-0

Bondar, T., & Medzhitov, R. (2010). p53-Mediated Hematopoietic Stem and Progenitor Cell Competition. *Cell Stem Cell*, *6*(4), 309–322. https://doi.org/10.1016/j.stem.2010.03.002

Böttner, M., Suter-Crazzolara, C., Schober, A., & Unsicker, K. (1999). Expression of a novel member of the TGF-β superfamily, growth/differentiation factor-15/macrophage-inhibiting cytokine-1 (GDF-15/MIC-1) in adult rat tissues. *Cell and Tissue Research*, *297*(1), 103–110. https://doi.org/10.1007/s004410051337

Bressac, B., Galvin, K. M., Liang, T. J., Isselbacher, K. J., Wands, J. R., & Ozturk, M. (1990). Abnormal structure and expression of p53 gene in human hepatocellular carcinoma (hepatocarcinogenesis/recessive oncogene/gene expression). In *Proc. Nati. Acad. Sci. USA* (Vol. 87).

Brooks, P. J., Enoch, M. A., Goldman, D., Li, T. K., & Yokoyama, A. (2009). The alcohol flushing response: An unrecognized risk factor for esophageal cancer from alcohol

consumption. *PLoS Medicine*, *6*(3), 0258–0263. https://doi.org/10.1371/journal.pmed.1000050

Burgos-Barragan, G., Wit, N., Meiser, J., Dingler, F. A., Pietzke, M., Mulderrig, L., Pontel, L. B., Rosado, I. V., Brewer, T. F., Cordell, R. L., Monks, P. S., Chang, C. J., Vazquez, A., & Patel, K. J. (2017). Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism. *Nature*, *548*(7669), 549–554. https://doi.org/10.1038/nature23481

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T., & Rodewald, H. R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, *518*(7540), 542–546. https://doi.org/10.1038/nature14242

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T. H. H., Sanders, M. A., Lawson, A. R. J., Harvey, L. M. R., Bhosle, S., Jones, D., Alcantara, R. E., Butler, T. M., Hooks, Y., Roberts, K., Anderson, E., Lunn, S., Flach, E., Spiro, S., Januszczak, I., … Martincorena, I. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature*, *604*(7906), 517–524. https://doi.org/10.1038/s41586-022-04618-z

Calabro, S., Tortoli, M., Baudner, B. C., Pacitto, A., Cortese, M., O'Hagan, D. T., De Gregorio, E., Seubert, A., & Wack, A. (2011). Vaccine adjuvants alum and MF59 induce rapid recruitment of neutrophils and monocytes that participate in antigen transport to draining lymph nodes. *Vaccine*, *29*(9), 1812–1823. https://doi.org/10.1016/j.vaccine.2010.12.090

Ceccaldi, R., Parmar, K., Mouly, E., Delord, M., Kim, J. M., Regairaz, M., Pla, M., Vasquez, N., Zhang, Q. S., Pondarre, C., Peffault De Latour, R., Gluckman, E., Cavazzana-Calvo, M., Leblanc, T., Larghero, J., Grompe, M., Socié, G., D'Andrea, A. D., & Soulier, J. (2012). Bone marrow failure in fanconi anemia is triggered by an exacerbated p53/p21 DNA damage response that impairs hematopoietic stem and progenitor cells. *Cell Stem Cell*, *11*(1), 36–49. https://doi.org/10.1016/j.stem.2012.05.013

Chambers, S. M., Shaw, C. A., Gatza, C., Fisk, C. J., Donehower, L. A., & Goodell, M. A. (2007). Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS Biology*, *5*(8), 1750–1762. https://doi.org/10.1371/journal.pbio.0050201

Chang, J., Wang, Y., Shao, L., Laberge, R. M., Demaria, M., Campisi, J., Janakiraman, K., Sharpless, N. E., Ding, S., Feng, W., Luo, Y., Wang, X., Aykin-Burns, N., Krager, K., Ponnappan, U., Hauer-Jensen, M., Meng, A., & Zhou, D. (2016). Clearance of senescent cells by ABT263 rejuvenates aged hematopoietic stem cells in mice. *Nature Medicine*, *22*(1), 78–83. https://doi.org/10.1038/nm.4010

Chang, X., Li, Z., Khac Thai, P. V., Minh Ha, D. T., Thuong Thuong, N. T., Wee, D., Binte Mohamed Subhan, A. S., Silcocks, M., Eng Chee, C. Bin, Quynh Nhu, N. T., Heng, C.-K., Teo, Y. Y., Singal, A., Oehlers, S. H., Yuan, J.-M., Koh, W.-P., Caws, M., Khor, C. C., Dorajoo, R., & Dunstan, S. J. (2024). *Genome-wide association study reveals a novel tuberculosis susceptibility locus in multiple East Asian and European populations*. https://doi.org/10.1101/2024.03.14.24304327

Che, R., Yuan, Y., Huang, S., & Zhang, A. (2014). Mitochondrial dysfunction in the pathophysiology of renal diseases. *Am J Physiol Renal Physiol*, *306*, 367–378. https://doi.org/10.1152/ajprenal.00571.2013.-Mi

Chen, C.-C., Lu, R.-B., Chen, Y.-C., Wang, M.-F., Chang, Y.-C., Li, T.-K., & Yin, S.-J. (1999). Interaction between the Functional Polymorphisms of the Alcohol-

Metabolism Genes in Protection against Alcoholism. *The American Journal of Human Genetics*, *65*(3), 795–807. https://doi.org/10.1086/302540

Chipchase, M. D., O'Neill, M., & Melton, D. W. (2003). Characterization of premature liver polyploidy in DNA repair (Ercc1)-deficient mice. *Hepatology*, *38*(4), 958–966. https://doi.org/10.1053/jhep.2003.50421

Christensen, J. L., Wright, D. E., Wagers, A. J., & Weissman, I. L. (2004). Circulation and chemotaxis of fetal hematopoietic stem cells. *PLoS Biology*, *2*(3). https://doi.org/10.1371/journal.pbio.0020075

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, *33*(5), 543–548. https://doi.org/10.1038/nbt.3198

Clejan, L. A., & Cederbaum, A. I. (1992). Role of cytochrome P450 in the oxidation of glycerol by reconstituted systems and microsomes. *The FASEB Journal*, *6*(2), 765–770. https://doi.org/10.1096/fasebj.6.2.1537467

Cong, P. K., Bai, W. Y., Li, J. C., Yang, M. Y., Khederzadeh, S., Gai, S. R., Li, N., Liu, Y. H., Yu, S. H., Zhao, W. W., Liu, J. Q., Sun, Y., Zhu, X. W., Zhao, P. P., Xia, J. W., Guan, P. L., Qian, Y., Tao, J. G., Xu, L., … Zheng, H. F. (2022). Genomic analyses of 10,376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30526-x

Crabb, D. W., Edenberg, H. J., Bosron, W. F., & Li, T. K. (1989). Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. The inactive ALDH22 allele is dominant. *Journal of Clinical Investigation*, *83*(1), 314–316. https://doi.org/10.1172/JCI113875

Crotty, S. (2014). T Follicular Helper Cell Differentiation, Function, and Roles in Disease. In *Immunity* (Vol. 41, Issue 4, pp. 529–542). Cell Press. https://doi.org/10.1016/j.immuni.2014.10.004

Daniels, P. S., Jeffries, S., Yates, P., Schild, G. C., Rogers, G. N., Paulson, J. C., Wharton, S. A., Douglas, A. R., Skehel, J. J., & Wiley, D. C. (1987). The receptor-binding and membrane-fusion properties of influenza virus variants selected using anti-haemagglutinin monoclonal antibodies. *The EMBO Journal*, *6*(5), 1459–1465. https://doi.org/10.1002/j.1460-2075.1987.tb02387.x

Darwin, K. H., & Stanley, S. A. (2022). *The aldehyde hypothesis: metabolic intermediates as antimicrobial effectors*. https://doi.org/10.1098/rsob.22.0010

Delaforest, A., Di Furio, F., Jing, R., Ludwig-Kubinski, A., Twaroski, K., Urick, A., Pulakanti, K., Rao, S., & Duncan, S. A. (2019). HNF4A regulates the formation of hepatic progenitor cells from human iPSC-derived endoderm by facilitating efficient recruitment of RNA pol II. *Genes*, *10*(1). https://doi.org/10.3390/genes10010021

Di Tommaso, A., De Magistris, M. T., Bugnoli, M., Marsili, I., Rappuoli, R., & Abrignani, S. (1994). Formaldehyde Treatment of Proteins Can Constrain Presentation to T Cells by Limiting Antigen Processing. In *INFECTION AND IMMUNITY* (Vol. 62, Issue 5). https://journals.asm.org/journal/iai

Dingler, F. A., Wang, M., Mu, A., Millington, C. L., Oberbeck, N., Watcham, S., Pontel, L. B., Kamimae-Lanning, A. N., Langevin, F., Nadler, C., Cordell, R. L., Monks, P. S., Yu, R., Wilson, N. K., Hira, A., Yoshida, K., Mori, M., Okamoto, Y., Okuno, Y., … Patel, K. J. (2020). Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Molecular Cell*, *80*(6), 996-1012.e9. https://doi.org/10.1016/j.molcel.2020.10.012

Dorokhov, Y. L., Shindyapina, A. V, Sheshukova, E. V, & Komarova, T. V. (2015). Metabolic Meth-anol: Molecular Pathways and Physiological Roles. *Physiol Rev*, *95*, 603–644. https://doi.org/10.1152/physrev.00034.2014.-Methanol

Dumble et al., 2007. (n.d.).

Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M., & Haan, G. De. (2011). Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *Journal of Experimental Medicine*, *208*(13), 2691–2703. https://doi.org/10.1084/jem.20111490

Edenberg, H. J. (2007). The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Research & Health : The Journal of the National Institute on Alcohol Abuse and Alcoholism*, *30*(1), 5–13. http://www.ncbi.nlm.nih.gov/pubmed/17718394

Elyahu, Y., Hekselman, I., Eizenberg-Magar, I., Berner, O., Strominger, I., Schiller, M., Mittal, K., Nemirovsky, A., Eremenko, E., Vital, A., Simonovsky, E., Chalifa-Caspi, V., Friedman, N., Yeger-Lotem, E., & Monsonego, A. (2019). Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Science Advances*, *5*(8). https://doi.org/10.1126/sciadv.aaw8330

Ema, H., & Nakauchi, H. (2000). Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood*, *95*(7), 2284–2288. http://www.ncbi.nlm.nih.gov/pubmed/10733497

Eng, M. Y., Luczak, S. E., & Wall, T. L. (n.d.). *ALDH2, ADH1B, AND ADH1C GENOTYPES IN ASIANS: A LITERATURE REVIEW*.

Fabre, M. A., de Almeida, J. G., Fiorillo, E., Mitchell, E., Damaskou, A., Rak, J., Orrù, V., Marongiu, M., Chapman, M. S., Vijayabaskar, M. S., Baxter, J., Hardy, C., Abascal, F., Williams, N., Nangalia, J., Martincorena, I., Campbell, P. J., McKinney, E. F., Cucca, F., … Vassiliou, G. S. (2022). The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature*, *606*(7913), 335–342. https://doi.org/10.1038/s41586-022-04785-z

Fen, Y., Chaw, M., Crane, L. E., Lange, P., & Shapiro, R. (1980). Isolation and Identification of Cross-Links from Formaldehyde-Treated Nucleic Acids"1". In *Biochemistry* (Vol. 19). https://pubs.acs.org/sharingguidelines

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M. I., & Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. *Science*, *354*(6313), 760–764. https://doi.org/10.1126/science.aag0776

Flach, J., Bakker, S. T., Mohrin, M., Conroy, P. C., Pietras, E. M., Reynaud, D., Alvarez, S., Diolaiti, M. E., Ugarte, F., Forsberg, E. C., Le Beau, M. M., Stohr, B. A., Méndez, J., Morrison, C. G., & Passegué, E. (2014). Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature*, *512*(7513), 198–202. https://doi.org/10.1038/nature13619

Flohr Svendsen, A., Yang, D., Kim, K., Lazare, S., Skinder, N., Zwart, E., Mura-Meszaros, A., Ausema, A., von Eyss, B., de Haan, G., & Bystrykh, L. (2021). A comprehensive transcriptome signature of murine hematopoietic stem cell aging. *Blood*, *138*(6), 439–451. https://doi.org/10.1182/blood.2020009729

Frisch, B. J., Hoffman, C. M., Latchney, S. E., LaMere, M. W., Myers, J., Ashton, J., Li, A. J., Saunders, J., Palis, J., Perkins, A. S., McCabe, A., Smith, J. N. P., McGrath, K. E., Rivera-Escalera, F., McDavid, A., Liesveld, J. L., Korshunov, V. A., Elliott, M. R., MacNamara, K. C., … Calvi, L. M. (2019). Aged marrow macrophages expand platelet-biased hematopoietic stem cells via interleukin-1B. *JCI Insight*, *4*(10). https://doi.org/10.1172/jci.insight.124213

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H., Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L., Lander, E. S., & Engreitz, J. M. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. In *Nature Genetics* (Vol. 51, Issue 12, pp. 1664–1669). Nature Research. https://doi.org/10.1038/s41588-019-0538-0

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, *7*(11). https://doi.org/10.1371/journal.pgen.1002355

Garaycoechea, J. I., Crossan, G. P., Langevin, F., Daly, M., Arends, M. J., & Patel, K. J. (2012). Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature*, *489*(7417), 571–575. https://doi.org/10.1038/nature11368

Garaycoechea, J. I., Crossan, G. P., Langevin, F., Mulderrig, L., Louzada, S., Yang, F., Guilbaud, G., Park, N., Roerink, S., Nik-Zainal, S., Stratton, M. R., & Patel, K. J. (2018). Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature*, *553*(7687), 171–177. https://doi.org/10.1038/nature25154

García-García, V. A., Alameda, J. P., Page, A., & Casanova, M. L. (2021). Role of nf-κb in ageing and age-related diseases: Lessons from genetically modified mouse models. In *Cells* (Vol. 10, Issue 8). MDPI. https://doi.org/10.3390/cells10081906

Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., … McCarroll, S. A. (2014). Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*, *371*(26), 2477–2487. https://doi.org/10.1056/nejmoa1409405

Gentric, G., Maillet, V., Paradis, V., Couton, D., L'Hermitte, A., Panasyuk, G., Fromenty, B., Celton-Morizur, S., & Desdouets, C. (2015). Oxidative stress promotes pathologic polyploidization in nonalcoholic fatty liver disease. *Journal of Clinical Investigation*, *125*(3), 981–992. https://doi.org/10.1172/JCI73957

Gioia, L., Siddique, A., Head, S. R., Salomon, D. R., & Su, A. I. (2018). A genome-wide survey of mutations in the Jurkat cell line. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4718-6

Glombt, M. A., & Monnier, V. M. (1995). *THE JOURNAL OF BIOLOGICAL CHEMISTRY Mechanism of Protein Modification by Glyoxal and Glycolaldehyde, Reactive Intermediates of the Maillard Reaction** (Vol. 270, Issue 17).

Gómez-López, S., Alhendi, A. S. N., Przybilla, M. J., Bordeu, I., Whiteman, Z. E., Butler, T., Rouhani, M. J., Kalinke, L., Uddin, I., Otter, K. E. J., Chandrasekharan, D. P., Lebrusant-Fernandez, M., Shurr, A. Y. L., Durrenberger, P. F., Moore, D. A., Falzon, M., Reading, J. L., Martincorena, I., Simons, B. D., … Janes, S. M. (2025). Aberrant basal cell clonal dynamics shape early lung carcinogenesis. *Science*, *388*(6752). https://doi.org/10.1126/science.ads9145

Green, L. S., Chun, L. E., Patton, A. K., Sun, X., Rosenthal, G. J., & Richards, J. P. (2012). Mechanism of inhibition for N6022, a first-in-class drug targeting S-nitrosoglutathione reductase. *Biochemistry*, *51*(10), 2157–2168. https://doi.org/10.1021/bi201785u

Guralnik, J. M., Eisenstaedt, R. S., Ferrucci, L., Klein, H. G., & Woodman, R. C. (2004). Prevalence of anemia in persons 65 years and older in the United States: Evidence

for a high rate of unexplained anemia. *Blood*, *104*(8), 2263–2268. https://doi.org/10.1182/blood-2004-05-1812

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927–930. https://doi.org/10.1038/s41591-018-0049-z

Hadjur, S., Ung, K., Wadsworth, L., Dimmick, J., Rajcan-Separovic, E., Scott, R. W., Buchwald, M., & Jirik, F. R. (2001). *Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and Cu/Zn superoxide dismutase*. http://ashpublications.org/blood/article-pdf/98/4/1003/1676353/h8160101003.pdf

Hall, T. D., Kim, H., Dabbah, M., Myers, J. A., Crawford, J. C., Morales-Hernandez, A., Caprio, C. E., Sriram, P., Kooienga, E., Derecka, M., Obeng, E. A., Thomas, P. G., & McKinney-Freeman, S. (2022). Murine fetal bone marrow does not support functional hematopoietic stem and progenitor cells until birth. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33092-4

Han, J.-W., Zheng, H.-F., Cui, Y., Sun, L.-D., Ye, D.-Q., Hu, Z., Xu, J.-H., Cai, Z.-M., Huang, W., Zhao, G.-P., Xie, H.-F., Fang, H., Lu, Q.-J., Xu, J.-H., Li, X.-P., Pan, Y.-F., Deng, D.-Q., Zeng, F.-Q., Ye, Z.-Z., … Zhang, X.-J. (2009). Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nature Genetics*, *41*(11), 1234–1237. https://doi.org/10.1038/ng.472

Hanlon, K., Thompson, A., Pantano, L., Hutchinson, J. N., Al-Obeidi, A., Wang, S., Bliss-Moreau, M., Helble, J., Alexe, G., Stegmaier, K., Bauer, D. E., & Croker, B. A. (2019). Single-cell cloning of human T-cell lines reveals clonal variation in cell death responses to chemotherapeutics. *Cancer Genetics*, *237*, 69–77. https://doi.org/10.1016/j.cancergen.2019.06.003

Harada, S., Agarwal, D. P., & Goedde, H. W. (1981). ALDEHYDE DEHYDROGENASE DEFICIENCY AS CAUSE OF FACIAL FLUSHING REACTION TO ALCOHOL IN JAPANESE. *The Lancet*, *318*(8253), 982. https://doi.org/10.1016/S0140-6736(81)91172-7

Harding, S. M., Benci, J. L., Irianto, J., Discher, D. E., Minn, A. J., & Greenberg, R. A. (2017). Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature*, *548*(7668), 466–470. https://doi.org/10.1038/nature23470

Hartman, J. H., Miller, G. P., & Meyer, J. N. (2017). Toxicological implications of mitochondrial localization of CYP2E1. In *Toxicology Research* (Vol. 6, Issue 3, pp. 273–289). Royal Society of Chemistry. https://doi.org/10.1039/c7tx00020k

Hazen, S. L., Hsu, F. F., & Heinecke, J. W. (1998). *Human Neutrophils Employ Myeloperoxidase To Convert R-Amino Acids to a Battery of Reactive Aldehydes: A Pathway for Aldehyde Generation at Sites of Inflammation †*. https://pubs.acs.org/sharingguidelines

Higuchi, S., Matsushita, S., Muramatsu, T., Murayama, M., & Hayashida, M. (1996). Alcohol and Aldehyde Dehydrogenase Genotypes and Drinking Behavior in Japanese. *Alcoholism: Clinical and Experimental Research*, *20*(3), 493–497. https://doi.org/10.1111/j.1530-0277.1996.tb01080.x

Hill, W., Lim, E. L., Weeden, C. E., Lee, C., Augustine, M., Chen, K., Kuan, F. C., Marongiu, F., Evans, E. J., Moore, D. A., Rodrigues, F. S., Pich, O., Bakker, B., Cha, H., Myers, R., van Maldegem, F., Boumelha, J., Veeriah, S., Rowan, A., … Swanton, C. (2023). Lung adenocarcinoma promotion by air pollutants. *Nature*, *616*(7955), 159–167. https://doi.org/10.1038/s41586-023-05874-3

Hira, A., Yabe, H., Yoshida, K., Okuno, Y., Shiraishi, Y., Chiba, K., Tanaka, H., Miyano, S., Nakamura, J., Kojima, S., Ogawa, S., Matsuo, K., Takata, M., & Yabe, M. (2013). *Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients*. https://doi.org/10.1182/blood-2013-06

Ho, J. E., Hwang, S. J., Wollert, K. C., Larson, M. G., Cheng, S., Kempf, T., Vasan, R. S., Januzzi, J. L., Wang, T. J., & Fox, C. S. (2013). Biomarkers of cardiovascular stress and incident chronic kidney disease. *Clinical Chemistry*, *59*(11), 1613–1620. https://doi.org/10.1373/clinchem.2013.205716

Ho, Y. H., del Toro, R., Rivera-Torres, J., Rak, J., Korn, C., García-García, A., Macías, D., González-Gómez, C., del Monte, A., Wittner, M., Waller, A. K., Foster, H. R., López-Otín, C., Johnson, R. S., Nerlov, C., Ghevaert, C., Vainchenker, W., Louache, F., Andrés, V., & Méndez-Ferrer, S. (2019). Remodeling of Bone Marrow Hematopoietic Stem Cell Niches Promotes Myeloid Cell Expansion during Premature or Physiological Aging. *Cell Stem Cell*, *25*(3), 407-418.e6. https://doi.org/10.1016/j.stem.2019.06.007

Hodskinson, M. R., Bolner, A., Sato, K., Kamimae-Lanning, A. N., Rooijers, K., Witte, M., Mahesh, M., Silhan, J., Petek, M., Williams, D. M., Kind, J., Chin, J. W., Patel, K. J., & Knipscheer, P. (2020). Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms. *Nature*, *579*(7800), 603–608. https://doi.org/10.1038/s41586-020-2059-5

Hoffmann, T., Meyer, R. J., Sorrell, M. F., & Tuma, D. J. (1993). Reaction of Acetaldehyde with Proteins: Formation of Stable Fluorescent Adducts. *Alcoholism: Clinical and Experimental Research*, *17*(1), 69–74. https://doi.org/10.1111/j.1530-0277.1993.tb00728.x

Hsu, J. I., Dayaram, T., Tovy, A., De Braekeleer, E., Jeong, M., Wang, F., Zhang, J., Heffernan, T. P., Gera, S., Kovacs, J. J., Marszalek, J. R., Bristow, C., Yan, Y., Garcia-Manero, G., Kantarjian, H., Vassiliou, G., Futreal, P. A., Donehower, L. A., Takahashi, K., & Goodell, M. A. (2018). PPM1D Mutations Drive Clonal Hematopoiesis in Response to Cytotoxic Chemotherapy. *Cell Stem Cell*, *23*(5), 700-713.e6. https://doi.org/10.1016/j.stem.2018.10.004

Huang, R., & Zhou, P. K. (2021). DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. In *Signal Transduction and Targeted Therapy* (Vol. 6, Issue 1). Springer Nature. https://doi.org/10.1038/s41392-021-00648-7

Im, P. K., Yang, L., Kartsonaki, C., Chen, Y., Guo, Y., Du, H., Lin, K., Kerosi, R., Hacker, A., Liu, J., Yu, C., Lv, J., Walters, R. G., Li, L., Chen, Z., & Millwood, I. Y. (2022). Alcohol metabolism genes and risks of site-specific cancers in Chinese adults: An 11-year prospective study. *International Journal of Cancer*, *150*(10), 1627–1639. https://doi.org/10.1002/ijc.33917

Imtiaz, S., Shield, K. D., Roerecke, M., Samokhvalov, A. V., Lönnroth, K., & Rehm, J. (2017). Alcohol consumption as a risk factor for tuberculosis: Meta-analyses and burden of disease. *European Respiratory Journal*, *50*(1). https://doi.org/10.1183/13993003.00216-2017

Israel, Y., Hurwitzt, E., Niemela, O., & Arnont, R. (1986). Monoclonal and polyclonal antibodies against acetaldehyde-containing epitopes in acetaldehyde-protein adducts (alcoholism/immunoglobulins/ethanol metabolites). In *Medical Sciences* (Vol. 83). https://www.pnas.org

Iwanami, S., Sato, T., Haeno, H., Xu, L., Imamura, K., Ooehara, J., Lan, X., Nakauchi, H., Iwami, S., & Yamamoto, R. (2025). Stable platelet production via the bypass

pathway explains long-term hematopoietic stem cell reconstitution. *IScience*, *28*(6). https://doi.org/10.1016/j.isci.2025.112547

Iyama, T., & Wilson, D. M. (2013). DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair*, *12*(8), 620–636. https://doi.org/10.1016/j.dnarep.2013.04.015

Jaiswal, S., & Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science*, *366*(6465). https://doi.org/10.1126/science.aan4673

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz, G., … Ebert, B. L. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine*, *371*(26), 2488–2498. https://doi.org/10.1056/nejmoa1408617

Jakobsen, N. A., Turkalj, S., Zeng, A. G. X., Stoilova, B., Metzner, M., Rahmig, S., Nagree, M. S., Shah, S., Moore, R., Usukhbayar, B., Angulo Salazar, M., Gafencu, G. A., Kennedy, A., Newman, S., Kendrick, B. J. L., Taylor, A. H., Afinowi-Luitz, R., Gundle, R., Watkins, B., … Vyas, P. (2024). Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging. *Cell Stem Cell*, *31*(8), 1127-1144.e17. https://doi.org/10.1016/j.stem.2024.05.010

Joseph, C., Quach, J. M., Walkley, C. R., Lane, S. W., Lo Celso, C., & Purton, L. E. (2013). Deciphering Hematopoietic Stem Cells in Their Niches: A Critical Appraisal of Genetic Models, Lineage Tracing, and Imaging Strategies. *Cell Stem Cell*, *13*(5), 520–533. https://doi.org/10.1016/j.stem.2013.10.010

Kamimae-Lanning, A. N., Goloviznina, N. A., & Kurre, P. (2013). *Fetal origins of hematopoietic failure in a murine model of Fanconi anemia Key Points*. https://doi.org/10.1182/blood-2012-06

Kamps, J. J. A. G., Hopkinson, R. J., Schofield, C. J., & Claridge, T. D. W. (2019). How formaldehyde reacts with amino acids. *Communications Chemistry*, *2*(1), 126. https://doi.org/10.1038/s42004-019-0224-2

Kapadia, C. D., Williams, N., Dawson, K. J., Watson, C., Yousefzadeh, M. J., Le, D., Nyamondo, K., Kodavali, S., Cagan, A., Waldvogel, S., Zhang, X., De La Fuente, J., Leongamornlert, D., Mitchell, E., Florez, M. A., Sosnowski, K., Aguilar, R., Martell, A., Guzman, A., … Nangalia, J. (2025). Clonal dynamics and somatic evolution of haematopoiesis in mouse. *Nature*, *641*(8063), 681–689. https://doi.org/10.1038/s41586-025-08625-8

Karikkineth, A. C., Scheibye-Knudsen, M., Fivenson, E., Croteau, D. L., & Bohr, V. A. (2017). Cockayne syndrome: Clinical features, model systems and pathways. In *Ageing Research Reviews* (Vol. 33, pp. 3–17). Elsevier Ireland Ltd. https://doi.org/10.1016/j.arr.2016.08.002

Karlsson, E. K., Kwiatkowski, D. P., & Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. In *Nature Reviews Genetics* (Vol. 15, Issue 6, pp. 379–393). Nature Publishing Group. https://doi.org/10.1038/nrg3734

Kast, J., & Klockenbusch, C. (2010). Optimization of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin β 1. *Journal of Biomedicine and Biotechnology*, *2010*. https://doi.org/10.1155/2010/927585

Kath, J., Du, W., Pruene, A., Braun, T., Thommandru, B., Turk, R., Sturgeon, M. L., Kurgan, G. L., Amini, L., Stein, M., Zittel, T., Martini, S., Ostendorf, L., Wilhelm, A., Akyüz, L., Rehm, A., Höpken, U. E., Pruß, A., Künkele, A., … Wagner, D. L. (2022). Pharmacological interventions enhance virus-free generation of TRAC-replaced CAR

T cells. *Molecular Therapy Methods and Clinical Development*, *25*, 311–330. https://doi.org/10.1016/j.omtm.2022.03.018

Kawanishi, S., Hiraku, Y., Pinlaor, S., & Ma, N. (2006). Oxidative and nitrative DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis. In *Biological chemistry* (Vol. 387, Issue 4, pp. 365–372). https://doi.org/10.1515/BC.2006.049

Kazak, L., Reyes, A., & Holt, I. J. (2012). Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nature Reviews Molecular Cell Biology*, *13*(10), 659–671. https://doi.org/10.1038/nrm3439

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellåker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., … Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, *477*(7364), 289–294. https://doi.org/10.1038/nature10413

Kerenyi, M. A., Shao, Z., Hsu, Y. J., Guo, G., Luc, S., O'Brien, K., Fujiwara, Y., Peng, C., Nguyen, M., & Orkin, S. H. (2013). Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. *ELife*, *2013*(2). https://doi.org/10.7554/eLife.00633

Khatoon, F., Moinuddin, Alam, K., & Ali, A. (2012). Physicochemical and immunological studies on 4-hydroxynonenal modified HSA: Implications of protein damage by lipid peroxidation products in the etiopathogenesis of SLE. *Human Immunology*, *73*(11), 1132–1139. https://doi.org/10.1016/j.humimm.2012.08.011

Kim, J., Kim, S. H., Kang, H., Lee, S., Park, S. Y., Cho, Y., Lim, Y. M., Ahn, J. W., Kim, Y. H., Chung, S., Choi, C. S., Jang, Y. J., Park, H. S., Heo, Y., Kim, K. H., & Lee, M. S. (2021). TFEB–GDF15 axis protects against obesity and insulin resistance as a lysosomal stress response. *Nature Metabolism*, *3*(3), 410–427. https://doi.org/10.1038/s42255-021-00368-w

Kirkwood, T. B. L. (2005). Understanding the odd science of aging. In *Cell* (Vol. 120, Issue 4, pp. 437–447). Elsevier B.V. https://doi.org/10.1016/j.cell.2005.01.027

Klose, R. J., Kallin, E. M., & Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics*, *7*(9), 715–727. https://doi.org/10.1038/nrg1945

Knockaert, L., Fromenty, B., & Robin, M. (2011). Mechanisms of mitochondrial targeting of cytochrome P450 2E1: physiopathological role in liver injury and obesity. *The FEBS Journal*, *278*(22), 4252–4260. https://doi.org/10.1111/j.1742-4658.2011.08357.x

Knudson, A. G. (1971). *Mutation and Cancer: Statistical Study of Retinoblastoma* (Vol. 68, Issue 4). https://www.pnas.org

Kobayashi, M., Wei, H., Yamanashi, T., Azevedo Portilho, N., Cornelius, S., Valiente, N., Nishida, C., Cheng, H., Latorre, A., Zheng, W. J., Kang, J., Seita, J., Shih, D. J., Wu, J. Q., & Yoshimoto, M. (2023). HSC-independent definitive hematopoiesis persists into adult life. *Cell Reports*, *42*(3). https://doi.org/10.1016/j.celrep.2023.112239

Koganebuchi, K., Haneji, K., Toma, T., Joh, K., Soejima, H., Fujimoto, K., Ishida, H., Ogawa, M., Hanihara, T., Harada, S., Kawamura, S., & Oota, H. (2017). The allele frequency of *ALDH2*Glu504Lys* and *ADH1B*Arg47His* for the Ryukyu islanders and their history of expansion among East Asians. *American Journal of Human Biology*, *29*(2). https://doi.org/10.1002/ajhb.22933

Körber, V., Jakobsen, N. A., Ansari-Pour, N., Moore, R., Claudino, N., Metzner, M., Thielecke, E., Esau, F., Usukhbayar, B., Salazar, M. A., Newman, S., Kendrick, B. J. L., Taylor, A. H., Afinowi-Luitz, R., Gundle, R., Watkins, B., Wheway, K., Beazley, D.,

Dakin, S. G., … Höfer, T. (2025). Detecting and quantifying clonal selection in somatic stem cells. *Nature Genetics*, *57*(7), 1718–1729. https://doi.org/10.1038/s41588-025-02217-y

Kralovics, R., Passamonti, F., Buser, A. S., Teo, S.-S., Tiedt, R., Passweg, J. R., Tichelli, A., Cazzola, M., & Skoda, R. C. (2005). A Gain-of-Function Mutation of *JAK2* in Myeloproliferative Disorders. *New England Journal of Medicine*, *352*(17), 1779–1790. https://doi.org/10.1056/NEJMoa051113

Kwak, K., Akkaya, M., & Pierce, S. K. (2019). B cell signaling in context. In *Nature Immunology* (Vol. 20, Issue 8, pp. 963–969). Nature Publishing Group. https://doi.org/10.1038/s41590-019-0427-9

Lagergård, T., Lundqvist, A., Wising, C., Gabrielsson, V., & Ahlman, K. (2007). Formaldehyde treatment increases the immunogenicity and decreases the toxicity of Haemophilus ducreyi cytolethal distending toxin. *Vaccine*, *25*(18), 3606–3614. https://doi.org/10.1016/j.vaccine.2007.01.061

Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J., & Patel, K. J. (2011). Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature*, *475*(7354), 53–59. https://doi.org/10.1038/nature10192

Larson, H. N., Weiner, H., & Hurley, T. D. (2005). Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase “Asian” variant. *Journal of Biological Chemistry*, *280*(34), 30550–30556. https://doi.org/10.1074/jbc.M502345200

Lau, T. Y., & Poon, R. Y. C. (2023). Whole-Genome Duplication and Genome Instability in Cancer Cells: Double the Trouble. *International Journal of Molecular Sciences*, *24*(4), 3733. https://doi.org/10.3390/ijms24043733

Lee, J., Lee, J., Jeon, S., Lee, J., Jang, I., Yang, J. O., Park, S., Lee, B., Choi, J., Choi, B. O., Gee, H. Y., Oh, J., Jang, I. J., Lee, S., Baek, D., Koh, Y., Yoon, S. S., Kim, Y. J., Chae, J. H., … Choi, M. (2022). A database of 5305 healthy Korean individuals reveals genetic and clinical implications for an East Asian population. *Experimental and Molecular Medicine*, *54*(11), 1862–1871. https://doi.org/10.1038/s12276-022-00871-4

Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R. J., Sanders, M. A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., Robinson, P., Coorens, T. H. H., O’Neill, L., Alder, C., Wang, J., Fitzgerald, R. C., Zilbauer, M., Coleman, N., Saeb-Parsy, K., … Stratton, M. R. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*, *574*(7779), 532–537. https://doi.org/10.1038/s41586-019-1672-7

Lehrer, R. I. (1969). Antifungal Effects of Peroxidase Systems. *Journal of Bacteriology*, *99*(2), 361–365. https://doi.org/10.1128/jb.99.2.361-365.1969

Li, D., Zhang, H., & Zhong, Y. (2018). Hepatic GDF15 is regulated by CHOP of the unfolded protein response and alleviates NAFLD progression in obese mice. *Biochemical and Biophysical Research Communications*, *498*(3), 388–394. https://doi.org/10.1016/j.bbrc.2017.08.096

Li, H., Borinskaya, S., Yoshimura, K., Kal’ina, N., Marusin, A., Stepanov, V. A., Qin, Z., Khaliq, S., Lee, M. Y., Yang, Y., Mohyuddin, A., Gurwitz, D., Mehdi, S. Q., Rogaev, E., Jin, L., Yankovsky, N. K., Kidd, J. R., & Kidd, K. K. (2009). Refined geographic distribution of the oriental ALDH2* 504Lys (nee 487Lys) variant. *Annals of Human Genetics*, *73*(3), 335–345. https://doi.org/10.1111/j.1469-1809.2009.00517.x

Li, H., Mukherjee, N., Soundararajan, U., Tárnok, Z., Barta, C., Khaliq, S., Mohyuddin, A., Kajuna, S. L. B., Mehdi, S. Q., Kidd, J. R., & Kidd, K. K. (2007). Geographically separate increases in the frequency of the derived ADH1B*47His allele in eastern

and western Asia. *American Journal of Human Genetics*, *81*(4), 842–846. https://doi.org/10.1086/521201

Li, S., Ali, S., Duan, X., Liu, S., Du, J., Liu, C., Dai, H., Zhou, M., Zhou, L., Yang, L., Chu, P., Li, L., Bhatia, R., Schones, D. E., Wu, X., Xu, H., Hua, Y., Guo, Z., Yang, Y., … Shen, B. (2018). JMJD1B Demethylates H4R3me2s and H3K9me2 to Facilitate Gene Expression for Development of Hematopoietic Stem and Progenitor Cells. *Cell Reports*, *23*(2), 389–403. https://doi.org/10.1016/j.celrep.2018.03.051

Li, X., Li, C., Zhang, W., Wang, Y., Qian, P., & Huang, H. (2023). Inflammation and aging: signaling pathways and intervention therapies. In *Signal Transduction and Targeted Therapy* (Vol. 8, Issue 1). Springer Nature. https://doi.org/10.1038/s41392-023-01502-8

Lin, M. T., & Beal, M. F. (2006). Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*, *443*(7113), 787–795. https://doi.org/10.1038/nature05292

Lin, Y.-P., & Cheng, T.-J. (2002). Why can't Chinese Han drink alcohol? Hepatitis B virus infection and the evolution of acetaldehyde dehydrogenase deficiency. *Medical Hypotheses*, *59*(2), 204–207. https://doi.org/10.1016/S0306-9877(02)00253-0

Lindahl, T. (1993). *Instability and decay of the primary structure of DNA*.

Little, S. F., Ivins, B. E., Webster, W. M., Norris, S. L. W., & Andrews, G. P. (2007). Effect of aluminum hydroxide adjuvant and formaldehyde in the formulation of rPA anthrax vaccine. *Vaccine*, *25*(15), 2771–2777. https://doi.org/10.1016/j.vaccine.2006.12.043

Liu, J., Kumar, S., Heinzel, A., Gao, M., Guo, J., Alvarado, G. F., Reindl-Schwaighofer, R., Krautzberger, A. M., Cippà, P. E., McMahon, J., Oberbauer, R., & McMahon, A. P. (2020). Renoprotective and immunomodulatory effects of GDF15 following AKI invoked by ischemia-reperfusion injury. *Journal of the American Society of Nephrology*, *31*(4), 701–715. https://doi.org/10.1681/ASN.2019090876

Liu, J., Tran, D., Xue, L., Wiley, B. J., Vlasschaert, C., Watson, C. J., MacGregor, H. A. J., Zong, X., Chan, I. C. C., Das, I., Uddin, M. M., Niroula, A., Griffin, G., Ebert, B. L., Mack, T., Pershad, Y., Sharber, B., Berger, M., Zehir, A., … Bolton, K. L. (2025). Germline genetic variation impacts clonal hematopoiesis landscape and progression to malignancy. *Nature Genetics*, *57*(8), 1872–1880. https://doi.org/10.1038/s41588-025-02250-x

Liu, L., Yan, Y., Zeng, M., Zhang, J., Hanes, M. A., Ahearn, G., McMahon, T. J., Dickfeld, T., Marshall, H. E., Que, L. G., & Stamler, J. S. (2004). Essential Roles of S-Nitrosothiols in Vascular Homeostasis and Endotoxic Shock. *Cell*, *116*(4), 617–628. https://doi.org/10.1016/S0092-8674(04)00131-X

Lockhart, S. M., Saudek, V., & O'Rahilly, S. (2020). Gdf15: A hormone conveying somatic distress to the brain. In *Endocrine Reviews* (Vol. 41, Issue 4, pp. 610–642). Endocrine Society. https://doi.org/10.1210/ENDREV/BNAA007

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. In *Cell* (Vol. 153, Issue 6, p. 1194). Elsevier B.V. https://doi.org/10.1016/j.cell.2013.05.039

Lu, K., Collins, L. B., Ru, H., Bermudez, E., & Swenberg, J. A. (2010). Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. *Toxicological Sciences*, *116*(2), 441–451. https://doi.org/10.1093/toxsci/kfq061

Lu, K., Ye, W., Zhou, L., Collins, L. B., Chen, X., Gold, A., Ball, L. M., & Swenberg, J. A. (2010). Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. *Journal of the American Chemical Society*, *132*(10), 3388–3399. https://doi.org/10.1021/ja908282f

Luo, H., Zhang, P., Zhang, W., Zheng, Y., Hao, D., Shi, Y., Niu, Y., Song, T., Li, Y., Zhao, S., Chen, H., Xu, T., & He, S. (2023). Recent positive selection signatures reveal phenotypic evolution in the Han Chinese population. *Science Bulletin*, *68*(20), 2391–2404. https://doi.org/10.1016/j.scib.2023.08.027

Majorek, K. A., Porebski, P. J., Dayal, A., Zimmerman, M. D., Jablonska, K., Stewart, A. J., Chruszcz, M., & Minor, W. (2012). Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Molecular Immunology*, *52*(3–4), 174–182. https://doi.org/10.1016/j.molimm.2012.05.011

Malong, L., Roskosch, J., Hager, C., Fortin, J. P., Schmucki, R., Callow, M. G., Weile, C., Romeo, V., Patsch, C., Martin, S., Costa, M., Modrusan, Z., Villaseñor, R., Koller, E., Haley, B., Spang, A., & Roudnicky, F. (2025). A CRISPR/Cas9 screen reveals proteins at the endosome-Golgi interface that modulate cellular anti-sense oligonucleotide activity. *Nature Communications* , *16*(1). https://doi.org/10.1038/s41467-025-61039-y

Mann, M., Mehta, A., de Boer, C. G., Kowalczyk, M. S., Lee, K., Haldeman, P., Rogel, N., Knecht, A. R., Farouq, D., Regev, A., & Baltimore, D. (2018). Heterogeneous Responses of Hematopoietic Stem Cells to Inflammatory Stimuli Are Altered with Age. *Cell Reports*, *25*(11), 2992-3005.e5. https://doi.org/10.1016/j.celrep.2018.11.056

Mansouri, A., Gattolliat, C. H., & Asselah, T. (2018). Mitochondrial Dysfunction and Signaling in Chronic Liver Diseases. In *Gastroenterology* (Vol. 155, Issue 3, pp. 629–647). W.B. Saunders. https://doi.org/10.1053/j.gastro.2018.06.083

Mao, Z., Bozzella, M., Seluanov, A., & Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair*, *7*(10), 1765–1771. https://doi.org/10.1016/j.dnarep.2008.06.018

Martincorena, I., Fowler, J. C., Wabik, A., Lawson, A. R. J., Abascal, F., Hall, M. W. J., Cagan, A., Murai, K., Mahbubani, K., Stratton, M. R., Fitzgerald, R. C., Handford, P. A., Campbell, P. J., Saeb-Parsy, K., & Jones, P. H. (2018). Somatic mutant clones colonize the human esophagus with age. *Science*, *362*(6417), 911–917. https://doi.org/10.1126/science.aau3879

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., & Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature Biotechnology*, *33*(5), 538–542. https://doi.org/10.1038/nbt.3190

Matsumoto, A., Hara, M., Ashenagar, M. S., Tokiya, M., Sawada, T., Iwasaka, C., Furukawa, T., Kitagawa, K., Miyake, Y., & Hirota, Y. (2022). Variant Allele of ALDH2, rs671, Associates with Attenuated Post-Vaccination Response in Anti-SARS-CoV-2 Spike Protein IgG: A Prospective Study in the Japanese General Population. *Vaccines*, *10*(7), 1035. https://doi.org/10.3390/vaccines10071035

McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. In *Cell Research* (Vol. 18, Issue 1, pp. 148–161). https://doi.org/10.1038/cr.2008.4

Meng, Y., Carrelha, J., Drissen, R., Ren, X., Zhang, B., Gambardella, A., Valletta, S., Thongjuea, S., Jacobsen, S. E., & Nerlov, C. (2023). Epigenetic programming defines haematopoietic stem cell fate restriction. *Nature Cell Biology*, *25*(6), 812–822. https://doi.org/10.1038/s41556-023-01137-5

Meng, Y., & Nerlov, C. (2025). Epigenetic regulation of hematopoietic stem cell fate. In *Trends in Cell Biology* (Vol. 35, Issue 3, pp. 217–229). Elsevier Ltd. https://doi.org/10.1016/j.tcb.2024.08.005

Metz, B., Jiskoot, W., Hennink, W. E., Crommelin, D. J. A., & Kersten, G. F. A. (2003). Physicochemical and immunochemical techniques predict the quality of diphtheria toxoid vaccines. *Vaccine*, *22*(2), 156–167. https://doi.org/10.1016/j.vaccine.2003.08.003

Mitchell, C. A., Verovskaya, E. V., Calero-Nieto, F. J., Olson, O. C., Swann, J. W., Wang, X., Hérault, A., Dellorusso, P. V., Zhang, S. Y., Svendsen, A. F., Pietras, E. M., Bakker, S. T., Ho, T. T., Göttgens, B., & Passegué, E. (2023). Stromal niche inflammation mediated by IL-1 signalling is a targetable driver of haematopoietic ageing. *Nature Cell Biology*, *25*(1), 30–41. https://doi.org/10.1038/s41556-022-01053-0

Mitchell, E., Pham, M. H., Clay, A., Sanghvi, R., Williams, N., Pietsch, S., Hsu, J. I., Jung, H., Vedi, A., Moody, S., Wang, J., Leonganmornlert, D., Spencer Chapman, M., Dunstone, E., Santarsieri, A., Cagan, A., Machado, H. E., Baxter, E. J., Follows, G., … Stratton, M. R. (2025). The long-term effects of chemotherapy on normal blood cells. *Nature Genetics*, *57*(7), 1684–1694. https://doi.org/10.1038/s41588-025-02234-x

Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K. J., Mende, N., Calderbank, E. F., Jung, H., Mitchell, T., Coorens, T. H. H., Spencer, D. H., Machado, H., Lee-Six, H., Davies, M., Hayler, D., Fabre, M. A., Mahbubani, K., Abascal, F., Cagan, A., Vassiliou, G. S., … Campbell, P. J. (2022). Clonal dynamics of haematopoiesis across the human lifespan. *Nature*, *606*(7913), 343–350. https://doi.org/10.1038/s41586-022-04786-y

Miyake, M., Zhang, J., Yasue, A., Hisanaga, S., Tsugawa, K., Sakaue, H., Oyadomari, M., Kiyonari, H., & Oyadomari, S. (2021). Integrated stress response regulates GDF15 secretion from adipocytes, preferentially suppresses appetite for a high-fat diet and improves obesity. *IScience*, *24*(12). https://doi.org/10.1016/j.isci.2021.103448

Mohrin, M., Bourke, E., Alexander, D., Warr, M. R., Barry-Holson, K., Le Beau, M. M., Morrison, C. G., & Passegué, E. (2010). Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell*, *7*(2), 174–185. https://doi.org/10.1016/j.stem.2010.06.014

Montecino-Rodriguez, E., Berent-Maoz, B., & Dorshkind, K. (2013). Causes, consequences, and reversal of immune system aging. In *Journal of Clinical Investigation* (Vol. 123, Issue 3, pp. 958–965). https://doi.org/10.1172/JCI64096

Moody, S., Senkin, S., Islam, S. M. A., Wang, J., Nasrollahzadeh, D., Cortez Cardoso Penha, R., Fitzgerald, S., Bergstrom, E. N., Atkins, J., He, Y., Khandekar, A., Smith-Byrne, K., Carreira, C., Gaborieau, V., Latimer, C., Thomas, E., Abnizova, I., Bucciarelli, P. E., Jones, D., … Stratton, M. R. (2021). Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. *Nature Genetics*, *53*(11), 1553–1563. https://doi.org/10.1038/s41588-021-00928-6

Morrison, S. J., Hemmati, H. D., Wandycz, A. M., & Weissmant, I. L. (1995). The purification and characterization of fetal liver hematopoietic stem cells. In *Immunology* (Vol. 92). https://www.pnas.org

Mulderrig, L., Garaycoechea, J. I., Tuong, Z. K., Millington, C. L., Dingler, F. A., Ferdinand, J. R., Gaul, L., Tadross, J. A., Arends, M. J., O'Rahilly, S., Crossan, G. P., Clatworthy, M. R., & Patel, K. J. (2021). Aldehyde-driven transcriptional stress triggers an anorexic DNA damage response. *Nature*, *600*(7887), 158–163. https://doi.org/10.1038/s41586-021-04133-7

N I Niemela, O. N., Klajner, F., Orrego, H., Vidins, E., Blendis, L., & Israel, Y. (1987). *Antibodies against Acetaldehyde-Modified Protein Epitopes in Human Alcoholics* (Vol. 7, Issue 6).

Nakabayashi, K., Tajima, A., Yamamoto, K., Takahashi, A., Hata, K., Takashima, Y., Koyanagi, M., Nakaoka, H., Akamizu, T., Ishikawa, N., Kubota, S., Maeda, S., Tsunoda, T., Kubo, M., Kamatani, N., Nakamura, Y., Sasazuki, T., & Shirasawa, S. (2011). Identification of independent risk loci for Graves' disease within the MHC in the Japanese population. *Journal of Human Genetics*, *56*(11), 772–778. https://doi.org/10.1038/jhg.2011.99

Neville, M. D., Lawson, A. R., Sanghvi, R., Abascal, F., Pham, M. H., Cagan, A., Nicola, P. A., Bayzetinova, T., Baez-Ortega, A., Roberts, K., Lensing, S. V., Widaa, S., Alcantara, R. E., García, M. P., Wadge, S., Stratton, M. R., Campbell, P. J., Small, K., Martincorena, I., … Rahbari, R. (2024). *Sperm sequencing reveals extensive positive selection in the male germline*. https://doi.org/10.1101/2024.10.30.24316414

Ng, S. W. K., Rouhani, F. J., Brunner, S. F., Brzozowska, N., Aitken, S. J., Yang, M., Abascal, F., Moore, L., Nikitopoulou, E., Chappell, L., Leongamornlert, D., Ivovic, A., Robinson, P., Butler, T., Sanders, M. A., Williams, N., Coorens, T. H. H., Teague, J., Raine, K., … Campbell, P. J. (2021). Convergent somatic mutations in metabolism genes in chronic liver disease. *Nature*, *598*(7881), 473–478. https://doi.org/10.1038/s41586-021-03974-6

Nguyen, T. T., Wei, S., Nguyen, T. H., Jo, Y., Zhang, Y., Park, W., Gariani, K., Oh, C. M., Kim, H. H., Ha, K. T., Park, K. S., Park, R., Lee, I. K., Shong, M., Houtkooper, R. H., & Ryu, D. (2023). Mitochondria-associated programmed cell death as a therapeutic target for age-related disease. In *Experimental and Molecular Medicine* (Vol. 55, Issue 8, pp. 1595–1619). Springer Nature. https://doi.org/10.1038/s12276-023-01046-5

Niedernhofer, L. J., Garinis, G. A., Raams, A., Lalai, A. S., Robinson, A. R., Appeldoorn, E., Odijk, H., Oostendorp, R., Ahmad, A., Van Leeuwen, W., Theil, A. F., Vermeulen, W., Van Der Horst, G. T. J., Meinecke, P., Kleijer, W. J., Vijg, J., Jaspers, N. G. J., & Hoeijmakers, J. H. J. (2006). A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature*, *444*(7122), 1038–1043. https://doi.org/10.1038/nature05456

Niedernhofer, L. J., Gurkar, A. U., Wang, Y., Vijg, J., Hoeijmakers, J. H. J., & Robbins, P. D. (2018). *Nuclear Genomic Instability and Aging*. *112*, 12. https://doi.org/10.1146/annurev-biochem

Nijnik, A., Woodbine, L., Marchetti, C., Dawson, S., Lambe, T., Liu, C., Rodrigues, N. P., Crockford, T. L., Cabuy, E., Vindigni, A., Enver, T., Bell, J. I., Slijepcevic, P., Goodnow, C. C., Jeggo, P. A., & Cornall, R. J. (2007). DNA repair is limiting for haematopoietic stem cells during ageing. *Nature*, *447*(7145), 686–690. https://doi.org/10.1038/nature05875

Nishimura, T., Kakiuchi, N., Yoshida, K., Sakurai, T., Kataoka, T. R., Kondoh, E., Chigusa, Y., Kawai, M., Sawada, M., Inoue, T., Takeuchi, Y., Maeda, H., Baba, S., Shiozawa, Y., Saiki, R., Nakagawa, M. M., Nannya, Y., Ochi, Y., Hirano, T., … Ogawa, S. (2023). Evolutionary histories of breast cancer and related clones. *Nature*, *620*(7974), 607–614. https://doi.org/10.1038/s41586-023-06333-9

Nygard, A. B., Jørgensen, C. B., Cirera, S., & Fredholm, M. (2007). Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. *BMC Molecular Biology*, *8*. https://doi.org/10.1186/1471-2199-8-67

O'Brien, P., Siraki, A., & Shangari, N. (2005). Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. In *Critical Reviews in*

Toxicology (Vol. 35, Issue 7, pp. 609–662). https://doi.org/10.1080/10408440591002183

O'Connell, K. E., Mikkola, A. M., Stepanek, A. M., Vernet, A., Hall, C. D., Sun, C. C., Yildirim, E., Staropoli, J. F., Lee, J. T., & Brown, D. E. (2015). Practical murine hematopathology: a comparative review and implications for research. *Comparative Medicine*, *65*(2), 96–113.

O'Hagan, D. T., Ott, G. S., De Gregorio, E., & Seubert, A. (2012). The mechanism of action of MF59 - An innately attractive adjuvant formulation. In *Vaccine* (Vol. 30, Issue 29, pp. 4341–4348). https://doi.org/10.1016/j.vaccine.2011.09.061

Oka, Y., Hamada, M., Nakazawa, Y., Muramatsu, H., Okuno, Y., Higasa, K., Shimada, M., Takeshima, H., Hanada, K., Hirano, T., Kawakita, T., Sakaguchi, H., Ichimura, T., Ozono, S., Yuge, K., Watanabe, Y., Kotani, Y., Yamane, M., Kasugai, Y., … Ogi, T. (2020). Digenic mutations in ALDH2 and ADH5 impair formaldehyde clearance and cause a multisystem disorder, AMeD syndrome. *Science Advances*, *6*(51). https://doi.org/10.1126/sciadv.abd7197

Oka, Y., Nakazawa, Y., Shimada, M., & Ogi, T. (2024). Endogenous aldehyde-induced DNA–protein crosslinks are resolved by transcription-coupled repair. *Nature Cell Biology*, *26*(5), 784–796. https://doi.org/10.1038/s41556-024-01401-2

Okada, Y., Momozawa, Y., Sakaue, S., Kanai, M., Ishigaki, K., Akiyama, M., Kishikawa, T., Arai, Y., Sasaki, T., Kosaki, K., Suematsu, M., Matsuda, K., Yamamoto, K., Kubo, M., Hirose, N., & Kamatani, Y. (2018). Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03274-0

Olivieri, M., Cho, T., Álvarez-Quilón, A., Li, K., Schellenberg, M. J., Zimmermann, M., Hustedt, N., Rossi, S. E., Adam, S., Melo, H., Heijink, A. M., Sastre-Moreno, G., Moatti, N., Szilard, R. K., McEwan, A., Ling, A. K., Serrano-Benitez, A., Ubhi, T., Feng, S., … Durocher, D. (2020). A Genetic Map of the Response to DNA Damage in Human Cells. *Cell*, *182*(2), 481-496.e21. https://doi.org/10.1016/j.cell.2020.05.040

Oota, H., Pakstis, A. J., Bonne-Tamir, B., Goldman, D., Grigorenko, E., Kajuna, S. L. B., Karoma, N. J., Kungulilo, S., Lu, R. B., Odunsi, K., Okonofua, F., Zhukova, O. V., Kidd, J. R., & Kidd, K. K. (2004). The evolution and population genetics of the ALDH2 locus: Random genetic drift, selection, and low levels of recombination. *Annals of Human Genetics*, *68*(2), 93–109. https://doi.org/10.1046/j.1529-8817.2003.00060.x

Osada, M., Park, H. L., Park, M. J., Liu, J. W., Wu, G., Trink, B., & Sidransky, D. (2007). A p53-type response element in the GDF15 promoter confers high specificity for p53 activation. *Biochemical and Biophysical Research Communications*, *354*(4), 913–918. https://doi.org/10.1016/j.bbrc.2007.01.089

Oshima, J., Sidorova, J. M., & Monnat, R. J. (2017). Werner syndrome: Clinical features, pathogenesis and potential therapeutic interventions. In *Ageing Research Reviews* (Vol. 33, pp. 105–114). Elsevier Ireland Ltd. https://doi.org/10.1016/j.arr.2016.03.002

Osman, A., Lindén, M., Österlund, T., Vannas, C., Andersson, L., Escobar, M., Ståhlberg, A., & Åman, P. (2023). Identification of genomic binding sites and direct target genes for the transcription factor DDIT3/CHOP. *Experimental Cell Research*, *422*(1). https://doi.org/10.1016/j.yexcr.2022.113418

Ou, H.-L., Bj¨, B., & Schumacher, B. (2018). *DNA damage responses and p53 in the aging process* (Vol. 131, Issue 5). http://ashpublications.org/blood/article-pdf/131/5/488/1465962/blood746396.pdf

Pal, A., & Kundu, R. (2020). Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy. In *Frontiers in Microbiology* (Vol. 10). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2019.03116

Panier, S., & Boulton, S. J. (2014). Double-strand break repair: 53BP1 comes into focus. In *Nature Reviews Molecular Cell Biology* (Vol. 15, Issue 1, pp. 7–18). https://doi.org/10.1038/nrm3719

Park, S. K., Park, C. S., Lee, H. S., Park, K. S., Park, B. L., Cheong, H. S., & Shin, H. D. (2014). Functional polymorphism in aldehyde dehydrogenase-2 gene associated with risk of tuberculosis. *BMC Medical Genetics*, *15*(1). https://doi.org/10.1186/1471-2350-15-40

Patel, S., Alvarez-Guaita, A., Melvin, A., Rimmington, D., Dattilo, A., Miedzybrodzka, E. L., Cimino, I., Maurin, A. C., Roberts, G. P., Meek, C. L., Virtue, S., Sparks, L. M., Parsons, S. A., Redman, L. M., Bray, G. A., Liou, A. P., Woods, R. M., Parry, S. A., Jeppesen, P. B., … O'Rahilly, S. (2019). GDF15 Provides an Endocrine Signal of Nutritional Stress in Mice and Humans. *Cell Metabolism*, *29*(3), 707-718.e8. https://doi.org/10.1016/j.cmet.2018.12.016

Patel, S. H., Christodoulou, C., Weinreb, C., Yu, Q., da Rocha, E. L., Pepe-Mooney, B. J., Bowling, S., Li, L., Osorio, F. G., Daley, G. Q., & Camargo, F. D. (2022). Lifelong multilineage contribution by embryonic-born blood progenitors. *Nature*, *606*(7915), 747–753. https://doi.org/10.1038/s41586-022-04804-z

Paus, D., Tri, G. P., Chan, T. D., Gardam, S., Basten, A., & Brink, R. (2006). Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. *Journal of Experimental Medicine*, *203*(4), 1081–1091. https://doi.org/10.1084/jem.20060087

Pei, W., Feyerabend, T. B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., Chen, W., Sauer, S., Wolf, S., Höfer, T., & Rodewald, H. R. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature*, *548*(7668), 456–460. https://doi.org/10.1038/nature23653

Pei, W., Shang, F., Wang, X., Fanti, A. K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., Feyerabend, T. B., Höfer, T., & Rodewald, H. R. (2020). Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding. *Cell Stem Cell*, *27*(3), 383-395.e8. https://doi.org/10.1016/j.stem.2020.07.018

Peng, G. S., Chen, Y. C., Wang, M. F., Lai, C. L., & Yin, S. J. (2014). ALDH2∗2 but not ADH1B∗2 is a causative variant gene allele for Asian alcohol flushing after a low-dose challenge: Correlation of the pharmacokinetic and pharmacodynamic findings. *Pharmacogenetics and Genomics*, *24*(12), 607–617. https://doi.org/10.1097/FPC.0000000000000096

Peng, G.-S., & Yin, S.-J. (2009). Effect of the allelic variants of aldehyde dehydrogenase ALDH2*2 and alcohol dehydrogenase ADH1B*2 on blood acetaldehyde concentrations. *Human Genomics*, *3*(2), 121. https://doi.org/10.1186/1479-7364-3-2-121

Pietras, E. M., Reynaud, D., Kang, Y. A., Carlin, D., Calero-Nieto, F. J., Leavitt, A. D., Stuart, J. A., Göttgens, B., & Passegué, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell*, *17*(1), 35–46. https://doi.org/10.1016/j.stem.2015.05.003

Pilzecker, B., Buoninfante, O. A., Van Den Berk, P., Lancini, C., Song, J. Y., Citterio, E., & Jacobs, H. (2017). DNA damage tolerance in hematopoietic stem and progenitor

cells in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(33), E6875–E6883. https://doi.org/10.1073/pnas.1706508114

Pinho, S., & Frenette, P. S. (2019). Haematopoietic stem cell activity and interactions with the niche. In *Nature Reviews Molecular Cell Biology* (Vol. 20, Issue 5, pp. 303–320). Nature Publishing Group. https://doi.org/10.1038/s41580-019-0103-9

Pishesha, N., Harmand, T. J., & Ploegh, H. L. (2022). A guide to antigen processing and presentation. *Nature Reviews Immunology*, *22*(12), 751–764. https://doi.org/10.1038/s41577-022-00707-2

Pontel, L. B., Rosado, I. V., Burgos-Barragan, G., Garaycoechea, J. I., Yu, R., Arends, M. J., Chandrasekaran, G., Broecker, V., Wei, W., Liu, L., Swenberg, J. A., Crossan, G. P., & Patel, K. J. (2015). Endogenous Formaldehyde Is a Hematopoietic Stem Cell Genotoxin and Metabolic Carcinogen. *Molecular Cell*, *60*(1), 177–188. https://doi.org/10.1016/j.molcel.2015.08.020

Porter, D. H., Cook, R. J., & Wagner2, C. (1985). Enzymatic Properties of Dimethylglycine Dehydrogenase and Sarcosine Dehydrogenase from Rat Liver'. In *ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS* (Vol. 243, Issue 2).

Reilly, S. K., Gosai, S. J., Gutierrez, A., Mackay-Smith, A., Ulirsch, J. C., Kanai, M., Mouri, K., Berenzy, D., Kales, S., Butler, G. M., Gladden-Young, A., Bhuiyan, R. M., Stitzel, M. L., Finucane, H. K., Sabeti, P. C., & Tewhey, R. (2021). Direct characterization of cis-regulatory elements and functional dissection of complex genetic associations using HCR–FlowFISH. *Nature Genetics*, *53*(8), 1166–1176. https://doi.org/10.1038/s41588-021-00900-4

Reingruber, H., & Pontel, L. B. (2018). Formaldehyde metabolism and its impact on human health. *Current Opinion in Toxicology*, *9*, 28–34. https://doi.org/10.1016/j.cotox.2018.07.001

Robert, F., Barbeau, M., Éthier, S., Dostie, J., & Pelletier, J. (2015). Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Medicine*, *7*(1). https://doi.org/10.1186/s13073-015-0215-6

Roberts, M. A., Deol, K. K., Mathiowetz, A. J., Lange, M., Leto, D. E., Stevenson, J., Hashemi, S. H., Morgens, D. W., Easter, E., Heydari, K., Nalls, M. A., Bassik, M. C., Kampmann, M., Kopito, R. R., Faghri, F., & Olzmann, J. A. (2023). Parallel CRISPR-Cas9 screens identify mechanisms of PLIN2 and lipid droplet regulation. *Developmental Cell*, *58*(18), 1782-1800.e10. https://doi.org/10.1016/j.devcel.2023.07.001

Rodier, F., Coppé, J. P., Patil, C. K., Hoeijmakers, W. A. M., Muñoz, D. P., Raza, S. R., Freund, A., Campeau, E., Davalos, A. R., & Campisi, J. (2009). Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. *Nature Cell Biology*, *11*(8), 973–979. https://doi.org/10.1038/ncb1909

Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., & Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature*, *553*(7687), 212–216. https://doi.org/10.1038/nature25168

Roost, H.-P., Bachmann, M. F., Haag, A., Kalinke, U., Pliskat, V., Hengartner, H., & Zinkernagel, R. M. (1995). *Early high-affinity neutralizing anti-viral IgG responses without further overall improvements of affinity* (Vol. 92). https://www.pnas.org

Ross, J. B., Myers, L. M., Noh, J. J., Collins, M. M., Carmody, A. B., Messer, R. J., Dhuey, E., Hasenkrug, K. J., & Weissman, I. L. (2024). Depleting myeloid-biased haematopoietic stem cells rejuvenates aged immunity. *Nature*, *628*(8006), 162–170. https://doi.org/10.1038/s41586-024-07238-x

Rossi, D. J., Bryder, D., Seita, J., Nussenzweig, A., Hoeijmakers, J., & Weissman, I. L. (2007). Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature*, *447*(7145), 725–729. https://doi.org/10.1038/nature05862

Rossi, D. J., Bryder, D., Zahn, J. M., Ahlenius, H., Sonu, R., Wagers, A. J., & Weissman, I. L. (2005). *Cell intrinsic alterations underlie hematopoietic stem cell aging*. www.pnas.orgcgidoi10.1073pnas.0503280102

Rossi, D. J., Seita, J., Czechowicz, A., Bhattacharya, D., Bryder, D., & Weissman, I. L. (2007). Hematopoietic stem cell quiescence attenuates DNA damage response and permits DNA damage accumulation during aging. In *Cell Cycle* (Vol. 6, Issue 19, pp. 2371–2376). Taylor and Francis Inc. https://doi.org/10.4161/cc.6.19.4759

Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., Ishigaki, K., Suzuki, A., Suzuki, K., Obara, W., Yamaji, K., Takahashi, K., Asai, S., Takahashi, Y., Suzuki, T., … Okada, Y. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics*, *53*(10), 1415–1424. https://doi.org/10.1038/s41588-021-00931-x

Sanjuan-Pla, A., Macaulay, I. C., Jensen, C. T., Woll, P. S., Luis, T. C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Jones, T. B., Chowdhury, O., Stenson, L., Lutteropp, M., Green, J. C. A., Facchini, R., Boukarabila, H., Grover, A., Gambardella, A., Thongjuea, S., … Jacobsen, S. E. W. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, *502*(7470), 232–236. https://doi.org/10.1038/nature12495

Savelyeva, I., & Dobbelstein, M. (2011). Infection with E1B-mutant adenovirus stabilizes p53 but blocks p53 acetylation and activity through E1A. *Oncogene*, *30*(7), 865–875. https://doi.org/10.1038/onc.2010.461

Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J., & Howley, P. M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell*, *63*(6), 1129–1136. https://doi.org/10.1016/0092-8674(90)90409-8

Schumacher, B., Pothof, J., Vijg, J., & Hoeijmakers, J. H. J. (2021). The central role of DNA damage in the ageing process. In *Nature* (Vol. 592, Issue 7856, pp. 695–703). Nature Research. https://doi.org/10.1038/s41586-021-03307-7

Sebert, M., Gachet, S., Leblanc, T., Rousseau, A., Bluteau, O., Kim, R., Ben Abdelali, R., Sicre de Fontbrune, F., Maillard, L., Fedronie, C., Murigneux, V., Bellenger, L., Naouar, N., Quentin, S., Hernandez, L., Vasquez, N., Da Costa, M., Prata, P. H., Larcher, L., … Soulier, J. (2023). Clonal hematopoiesis driven by chromosome 1q/MDM4 trisomy defines a canonical route toward leukemia in Fanconi anemia. *Cell Stem Cell*, *30*(2), 153-170.e9. https://doi.org/10.1016/j.stem.2023.01.006

Sedelnikova, O. A., Horikawa, I., Redon, C., Nakamura, A., Zimonjic, D. B., Popescu, N. C., & Bonner, W. M. (2008). Delayed kinetics of DNA double-strand break processing in normal and pathological aging. *Aging Cell*, *7*(1), 89–100. https://doi.org/10.1111/j.1474-9726.2007.00354.x

Segal, A. W. (2005). How neutrophils kill microbes. In *Annual Review of Immunology* (Vol. 23, pp. 197–223). https://doi.org/10.1146/annurev.immunol.23.021704.115653

Seltzer, M. H., Bastidas, J. A., Cooper, D. M., Engler, P., Slocum, B., & Fletcher, H. S. (1979). Instant Nutritional Assessment. *Journal of Parenteral and Enteral Nutrition*, *3*(3), 157–159. https://doi.org/10.1177/014860717900300309

Sender, R., & Milo, R. (2021). The distribution of cellular turnover in the human body. *Nature Medicine*, *27*(1), 45–48. https://doi.org/10.1038/s41591-020-01182-9

Shang, W., Jiang, Y., Boettcher, M., Ding, K., Mollenauer, M., Liu, Z., Wen, X., Liu, C., Hao, P., Zhao, S., McManus, M. T., Wei, L., Weiss, A., & Wang, H. (2018). Genome-

wide CRISPR screen identifies FAM49B as a key regulator of actin dynamics and T cell activation. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), E4051–E4060. https://doi.org/10.1073/pnas.1801340115

Shih, T. A. Y., Meffre, E., Roederer, M., & Nussenzweig, M. C. (2002). Role of BCR affinity in T cell-dependent antibody responses in vivo. *Nature Immunology*, *3*(6), 570–575. https://doi.org/10.1038/ni803

Shimamura, A., & Alter, B. P. (2010). Pathophysiology and management of inherited bone marrow failure syndromes. In *Blood Reviews* (Vol. 24, Issue 3, pp. 101–122). https://doi.org/10.1016/j.blre.2010.03.002

Singh, A., Chia, J. J., Rao, D. S., & Hoffmann, A. (2025). *Population dynamics modeling reveals that myeloid bias involves both HSC differentiation and progenitor proliferation biases*. *Blood, 145*(12), 1293-1308. https://doi.org/10.1182/blood.2024025598

Skarnes, W. C., Rosen, B., West, A. P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A. O., Thomas, M., Harrow, J., Cox, T., Jackson, D., Severin, J., Biggs, P., Fu, J., Nefedov, M., De Jong, P. J., Stewart, A. F., & Bradley, A. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*, *474*(7351), 337–344. https://doi.org/10.1038/nature10163

Slifka, M. K., Antia, R., Whitmire, J. K., & Ahmed, R. (1998). Humoral Immunity Due to Long-Lived Plasma Cells. In *Immunity* (Vol. 8).

Soerens, A. G., Künzli, M., Quarnstrom, C. F., Scott, M. C., Swanson, L., Locquiao, J., Ghoneim, H. E., Zehn, D., Youngblood, B., Vezys, V., & Masopust, D. (2023). Functional T cells are capable of supernumerary cell division and longevity. *Nature*, *614*(7949), 762–766. https://doi.org/10.1038/s41586-022-05626-9

Soeters, P. B., Wolfe, R. R., & Shenkin, A. (2019). Hypoalbuminemia: Pathogenesis and Clinical Significance. In *Journal of Parenteral and Enteral Nutrition* (Vol. 43, Issue 2, pp. 181–193). John Wiley and Sons Inc. https://doi.org/10.1002/jpen.1451

Spencer Chapman, M., Mitchell, E., Yoshida, K., Williams, N., Fabre, M. A., Ranzoni, A. M., Robinson, P. S., Kregar, L. D., Wilk, M., Boettcher, S., Mahbubani, K., Saeb Parsy, K., Gowers, K. H. C., Janes, S. M., Ng, S. W. K., Hoare, M., Green, A. R., Vassiliou, G. S., Cvejic, A., … Campbell, P. J. (2025). Prolonged persistence of mutagenic DNA lesions in somatic cells. *Nature*, *638*(8051), 729–738. https://doi.org/10.1038/s41586-024-08423-8

Spencer Chapman, M., Wilk, C. M., Boettcher, S., Mitchell, E., Dawson, K., Williams, N., Müller, J., Kovtonyuk, L., Jung, H., Caiado, F., Roberts, K., O'Neill, L., Kent, D. G., Green, A. R., Nangalia, J., Manz, M. G., & Campbell, P. J. (2024). Clonal dynamics after allogeneic haematopoietic cell transplantation. *Nature*, *635*(8040), 926–934. https://doi.org/10.1038/s41586-024-08128-y

Spracklen, C. N., Horikoshi, M., Kim, Y. J., Lin, K., Bragg, F., Moon, S., Suzuki, K., Tam, C. H. T., Tabara, Y., Kwak, S.-H., Takeuchi, F., Long, J., Lim, V. J. Y., Chai, J.-F., Chen, C.-H., Nakatochi, M., Yao, J., Choi, H. S., Iyengar, A. K., … Sim, X. (2020). Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*, *582*(7811), 240–245. https://doi.org/10.1038/s41586-020-2263-3

Stagos, D., Chen, Y., Brocker, C., Donald, E., Jackson, B. C., Orlicky, D. J., Thompson, D. C., & Vasiliou, V. (2010). Aldehyde dehydrogenase 1B1: Molecular cloning and characterization of a novel mitochondrial acetaldehyde-metabolizing enzyme. *Drug Metabolism and Disposition*, *38*(10), 1679–1687. https://doi.org/10.1124/dmd.110.034678

Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P., & Ebert, B. L. (2015). Clonal hematopoiesis of indeterminate potential and its

distinction from myelodysplastic syndromes. In *Blood* (Vol. 126, Issue 1, pp. 9–16). American Society of Hematology. https://doi.org/10.1182/blood-2015-03-631747

Steinberg, S. E. (1984). Mechanisms of folate homeostasis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *246*(4), G319–G324. https://doi.org/10.1152/ajpgi.1984.246.4.G319

Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K. F., Chen, R., Gu, H., Bock, C., Meissner, A., Göttgens, B., Darlington, G. J., Li, W., & Goodell, M. A. (2014). Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell*, *14*(5), 673–688. https://doi.org/10.1016/j.stem.2014.03.002

Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y. J., Klein, A., Hofmann, O., & Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. *Nature*, *514*(7522), 322–327. https://doi.org/10.1038/nature13824

Sun, N., Cai, Q., Zhang, Y., Zhang, R.-R., Jiang, J., Yang, H., Qin, C.-F., & Cheng, G. (2024). The aldehyde dehydrogenase ALDH1B1 exerts antiviral effects through the aggregation of the adaptor MAVS. In *Sci. Signal* (Vol. 17). https://www.science.org

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S. been, Tian, X., Browning, B. L., Das, S., Emde, A. K., Clarke, W. E., Loesch, D. P., … Abecasis, G. R. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, *590*(7845), 290–299. https://doi.org/10.1038/s41586-021-03205-y

Tamura, H., Kitta, K., & Shibamoto, T. (1991). Formation of reactive aldehydes from fatty acids in a iron(2+)/hydrogen peroxide oxidation system. *Journal of Agricultural and Food Chemistry*, *39*(3), 439–442. https://doi.org/10.1021/jf00003a002

Tanaka, T., Biancotto, A., Moaddel, R., Moore, A. Z., Gonzalez-Freire, M., Aon, M. A., Candia, J., Zhang, P., Cheung, F., Fantoni, G., Semba, R. D., & Ferrucci, L. (2018). Plasma proteomic signature of age in healthy humans. *Aging Cell*, *17*(5). https://doi.org/10.1111/acel.12799

Theofilopoulos, A. N., Kono, D. H., & Baccala, R. (2017). The multiple pathways to autoimmunity. In *Nature Immunology* (Vol. 18, Issue 7, pp. 716–724). Nature Publishing Group. https://doi.org/10.1038/ni.3731

Thiele, G. M., Tuma, D. J., Willis, M. S., Miller, J. A., McDonald, T. L., Sorrell, M. F., & Klassen, L. W. (1998). Soluble proteins modified with acetaldehyde and malondialdehyde are immunogenic in the absence of adjuvant. *Alcoholism: Clinical and Experimental Research*, *22*(8), 1731–1739. https://doi.org/10.1111/j.1530-0277.1998.tb03973.x

Trapp, C., Reite, K., Klungland, A., & Epe, B. (2007). Deficiency of the Cockayne syndrome B (CSB) gene aggravates the genomic instability caused by endogenous oxidative DNA base damage in mice. *Oncogene*, *26*(27), 4044–4048. https://doi.org/10.1038/sj.onc.1210167

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J. N., Rovio, A. T., Bruder, C. E., Bohlooly-Y, M., Gidlöf, S., Oldfors, A., Wibom, R., Törnell, J., Jacobs, H. T., & Larsson, N.-G. (2004). Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature*, *429*(6990), 417–423. https://doi.org/10.1038/nature02517

Tubo, N. J., Pagán, A. J., Taylor, J. J., Nelson, R. W., Linehan, J. L., Ertelt, J. M., Huseby, E. S., Way, S. S., & Jenkins, M. K. (2013). Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell*, *153*(4), 785–796. https://doi.org/10.1016/j.cell.2013.04.007

Tuma, D. J. (2002). Role of malondialdehyde-acetaldehyde adducts in liver injury1,2 1Guest editor: Arthur Cederbaum 2This article is part of a series of reviews on "Alcohol, Oxidative Stress and Cell Injury." The full list of papers may be found on the homepage of the journal. *Free Radical Biology and Medicine*, *32*(4), 303–308. https://doi.org/10.1016/S0891-5849(01)00742-0

Tuma, D. J., Newman, M. R., Donohue, T. M., & Sorrell, M. F. (1987). Covalent Binding of Acetaldehyde to Proteins: Participation of Lysine Residues. *Alcoholism: Clinical and Experimental Research*, *11*(6), 579–584. https://doi.org/10.1111/j.1530-0277.1987.tb00178.x

Van Steeg, H., & Kraemer, K. H. (1999). *Xeroderma pigmentosum and the role of UV-induced DNA damage in skin cancer*.

Vasillou, V., Pappa, A., & Estey, T. (2004). Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism. *Drug Metabolism Reviews*, *36*(2), 279–299. https://doi.org/10.1081/DMR-120034001

Veach, R. A., & Wilson, M. H. (2018). CRISPR/Cas9 engineering of a KIM-1 reporter human proximal tubule cell line. *PLoS ONE*, *13*(9). https://doi.org/10.1371/journal.pone.0204487

Vollmer, C. M., Ribas, A., Butterfield, L. H., Dissette, V. B., Andrews, K. J., Eilber, F. C., Montejo, L. D., Chen, A. Y., Hu, B., Glaspy, J. A., Mcbride, W. H., & Economou, J. S. (1999). p53 Selective and Nonselective Replication of an E1B-deleted Adenovirus in Hepatocellular Carcinoma 1. In *CANCER RESEARCH* (Vol. 59). http://aacrjournals.org/cancerres/article-pdf/59/17/4369/3243566/ch179904369p.pdf

Walter, D., Lier, A., Geiselhart, A., Thalheimer, F. B., Huntscha, S., Sobotta, M. C., Moehrle, B., Brocks, D., Bayindir, I., Kaschutnig, P., Muedder, K., Klein, C., Jauch, A., Schroeder, T., Geiger, H., Dick, T. P., Holland-Letz, T., Schmezer, P., Lane, S. W., … Milsom, M. D. (2015). Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature*, *520*(7548), 549–552. https://doi.org/10.1038/nature14131

Wang, D., Day, E. A., Townsend, L. K., Djordjevic, D., Jørgensen, S. B., & Steinberg, G. R. (2021). GDF15: emerging biology and therapeutic applications for obesity and cardiometabolic disease. In *Nature Reviews Endocrinology* (Vol. 17, Issue 10, pp. 592–607). Nature Research. https://doi.org/10.1038/s41574-021-00529-7

Wang, J., Hilchey, S. P., DeDiego, M., Perry, S., Hyrien, O., Nogales, A., Garigen, J., Amanat, F., Huertas, N., Krammer, F., Martinez-Sobrido, L., Topham, D. J., Treanor, J. J., Sangster, M. Y., & Zand, M. S. (2018). Broad cross-reactive IgG responses elicited by adjuvanted vaccination with recombinant influenza hemagglutinin (rHA) in ferrets and mice. *PLoS ONE*, *13*(4). https://doi.org/10.1371/journal.pone.0193680

Wang, M., Brandt, L. T. L., Wang, X., Russell, H., Mitchell, E., Kamimae-Lanning, A. N., Brown, J. M., Dingler, F. A., Garaycoechea, J. I., Isobe, T., Kinston, S. J., Gu, M., Vassiliou, G. S., Wilson, N. K., Göttgens, B., & Patel, K. J. (2023). Genotoxic aldehyde stress prematurely ages hematopoietic stem cells in a p53-driven manner. *Molecular Cell*, *83*(14), 2417-2433.e7. https://doi.org/10.1016/j.molcel.2023.05.035

Wang, M., Dingler, F. A., & Patel, K. J. (2022). Genotoxic aldehydes in the hematopoietic system. *Blood*, *139*(14), 2119–2129. https://doi.org/10.1182/blood.2019004316

Wang, M., McIntee, E. J., Cheng, G., Shi, Y., Villalta, P. W., & Hecht, S. S. (2000). Identification of DNA adducts of acetaldehyde. *Chemical Research in Toxicology*, *13*(11), 1149–1157. https://doi.org/10.1021/tx000118t

Wang, R., Lan, C., Benlagha, K., Camara, N. O. S., Miller, H., Kubo, M., Heegaard, S., Lee, P., Yang, L., Forsman, H., Li, X., Zhai, Z., & Liu, C. (2024). The interaction of

innate immune and adaptive immune system. In *MedComm* (Vol. 5, Issue 10). John Wiley and Sons Inc. https://doi.org/10.1002/mco2.714

Wang, T., Ma, X., Peng, D., Zhang, R., Sun, X., Chen, M., Yan, J., Wang, S., Yan, D., He, Z., Jiang, F., Bao, Y., Hu, C., & Jia, W. (2016). Effects of obesity related genetic variations on visceral and subcutaneous fat distribution in a Chinese population. *Scientific Reports*, *6*. https://doi.org/10.1038/srep20691

Watson, C. J., Papula, A. L., Poon, G. Y. P., Wong, W. H., Young, A. L., Druley, T. E., Fisher, D. S., & Blundell, J. R. (2020). The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*, *367*(6485), 1449–1454. https://doi.org/10.1126/science.aay9333

Wei, W., Li, B., Hanes, M. A., Kakar, S., Chen, X., & Liu, L. (2010). *S-Nitrosylation from GSNOR Deficiency Impairs DNA Repair and Promotes Hepatocarcinogenesis*. https://www.science.org

Wilkinson, A. C., Ishida, R., Kikuchi, M., Sudo, K., Morita, M., Crisostomo, R. V., Yamamoto, R., Loh, K. M., Nakamura, Y., Watanabe, M., Nakauchi, H., & Yamazaki, S. (2019). Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. *Nature*, *571*(7763), 117–121. https://doi.org/10.1038/s41586-019-1244-x

Wit, N., Gogola, E., West, J. A., Vornbäumen, T., Seear, R. V, J Bailey, P. S., Burgos-Barragan, G., Wang, M., Krawczyk, P., E W Huberts, D. H., Gergely, F., Matheson, N. J., Kaser, A., Nathan, J. A., & Patel, K. J. (2023). *A histone deacetylase 3 and mitochondrial complex I axis regulates toxic formaldehyde production*. https://www.science.org

Wong, W. J., Emdin, C., Bick, A. G., Zekavat, S. M., Niroula, A., Pirruccello, J. P., Dichtel, L., Griffin, G., Uddin, M. M., Gibson, C. J., Kovalcik, V., Lin, A. E., McConkey, M. E., Vromman, A., Sellar, R. S., Kim, P. G., Agrawal, M., Weinstock, J., Long, M. T., … Natarajan, P. (2023). Clonal haematopoiesis and risk of chronic liver disease. *Nature*, *616*(7958), 747–754. https://doi.org/10.1038/s41586-023-05857-4

WORRALL, S., DE JERSEY, J., SHANLEY, B. C., & WILCE, P. A. (1991). Antibodies against acetaldehyde-modified epitopes: an elevated IgA response in alcoholics. *European Journal of Clinical Investigation*, *21*(1), 90–95. https://doi.org/10.1111/j.1365-2362.1991.tb01364.x

Xiao, Q., Weiner, H., & Crabb, D. W. (1996). The mutation in the mitochondrial aldehyde dehydrogenase (ALDH2) gene responsible for alcohol-induced flushing increases turnover of the enzyme tetramers in a dominant fashion. *Journal of Clinical Investigation*, *98*(9), 2027–2032. https://doi.org/10.1172/JCI119007

Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dipersio, J. F., Chen, F., Wilson, R. K., Ley, T. J., & Ding, L. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature Medicine*, *20*(12), 1472–1478. https://doi.org/10.1038/nm.3733

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H., & Nakauchi, H. (2013). Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells. *Cell*, *154*(5), 1112–1126. https://doi.org/10.1016/j.cell.2013.08.007

Yamamoto, R., Wilkinson, A. C., & Nakauchi, H. (2018). Changing concepts in hematopoietic stem cells. In *Science* (Vol. 362, Issue 6417, pp. 895–896). American Association for the Advancement of Science. https://doi.org/10.1126/science.aat7873

Yamamoto, R., Wilkinson, A. C., Ooehara, J., Lan, X., Lai, C. Y., Nakauchi, Y., Pritchard, J. K., & Nakauchi, H. (2018). Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. *Cell Stem Cell*, *22*(4), 600-607.e4. https://doi.org/10.1016/j.stem.2018.03.013

Yang, D., Scavuzzo, M. A., Chmielowiec, J., Sharp, R., Bajic, A., & Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Scientific Reports*, *6*. https://doi.org/10.1038/srep21264

Yao, X., Zhang, M., Wang, X., Ying, W., Hu, X., Dai, P., Meng, F., Shi, L., Sun, Y., Yao, N., Zhong, W., Li, Y., Wu, K., Li, W., Chen, Z. jiang, & Yang, H. (2018). Tild-CRISPR Allows for Efficient and Precise Gene Knockin in Mouse and Human Cells. *Developmental Cell*, *45*(4), 526-536.e5. https://doi.org/10.1016/j.devcel.2018.04.021

Yi, W., Zhang, Y., Liu, B., Zhou, Y., Liao, D., Qiao, X., Gao, D., Xie, T., Yao, Q., Zhang, Y., Qiu, Y., Huang, G., Chen, Z., Chen, C., & Ju, Z. (2021). Protein S-nitrosylation regulates proteostasis and viability of hematopoietic stem cell during regeneration. *Cell Reports*, *34*(13). https://doi.org/10.1016/j.celrep.2021.108922

Yin, X., Kim, K., Suetsugu, H., Bang, S.-Y., Wen, L., Koido, M., Ha, E., Liu, L., Sakamoto, Y., Jo, S., Leng, R.-X., Otomo, N., Laurynenka, V., Kwon, Y.-C., Sheng, Y., Sugano, N., Hwang, M. Y., Li, W., Mukai, M., … Bae, S.-C. (2021). Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, *80*(5), 632–640. https://doi.org/10.1136/annrheumdis-2020-219209

Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., Shiozawa, Y., Sato, Y., Aoki, K., Kim, S. K., Fujii, Y., Yoshida, K., Kataoka, K., Nakagawa, M. M., Inoue, Y., Hirano, T., Shiraishi, Y., Chiba, K., Tanaka, H., … Ogawa, S. (2019). Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*, *565*(7739), 312–317. https://doi.org/10.1038/s41586-018-0811-x

Yokoyama, A., Muramatsu, T., Ohmori, T., Higuchi, S., Hayashida, M., & Ishii, H. (1996). Esophageal cancer and aldehyde dehydrogenase-2 genotypes in Japanese males. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *5*(2), 99–102. http://www.ncbi.nlm.nih.gov/pubmed/8850269

Yokoyama, T., Yokoyama, A., Kato, H., Tsujinaka, T., Muto, M., Omori, T., Haneda, T., Kumagai, Y., Igaki, H., Yokoyama, M., Watanabe, H., & Yoshimizu, H. (2003). Alcohol flushing, alcohol and aldehyde dehydrogenase genotypes, and risk for esophageal squamous cell carcinoma in Japanese men. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *12*(11 Pt 1), 1227–1233. http://www.ncbi.nlm.nih.gov/pubmed/14652286

Yoshida, A., Huang, I.-Y., & Ikawa, M. (1984). Molecular Abnormality of an Inactive Aldehyde Dehydrogenase Variant Commonly Found. In *Source* (Vol. 81, Issue 1).

Yoshida, K., Gowers, K. H. C., Lee-Six, H., Chandrasekharan, D. P., Coorens, T., Maughan, E. F., Beal, K., Menzies, A., Millar, F. R., Anderson, E., Clarke, S. E., Pennycuick, A., Thakrar, R. M., Butler, C. R., Kakiuchi, N., Hirano, T., Hynds, R. E., Stratton, M. R., Martincorena, I., … Campbell, P. J. (2020). Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature*, *578*(7794), 266–272. https://doi.org/10.1038/s41586-020-1961-1

Yousefzadeh, M. J., Flores, R. R., Zhu, Y., Schmiechen, Z. C., Brooks, R. W., Trussoni, C. E., Cui, Y., Angelini, L., Lee, K. A., McGowan, S. J., Burrack, A. L., Wang, D.,

Dong, Q., Lu, A., Sano, T., O'Kelly, R. D., McGuckian, C. A., Kato, J. I., Bank, M. P., … Niedernhofer, L. J. (2021). An aged immune system drives senescence and ageing of solid organs. *Nature*, *594*(7861), 100–105. https://doi.org/10.1038/s41586-021-03547-7

Yu, V. W. C., Yusuf, R. Z., Oki, T., Wu, J., Saez, B., Wang, X., Cook, C., Baryawno, N., Ziller, M. J., Lee, E., Gu, H., Meissner, A., Lin, C. P., Kharchenko, P. V., & Scadden, D. T. (2016). Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell*, *167*(5), 1310-1322.e17. https://doi.org/10.1016/j.cell.2016.10.045

Zhang, R., Brennan, M. L., Shen, Z., MacPherson, J. C., Schmitt, D., Molenda, C. E., & Hazen, S. L. (2002). Myeloperoxidase functions as a major enzymatic catalyst for initiation of lipid peroxidation at sites of inflammation. *Journal of Biological Chemistry*, *277*(48), 46116–46122. https://doi.org/10.1074/jbc.M209124200

Zhang, W., Cui, H., Xu, J., Shi, M., Bian, L., Cui, L., Jiang, C., & Zhang, Y. (2025). Biodistribution and mechanisms of action of MF59 and MF59-like adjuvants. In *Journal of Controlled Release* (Vol. 378, pp. 573–587). Elsevier B.V. https://doi.org/10.1016/j.jconrel.2024.12.044

Zheng, R., Li, Z., He, F., Liu, H., Chen, J., Chen, J., Xie, X., Zhou, J., Chen, H., Wu, X., Wu, J., Chen, B., Liu, Y., Cui, H., Fan, L., Sha, W., Liu, Y., Wang, J., Huang, X., … Ge, B. (2018). Genome-wide association study identifies two risk loci for tuberculosis in Han Chinese. *Nature Communications*, *9*(1), 4072. https://doi.org/10.1038/s41467-018-06539-w

Zhou, J., & Weiner, H. (2000). Basis for half-of-the-site reactivity and the dominance of the K487 oriental subunit over the E487 subunit in heterotetrameric human liver mitochondrial aldehyde dehydrogenase. *Biochemistry*, *39*(39), 12019–12024. https://doi.org/10.1021/bi001221k

Zimmermann, M., Murina, O., Reijns, M. A. M., Agathanggelou, A., Challis, R., Tarnauskaite, Ž., Muir, M., Fluteau, A., Aregger, M., McEwan, A., Yuan, W., Clarke, M., Lambros, M. B., Paneesha, S., Moss, P., Chandrashekhar, M., Angers, S., Moffat, J., Brunton, V. G., … Durocher, D. (2018). CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature*, *559*(7713), 285–289. https://doi.org/10.1038/s41586-018-0291-z

Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., Thorgeirsson, T. E., Sigurdsson, A., Gudjonsson, S. A., Gudmundsson, J., Jonasson, J. G., Tryggvadottir, L., Jonsson, T., Helgason, A., Gylfason, A., Sulem, P., Rafnar, T., Thorsteinsdottir, U., Gudbjartsson, D. F., … Stefansson, K. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*, *130*(6), 742–752. https://doi.org/10.1182/blood-2017-02-769869

Zou, X., Koh, G. C. C., Nanda, A. S., Degasperi, A., Urgo, K., Roumeliotis, T. I., Agu, C. A., Badja, C., Momen, S., Young, J., Amarante, T. D., Side, L., Brice, G., Perez-Alonso, V., Rueda, D., Gomez, C., Bushell, W., Harris, R., Choudhary, J. S., … Nik-Zainal, S. (2021). A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nature Cancer*, *2*(6), 643–657. https://doi.org/10.1038/s43018-021-00200-0