



Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin

Ollscoil Átha Cliath | The University of Dublin

Clonal haematopoiesis in telomere biology disorders influences risk of haematological malignancy and overall survival

A dissertation submitted to Trinity College Dublin, the
University of Dublin for the degree of Doctor of Medicine
(MD)

Supervised By:
Professor Paul Browne
&
Dr. Neal S. Young

2025

Declaration

The work presented in this thesis, except where duly acknowledged, was carried out by me and has not been submitted as an exercise for a degree at this or any other university.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

All patients included in this thesis provided written informed consent.

Peer-reviewed papers and conference abstracts involving this thesis research

Gutierrez-Rodrigues F*, **Groarke EM***, Thongon N*, Rodriguez-Sevilla JJ, Bazzo Catto LF, Niewisch MR, Shalhoub RN, McReynolds LJ, Clé DV, Patel BA, Ma X, Hironaka D, Donaires FS, Spitofsky NR, Santana BA, Lai TP, Alemu L, Kajigaya S, Darden I, Zhou W, Browne PV, Paul S, Lack J, Young DJ, DiNardo CD, Aviv A, Ma F, Michels de Oliveira M, Azambuja AP, Dunbar CE, Olszewska M, Olivier E, Papapetrou EP, Giri N, Alter BP, Bonfim CMS, Wu CO, Garcia-Manero G, Savage SA, Young NS, Colla S, Calado RT. Clonal landscape and clinical outcomes of telomere biology disorders: somatic rescuing and cancer mutations. *Blood*. 2024 Sep 24;blood.2024025023. doi: 10.1182/blood.2024025023. Epub ahead of print. PMID: 39316766

Groarke EM, Calado RT, Liu JM. Cell senescence and malignant transformation in the inherited bone marrow failure syndromes: Overlapping pathophysiology with therapeutic implications. *Semin Hematol*. 2022 Jan;59(1):30-37. doi: 0.1053/j.seminhematol.2022.01.003. Epub 2022 Jan 31. PMID: 35491056

Groarke EM, Gutierrez-Rodrigues F, Ma, Xioyang, Patel BA, Spitofsky N, Browne PV, Chien KS, Wu CO, DiNardo C, Colla S, Calado RT, Young NS. *U2AF1* and other splicing factor gene mutations in telomere biology disorders are associated with hematologic neoplasia and worse overall survival. (Abstract 862) Presented as an oral abstract at the American Society of Hematology Annual Meeting in Atlanta 2021

Gutierrez-Rodrigues F, **Groarke EM**, Cle DV, Patel BA, Donaires FS, Spitofsky N, Alemu L, Santana BA, Kajigaya S, Colla S, De Oliveira MM, Bonfim C, Calado RT, Young NS. Clonal hematopoiesis in telomere biology disorders associates with the underlying germline defect and somatic mutations in *POT1*, *PPM1D* and *TERT* promotor. (Abstract 1111) Presented as a poster at the American Society of Hematology Annual Meeting in Atlanta 2021

Acknowledgement

I wish to acknowledge my supervisors Drs. Paul Browne and Neal S. Young.

I would like to acknowledge my lab colleagues who helped with this work – in particular for Dr. Fernanda Gutierrez-Rodrigues for working with me on that project and for her expertise and help in DNA sequencing & interpretation. I would also like to thank statisticians Drs. Colin Wu and Xiaoyang Ma who provided invaluable statistical assistance and helped in generating KM plots.

I also want to thank all my colleagues including my co-investigators (in particular Neal S. Young), research nurses, data managers, and others involved in the management of the National Heart Lung and Blood Institute clinical trials and protocols our patients were enrolled on. I would like to thank our collaborators in The National Cancer Institute (in particular Sharon Savage and Lisa McReynolds), Brazil (in particular Rodrigo Calado), and MD Anderson (in particular Simona Colla) for sharing their patient samples and data.

Finally, I would like to thank all the patients who participated in this research.

Abbreviations

AD: autosomal dominant
ACMG: American College of Medical Genetics
AKC: Ammonium-Chloride-Potassium
AML: acute myeloid leukaemia
ANC: absolute neutrophil count
AR: autosomal recessive
ATM: ataxia telangiectasia mutated
ATR: ataxia telangiectasia and Rad3 related
AVM: arteriovenous malformation
BMF: bone marrow failure
BMT: bone marrow transplant
CAP: controlled attenuation parameter
CCUS: clonal cytopenia of undetermined significance
CHIP: clonal haematopoiesis of indeterminate potential
CH: clonal haematopoiesis
CLL: chronic lymphocytic leukaemia
CLIA: Clinical Laboratory Improvement Amendments
CNV: copy number variation
CP: Coats Plus (syndrome)
CT: computed tomography
DBA: Diamond-Blackfan anaemia
DC: dyskeratosis congenita
DDR: DNA damage response
DLCO: diffusing capacity for carbon monoxide
DMSO: Dimethyl sulfoxide
DNA: deoxyribonucleic acid
EDTA: Ethylenediaminetetraacetic acid
ECS: error corrected sequencing
FA: Fanconi Anaemia
FEV1: forced expiratory volume
FVC: forced vital capacity
FISH: fluorescent *in-situ* hybridisation
HH: Hoyeraal-Hreidarsson (syndrome)
HSCT: haematopoietic stem cell transplant
IBMFS: inherited bone marrow failure syndromes
ICUS: Idiopathic cytopenia of undetermined significance
MDS: myelodysplastic syndrome
MGUS: monoclonal gammopathy of undetermined significance
MPN: myeloproliferative neoplasm
NGS: next generation sequencing
OS: overall survival
PCR: polymerase chain reaction
PF: pulmonary fibrosis
PFTs: pulmonary function testing
R-IPSS: revised international prognostic scoring system
RS: Revesz syndrome
SDS: Shwachman Diamond Syndrome
SNP: single nucleotide polymorphism
SNV: single nucleotide variant
SR: somatic rescue
TBD: telomere biology disorder
UMAP: uniform manifold approximation and projection

US: ultrasound
VAF: variant allele frequency
VUS: variant of uncertain significance
WHO: World Health Organization
XLR: x-linked recessive

Abstract

Telomere biology disorders (TBD) are a spectrum of diseases caused by germline variants in telomere repair and maintenance. They comprise one of the inherited bone marrow failure syndromes, a heterogeneous group of disorders characterised by bone marrow failure, increased risk of haematological and solid malignancy, and other disease specific organ dysfunction. Our aims were to characterise myeloid malignancy in our cohort, perform somatic mutation testing in genes related to telomere maintenance and myeloid malignancy to elucidate the clonal landscape of TBD, and to see whether somatic mutations had clinical implications for TBD patients. We assessed a large cohort of TBD patients (n=208) from 3 institutions and found there was a cumulative incidence of 16% by age 16; however, this was significantly impacted by the patients' specific germline variant with *TERC* and *DKC1* patients having the highest risk of MDS/AML, and *TINF2* and biallelic *RTEL1* patients having very low risk. When we sequenced patient samples, we found the clonal landscape was dominated by 4 somatic mutations, 2 in telomere related genes (*POT1* and *TERTp*) and 2 related to myeloid malignancy (*U2AF1* and *PPM1D*). We grouped mutations into: typical MDS/AML mutations (with a subgroup for splicing factors given their predominance and *TP53*), typical CH mutations, *PPM1D*, *POT1* and *TERTp*. We used logistic regression and Kaplan Meier curves and found that MDS/AML associated mutations (including splicing factors) predicted both for MDS/AML and worse overall survival. *TERTp* mutations also predicted for worse OS but not myeloid malignancy. *TP53* also strongly predicted for worse OS and the development of solid but not haematological malignancy. *PPM1D* and *POT* appeared to have an overall neutral effect on survival and did not positively influence blood counts. Using single cell analysis, we saw that *U2AF1* acts as a founder clone for the development of subsequent malignant subclones. Our description of the clonal landscape of TBD help to drive our understanding of these disorders; we believe these somatic mutations arise as a compensatory mechanism to improve cell fitness. Splicing factor mutations and *TP53* were identified as important biomarkers for cancer. Work is ongoing to better elucidate the functional consequences of these somatic mutations in TBD.

Table of Contents

1. Chapter 1: Introduction and project aims	1
1.1 The end replication problem and discovery of the telomere.....	1
1.2 Mosaicism and clonality.....	1
1.3 An inherited disorder with a heterogenous clinical phenotype.....	4
1.4 Haematological malignancy in TBD.....	7
1.5 Telomere length in cancer.....	8
1.6 Potential mechanisms underlying malignant transformation due to long telomeres.....	8
1.7 Potential mechanisms underlying malignant transformation due to short telomeres.....	9
1.8 Current therapeutic approaches in TBD.....	11
1.9 Somatic mosaicism in haematologic malignancy and disease.....	12
1.10 Somatic mosaicism in germline marrow failure disorders.....	14
1.11 Overall aims and objectives.....	16
 2. Chapter 2: Clinical characteristics of the TBD cohort.....	 188
2.1 Introduction.....	18
2.2 Methods.....	18
2.2.1 Definition of TBD used for inclusion in the study.....	19
2.2.2 Definition of bone marrow failure and myeloid malignancy used in the study.....	20
2.2.3 Definition of lung disease used in the study.....	21
2.2.4 Definition of liver disease used in the study.....	21
2.2.5 Other clinical definitions used in the study.....	22
2.2.6 Statistical methods.....	22
2.3 Results.....	22
2.3.1 General characteristics of the cohort.....	22
2.3.2 General characterization of haematological and solid malignancy in the cohort.....	24
2.3.3 Cumulative incidence of haematological malignancy in the cohort.....	28

2.4 Conclusions.....	32
3. Chapter 3: The clonal landscape of somatic mutations in TBD	33
3.1 Introduction.....	33
3.2 Methods.....	34
3.2.1 Protocol for error corrected sequencing.....	34
3.2.2 Statistical methods.....	36
3.3 Results.....	36
3.3.1 Association of somatic mutations and age.....	37
3.3.2 Oncoprints showing spectrum of somatic mutations for the cohort organized by germline genotype.....	37
3.3.3 Association with the presence or absence of somatic mutations and peripheral blood counts and clinical phenotype.....	43
3.3.4 Variant allele frequency of somatic mutations.....	45
3.3.5 Chemotherapy exposure.....	46
3.3.6 Somatic variants in the <i>TINF2</i> cohort.....	46
3.3.7 Somatic variants in the solid malignancy cohort.....	48
3.4 Conclusions.....	48
4. Chapter 4: Association of somatic single nucleotide variants and development of cancer and other clinical outcomes in TBD patients.....	551
4.1 Introduction.....	51
4.2 Methods.....	52
4.2.1 Grouping of somatic mutations.....	51
4.2.2 Statistics.....	54
4.3 Results.....	54
4.3.1 Association with somatic mutation groups and age.....	54
4.3.2 Association with somatic mutation groups and blood counts.....	54
4.3.3 Predictors of MDS/AML using logistic regression.....	56
4.3.4 Causes of death in the cohort.....	57
4.3.5 Predictors of survival using logistic regression.....	58
4.3.6 Association of somatic mutations with overall survival.....	60
4.3.7 Association of somatic mutations with overall survival in patients <20 years old.....	62

4.3.8 Association of somatic mutations with overall survival in patients ≥ 20 years but < 40 years old.....	64
4.3.9 Association of somatic mutations with overall survival in patients ≥ 40 years but less than 60 years old.....	66
4.3.10 Predictors of solid malignancy using logistic regression.....	69
4.3.11 Predictors of lung and liver disease using logistic regression.....	69
4.3.12 Effect of androgens on somatic mutations.....	71
4.4 Conclusions.....	73
5. Chapter 5: Single cell analysis of somatic variants in TBD patients	75
5.1 Introduction.....	75
5.2 Methods.....	75
5.2.1 Mission Bio Tapestri protocol overview.....	75
5.2.2 Mission Bio Tapestri single cell analysis.....	76
5.3 Results.....	77
5.3.1 Patient 1 with <i>U2AF1</i> and other mutations who developed MDS-EB2.....	77
5.3.2 Patient 2 with <i>U2AF1</i> and other MDS mutations who did not develop myeloid malignancy.....	83
5.4 Conclusions.....	89
6. Chapter 6: Final conclusions and discussion	91
6.1 Clinical and biological implications of the work.....	91
6.2 Implications for future research.....	93
6.2.1 Cancer screening.....	93
6.2.2 Androgen therapy.....	93
6.2.3 Spliceosome inhibitors.....	94
6.2.4 Gene therapy.....	94
6.3 Final conclusions.....	95
7. Appendices: A, B and C	
Appendix A: Description of clinical protocols.....	104
Appendix B: Archer NGS protocol.....	109
Appendix C: MissionBio single cell sequencing protocol.....	114

List of Figures

Figure 1. Somatic mosaicism	2
Figure 2. Concept of clonality.....	3
Figure 3. Clonal haematopoiesis in the marrow environment.....	3
Figure 4. Effect of short telomeres and DNA damage on cells.....	10
Figure 5. CHIP as compared to MDS.....	13
Figure 6. Somatic reversion.....	15
Figure 7. Cumulative risk of developing MDS/AML in all patients with TBD.....	28
Figure 8A. Cumulative risk of developing MDS/AML in patients with TBD depending on specific germline variant.....	29
Figure 8B. Cumulative risk of developing MDS/AML in patients with TBD depending on specific germline variant.....	30
Figure 9. Cumulative incidence of MDS/AML in <i>TERT</i> or <i>TERC</i> mutated patients....	31
Figure 10. Protocol overview for ArcherDx ECS platform.....	35
Figure 11. Association of somatic mutations and age.....	37
Figure 12. Oncoprint showing spectrum of CH in TBD including <u>all</u> germline variants.....	38
Figure 12A. Oncoprint showing spectrum of CH in TBD including <u><i>TERT</i></u> germline variants.....	39
Figure 12B. Oncoprint showing spectrum of CH in TBD including <u><i>TERC</i></u> germline variants.....	40
Figure 12C. Oncoprint showing spectrum of CH in TBD including <u><i>RTEL1</i></u> germline variants.....	41
Figure 12D. Oncoprint showing spectrum of CH in TBD including <u><i>TINF2</i></u> germline variants.....	42
Figure 12E. Oncoprint showing spectrum of CH in TBD including <u><i>DKC1</i></u> germline variants.....	42
Figure 13. Association with any somatic mutation and haemoglobin.....	43
Figure 14. Association with any somatic mutation and platelets.....	44
Figure 15. Association with any clinical phenotype and mutation status.....	44
Figure 16. Variant allele frequency per mutation.....	45
Figure 17. Association between blood counts and <i>POT1</i> mutation status in <i>TINF2</i> patients.....	47

Figure 18. Overall survival comparing <i>TINF2</i> patients with and without <i>POT1</i> mutations.....	48
Figure 19. Age distribution of mutation groups.....	54
Figure 20. Haemoglobin distribution of mutation groups.....	55
Figure 21. Platelet distribution of mutation groups.....	55
Figure 22. Overall survival comparing the presence and absence of any somatic mutation.....	60
Figure 23. Overall survival comparing the presence and absence of any somatic mutation by somatic mutation grouping.....	61
Figure 24. Overall survival of patients with or without mutations stratified by age.....	62
Figure 25. Overall survival of patients with or without mutations in patients <20 years.....	63
Figure 26. Overall survival of patients with <i>POT1</i> mutations compared to those without <i>POT1</i> mutations in patients <20 years.....	63
Figure 27. Overall survival of patients with <i>TP53</i> mutations compared to those without <i>TP53</i> mutations in patients <20 years.....	64
Figure 28. Overall survival of patients with or without mutations in patients ≥ 20 years but <40 years.....	65
Figure 29. Overall survival of patients with splicing factor mutations compared to those without splicing factor mutations in patients ≥ 20 years but <40 years.....	65
Figure 30. Overall survival of patients with <i>TERTp</i> mutations compared to those without <i>TERTp</i> mutations in patients ≥ 20 years but <40 years.....	66
Figure 31. Overall survival of patients with <i>TP53</i> mutations compared to those without <i>TP53</i> mutations in patients ≥ 20 years but <40 years.....	66
Figure 32. Overall survival of patients with MDS/AML associated mutations compared to those without MDS/AML associated mutations in patients ≥ 40 years but <60 years.....	67
Figure 33. Overall survival of patients with <i>PPMID</i> mutations compared to those without <i>PPMID</i> mutations in patients ≥ 40 years but <60 years.....	68
Figure 34. Overall survival of patients with <i>TP53</i> mutations compared to those without <i>TP53</i> mutations in patients ≥ 40 years but <60 years.....	68
Figure 35. Overall survival of patients with or without mutations in patients >60 years.....	68
Figure 36. Effect of androgen therapy on clonal trajectories.....	72

Figure 37. Proteins used in MissionBio for cell identification.....	77
Figure 38. Clonal trajectory of each mutation and VAF (Patient 1; MDS/AML).....	78
Figure 39. Single cell DNA clones and cell counts (Patient 1; MDS/AML).....	79
Figure 40. Single cells according to cell type using protein coding (immunophenotype) (Patient 1; MDS/AML).....	80
Figure 41. Cell populations classified by UMAP clustering (Patient 1; MDS/AML)...	81
Figure 42. Cell populations with DNA genotype superimposed over protein UMAP clustering (Patient 1; MDS/AML).....	81
Figure 43. Mutant subclones detected by single cell superimposed over protein UMAP (Patient 1; MDS/AML).....	82
Figure 44. Clonal dynamics by single cell analysis (Patient 2; No MDS/AML).....	82
Figure 45. Clonal trajectory of each mutation and VAF (Patient 2; No MDS/AML)...	83
Figure 46. Single cell DNA clones and cell counts (Patient 2; No MDS/AML).....	84
Figure 47. Single cells according to cell type using protein coding (immunophenotype) (Patient 2. No MDS/AML).....	85
Figure 48. Cell populations classified by UMAP clustering (Patient 2; No MDS/AML).....	86
Figure 49. Cell populations with DNA genotype superimposed over protein UMAP clustering (Patient 2; No MDS/AML).....	87
Figure 50. Mutant subclones detected by single cell superimposed over protein UMAP (Patient 2; No MDS/AML).....	88
Figure 51. Clonal dynamics by single cell analysis (Patient 2; No MDS/AML).....	89
Figure 52. Our conclusions regarding the factors influencing the clonal landscape of TBD.....	96

List of Tables

Table 1. The inherited bone marrow failure syndromes.....	5
Table 2. Disease characteristics of the TBD cohort.....	23
Table 3. Malignancy in the TBD cohort.....	24
Table 4. MDS in the TBD cohort.....	25
Table 5. AML in the TBD cohort.....	26
Table 6. Somatic variants at the time of, preceding, or subsequent to MDS/AML diagnosis.....	27
Table 7. Error-corrected DNA sequencing customized panel with genes related to telomere biology, clonal haematopoiesis, and myeloid malignancies	34
Table 8. Univariate logistic regression of development of MDS/AML.....	56
Table 9. Multivariate logistic regression of development of MDS/AML.....	57
Table 10. Causes of death in the TBD cohort.....	57
Table 11. Univariate logistic regression for development of death.....	58
Table 12. Multivariate logistic regression for development of death.....	59
Table 13. Univariate logistic regression for development of solid malignancy.....	69
Table 14. Multivariate logistic regression for development of solid malignancy.....	69
Table 15. Univariate logistic regression for development of liver fibrosis.....	70
Table 16. Multivariate logistic regression for development of liver fibrosis.....	70
Table 17. Univariate logistic regression for development of lung fibrosis.....	70
Table 18. Multivariate logistic regression for development of lung fibrosis.....	71

1. Chapter 1: Introduction and project aims

1.1 The end replication problem and discovery of the telomere

The “end replication problem” refers to the inability of the lagging strand of DNA to be fully copied during replication, conferring genomic instability, and triggering cell senescence once a critical point is reached. However, it had been long observed that while broken stands of DNA were unstable, chromosome ends appeared to be protected, potentially due to chromosome caps called telomeres.

The mechanisms underlying this protection remained a mystery until work pioneered by scientist Elizabeth Blackburn in the late 1970s and 1980s elucidated the presence, structure, and function of the telomere. First, using the aquatic single cell organism *Tetrahymena Thermophila*, Blackburn discovered that tandem repeats (TTGGGG) were present at the end of each chromosome². Subsequently, in collaboration with Jack Szostak, telomere DNA from *Tetrahymena Thermophila* was successfully introduced into yeast cells, confirming the preservation of the telomere structure in two different organisms³. Soon thereafter, telomerase, the enzyme that synthesizes telomeres, was discovered by Blackburn in collaboration with Carol Greider⁴.

The discovery of the telomere and the enzyme telomerase has had widespread implications for understanding of human disease, in particular ageing, cancer, and inherited telomere and telomerase defects. In 2009 Blackburn, Szostak, and Greider received the Nobel Prize for their discoveries.

1.2 Mosaicism and clonality

A germline variant is defined as a gene change in a reproductive cell (egg or sperm) that becomes incorporated into the DNA of every cell in the body of the offspring. Such variants can be passed onto offspring and as such are inherited.

A somatic variant or is the presence of post-zygotic DNA alterations acquired after the first mitotic division in the zygote (as opposed to germline which are derived from the reproductive

cells). Such alterations are often referred to as “mutations” or “clones” and are not passed to offspring⁵.

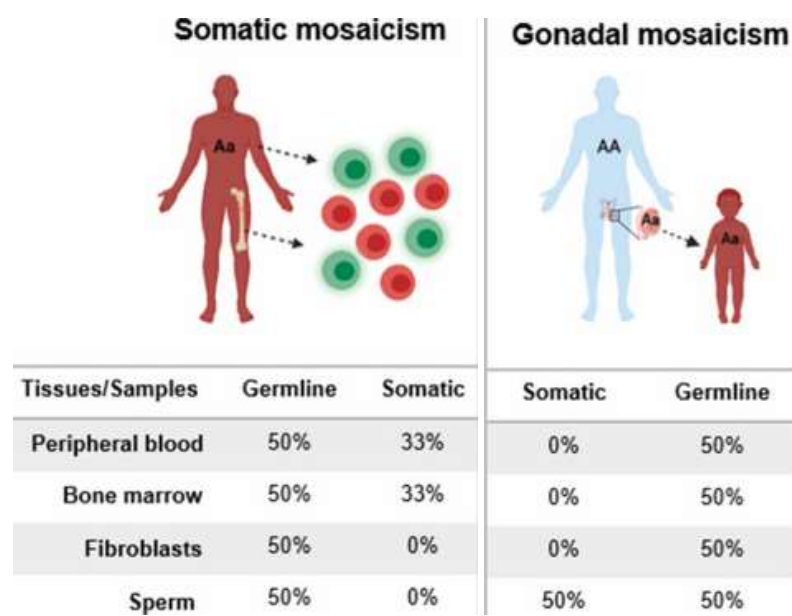


Figure 1. Somatic versus germline mosaicism: *Germline mosaicism occurs when mutations occur in and are confined to germ cells (egg or sperm); such mutations will be passed to offspring and will be present in every cell of the offspring’s body. Somatic mosaicism occurs in post-zygotic (non-germline cells) and will occur in particular cells in the body (i.e. in the marrow); such mutations will not be passed to offspring. Image taken from Gutierrez-Rodrigues et al. Best Practice and Research Clinical Haematology 2021⁵.*

Humans continuously acquire mutations as cells divide, and most are inconsequential and do not leave a detectable genomic footprint because they are neutral (do not alter the cell’s phenotype) or are so harmful that they lead to cell senescence or apoptosis. However, some mutations gain an advantage, continue to propagate, and can have serious functional consequences including development of malignancy, disrupted or altered organ development, or other system specific disease⁶ (Figure 2). Clones may develop due to intrinsic cell factors leading to evasion of cell death pathways and cellular immortality, or may be selected for as a result of changes in the marrow environment⁷ (Figure 3).

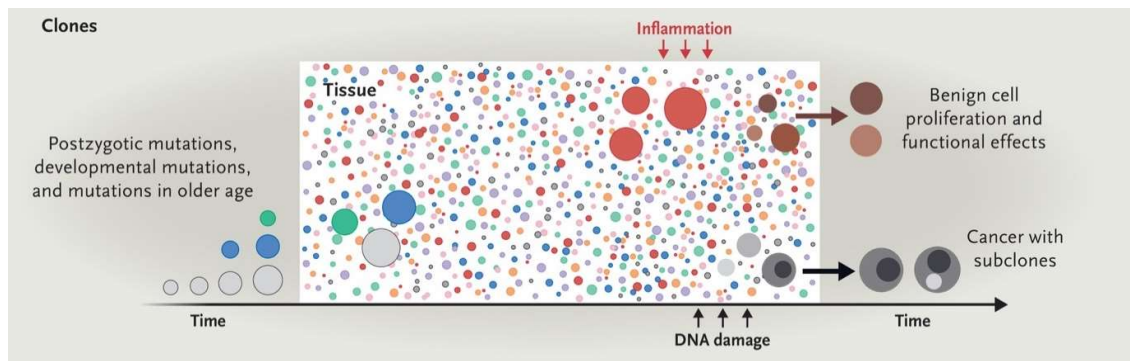


Figure 2. Concept of clonality. *As cells divide, mutations occur and may be selected by the cell's environment; some clones may provide positive effects, but others may be maladaptive and drive malignancy. Image taken from Mustjoki and Young, NEJM 2021⁶.*

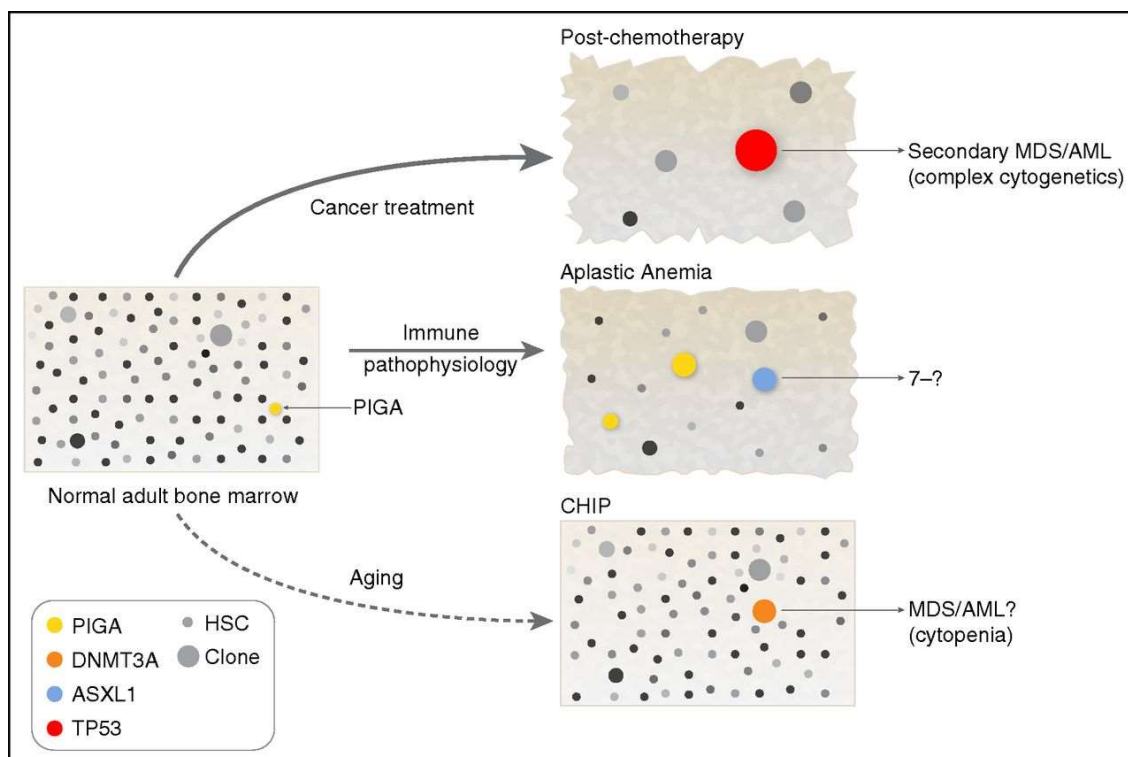


Figure 3. Clonal haematopoiesis in the marrow environment. *Depending on the specific bone marrow environment, certain clones may have a proliferative advantage. For example, in immune bone marrow failure, cells that are not recognized or resistant to immune destruction would be favoured, such as PIGA-mutated cells or cells that have partially lost HLA genes. In an environment where a marrow is exposed to cytotoxic agents, cells that are resistant to apoptosis or undergo more rapid cell cycling may be positively selected. Therefore, it is likely that an environment characterized by increased telomere attrition will also select for specific clones. Image taken from Cooper and Young, Blood 2017⁷.*

Clones can take the form of large structural variations (including copy number variations [CNV]), genomic alterations without copy number change, small insertions and deletions (indels), or single nucleotide variants (SNV). Large structural variations (>3 mega-base pairs) include chromosomal loss or aneuploidy and can be detected by traditional techniques like karyotyping or fluorescence *in situ* hybridization (FISH). Detection of SNVs and indels, which occur over a much smaller section of the genome, requires more sensitive techniques such as single nucleotide polymorphism (SNP) arrays, polymerase chain reaction (PCR), and next generation sequencing (NGS)⁵.

1.3 An inherited disorder with a heterogenous clinical phenotype

Telomere biology disorders (TBD) are a spectrum of diseases caused by germline variants in telomere repair and maintenance⁸. They are part of the inherited bone marrow failure syndromes (IBMFS), a group of disorders characterized by a germline position to BMF, other disease-specific organ involvement, and most with a pre-disposition to cancer; these are briefly summarized in Table 1. Dyskeratosis congenita (DC), the prototype TBD, was first described by Zinsser in 1910, where he outlined the case of two brothers with skin pigmentation, oral leucoplakia, and nail dystrophy; this later became the classical DC clinical triad⁹. Subsequent reports of DC running in families lead to the conclusion that it was likely a genetic disorder. Given its mode of inheritance and male predominance, it was deemed early on to be likely X-linked¹⁰. The genetic basis of DC was first proven in 1998 with the discovery of the *DKC1* gene on the X chromosome¹¹. In patients with *DKC1* variants, telomere length was subsequently found to be short in both peripheral blood cells and fibroblasts, providing a mechanism of disease¹². Since then, variants in several other genes (now a total of 15) related to telomere repair and maintained have been associated with TBD, most notably *TERC* and *TERT*, both of which cause autosomal dominant (AD) disease^{13,14}. In general, autosomal recessive (AR) and X-linked recessive (XLR) forms of TBD result in a more severe disease phenotype with earlier age of onset¹⁵.

Inheritance and common clinical findings of constitutional BMF.

	Fanconi anemia	Telomere disease	Schwachman diamond syndrome	Diamond-blackfan anemia	GATA2	SAMD9/L	Platelet disorders
Genes	FANC genes, BRCA2	DKC1, TERT, TERC, PARN, RTEL1, TINF2, CTC + others	SBDS	RP genes and TSR2	GATA2	SAMD9/L	c-MPL (CAMT) RBM8A (TAR) RUNX1, ETV6, ANKRD26
Inheritance	AR except for FANCB (XLR) and FANCR (AD)	AD: TERT, TERC, TINF2 AR: CTC, RTEL1 XR: DKC1	AR	AD or sporadic	AD	AD	AR: (CAMT, TAR) AD: RUNX1, ETV6, ANKRD26
Common non hematologic clinical findings	Limb abnormalities (absent radii/short thumbs) Short stature Renal anatomical defects Café au lait spots Microcephaly/microphthalmia	DC triad: (oral leukoplakia, dyskeratotic nails, reticulated skin), Pulmonary fibrosis, Liver disease (fibrosis, fatty liver), AVM, Early grey hair, Immunodeficiency, Osteoporosis	Failure to thrive/poor feeding, Steatorrhea, Recurrent infections, Skeletal abnormalities, Hepatomegaly Intellectual disability, Congenital cardiac defects, Endocrinopathy	Short stature/TUGR, Limb abnormalities (triphalangeal thumb), Cardiac defects (VSD, ASD), Cephalic malformation (microcephaly), Developmental delay	Immunodeficiency (atypical mycobacteria, recurrent warts from HPV), Lymphedema, Thrombosis, Pulmonary alveolar proteinosis (dyspnea and cough)	MIRAGE: (myelodysplasia, infection, growth restriction, adrenal hypoplasia, genital problems, enteropathy), Ataxia Pancytopenia: (cerebellar symptoms and pancytopenia), SAAD: (nodular neutrophilic panniculitis, ILD, basal ganglia calcifications, cytopenia)	CAMT: some neurological associations, possibly related to ICH TAR: skeletal defects (absent radii), cow's milk intolerance, renal tract abnormalities, cardiac defects) RUNX1: platelet function defect
Malignancy risk	MDS/AML, SCC (skin, head/neck, anogenital region)	MDS/AML SCC (skin, head/neck, anogenital region), BCC (skin)	MDS/AML	MDS/AML, Colon cancer, Osteogenic sarcoma	MDS/AML SCC (skin, anogenital region), BCC (skin)	MDS/AML	RUNX1/ETV6/ ANKRD26: MDS/AML or ALL ALL > ETV6, MDS/AML > RUNX1/ANKRD26

Abbreviations: AR; autosomal recessive, AD; autosomal dominant, XR; x-linked recessive, MDS: myelodysplastic syndrome, AML; acute myeloid leukemia, SCC; squamous cell carcinoma, BCC; basal cell carcinoma, AVM; arteriovenous malformation, VSD; ventricular septal defect, ASD; atrial septal defect, HPV; human papilloma virus, CAMT; congenital amegakaryocytic thrombocytopenia, TAR: thrombocytopenia absent radii, ICH; intracranial hemorrhage, RP; ribosomal protein.

Table 1. The inherited bone marrow failure syndromes. *Description of the common IBMFS. Taken from Groarke et. al.*¹

Telomere biology disorders clinically present most notably with bone marrow failure, pulmonary fibrosis, liver disease, and a predisposition to both haematological and solid malignancy. Patients with TBD exist along a clinical spectrum with DC phenotypes typically presenting in childhood with a more severe presentation. Biallelic gene mutations cause the most severe DC subsets; Hoyeraal-Hreidarsson (HH) is characterized by growth retardation, cerebellar hypoplasia, ataxia and microcephaly, Revesz syndrome (RS) is defined as HH with exudative retinopathy, and Coats plus (CP) is characterized by retinal vasculopathy, cerebral calcification and leukodystrophy¹. DC classically presents with the triad of oral leucoplakia, rash, and nail dystrophy, and is associated with germline variants in *DKC1* and *TINF2*. Patients with variants in *TERT* and *TERC* typically present later in life with marrow failure and pulmonary fibrosis, and the DC triad is often absent. TBD have also been associated with a number of other clinical manifestations including immunodeficiency, pulmonary arteriovenous malformations, gastrointestinal strictures and bleeding, and low bone mineral density¹⁶⁻²¹.

The risk of marrow failure in patients with DC accumulates to approximately 94% by the age of 40⁸. Patients with TBD may have chronic or progressive moderate cytopenia or may develop severe marrow failure requiring transfusion support and ultimately haematopoietic stem cell transplant (HSCT). The commonest pulmonary manifestation is pulmonary fibrosis (PF), which typically occurs in later life and causes significant morbidity and mortality²². In fact, mutations in telomere genes account for up to 30% of all hereditary pulmonary fibrosis diagnoses²³. Even prior to the development of overt pulmonary fibrosis, patients with TBD exhibit subtle abnormalities in pulmonary function testing such as reduced FEV1 and DLCO suggestive of subclinical lung disease²⁴. Liver abnormalities are present in approximately 40% of patients and may comprise abnormal liver function testing, hepatic steatosis or liver fibrosis or cirrhosis²⁵.

Recently, Niewisch et al. reported on the long-term clinical outcomes of 231 TBD patients. Patients with AD disease had a superior OS to those with AR/XLR disease (64.9 years versus 31.8 years respectively). The commonest overall reported cause of death in TBD patients was lung disease (PF or arteriovenous malformations [AVM]), followed by cancer and bone marrow failure. In those with AR/XLR disease, pulmonary disease and BMF contributed equally to mortality¹⁵.

1.4 Haematological malignancy in TBD

Cancer, both solid and haematological, is prevalent in TBD. Observed versus expected mortality (O/E) ratio (a measurement of observed cases compared to what would be expected in a general cohort of the same age and sex) for patients with DC followed long term at the National Cancer Institute (197 TBD patients) was 4.2. Commonest cancer types were myelodysplastic syndrome (MDS), acute myeloid leukaemia (AML), squamous cell carcinoma of the head and neck, and non-Hodgkin lymphoma. The O/E for MDS was very high at 578; patients had a cumulative risk of developing MDS of 15-25% by the age of 50. Acute myeloid leukaemia had a comparatively lower cumulative risk of <10%²⁶. In the Hopkins cohort of 180 TBD patients, 12.8% developed cancer. Of these, the majority were MDS or AML (75%). In this cohort, the O/E ratio for MDS was 145 and for AML was 21²⁷.

In Schratz et al. 27 patients (11.7% of the cohort) developed a malignancy – most commonly MDS or AML²⁷. MDS was commoner in AD rather than AR/XLR patients with 10.7% of AD patients, predominantly *TERC*, and 6.3% of AR/XLR. The median age at MDS diagnosis was 53.2 years in AD disease compared with 16.8 years for AR/XLR patients.

To date, the clinical, cytogenetic, and molecular profile of MDS and AML in TBD has been sparsely reported. Schratz et al. examined the characteristics of 18 TBD patients with MDS (TBD-MDS) or AML (TBD-AML)²⁷. While patients with TBD-MDS did not appear to fall under a common WHO MDS/AML subtype, some morphological commonalities were identified. The majority of TBD-MDS patients had a hypocellular marrow (compared to the approximately 10%-15% seen in standard MDS)²⁸ and for those who had normocellular or hypercellular marrows (n=7), all had ringed sideroblasts (as defined by WHO with >15% ringed sideroblasts in erythroid cells), but only 2 of 7 of these had the classically associated *SF3B1* mutation. Monosomy 7 was the commonest cytogenetic abnormality seen, and patients had a similar somatic mutation landscape molecular to that seen in standard MDS/AML.

Overall survival in TBD patients who develop MDS/AML is poor with Schratz et al. reporting a median OS of 15 months and a 2-year survival of 39%. Non-haematologic complications, in particular pulmonary disease, were the leading cause of death in this cohort rather than progression of MDS/AML²⁷.

1.5 Telomere length in cancer

Both long and short telomeres are known to increase the risk of cancer. In large population studies, polymorphisms associated with long telomere length were correlated with increased cancer risk, in particular melanoma, lung cancer, and glioma²⁹⁻³¹. In addition, long telomeres have also been seen in patients with haematological malignancies such as CLL and B-cell lymphomas^{32,33}. Short telomeres have been described in both (both TBD and non-TBD related) MDS and AML, prostate cancer, and in aplastic anaemia as a risk factor for clonal evolution to secondary MDS or AML^{34,35}.

1.6 Potential mechanisms underlying malignant transformation due to long telomeres

Telomerase is the enzyme that is responsible for telomere maintenance. Repression of telomerase leads to progressive telomere shortening and prevents uncontrolled cell proliferation. In general, cells that require ongoing replication such as stem cells, haematopoietic progenitor cells, and some epithelial cells (i.e. skin and intestines) will have high telomerase activity, while telomerase activity will be suppressed and low in other somatic cells. In the majority of cancer cells, telomerase activity is abnormally increased, which allows for ongoing proliferation and immortalisation as telomeres do not shorten. Telomerase is encoded by the reverse transcriptase TERT and telomerase RNA component (TERC)³⁶⁻³⁸. Evidence of TERT expression or telomerase activation is present in >90% of cancers.

Somatic mutations in telomere related genes, in particular, *TERT* or *POT1*, may account for the association between elongated telomeres and cancer. Activating mutations in the *TERT* promoter (*TERTp*) region were identified in >70% of sporadic melanoma cases; this mutation created a new binding motif for transcription factors, which was reported to cause a two-fold increase in telomerase transcription^{30,36}. Subsequently, *TERTp* variants have also been identified in a number of other cancers including bladder cancer, liver cancer, melanoma, and glioma³⁸. These *TERTp* variants are associated with increased expression of telomerase and elongated telomeres³⁹.

POT1 is part of the telomere shelterin complex, a conglomeration of 6 proteins which is responsible for protecting telomeres from DNA repair mechanisms, and which also play a role

in telomerase regulation. Both germline and somatic mutations in *POT1* have been associated with cancer. Somatic mutations have been seen in CLL, and in lymphoma, while germline variants have also been seen in familial cancer syndromes in CLL, as well as melanomas and gliomas^{33,40-42}. Mutations in *POT1* have been shown to lead to elongated telomeres but also increased genomic instability^{32,33,36}.

1.7 Potential mechanisms underlying malignant transformation due to short telomeres.

One of the major functions of telomeres is to ensure that the ends of chromosomes are not mistaken as DNA damage and therefore inadvertently activate the DNA damage response (DDR) pathways within the cell. The DDR is directly suppressed by the telomere shelterin complex. Key DDR-signalling components include protein kinases ATM (ataxia telangiectasia mutated) and ATR (ataxia telangiectasia and Rad3 related), which are activated by DNA double-stranded breaks and single-stranded DNA, respectively; Shelterin prevents activation of ATM and ATR⁴³.

If cell senescence signalling induced by short telomeres is overridden and cell proliferation continues, telomeres may become excessively short and dysfunctional (termed “telomere crisis”) and may be recognized as DNA breaks and thus activate the DDR. When this occurs, most cells will undergo apoptosis (thus leading to bone marrow failure), but in some cells end-to-end chromosome fusions may occur, resulting in aneuploidy, loss of heterozygosity, and chromosomal translocations⁴⁴; those cells that survive may then propagate and drive genomic instability and malignant transformation^{36,45,46}. Even a single chromosomal dysfunctional telomere in an otherwise stable genome has been shown to fuse with multiple genomic loci, a risk that increases in the absence of adequate p53 checkpoint control⁴³. Cells with DNA checkpoint damage (i.e. due to TP53 dysregulation) may be more vulnerable to genomic instability induced by short telomeres⁴⁵.

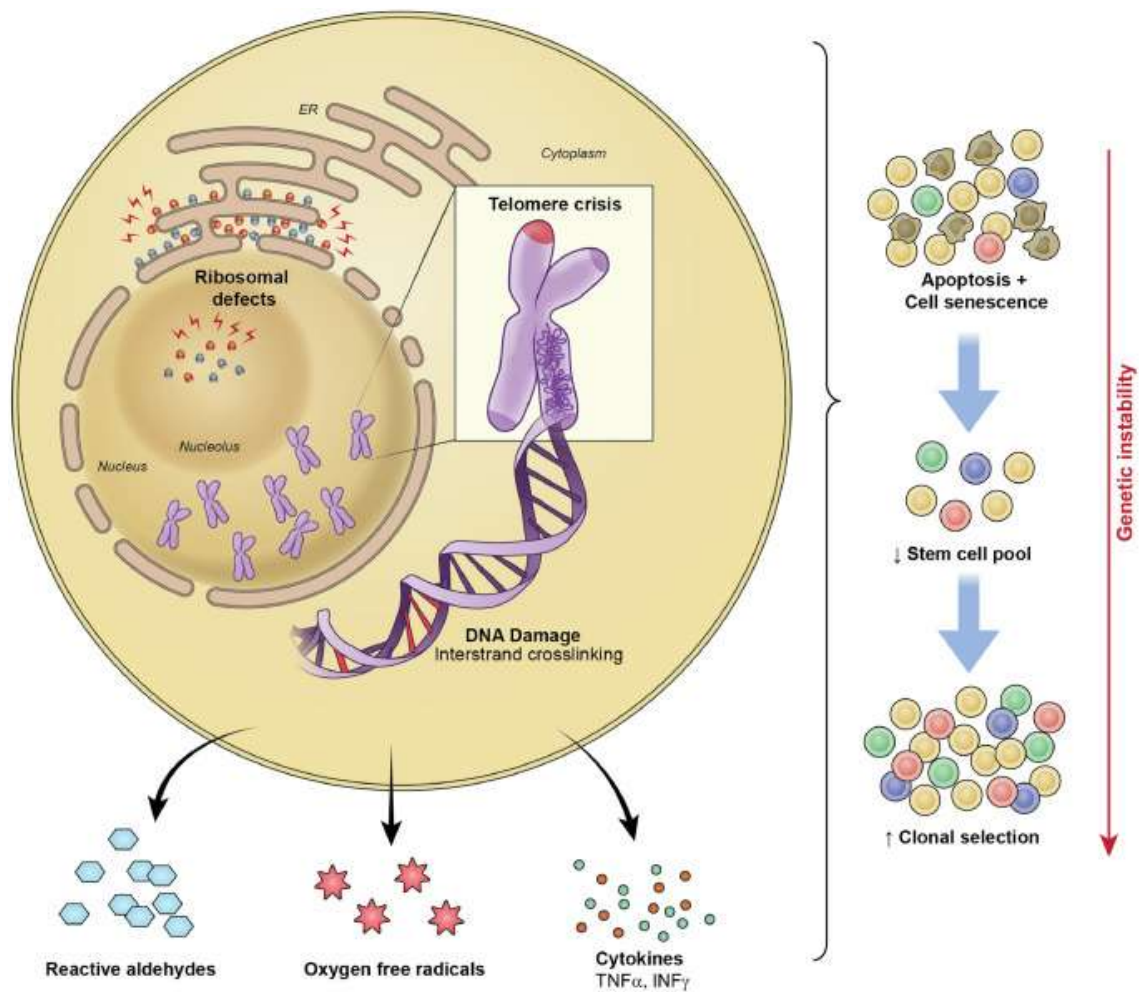


Figure 4. Effect of short telomeres and DNA damage on cells. *Ongoing DNA damage chronically activates safety pathways to halt the proliferation of damaged cells and leads to an increase in cell apoptosis and senescence. Due to a depleted stem cell pool and consequent increased replicative stress, a selective advantage for mutant clones may occur. In TBD this is driven by critically short telomeres, while in other inherited bone marrow failure syndromes with predisposition to haematological malignancy, other factors such as ribosomal defects (Shwachman Diamond Syndrome) or interstrand crosslinking leading to DNA damage (Fanconi anaemia) contribute to increased cancer risk. Taken from Groarke, EM et al. Seminars in Hematology, 2022⁴³.*

1.8 Current therapeutic approaches in TBD

The mainstay of treatment for TBD includes watch and wait, androgen therapy, or transplant. Patients with stable blood counts and little to no evidence of other organ disease are expectantly managed with appropriate follow up patient-dependent follow up per guidelines and cancer screening⁴⁷.

Patients who develop more severe cytopenia and either do not meet criteria for severe marrow failure or cannot undergo HSCT due to medical co-morbidities or lack of suitable donor may be treated with androgen therapy. Androgens may elongate telomeres and may mitigate rate of progression of pulmonary fibrosis in PF due to TBD. *In vitro* studies of peripheral blood and bone marrow samples suggest that telomere length can be modulated by sex hormones. Exposure of normal peripheral blood lymphocytes and human bone marrow derived CD34+ stem cells, as well as those from those patients with *TERT* variants, to androgens increased telomerase activity^{48,49}.

In an analysis of 16 patients on androgen therapy (oxymetholone, fluoxymesterone, and nandrolone) enrolled on a long term cohort study, which included mostly children with DC, 11 achieved clinically significant haematological responses⁵⁰. Additionally, a recent case report described marked improvement in the clinical and laboratory parameters of a child with *TINF2* mutation causing DC-associated pulmonary fibrosis after initiating danazol⁵¹.

Townsley et. al demonstrated in a prospective clinical trial that patients with telomere diseases who were treated with a high dose of danazol showed telomere elongation⁵². Twenty-seven patients were included with a median age of 41 years (17-66 years). The mutations were as follows: *TERT* (n=10), *TERC* (n=7), *DKCI* (n=3), *RTEL1* (n=1), and no mutation identified (n=6). Increases in telomere length, assessed by polymerase chain reaction (PCR) assays, were observed at 6 months (16 of 21 patients), 12 months (16 of 18 patients), and 24 months (11 of 12 patients). Haematological responses occurred in 19 of 24 patients (79%) evaluated at 3 months and in 10 of 12 patients (83%) evaluated at 24 months. This study used the highest dose of danazol, 800 mg daily, and known adverse effects, such as elevated liver enzyme levels and muscle cramps, occurred in 41% and 33% of patients, respectively. Overall, the treatment was well tolerated, but some patients did require dose reduction. The study was halted early, after all 12 patients who could be evaluated for the primary endpoint, demonstrated reduced

telomere attrition. Because of the limited power, investigators were unable to draw definitive conclusions regarding further clinical effect of danazol on lung or liver function, but stabilization or improvement was observed in a few cases of DLCO for pulmonary fibrosis and Fibroscan for cirrhosis.

A subsequent retrospective observational study comparing the rate of telomere attrition in younger patients with clinical DC who did or did not receive androgens demonstrated no significant difference in lymphocyte telomere attrition between groups. The median age of patients treated with androgens was 7 years (range 3-46 years), while the median age of patients never treated with androgens was 21 years (range 0-48 years). The majority of androgen treated patients had non *TERT/TERC* mutations: *DKC1* (n=3), *TERC* (n=1), *TINF2* (n=2), *RTEL1* (n=2), *PARN* (n=1), unknown (n=1)⁵³. Another retrospective cohort study of seven patients with TBD treated with androgen therapy (6 with danazol, 1 with low dose oxymetholone) demonstrated improvement in cytopenias and increase in lymphocyte telomere length in all patients. Duration of therapy ranged from 3 to 29 months and median age was 40 years (range 21-53 years). Three patients had *TERC* mutation, and 4 had *TERT* mutation. Median duration of therapy was 14 months (range 3-29 months)⁵⁴.

Bone marrow transplant is reserved for patients with severe marrow failure or who develop myeloid neoplasia. Outcomes had traditionally been poor with less than one third overall survival due to post transplant organ toxicity related to conditioning regimens, most commonly lung or endovascular^{55,56}. More recently, outcomes have improved with overall survival in the 60-70% range due to reduction or elimination of alkylating agents and radiation in preparative regimens, and an increasing use of fludarabine and antibody-based immunosuppressive conditioning^{57,58}. Lung transplant is increasingly utilized in TBD patients with end stage pulmonary fibrosis, though is associated with increased complications particularly cytopenias and graft dysfunction⁵⁹.

1.9 Somatic mosaicism in haematologic malignancy and disease

Somatic mosaicism that occurs in haematopoietic cells is termed clonal haematopoiesis (CH); CH has been found to be almost ubiquitous in the healthy aged population⁶⁰. Mutations in known driver genes associated with myeloid cancers have also been found in the ageing healthy population with normal blood counts – in approximately 10% of patients >60 years⁶¹. This is

termed clonal haematopoiesis of indeterminate potential (CHIP), has been associated with increased risk of both haematological and non-haematological diseases^{61,62}. It is increasingly being recognized as a pre-malignant condition predisposing to myeloid neoplasia – somewhat synonymous to MGUS and multiple myeloma, as well as monoclonal B-cell lymphocytosis and CLL. Healthy individuals with CHIP have a significantly increased risk of developing haematological malignancy, primarily myeloid neoplasia.^{61,63,64} These same driver mutations that occur in CHIP are the ones found in MDS, AML, and myeloproliferative neoplasms (MPNs).

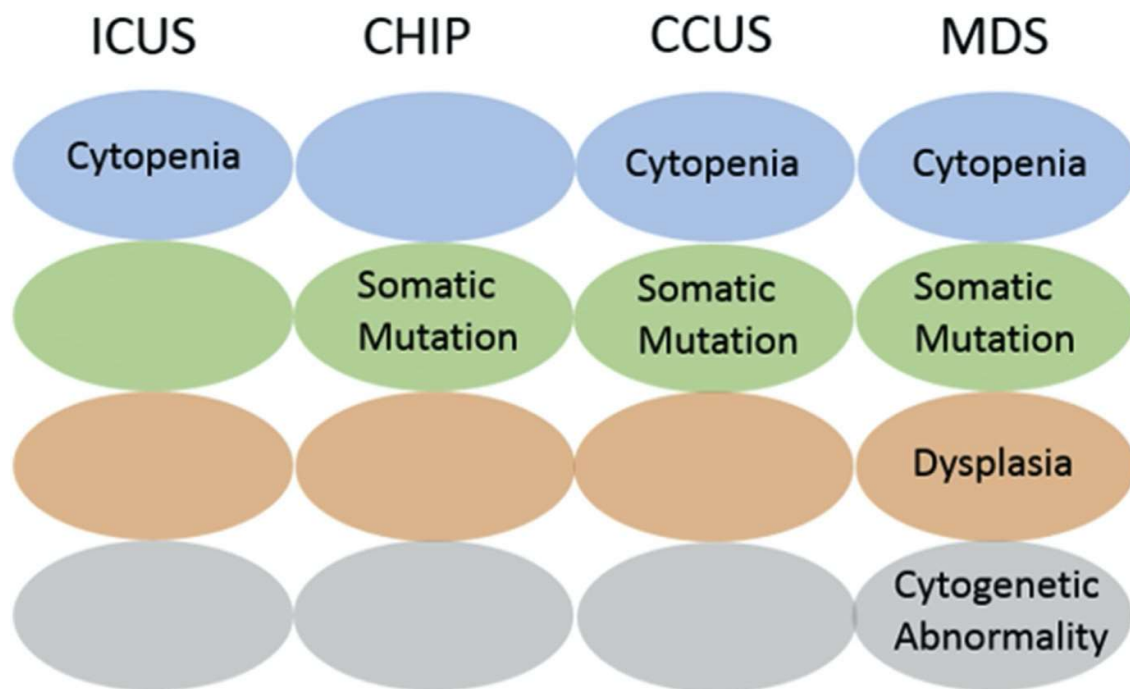


Figure 5. CHIP as compared to MDS. *Distinction between ICUS (idiopathic cytopenia of undetermined significance), CHIP (clonal haematopoiesis of indeterminate potential), CCUS (clonal cytopenia of undetermined significance) and MDS (myelodysplastic syndrome). Patients with cytopenias and an isolated cytogenetic abnormality not meeting MDS criteria may also be considered to have CCUS. Taken from Groarke EM et al. 2019⁶⁵*

Driver mutations occurring in patients who do have cytopenias but do not have a diagnosis of MDS / AML or other haematological disease is termed clonal cytopenia of undetermined significance (CCUS). In Malcovati et al., patients with CCUS were found to have a significantly higher risk of developing myeloid malignancy than cytopenic patients without

mutations⁶⁴ Overall survival of patients with CCUS who had MDS specific mutations is comparable to patients with a low-risk MDS diagnosis. Recently, the IPSS-M has been developed, incorporating mutation information into a predictor model for MDS prognosis¹¹⁶.

The risk of MDS/AML development has been shown to vary depending on which specific mutations are found. In age-related CHIP, *DNMT3A*, *ASXL1*, and *TET2*, are the commonest mutations seen, and appear to confer less risk of transformation than *TP53*, *JAK2*, and splicing factor mutations⁶⁶. In patients with AML, somatic mutations found at diagnosis, primarily *FLT3* and *NPM1*, have long played a prognostic role and guided therapy⁶⁷. In patients with MDS, somatic mutations are also increasingly being used as prognostic markers, and it is likely that in time they will be incorporated into the R-IPSS prognostic score⁶⁸. Clonal haematopoiesis is also common in immune aplastic anaemia, where somatic mutations in *BCOR/BCORL1* have been associated with longer overall survival and better response to immunosuppressive therapy⁶⁹, while splicing factor mutations (including *SRSF2*, *U2AF1* and *ZRSR2*), *ASXL1*, and *RUNX1* have been associated with progression to high risk clonal evolution when detected 6 months after immunosuppressive therapy (Groarke EM et al, *Leukemia*, In Press)⁷⁰.

1.10 Somatic mosaicism in germline marrow failure disorders

Somatic mosaicism in IBMFS is dependent on the specific germline defect and its downstream effects. When the change directly influences the effects of germline variant, it is called somatic rescue (SR). Somatic rescue can take two forms; direct somatic genetic rescue, which is direct correction or compensation of the germline variant, or indirect somatic genetic rescue, where the compensation occurs in a different gene within the same molecular network⁵. The commonest mechanism underlying SR is creation of a single nucleotide variants, with reversion being the commonest (by direct correction or the acquisition of uniparental disomy [two copies of DNA from a single parent]). Different methods of SR are outlined in Figure 6. The development of malignant clones may also occur, possibly driven by the environment or functional consequences of the germline defect.

	Direct SGR				Indirect SGR
	Single nucleotide substitution	Revertant variant	CN-LOH	Second-site variant	Second-gene variant
Autosomal dominant					
Autosomal Recessive					
Compound heterozygous					
X-linked					

Figure 6. Somatic reversion. *Types of mosaicism seen in IBMFS. SGR: somatic genetic rescue. Image taken from Gutierrez-Rodriguez et al. Best Practice and Research Clinical Haematology 2021⁵.*

One of the first examples of mosaicism in IBMFS was seen in Fanconi anaemia (FA), the prototype IBMFS caused by germline defects in DNA repair pathways (*FANC* or *BRCA* genes), when chromosome breakage assays were negative in the peripheral blood of some FA patients. This was ultimately found to be due to the presence of two lymphocytic populations in patients, one sensitive to DEB (diepoxybutane; a DNA cross-link agent used to assess for increased susceptibility to DNA damage) while the other was not due to somatic reversion. Since then, it has been realized SR is a common event in FA occurring in 15–25% of FA patients.⁷¹ SR has been associated with improved blood counts in FA but may also lead to misdiagnosis due to negative chromosome breakage testing in peripheral blood⁷².

Shwachman-Diamond Syndrome (SDS), another classical IBMFS due to defects in ribosome function, has a known high predisposition to MDS/AML. However, acquisition of 20q deletion on karyotyping was associated with a good prognosis, and while *TP53* was commonly mutated

in SDS, many patients had stable *TP53* clones over several years and did not develop MDS/AML^{73,74}. Later, it was discovered that *EIF6* mutations (The *EIF6* gene is located on chromosome 20q) were found to be recurrent in SDS and found to be an indirect SR, always found independent of *TP53*, that increased cell fitness. Additionally, while *TP53* clones could be stable over years, malignant transformation was driven by acquisition of a second *TP53* clone⁷⁵.

SAMD9/L syndrome is a newly described IBMFS associated BMF and MDS with monosomy 7. An interesting feature of this disease is a non-random loss of chromosome 7 containing the mutated *SAMD9/9L* allele⁷⁶. As SAMD9/L syndrome is due to a gain of function mutation, loss of the mutated allele is considered a rescue event. However, acquisition of monosomy 7 causes genomic instability and drives malignant transformation; this is considered a maladaptive form of rescue⁷⁷.

Described somatic mosaicism events in TBD have been primarily direct somatic or indirect reversion of the germline variant, primarily somatic mutations seen in the *TERT* promotor, which are present in >10% of TBD patients^{78,79}. Other somatic mutations in telomere related genes have also recently been described by Schratz et al. and include *POT1* and *TERF2IP*⁸⁰. Typical CHIP mutations appear to be rare in TBD, but few patients overall have been sequenced²⁷. Overall, no large-scale studies have occurred to date looking at the landscape of clonal haematopoiesis in TBD, and no clinical associations have been described.

1.11 Overall aims and objectives

The overall aim of this research is: 1) to determine the clonal landscape of somatic mutations in TBD 2) to understand whether somatic mosaicism contributes to disease outcomes in TBD patients as has been shown in other IBMFS and 3) see if somatic mutations predict for the development of haematological malignancy. If somatic mutations are found to be present, we plan to identify whether specific mutations, in particular those typically associated with myeloid malignancy, can predict for the presence or future development of haematological cancer, or if they impact overall survival.

These aims will be carried out as follows:

1. Clinically characterize haematological malignancy in the TBD cohort including the baseline rate of evolution, risk factors of evolution, and description of MDS/AML cases
2. Screen whole blood leukocytes from patients with TBD for recurrent somatic mutations in myeloid cancer genes and telomere-related genes using a targeted ECS panel
3. Correlate the presence of specific somatic mutations with development of haematological cancer and with overall survival.
4. Use single-cell analysis to determine the clonal trajectory of multiply mutated patients, in particular those who developed MDS or AML.
5. Investigate whether androgen therapy alters somatic mutation patterns in patients with TBD.

2. Chapter 2: Clinical characteristics of the TBD cohort

2.1 Introduction

As only a few publications exist characterizing the progression to MDS/AML in TBD patients, we first planned to clinically phenotype and determine the incidence of MDS/AML in our TBD study cohort. Two of the major publications that have looked at the incidence of MDS/AML patients in TBD (and in other IBMFS), from Alter et al in *Haematologica* 2018²⁶ and Alter et al. *British Journal of Haematology* 2010⁸¹, contain patients from one of the cohorts we have included in this study (the National Cancer Institute Inherited Bone Marrow Failure Syndrome Cohort). However, in these prior studies including this cohort, specific MDS or AML features such as karyotype, blast counts, and molecular markers were not reported.

In the report by Schratz et al. of the Johns Hopkins Telomere Syndrome Registry²⁷, 18 MDS/AML cases (14 MDS, 3 AML and one treatment related AML [prior ALL]) were characterized; key findings included low marrow cellularity, presence of ringed sideroblasts, and predominance of the monosomy 7 karyotype. Somatic mutations in myeloid cancer genes were present in 11/14 patients but no one gene mutation predominated. There was no association seen between specific germline variants and development of MDS/AML in this study (though an association was seen between *DKC1* and solid malignancy). However, in the NCI cohort, it was previously noted that the majority of patients who had developed MDS or AML had germline *TERC* variants¹⁵.

Given the overall rarity of TBD and sparse data in the literature, we wished to first confirm whether our incidence of MDS/AML was similar to previous reports and to determine the detailed clinical phenotype of those who developed haematological malignancy.

2.2 Methods

A total of 208 patients with a diagnosis of TBD were enrolled from 3 sites: National Institutes of Health Clinical Center, Bethesda, Maryland (n=163; [n=79 from the National Heart, Lung and Blood Institute, and n=84 from the National Cancer Institute]), MD Anderson Cancer Center, Houston Texas (n=2), and University of Sao Paulo, Brazil (n=43). Patients were enrolled on either a clinical trial using androgen therapy (Clinicaltrials.gov NCT03312400,

NCT02055456, NCT01441037), or on long-term cohort studies (Clinicaltrials.gov NCT05012111, NCT00027274); otherwise, specimens were collected as part of an institutional biorepository protocol. Healthy controls (n=140) at a variety of age ranges were sequenced using the same panel as a control group. Full details regarding registered protocols used in this research are available in Appendix A. All patients or their guardians gave written informed consent for use of their samples and data for research. Ethical approval was granted by the Institutional Review Boards of the 3 institutions.

2.2.1 Definition of TBD used for inclusion in the study

Patients with a diagnosis of TBD were included. This was defined as either:

1. Presence of a pathogenic or likely pathogenic mutation according to ACMG criteria in a telomere related gene associated with TBD (including *TERT*, *TERC*, *TINF2*, *NOP10*, *NHP2*, *WRAP53*, *RTEL1*, *PARN*, *ACD*, *POT1*, *CTC1*, *STN1*, *NAF1*, and *ZCCHC8*).

Or

2. Presence of a variant of uncertain significance (VUS) according to ACMG criteria in a telomere related gene associated with TBD or mutation negative TBD with telomere length <10th percentile for age using Flow-FISH and a strong clinical suspicion of telomere by having at least **2** of the following criteria:
 - a. Cytopenia (Hb <10g/dL or ANC <1.5 x10⁹/L or platelets <100 x10⁹/L)
 - b. Documented fatty liver or liver fibrosis on histology or abnormal liver US or fibro scan (>7.0 kPa)
 - c. Pulmonary fibrosis confirmed by histology or CT chest or PFTs showing FEV1 <80% or DLCO <60%
 - d. Family history of: marrow failure, myeloid malignancy, liver fibrosis or lung fibrosis in a first or second degree relative

Clinical data were collected and included: Age (date of birth and age of first institutional assessment), sex, clinical disease status, presence of haematological or solid organ neoplasia, treatment and duration of therapy with androgens, need for transplant (haematopoietic stem cell, liver or lung transplant), and survival.

Clinical disease status assessment included the presence and degree of bone marrow failure, clinical triad of dyskeratosis congenita (oral leucoplakia, nail dystrophy, reticular skin pigmentation), presence of family history, and presence and degree of lung or liver disease.

2.2.2 Definition of bone marrow failure and myeloid malignancy used in the study

Bone marrow failure was defined by the following criteria:

Severe Aplastic Anaemia:

Hypocellular bone marrow (<30% cellularity) with no evidence for other haematological disease processes causing marrow failure, and cytopenias in at least **two out of three** blood counts as follows:

- Neutrophil count < $0.5 \times 10^9/L$
- Platelet count < $20 \times 10^9/L$
- Absolute reticulocyte count < $60 \times 10^9/L$

Moderate Aplastic Anaemia:

Hypocellular bone marrow for age with no evidence for other haematological disease processes causing marrow failure, and severe cytopenia in at least **two out of three** blood counts as follows:

- Neutrophil count < $1.2 \times 10^9/L$
- Platelet count < $70 \times 10^9/L$
- Anaemia with haemoglobin < 8.5 g/dL or absolute reticulocyte count < $60 \times 10^9/L$ in transfusion-dependent patients

BMF not meeting criteria for AA:

Presence of a hypocellular marrow for age with at least one lineage cytopenia below normal reference values.

Myeloid neoplasia was defined using the 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukaemia⁸².

2.2.3 Definition of lung disease used in the study

Pulmonary involvement was defined as:

Pulmonary fibrosis diagnosed by either a lung biopsy or computed tomography scan of the chest according to guidelines from the American Thoracic Society and European Respiratory Society⁸³.

Pulmonary dysfunction without confirmed pulmonary fibrosis was defined as pulmonary function testing showing:

- FEV1 <80% or
- DLCO <60%

2.2.4 Definition of liver disease used in the study

Liver involvement was defined as follows:

Hepatic fibrosis diagnosed by:

- Transient Elastography > 10 kPa by Fibroscan
Or
- US evidence of cirrhotic liver
Or
- Liver biopsy demonstrating fibrosis

Or non-fibrotic liver abnormalities defined as:

- Documented fatty liver with abnormal liver US or fibro scan (>7.0 kPa) or CAP score >260 dB/m
- Liver function testing abnormalities persistently above the upper limit of normal with no evidence for other disease processes causing liver dysfunction

2.2.5 Other clinical definitions used in the study

The family screening cohort included patients who did not have overt clinical features of TBD but were screened due to having a known affected family member with a known gene mutation.

Pathology review of bone marrow aspirate and biopsy specimens occurred at the primary clinical site with clinical reports used to extract data on diagnosis of myeloid neoplasia. Chromosomal karyotyping was performed at a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory (or international equivalent). Data on solid malignancy were obtained from medical records at the participating clinical sites or from the database of an NIH natural history protocol as previously mentioned.

Germline variant testing in telomere related genes was performed using a CLIA certified sequencing assay (or international equivalent) and collected from clinical reports from patients' medical records or from the database of an NIH natural history protocol.

Somatic variant testing was performed at NIH using research ECS (full details available in Chapter 2). Germline variants were also confirmed using research ECS performed at NIH.

2.2.6 Statistical methods

Summary statistics of baseline characteristics were calculated using median and range for continuous variables. Survival probabilities were evaluated using Kaplan-Meier estimates. Cumulative incidence curves were used to display the probability of clonal evolution at timepoints. Analysis was performed and figures generated using R “ggplot2” package.

2.3 Results:

2.3.1 General characteristics of the cohort

A total of 208 TBD patients were included in clinical analysis; 173 probands and 35 family members who were found to have a pathogenic mutation. No patients had disease-causing mutations in other IBMFS genes. Median age of the entire cohort was 26 years (range 1 – 76 years) including both probands and family members. Telomere length was <1st percentile in

170 patients and <10th percentile but ≥1st percentile in 21 patients; TL was not documented in 17 patients (primarily family members with known pathogenic variants). Disease characteristics of the entire cohort are listed in Table 2.

Table 2. Disease characteristics of the TBD cohort									
	Total	Proband			Family Members			Inheritance	
	<i>n</i>	<i>n</i>	Median Age	Range	<i>n</i>	Median Age	Range	AD (<i>n</i>)	AR (<i>n</i>)
All cohort	208	173	17	1 - 76	35	36	4 - 62	156	36
Genotype									
<i>TERT</i>	62	51	35	3 - 72	11	38	11 - 62	59	3
<i>TERC</i>	39	35	34	12 - 59	4	16	11 - 60	39	0
<i>RTEL1</i> *	38	24	15	1 - 76	14	35	3 - 61	23	15
<i>TINF2</i>	24	24	11	2 - 71	0	--		24	0
<i>DKC1</i>	13	13	13	3 - 47	0	--		4	9
<i>PARN</i>	8	4	18	6 - 28	4	51	45 - 53	5	3
<i>CTC1</i>	2	2	25	16 - 33	0	--		0	2
<i>ACD</i>	3	1	3	--	2	20	4 - 36	2	1
<i>WRAP53</i>	3	3	15	15 - 32	0	--		0	3
Unknown	16	16	30	2 - 67	0	--		--	--
Phenotype									
DC/HH	26	26	11	2 - 41	0	--		16	5
BMF									
SAA	30	30	13	1 - 45	0	--		18	10
MAA	66	65	25	1 - 68	1	58	--	40	16
Cytopenia	65	64	27	2 - 72	1	17	--	50	8
MDS/AML	17	17	40	6 - 68	0	--		12	3
Lung fibrosis	47	46	43	3 - 75	1	58	--	38	7
Liver	20	19	35	17 - 61	1	58	--	16	3
Solid Cancer	15	14	45	19 - 55	1	61	--	10	4
Abnormal cytogenetics	25	24	37	6 - 68	1	3	--	16	4
AD; autosomal dominant, AR; autosomal recessive, DC; dyskeratosis congenita, HH; Hoyeraal Hreidarsson, BMF; bone marrow failure, SAA; severe aplastic anaemia, MAA; moderate aplastic anaemia, MDS; myelodysplastic syndrome, AML; acute myeloid leukaemia. *26 patients with RTEL1 had biallelic disease.									

The majority of patients had *TERT* (n=62) or *TERC* (n=39) germline variants; these patients had an older median age than the rest of the cohort (35 years for *TERT* mutated patients and 34 years for *TERC*). In paediatric patients, *TINF2*, *DKC1*, and biallelic *RTEL1*, were the commonest variants observed (Table 2). The median overall survival for the cohort was 60 years.

The commonest clinical manifestation was BMF; this was present in 163 patients (77% of the entire cohort). Lung fibrosis was also common, present in 48 patients (23% of the entire cohort). The median age of patients with features of the DC triad or severe BMF (SAA) was younger than patients with milder BMF, lung or liver fibrosis, or haematological or solid malignancy.

2.3.2 General characterization of haematological and solid malignancy in the cohort

Haematological malignancy was present in 8% (n=17) of the cohort; a total 14 patients developed MDS and 3 patients developed AML (Tables 3 + 4). The median age of MDS onset was 38 years and AML onset was 29 years. Of those who developed MDS/AML, germline variants were: *TERC* (n=7), *TERT* (n=3), and *DKC1* (n=3), none (n=2), *RTEL1* (n=1) and *PARN* (n=1). Of those who developed MDS, eight had an abnormal karyotype, four a normal karyotype, and two were unknown. Abnormal karyotypes included: Del 20q (n=3), and chromosome 1 abnormalities (n=5). Of those who developed AML, two had an abnormal karyotype, both complex, and one had a normal karyotype. Of note, one patient with chromosome 1 abnormalities had monosomy 7 detected by FISH that was absent on karyotype. We had iron stains for review in 8/17 patients; none had evidence of ringed sideroblasts.

Fifteen patients (7% of total cohort) developed solid organ malignancy, most commonly head and neck squamous cell cancer (Table 3).

Table 3. Malignancy in the TBD cohort		
Haematologic Malignancy		
	Number of patients	% of cohort
Total	17	
Myelodysplastic Syndrome	14	

Blasts >5%		
<i>Yes</i>	2	14%
<i>No</i>	12	86%
Abnormal karyotype		
<i>Yes</i>	8	57%
<i>No</i>	4	29%
<i>Unknown</i>	2	14%
Acute Myeloid Leukaemia	3	
AML n, (%)	3	100%
Solid Organ Malignancy		
Total	15	
HNSCC	7	47%
Lower GI	2	13%
Skin	3	2%
Gynaecological	3	2%
HNSCC; head & neck squamous cell cancer, AML; acute myeloid leukaemia		

Table 4. MDS in the TBD cohort					
UPN	Age at MDS onset	Germline variant	Blasts ≥ 5	Karyotype at time of MDS	Somatic mutations
MDS-1	41	<i>TERC</i>	No	Normal	<i>U2AF1</i> , <i>PPM1D</i> , <i>RUNX1</i>
MDS-2	47	<i>DKC1</i>	No	46,XY,del(20)(q11.2;q13.3)[2]	<i>U2AF1</i>
MDS-3 ^u	53	<i>TERT</i>	Yes	46,XY,dup(1)(q21q32)[5]/47,idem,+8[6]	<i>U2AF1</i>
MDS-4*	26	<i>TERC</i>	No	46,XX+1,(der1;22)(q10;q10)[20]	<i>U2AF1</i> , <i>ETV6</i>
MDS-5	40	<i>TERC</i>	No	47,XY,+i(1)(q10)[14]/48,idem,+21[5]/46,XY[1]	<i>ZRSR2</i> , <i>ETV6</i> , <i>RUNX1</i>
MDS-6	26	<i>TERC</i>	No	47,XY,+i(1)(q10)[2]/46,XY[18]	<i>KIT</i>

MDS-7	29	<i>TERC</i>	No	Normal	None
MDS-8**	40	None	Yes	46,XY,del(20)(q11.2)[20]	<i>U2AF1</i>
MDS-9	42	<i>TERC</i>	No	Unknown	<i>U2AF1</i> , <i>SF3B1</i> , <i>PPM1D</i> , <i>ETV6</i>
MDS-10	33	<i>TERC</i>	No	Unknown	None
MDS-11	13	<i>PARN</i>	No	47, XY, +der(1)i(1)(q10)inv(q10;q25)[20]	<i>RUNX1</i>
MDS-12	6	<i>DKC1</i>	No	Normal	None
MDS-13	65	<i>TERT</i>	No	46 XY,del(20)(q11.2)[13]/46,XY[7]	None
MDS-14	35	<i>TERT</i>	No	Normal	None

Table shows data from time of MDS diagnosis.

UPN; unique patient number, MDS; myelodysplastic syndrome

*Subsequent to MDS diagnosis

**Prior to MDS/AML diagnosis

μMDS-3 later progressed to MDS-EB2, acquired subsequent clones in *GATA2*, *ASXL1*, *KRAS*, and in *SETBP1*. At time of progression to MDS-EB2 monosomy 7 was detected by FISH but not karyotype.

Table 5. AML in the TBD cohort

UPN	Age at AML onset	Germline variant	Blasts %	Karyotype at time of AML	Somatic variants
AML-1	66	<i>RTEL1</i> (heterozygous)	No	Complex	TP53*
AML-2	27	<i>DKC1</i>	No	Complex	None**
AML-3	28	None	Yes	Unknown	PPM1D**

UPN; unique patient number, AML; acute myeloid leukaemia

*Subsequent to AML diagnosis

**Prior to AML diagnosis

Table 6. Somatic variants at the time of, preceding, or subsequent to MDS/AML diagnosis

UPN	Variant	HGVSp	VAF
MDS-1	<i>U2AF1</i>	p.Ser34Phe	0.2096
	<i>PPM1D</i>	p.Ser468GlnfsTer15	0.0602
	<i>RUNX1</i>	p.Arg204Ter	0.1633
MDS-2	<i>U2AF1</i>	p.Ser34Phe	0.4644
MDS-3	<i>U2AF1</i>	p.Gln157Arg	0.2883
MDS-4*	<i>U2AF1</i>	p.Gln157Arg	0.0233
	<i>ETV6</i>	p.Arg353Pro	0.0256
MDS-5	<i>ZRSR2</i>	p.Glu93LysfsTer15	0.0455
	<i>ZRSR2</i>	p.Leu237Ter	0.2204
	<i>ETV6</i>	p.Arg369Gln	0.1797
	<i>RUNX1</i>	p.Ser141Ter	0.2919
MDS-6	<i>KIT</i>	p.Asp737Tyr	0.0274
MDS-7	No mutations		
MDS-8**	<i>U2AF1</i>	p.Ser34Phe	0.2613
MDS-9	<i>ETV6</i>	p.Gln57Ter	0.1526
	<i>ETV6</i>	p.Arg359Leu	0.0539
	<i>PPM1D</i>	p.Ala481ProfsTer2	0.0506
	<i>SF3B1</i>	p.Lys700Glu	0.0087
	<i>U2AF1</i>	p.Ser34Tyr	0.0053
MDS-10	No mutations		
MDS-11	<i>RUNX1</i>	p.Arg166Ter	0.0152
MDS-12	No mutations		
MDS-13	No mutations		
MDS-14	No mutations		
AML-1***	<i>TP53</i>	p.Pro75LeufsTer48	0.048
	<i>TP53</i>	p.His193LeufsTer53	0.044
AML-2**	No mutations		
AML-3**	<i>PPM1D</i>	p.Glu475IlefsTer6	0.0112
	<i>PPM1D</i>	p.Asn477ProfsTer2	0.0834
	<i>PPM1D</i>	p.Lys480AsnfsTer2	0.0393
	<i>PPM1D</i>	p.Glu525LysfsTer14	0.0398

Mutational data displayed at time of MDS or AML diagnosis unless data were not available in which case available data prior or subsequent to diagnosis are shown

UPN; unique patient number, MDS; myelodysplastic syndrome, AML; acute myeloid leukaemia

*Subsequent to MDS diagnosis

**Prior to MDS/AML diagnosis

*** Post AML induction; baseline also showed TP53 but VAF not available.

Somatic variants in the table show only those assessed using ECS at NIH.

Somatic mutation testing at the time of MDS/AML diagnosis (Table 6) was available in 12 patients while the other 5 patients had testing prior to (n=3) or subsequent to (n=2) diagnosis. The commonest mutations seen in patients with MDS were: *U2AF1* (n=6), *ETV6* (n=3), and *RUNX1* (n=3). MDS-3 later progressed to MDS-EB2, acquired subsequent clones in *GATA2*, *ASXL1*, *KRAS*, and in *SETBP1*. MDS-4 had a diagnosis of MDS 17 years prior to somatic mutation testing, while AML-1 had testing performed 6 months after diagnosis (post AML induction chemotherapy) which showed 2 *TP53* clones; in this case baseline testing done externally had also shown *TP53* though VAF is not available. MDS-8 had *U2AF1* detected 6 months prior to a diagnosis of MDS-EB2, while AML-2 had no clones detected 3 months prior to AML diagnosis and AML-3 had just a small *PPM1D* clones detected 1 month prior to AML.

2.3.3 Cumulative incidence of haematological malignancy in the cohort

The cumulative incidence of MDS/AML for the entire cohort was 1.03% by age 20, 6.99% by age 40, and 16% by age 60 (Figure 7).

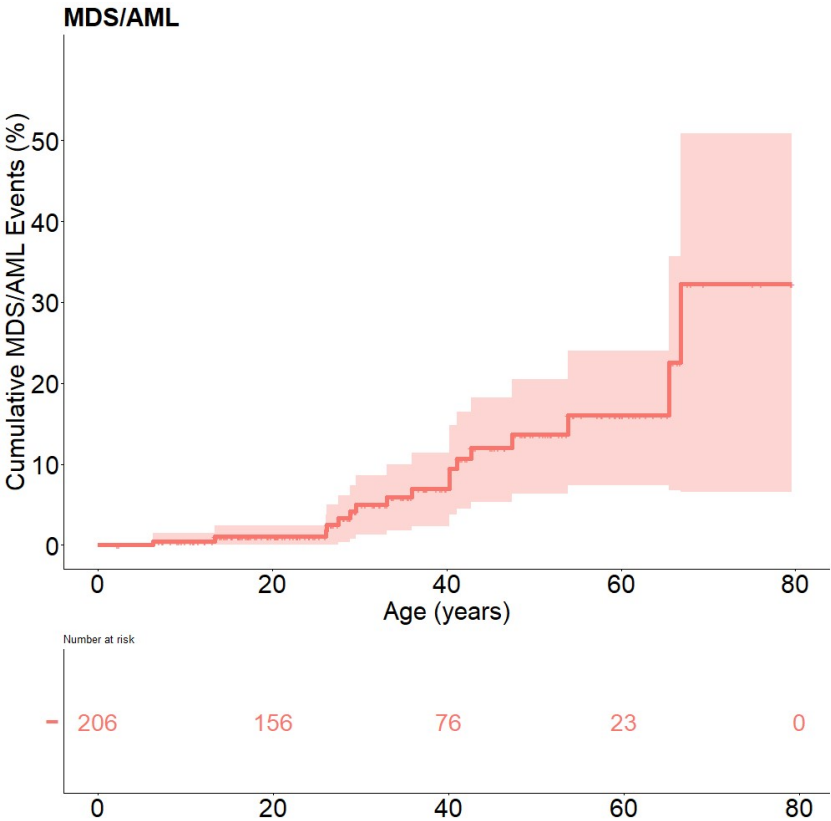


Figure 7. Cumulative risk of developing MDS/AML in all patients with TBD

Risk of MDS/AML was different according to the germline variant the patient had with the highest cumulative incidence seen in patients with *DKC1*, *TERC*, and heterozygous *RTEL1*. Patients with *DKC1* had a cumulative incidence of MDS/AML of 7.7% at age 20 and 23% at age 40. Patients with *TERC* had a cumulative incidence of 0% by age 20, 15.6% by age 40 and 32.97% by age 60. Only one patient with heterozygous *RTEL1* developed MDS/AML who was >60 years.

In contrast, patients with *TINF2* and biallelic *RTEL* germline variants had a very low incidence of MDS/AML; no one with *TINF2* or biallelic *RTEL1* developed MDS/AML in our cohort.

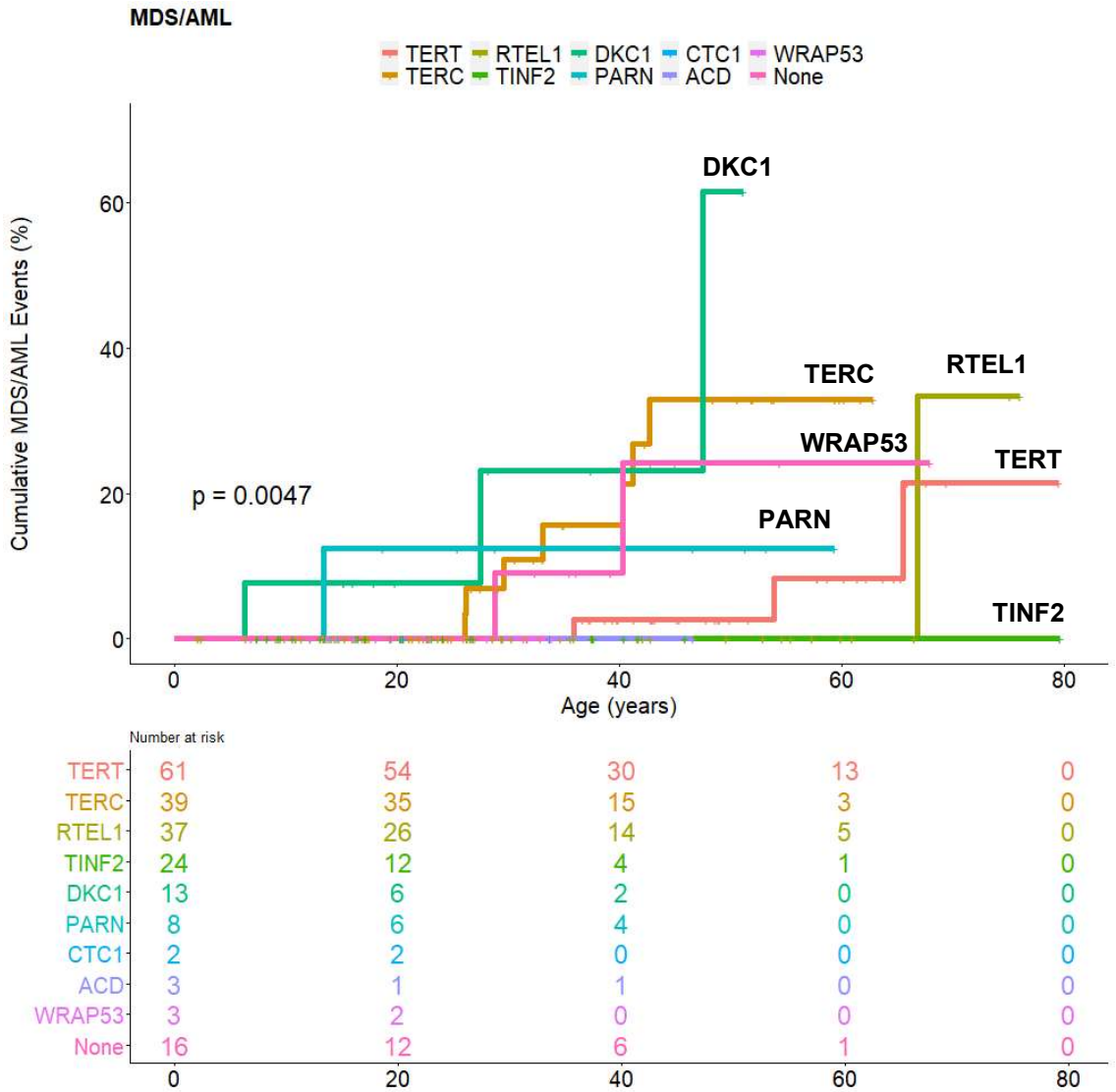


Figure 8A. Cumulative risk of developing MDS/AML in patients with TBD depending on specific germline variant. Germline variants are segregated by colour as listed and number at risk is given for each timepoint. Y axis represents cumulative incidence of MDS/AML and X axis represents time (age in years). P value of 0.0047 denotes that there was a significant difference in MDS/AML incidence between at least one pair of germline variants.

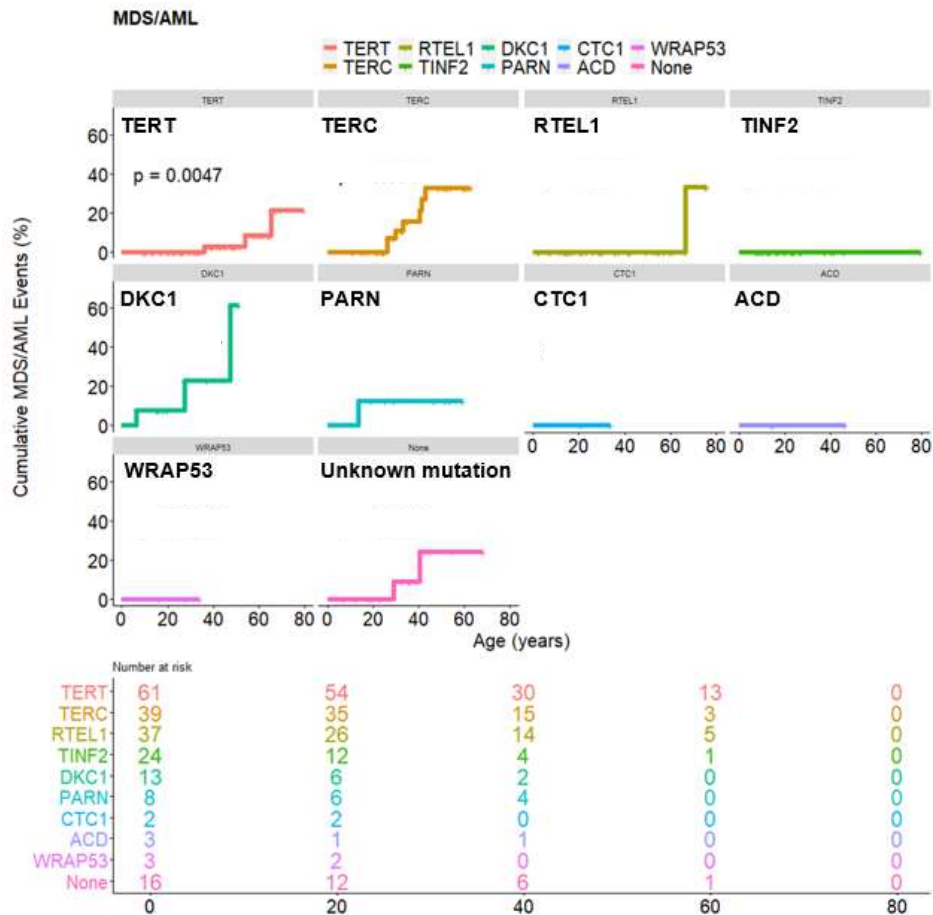


Figure 8B. Cumulative risk of developing MDS/AML in patients with TBD depending on specific germline variant. Germline variants are segregated by colour as listed and number at risk is given for each timepoint. Graphs are separated compared to figure 8A for ease of interpretation. Y axis represents cumulative incidence of MDS/AML and X axis represents time (age in years). P value of 0.0047 denotes that there was a significant difference in MDS/AML incidence between at least one pair of germline variants.

When the two commonest germline variants in our cohort *TERC* and *TERT* were compared, patients with *TERC* had a significantly increased risk of developing MDS/AML compared to those with *TERT* variants ($p=0.003$) (Figure 9). In addition, MDS/AML events occurred at a younger age in those with *TERC* variants with median age of onset of 33 years compared to 54 years in those with *TERT*.

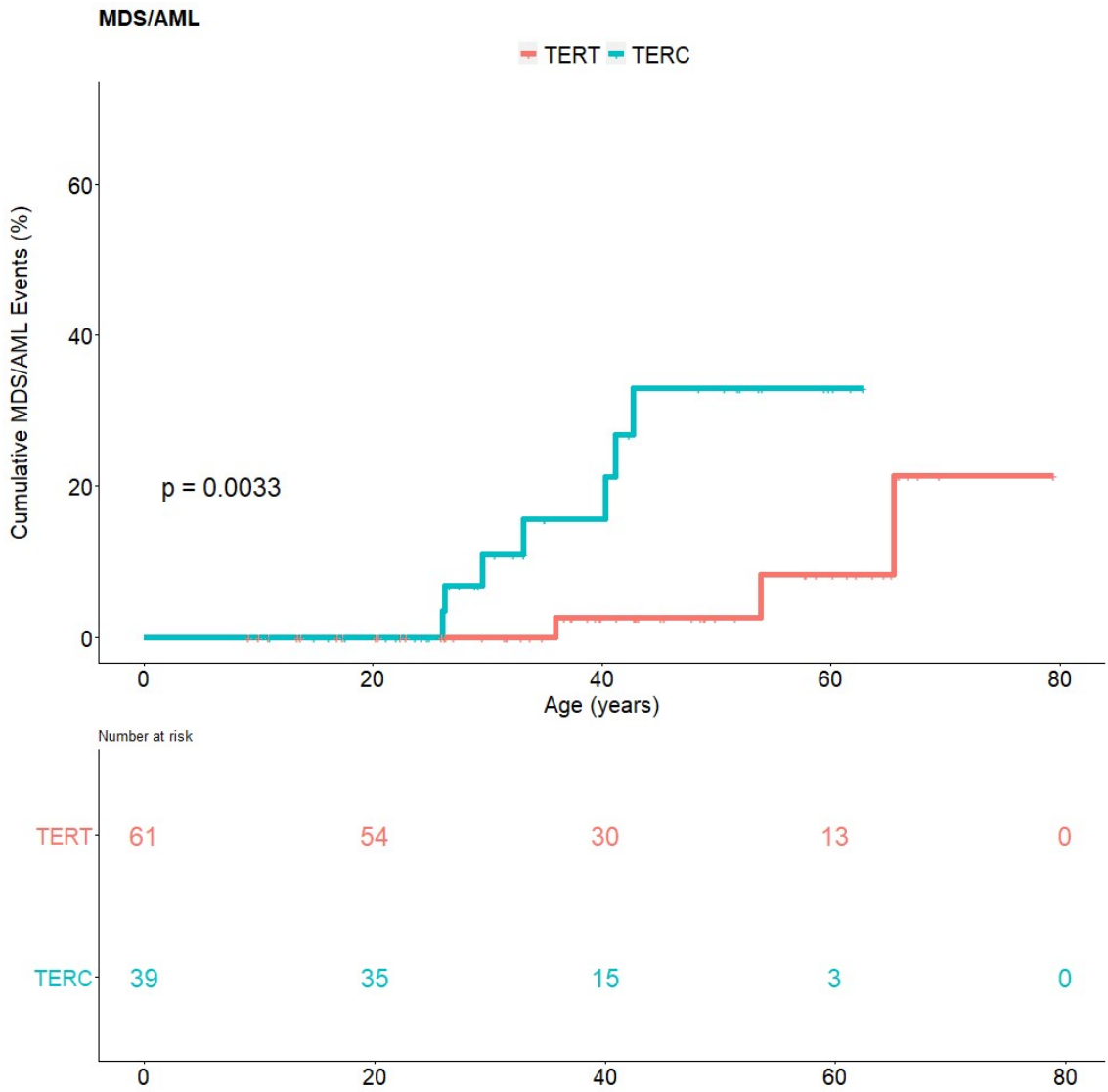


Figure 9. Cumulative incidence of MDS/AML in *TERT* or *TERC* mutated patients. *Cumulative risk of developing MDS/AML in patients with TBD in TERT compared with TERC variants*

2.4 Chapter Conclusions

We demonstrated an overall similar cumulative incidence of MDS/AML in our cohort to what has been described previously. However, the clinical phenotyping of our MDS/AML cases differs to what was described in the Hopkins cohort. In our study, we did not observe any patients with a monosomy 7 karyotype (though one had monosomy 7 detected by FISH at time of evolution to MDS-EB2). Instead, we observed a high rate of chromosome 1 abnormalities (mostly duplications in 1q [1q+] and translocations) and, also deletion of 20q. Duplication of 1q is also commonly found in Fanconi anaemia and there is hypothesized to relate to either a susceptibility of the chromosome to DNA damage or possibly is due to a compensatory mechanism driven by the marrow environment⁸⁴. Recently, the *MDM4* oncogene, which a negative modulator of p53 response located in the minimal 1q duplicated region, was found to be overexpressed in Fanconi anaemia cells with 1q+ but not in those with clonal non-1q FA cells, suggesting that 1q+ may attenuate FA-associated p53 activation through increased *MDM4*⁸⁵. As patients with TBD also have activated DDR triggered by critically short telomeres, 1q+ may result as a consequence of this activation.

It is unclear why few patients in our cohort had monosomy 7 but may be due to a difference in clinical phenotype between the two cohorts. The median age of our cohort was much younger than was reported in Schratz et al. (26 years compared to 50 years), and our cohort included paediatric patients as young as 1 year old compared to Schratz et al. where the youngest was 12 years²⁷. Additionally, their cohort contained primarily patients with either *TERT* variants or without known germline variants while in our cohort we had a more mixed picture with a high proportion of patients with *TERC*, *TINF2*, *DKC1*, and biallelic *RTEL1*.

There was a striking germline genotype association with progression to MDS/AML with *TERC* and *DKC1* patients incurring the highest risk. When *TERC* and *TERT*, the two commonest TBD genotypes, were assessed, patients with *TERC* variants had a significantly increased risk of haematological malignancy. Conversely, no patients with *TINF2* or biallelic *RTEL1* developed MDS or AML. This is the first time such a significant genotype association has been correlated with haematological cancer risk and this finding has implications for both cancer surveillance and genetic counselling for patients with *TERC* and *DKC1* germline variants.

3. Chapter 3: The clonal landscape of somatic mutations in TBD

3.1 Introduction

Expansion of clones likely represents fitness of a particular clone in a competitive environment. Clonal haematopoiesis in the IBMFS includes direct somatic reversion of the germline variant, indirect compensation for the germline defect, or somatic mutations in other genes that ultimately result in subversion of cell pathways, such as senescence. Such clones may either result in phenotypic rescue or may be maladaptive and lead to malignant clonal evolution.

Examples of beneficial clonal haematopoiesis in IBMFS include the well-established direct somatic reversion that occurs in Fanconi anaemia resulting in improved blood counts^{72,86}, as well as the *EIF6* clones recently described in SDS that improve ribosomal maturation and do not result in increased risk of MDS/AML⁷⁵. Conversely, the classic maladaptive response seen in the IBMFS is the non-random loss of chromosome 7 in *SAMD9/9L* syndromes that occurs in the mutated *SAMD9/9L* allele⁷⁷.

In TBD, clonal haematopoiesis has been demonstrated both in telomere related genes and in typical CHIP and myeloid cancer genes. The *TERT* promoter mutation (*TERTp*) is present in >10% of TBD patients; it has not been associated with either telomere elongation or malignant transformation^{78,79}. Other somatic mutations telomere related genes also recently described by Schratz et al. include *POT1* and *TERF2IP*, as well as the previously described *TERTp*; these somatic reversion events as a group were not associated with progression to MDS/AML^{79,80}.

More typical CH has also been seen in TBD patients; 30% of the non-MDS/AML cohort sequenced (n=20) in Schratz et al had mutations in CH or myeloid cancer genes, most commonly in *TP53* and *TET2*²⁷. In the MDS/AML TBD patients, CH or myeloid cancer gene mutations were present in 11 of 14 (79%) assessed patients though there was no discernible pattern detected. In a separate cohort of 13 non-MDS/AML affected and 2 MDS/AML affected TBD patients reported by Kirchner et al., only one had a somatic mutation, in *U2AF1*, which occurred in patient without MDS or AML⁸⁷.

However, to date very few TBD patients overall have undergone sequencing for somatic mutations. Our aim is to sequence a large cohort of TBD patients for mutations in both telomere

related genes and in CH / myeloid cancer genes to fully delineate the spectrum of somatic mutations in this disorder.

3.2 Methods

All patients included in the clinical analysis who had samples with adequate available DNA were included (n=208). Fifty-six (27%) patients had serial (>1 timepoint) samples available.

3.2.1 Protocol for error corrected sequencing

Error-corrected DNA sequencing:

ECS was performed on DNA extracted from peripheral blood leukocytes. The following 60 genes were sequenced:

Table 7. Error-corrected DNA sequencing customized panel with genes related to telomere biology, clonal haematopoiesis, and myeloid malignancies			
Gene	Accession	Genomic Location	Target Exons
<i>ACD</i>	NM_022914	chr16:67691414-67694718	All coding region
<i>ASXL1</i>	NM_015338	chr20:30946146-31027122	11, 12, 13
<i>BCOR</i>	NM_001123385	chrX:39910498-39956719	All coding region
<i>BCORL1</i>	NM_021946	chrX:129139163-129192058	All coding region
<i>BRAF</i>	NM_004333	chr7:140433812-140624564	6, 10, 11, 12, 13, 15, 16, 17
<i>CALR</i>	NM_004343	chr19:13049413-13055304	8, 9
<i>CBL</i>	NM_005188	chr11:119076985-119178859	8, 9
<i>CEBPA</i>	NM_004364	chr19:33790839-33793430	All coding region
<i>CSF3R</i>	NM_172313	chr1:36931643-36948915	10, 14, 15, 16, 17, 18
<i>CTC1</i>	NM_025099	chr17:8128138-8151413	All coding region
<i>CUX1</i>	NM_001913	chr7:101459183-101927250	All coding region
<i>DNMT3A</i>	NM_022552	chr2:25455829-25564784	All coding region
<i>DKC1</i>	NM_001363	chrX:153991030-154005964	All coding region
<i>ETV6</i>	NM_001987	chr12:11802787-12048325	All coding region
<i>EZH2</i>	NM_004456	chr7:148504463-148581441	All coding region
<i>FLT3</i>	NM_004119	chr13:28577410-28674729	8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21
<i>GATA1</i>	NM_002049	chrX:48644981-48652717	2, 4
<i>GATA2</i>	NM_032638	chr3:128198264-128212030	All coding region
<i>GNAS</i>	NM_000516	chr20:57466425-57486250	8, 9, 10, 11
<i>IDH1</i>	NM_005896	chr2:209100952-209118910	3, 4
<i>IDH2</i>	NM_002168	chr15:90627211-90645708	4, 6
<i>JAK2</i>	NM_004972	chr9:4985244-5128183	12, 13, 14, 15, 16, 19, 20, 21, 22
<i>JAK3</i>	NM_000215	chr19:17935592-17958841	3, 11, 15, 16
<i>KIT</i>	NM_000222	chr4:55524094-55606881	1, 2, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18
<i>KRAS</i>	NM_004985	chr12:25358179-25403854	All coding region
<i>MDM4</i>	NM_002393	chr1:204485506-204527248	All coding region
<i>MECOM</i>	NM_004991	chr3:168801286-169381563	All coding region
<i>MPL</i>	NM_005373	chr1:43803474-43820135	All coding region
<i>NAF1</i>	NM_138386	chr4:164049822-164088073	All coding region
<i>NHP2</i>	NM_017838	chr5:177576464-177580961	All coding region
<i>NOP10</i>	NM_018648	chr15:34633916-34635362	All coding region
<i>NPM1</i>	NM_002520	chr5:170814707-170837888	11
<i>NRAS</i>	NM_002524	chr1:115247084-115259515	All coding region
<i>OBFC1</i>	NM_024928	chr10:105637317-105678045	All coding region
<i>PARN</i>	NM_002582	chr16:14529556-14724128	All coding region

<i>PHF6</i>	NM_032335	chrX:133507341-133549321	All coding region
<i>POT1</i>	NM_015450	chr7:124462439-124570037	All coding region
<i>PPM1D</i>	NM_003620	chr17:58677543-58743640	6
<i>PTPN11</i>	NM_002834	chr12:112856535-112947717	3, 4, 7, 8,11, 12, 13
<i>RPA1</i>	NM_002945	chr17:1733272-1802848	All coding region
<i>RTEL1</i>	NM_032957	chr20:62289162-62328544	All coding region
<i>RTEL1</i>	NM_001283009	chr20:62326833-62327003	All coding region
<i>RUNX1</i>	NM_001754	chr21:36160097-36421462	All coding region
<i>SAMD9</i>	NM_017654	chr7:92728825-92747336	All coding region
<i>SAMD9L</i>	NM_152703	chr7:92759367-92777680	All coding region
<i>SETBP1</i>	NM_015559	chr18:42531699-42532155	4
<i>SF3B1</i>	NM_012433	chr2:198256697-198299771	13, 14, 15, 16
<i>SRSF2</i>	NM_003016	chr17:74730196-74733493	All coding region
<i>STAG2</i>	NM_006603	chrX:123095555-123236505	All coding region
<i>TERC</i>	NR_001566	chr3:169482397-169482848	All coding region
<i>TERT</i>	NM_198253	chr5:1253286-1295162	All coding region
<i>TET2</i>	NM_001127208	chr4:106067841-106200960	All coding region
<i>TINF2</i>	NM_001099274	chr14:24708850-24711880	All coding region
<i>TP53</i>	NM_001276696	chr17:7571719-7590868	All coding region
<i>U2AF1</i>	NM_006758	chr21:44513065-44527688	2, 6, 7
<i>UBA1</i>	NM_003334	chrX:47053200-47074527	3
<i>WRAP53</i>	NM_018081	chr17:7591666-7606820	All coding region
<i>WT1</i>	NM_000378	chr11:32409321-32457081	All coding region
<i>ZCCHC8</i>	NM_017612	chr12:122956145-122985543	All coding region
<i>ZRSR2</i>	NM_005089	chrX:15808573-15841382	All coding region
Genes were classified : grey= telomere related genes; blue = MDS/AML related genes; green = CH genes			

DNA libraries were prepared with 50ng of DNA for ECS using the ArcherDX platform for Illumina. See Appendix B for full step-by-step DNA library preparation protocol.

Protocol Overview

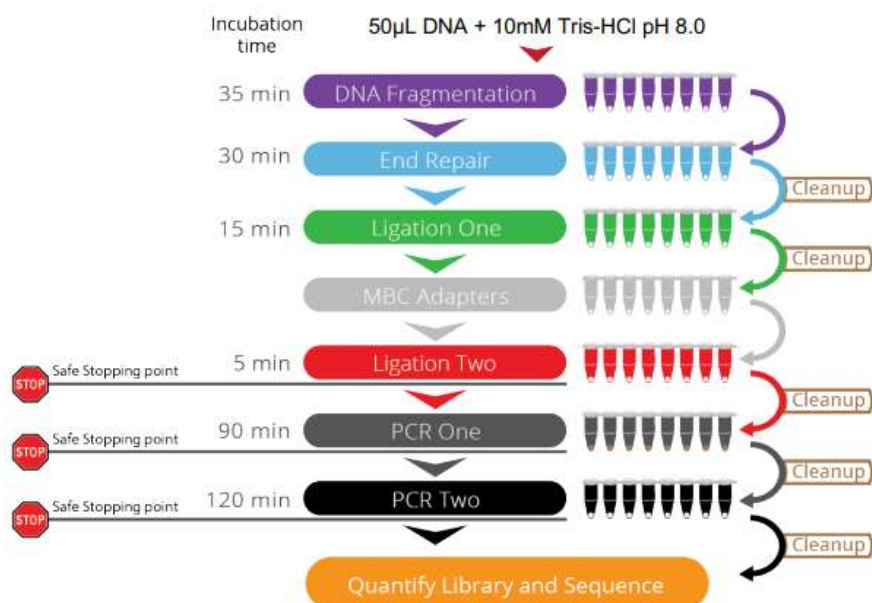


Figure 10. Protocol overview for ArcherDx ECS platform. Taken from: ArcherDX website⁸⁸

The DNA libraries were sequenced in the NovaSeq 6000 (Illumina, CA) at average coverage of 6000x.

Data were analysed using the Archer Analysis software Version 6.2.3, which used Bowtie2 to align de-duplicated reads to the reference genome and FreeBayes, Lofreq, and Vision for variant calling. Somatic variants were included in the analysis if at variant allele frequency (VAF) > 0.5%, with de-duplication ratio > 3:1, and at VAF above the minimum detectable allele fraction at which a variant could be distinguished from the background noise at statistical power of 0.95 (95 MDAF provided by the Archer analysis). Variants were next prioritised according to their predicted functional consequences and quality control statistics regarding unique molecular identifiers (UMI), strand bias, and outlier detection.

Patients were determined to be mutated or unmutated depending on the presence or absence of any somatic mutation in a tested gene (Table 7). A VAF of 0.5% was used as a cut off as this was the sensitivity of the ECS technique used.

3.2.2 Statistical methods

Age and clinical phenotype data were generated using histograms and density lines. The Wilcoxon signed-rank test was used to determine significance as data were non-parametric and population based. Survival probabilities were evaluated using Kaplan-Meier estimates. Analysis was performed and figures generated using R “ggplot2” package.

3.3 Results

A total of 84 patients (40%) had somatic mutations while 124 (60%) did not. The commonest mutations seen were: *PPM1D* (n=25 patients), *TERTp* (n=22 patients), *U2AF1* (n=17 patients), and *POT1* (n=16 patients). *PPM1D* mutations were all truncating and in exon 6, and the *TERTp* and *U2AF1* were in recurrent hotspots (*TERTp* c.-124G>A and c.-57A>C; *U2AF1* p.Ser34Phe and p.Gln157Arg). Germline variants were also confirmed. The full spectrum of somatic mutations in the TBD cohort is presented in the oncoprint displayed in Figure 12.

3.3.1 Association of somatic mutations and age

Patients who had mutations were older, with a median age of 38, compared to 27 years in those without mutations (Figure 11A). CH frequency in TBD at any age range was higher than that in HI (from 0 to 20 years: 22% vs. 0, $P=0.0359$; from 21 to 40 years: 44% vs. 4%, $P=0.0001$; from 41 to 70 years: 55% vs. 16%, $P<0.001$; older than 70 years: 75% vs. 42%, $P=0.31$) (Figure 11B).

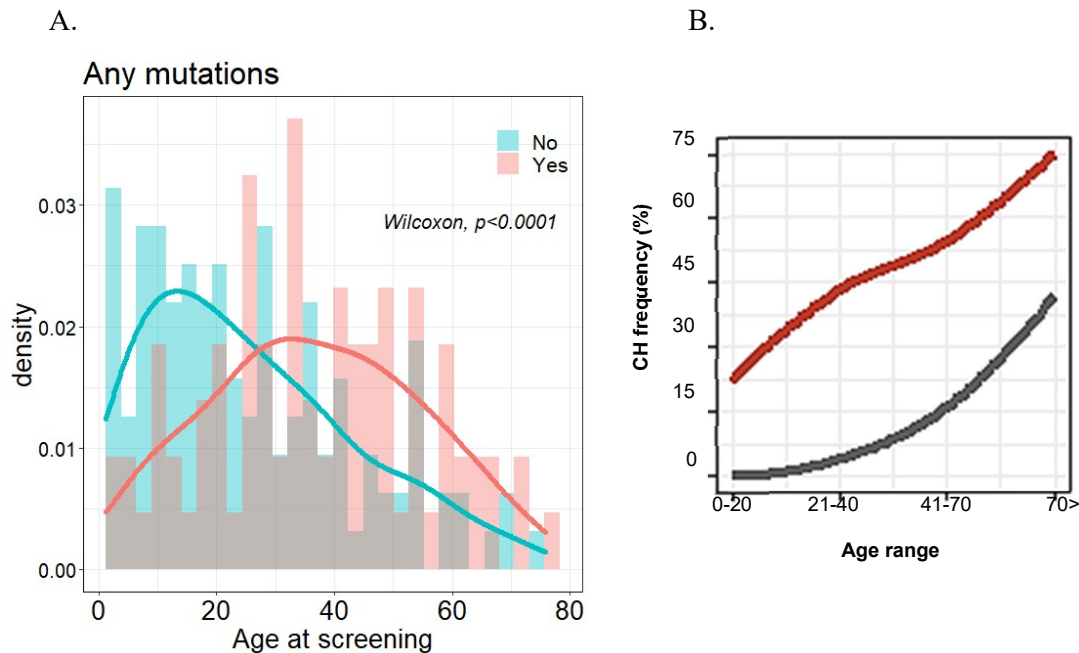


Figure 11. Association of somatic mutations and age. A) Histogram with overlaying density lines showing significant difference in age of patients with and without mutations ($P<0.0001$) B) CH frequency (at minimum VAF of 0.5%) relative to age ranges in TBD patients (red curve) and 140 age-matched healthy individuals (HI, grey curve) screened using the same panel.

3.3.2 Oncoprints showing spectrum of somatic mutations for the cohort organized by germline genotype

Using an oncoprint, it became clear that certain germline variants had different clonal landscapes (Figure 12).

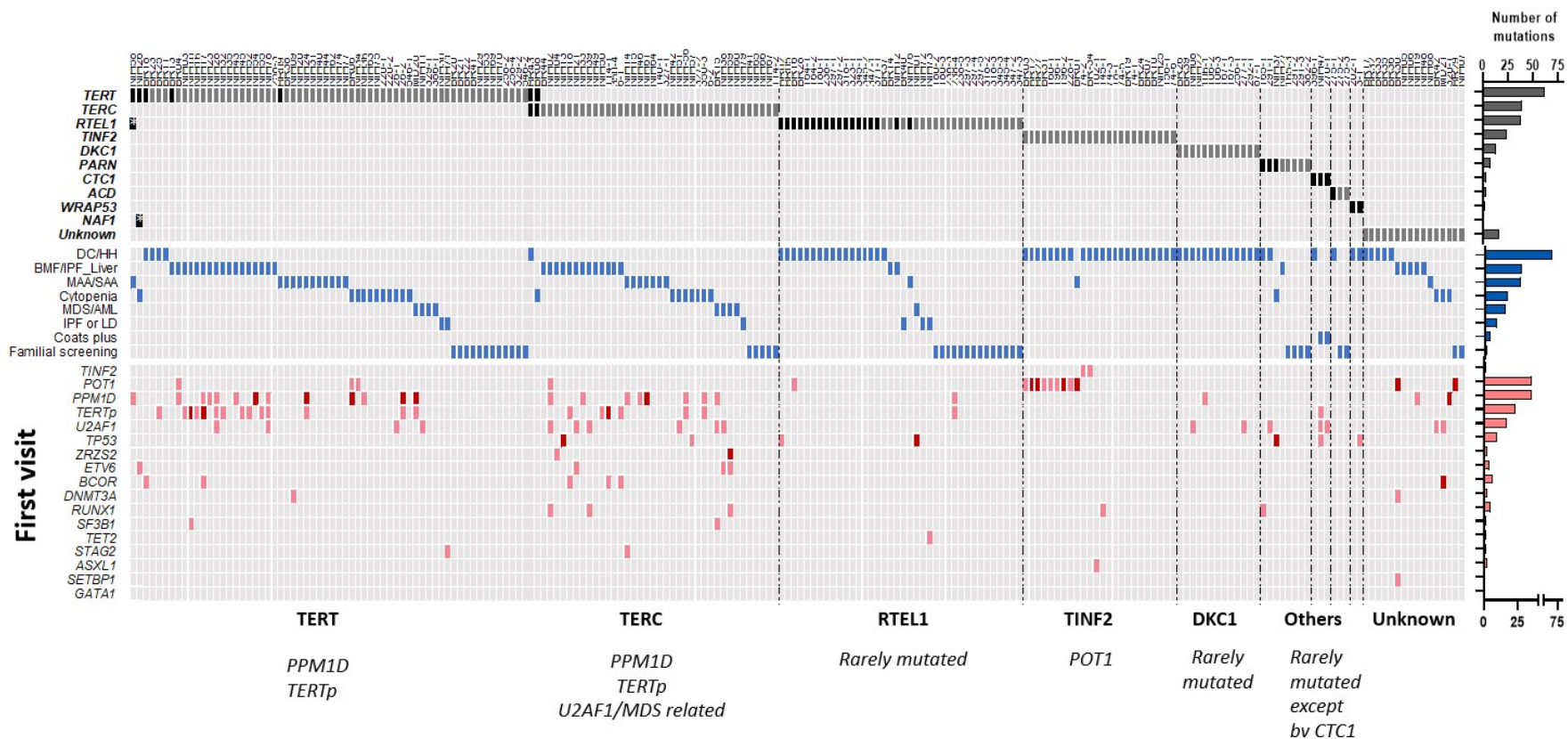


Figure 12. Oncoprint showing spectrum of CH in TBD including all germline variants. *Oncoprint showing the spectrum of somatic mutations by germline variant in patients at first assessment. Individual patients are listed in columns with associated germline mutation(s), clinical phenotype, and somatic mutations. Germline mutations are denoted in black (homozygous variants) or grey (heterozygous). Patients' clinical phenotype is denoted in blue. Somatic mutations are denoted in pink (single variant) or red (>1 variant). * denotes that the germline variant is of uncertain significance.*

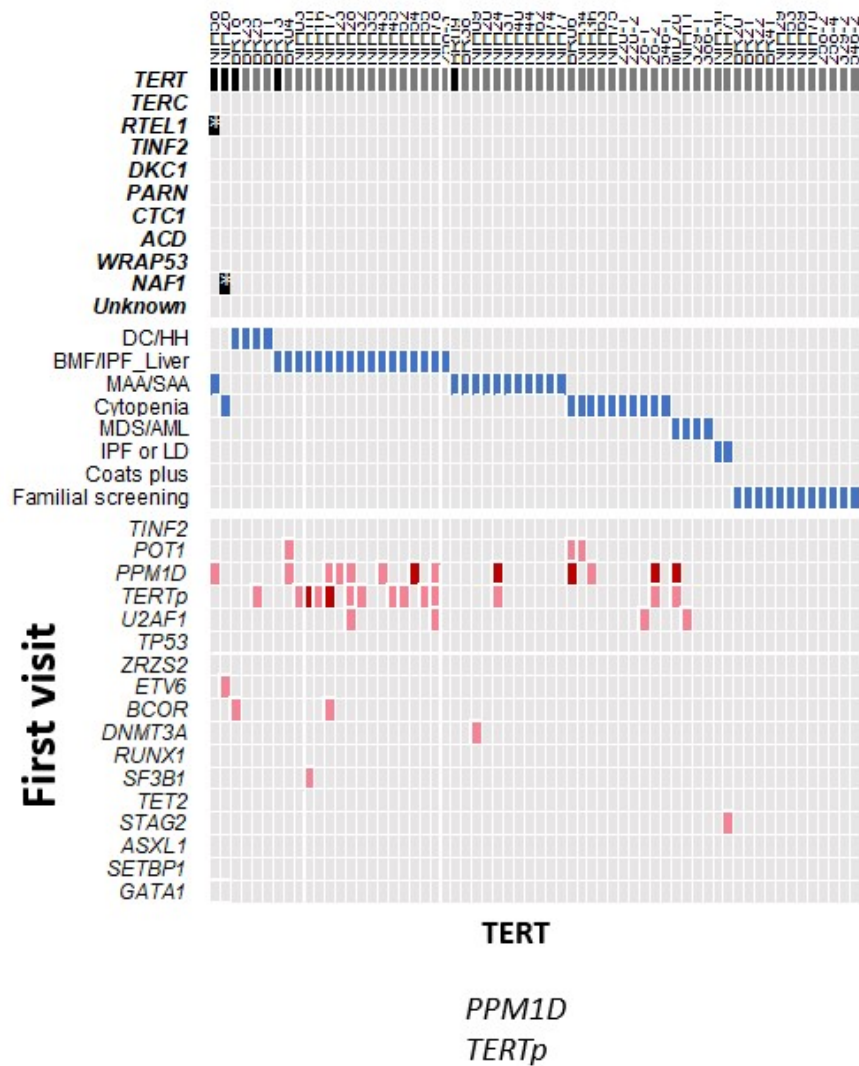


Figure 12A. Oncoprint showing the clonal landscape of *TERT* mutated TBD patients at time of first visit. Patients with germline *TERT* predominantly had *PPM1D* and *TERTp* somatic mutations. Germline variants, clinical presentation and somatic variants are listed on X-axis.

In patients with *TERT* germline variants, the commonest somatic mutations seen were in *TERTp*, (n=14 patients [23%]), *PPM1D* (n=13 patients [21%]), and less commonly in *U2AF1* (n=4 patients [6%]) (Figure 12A). Patients with *TERT* had a mixed clinical phenotype.

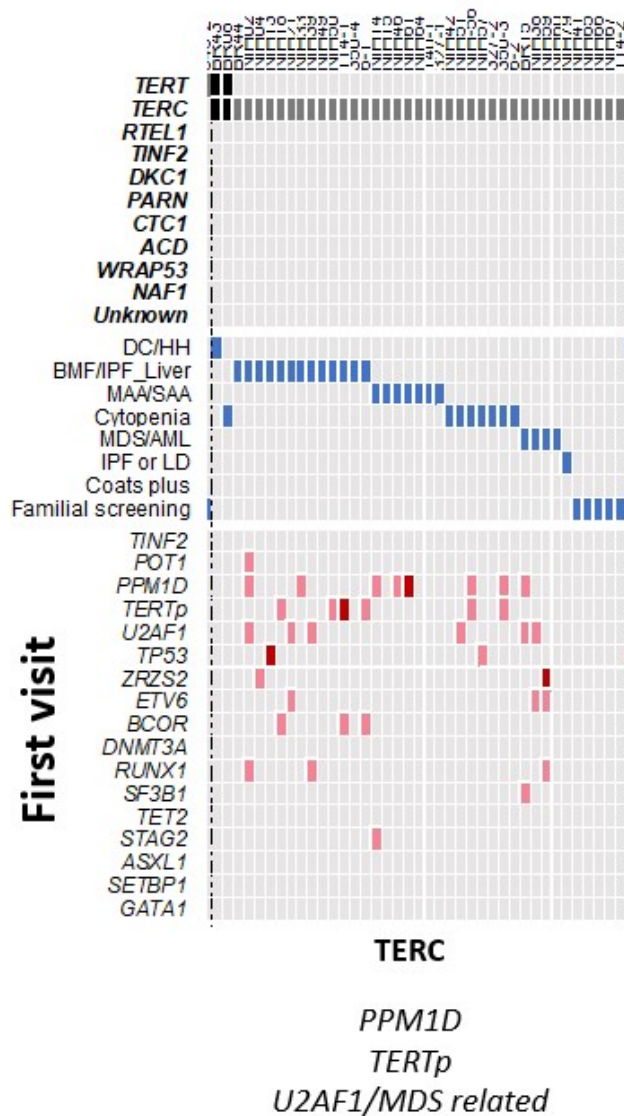


Figure 12B. Oncoprint showing the clonal landscape of *TERC* mutated TBD patients at time of first visit. The landscape is predominated by *PPM1D*, *TERTp* and *U2AF1* and MDS related genes. Germline variants, clinical presentation and somatic variants are listed on the X-axis.

In patients with *TERC* germline variants, the commonest somatic mutations seen were in *TERTp*, (n=6 patients [15%]), *PPM1D* (n=8 patients [20%]), and *U2AF1* (n=6 patients [15%]). Patients with *TERC* also had a mixed clinical phenotype (Figure 12B).

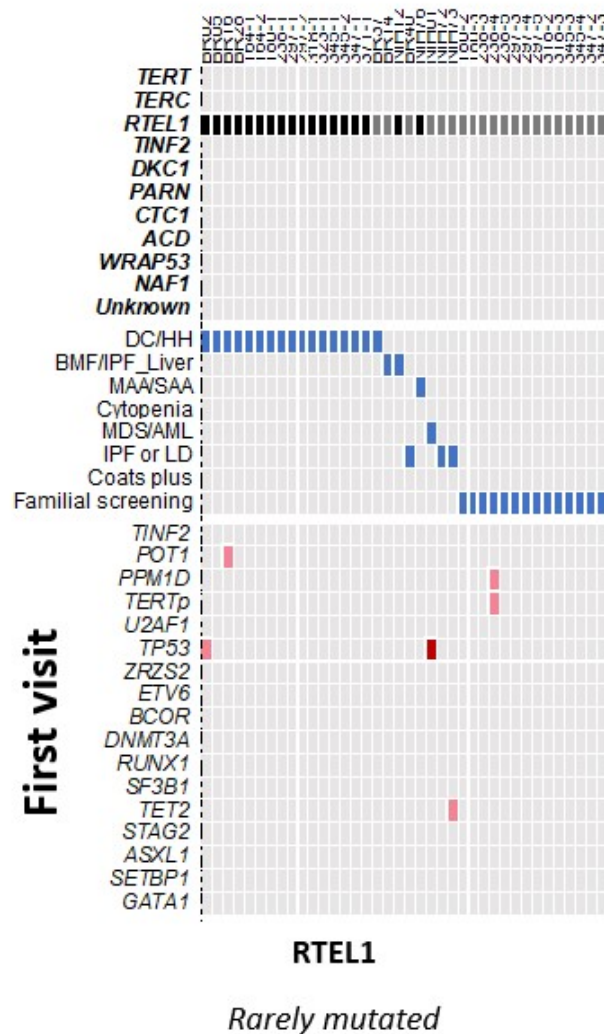


Figure 12C. Oncoprint showing the clonal landscape of *RTEL1* mutated TBD patients at time of first visit. Black denotes biallelic *RTEL1* and grey heterozygous *RTEL1*. Patients with *RTEL1* rarely had somatic mutations. Germline variants, clinical presentation and somatic variants are listed on the X-axis.

In general, somatic mutations were rare in *RTEL1* patients; in those with biallelic *RTEL1* germline variants, 2 patients had somatic mutations, one in *TP53* and one in *POT1*. In heterozygous *RTEL1*, 3 patients had somatic mutations. One had 2 *TP53* clones (had a diagnosis of AML), one had *TET2*, and one had both *PPM1D* and *TERTp* (Figure 12C).

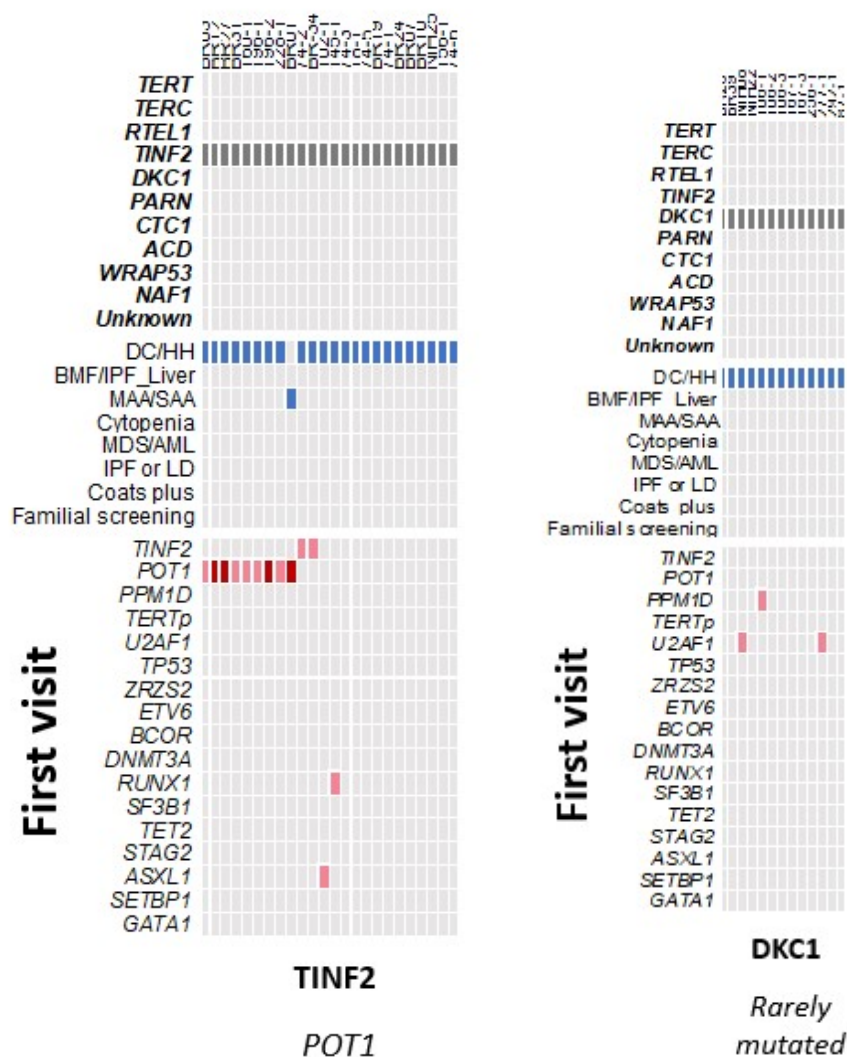


Figure 12D. Oncoprint showing the clonal landscape of *TINF2* mutated TBD patients at time of first visit and **Figure 12E.** Oncoprint showing closer view of the clonal landscape of *DKC1* mutated TBD patients at time of first visit. Patients with *TINF2* variants commonly had *POT1* somatic variants, while those with *DKC1* rarely had somatic variants. Germline variants, clinical presentation and somatic variants are listed on the X-axis.

The landscape of patients with *TINF2* germline variants was dominated was somatic *POT1* mutations which were present in 9 patients (38%) (Figure 12D). Two patients also had direct

somatic reversion of *TINF2*. Other somatic mutations were rare. The clinical phenotype of these patients was primarily the classic DC features. Patients with *DKC1* germline variants rarely had somatic mutations. One patient had *PPM1D* and 2 had *U2AF1*. The clinical phenotype of these patients was primarily the classic DC features (Figure 12E).

Other germline variants had low numbers of patients and showed no clear pattern of somatic mutations on the oncoprint (Figure 12).

3.3.3 Association with the presence or absence of somatic mutations and peripheral blood counts and clinical phenotype

Platelets (Figure 13) and absolute lymphocyte (Figure 14) were significantly correlated with mutation status while haemoglobin ($p=0.0667$) and neutrophil counts ($p=0.14$) were not. The presence of a somatic mutation was associated with a lower platelet count ($P=0.01$) and absolute lymphocyte count ($p<0.001$).

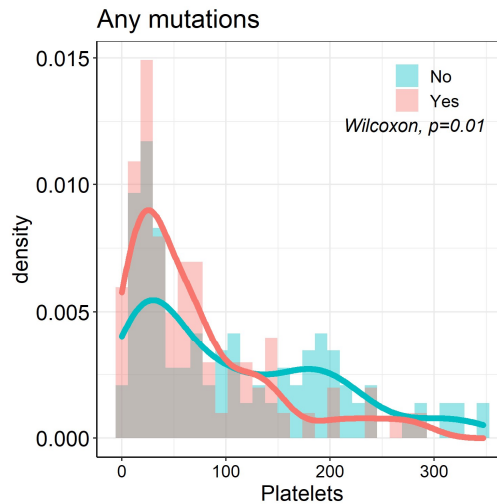


Figure 13. Association with any somatic mutation and platelets. *Histogram with overlaying density lines showing a significant difference in platelets of patients with and without mutations ($P=0.01$)*

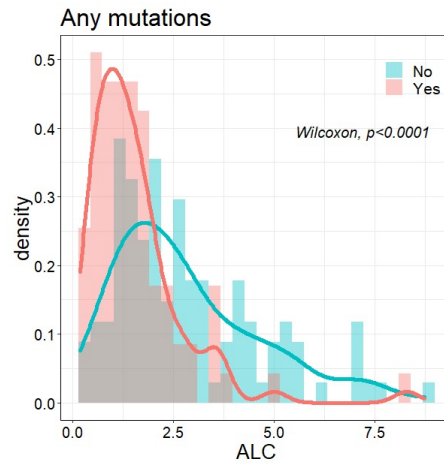
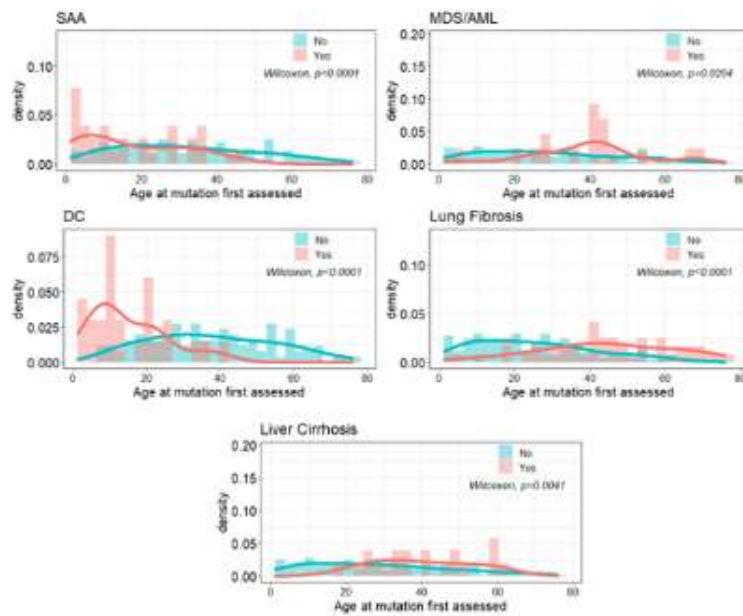


Figure 14. Association with any somatic mutation and absolute lymphocyte count. *Histogram with overlaying density lines showing a significant difference in absolute lymphocyte count of patients with and without mutations ($P < 0.0001$)*

Initial presentation with SAA, MDS/AML, DC, or liver or lung fibrosis was significantly associated with having a somatic mutation while presenting with milder cytopenias was not (Figure 15A&B).

A



B

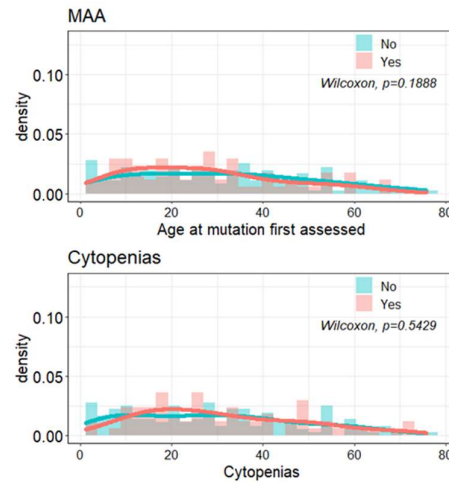


Figure 15A&B. Association with any clinical phenotype and mutation status. *Age distribution for different clinical phenotypes depending on mutation status using a histogram with overlaying density lines. A) Significant clinical phenotypes included: SAA, MDS/AML, DC, lung fibrosis, liver fibrosis. B) MAA and other cytopenias were non-significant.*

3.3.4 Variant allele frequency of somatic mutations

The VAF of 3 common mutations, *POT1*, *PPM1D*, and *TERTp*, were all low at <10%, with many very small clones <2% seen. In contrast, the median VAF of *U2AF1* was >10% with many clones >20% present (Figure 16).

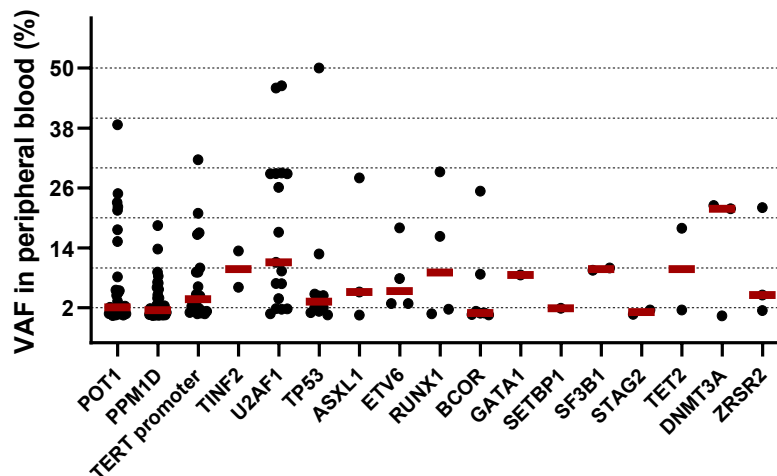


Figure 16. Variant allele frequency per mutation expressed as median.

3.3.5 Chemotherapy exposure

Of a total 208 patients, 8 had known chemotherapy exposure; 6 were prior to ECS and were 2 subsequent. Of the 6 patients who had received chemotherapy prior to sequencing, 2 had mutations. The first was an AML patient who had received induction chemotherapy, but external reports did say was *TP53* mutated at diagnosis. The second was a patient with multiple clones; 2 *PPMID*, 1 *POT1*, 1 *TERTp* and 1 *BCOR* who had received therapy for head and neck squamous cell carcinoma.

3.3.6 Somatic variants in the *TINF2* cohort

Patients with *TINF2* germline variants rarely had somatic mutations except for in *POT1* and in *TINF2* (truncated variants). A total of 9 *TINF2* patients had *POT1* mutations and 2 had somatic *TINF2* mutations. We assessed whether these mutations may be compensatory with positive effects on blood counts or survival. However, patients with somatic *POT1* or *TINF2* actually had no improvement in blood counts (Figure 17) and a trend towards poorer overall survival ($p=0.055$) (Figure 18).

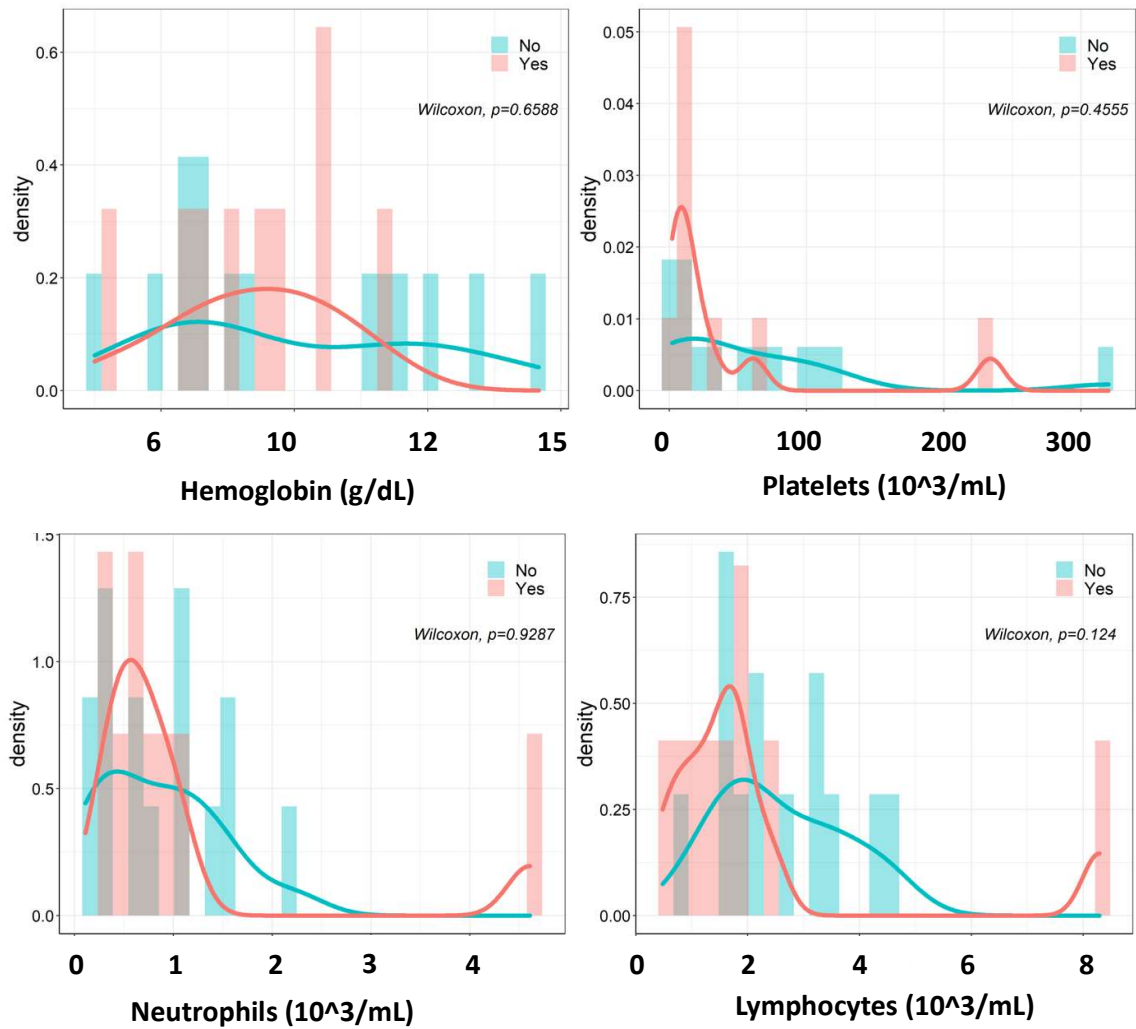


Figure 17. Association between blood counts and POT1 mutation status in TINF2 patients using a histogram with overlaying density lines.

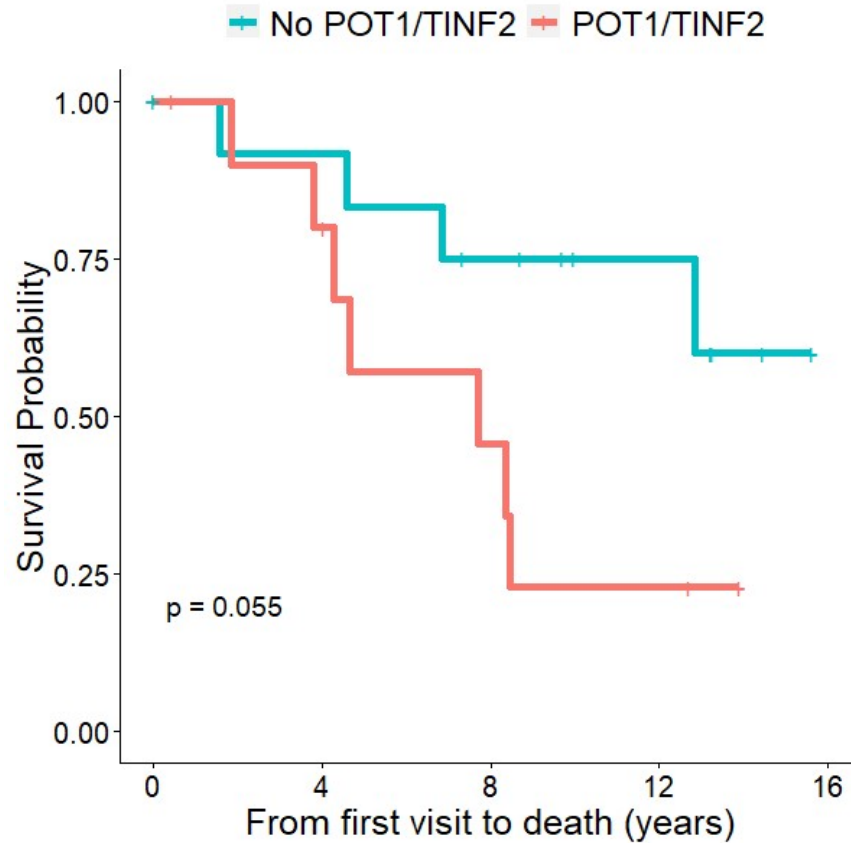


Figure 18. Overall survival comparing *TINF2* patients with and without *POT1* mutations ($P=0.055$).

3.3.7 Somatic variants in the solid malignancy cohort

Of the 15 patients who developed solid malignancy, 10 had somatic mutations in peripheral blood. Three patients had *TP53* (1 biallelic), four had *PPM1D* (2 of whom had received chemotherapy), and the 3 others had other MDS/AML associated mutations.

3.4 Conclusions

Clonal haematopoiesis has been described both in healthy individuals and in various disease states; in most cases it is related to increasing age⁸⁹. The clonal landscape of CH is likely driven by the environment. We showed that the presence somatic mutations in TBD was significantly associated with older age as well as the particular germline variant. However, overall patients with TBD appear to develop CH at a younger age than the healthy population. Patients with *TERT* and *TERC* had a predominance of *PPM1D* and *TERTp*; patients with *TERC* also were

enriched for *U2AF1*. In contrast, patients with *TINF2* germline variants had very few somatic mutations in typical clonal haematopoiesis or myeloid cancer genes, and instead had mutations in the telomere related genes such as *POT1* and direct somatic reversion of *TINF2*. Patients with *DKC1* and *RTEL1* had very few somatic mutations overall. Numbers of patients with other germline variants was low and it was not possible to discern any clonal patterns.

The *TERTp* has previously been described in TBD patients, mainly in those with *TERT* germline variants, but have also been seen in those with *TERC*, *PARN*, and *NHP2*^{78,79}. Cell lines derived from subjects with *TERT* promoter mutations showed increased telomerase expression and activity compared with cell lines from family members with the same germline *TERT* variant⁷⁹. While *TERTp* has been associated with several solid malignancies (most notably malignant melanoma), and associated with telomere elongation in that setting, to date, the *TERTp* mutation has not been associated with telomere elongation or malignant transformation in TBD patients³⁰. It is hypothesized that *TERTp* may act as a compensatory mechanism in TBD by upregulating telomerase expression but given the baseline high rate of telomere attrition, telomere elongation may not be possible as it is in malignancies⁵.

PPM1D is a negative regulator of *TP53* that controls cellular responses to environmental stress, by promoting termination of the DDR pathway⁹⁰. It has been strongly associated with exposure to chemotherapy in patients with solid cancers and is enriched in therapy related MDS⁹⁰⁻⁹². In our cohort, *PPM1D* was present in patients who had not received chemotherapy, suggesting an alternative driver. Given the chronic activation of the DDR induced by telomere shortening that drives cellular senescence, it is possible that *PPM1D* acts as a compensatory mechanism to counter BMF driven by *TP53* activation.

U2AF1 is an RNA spliceosome gene; it, and other RNA splicing factors, are recurrently mutated in MDS⁹³. Splicing factor mutations (most common being *SF3B1*, *U2AF1*, and *SRSF2*), with the exception *SF3B1*, have been associated with a poorer prognosis in myeloid malignancies, in particular mutations in *SRSF2*⁹⁴⁻⁹⁶. *U2AF1* mutations have been shown to associate with abnormal splicing of genes involved in functionally important pathways such including cell cycle progression and RNA processing, however, it remains unclear how they directly contribute to the pathophysiology of myeloid and solid malignancy⁹⁷. Interestingly, *U2AF1* mutations are known to be recurrent at two hotspots in myeloid malignancy; S34F and Q157P both of which account for >75% of mutations⁹⁸ and occur in the two zinc finger domains for RNA binding. One of the affected genes downstream due to mutated *U2AF1* is *ATR*, a key

components of the DDR, which may explain its predominance in the short-telomere environment⁹⁷. In patients without myeloid neoplasia, *U2AF1* and other splicing variants have been shown to pre-dispose towards the development of AML with an odds ratio of 7.4 for all splicing factor gene mutations combined⁹⁹. Therefore, it is possible that their presence in TBD patients without haematological malignancy may predispose to the development of MDS/AML.

POT1 variants were almost exclusively seen in patients with germline *TINF2*. Patients with *TINF2* TBD tend to have a severe clinical phenotype with DC and marrow failure. In cancers such as CLL and melanoma, somatic *POT1* mutations have been associated with telomere elongation. As *POT1* is a telomere related gene in the shelterin complex, as is *TINF2*, it is possible that these mutations represent indirect somatic reversion to attempt to offset telomere shortening⁴². However, in our patients *POT1* did not associate with improved blood counts.

We have demonstrated that the clonal landscape of TBD is dominated by 4 mutations: *PPM1D*, *U2AF1*, *TERTp*, and *POT1*. This is the first time the clonal landscape of TBD had been elucidated, and that *PPM1D*, *U2AF1*, and *POT1* variants have been identified as being recurrent in TBD; *TERTp* had been previously reported. We also demonstrated that the clonal landscape is influenced by age, as it is in healthy individuals and in other diseases, with older TBD patients more likely to have somatic mutations than younger patients. This predominance of specific mutations means these clones are likely being selected in the TBD marrow environment.

4. Chapter 4: Association of specific somatic mutations and the development of cancer and other clinical outcomes in TBD patients

4.1 Introduction

Clonal haematopoiesis has been linked to clinical outcomes in several circumstances. Firstly, in healthy individuals, clonal haematopoiesis has been associated an increased risk of haematological malignancy, cardiovascular disease, and decreased overall survival^{61-63,89}. CHIP, is present in approximately 10% of people >65 years old¹⁰⁰. The most commonly mutated genes in CHIP, *DNMT3A*, *TET2*, and *ASXL1*, are also commonly mutated in MDS, though in CHIP they typically appear as single mutations at lower VAF^{63,101}.

Specific mutations appear to influence the risk of developing myeloid neoplasia. For instance, in one study where the authors looked for somatic mutations in patients who ultimately developed AML years prior to their diagnosis, an increase in mutations was seen and the highest risk mutations were *TP53* (OR, 47.2) and *IDH1/2* (OR, 28.5), intermediate risk were spliceosome genes (including *SF3B1*, *SRSF2* and *U2AF1*) (OR 7.4), and lower risk were *TET2* (OR 5.8) and *DNMT3A* (OR 2.6)⁹⁹. In patients with clonal cytopenia of uncertain significance (presence of a somatic mutation but normal karyotype and no morphological dysplasia), mutations in *RUNX1*, *JAK2*, and splicing factors were the most predictive of developing MDS⁶⁴. In patients diagnosed with MDS, mutations have also been shown to associate with overall survival, with *U2AF1*, *SRSF2*, and *TP53* predicting a particularly poor prognosis⁶⁸.

Given the already known predictive value of mutations, in particular *U2AF1* and other splicing factors, in other clinical settings, we wished to determine whether specific mutations enriched in the TBD cohort could predict for the development of myeloid malignancy or confer poorer overall survival.

4.2 Methods

4.2.1 Grouping of somatic mutations

To allow for association analysis, somatic mutations were grouped according to function: telomere related genes, genes commonly mutated in age related clonal haematopoiesis, or in myeloid malignancy. CH and myeloid malignancy associated mutations were further

segregated into groupings to allow adequate numbers for clinical predictor analysis. These groupings were based on prior publications predicting the risk of specific mutations as well as the frequency of specific mutations in our cohort.

Groups were as follows:

Clonal haematopoiesis (CH) associated:

- *DNMT3A*, *TET2*, *ASXL1*, or *BCOR/L1* **without** a somatic variant that was MDS-associated

We grouped the typical CHIP mutations based on their prevalence in the healthy population and prior evidence that they are less predictive of MDS or AML than other myeloid cancer genes. We excluded patients with other more typical non-CHIP MDS/AML mutations from this group.

PPM1D (all exon 6 - truncating)

- Containing *PPM1D* **without** a MDS-associated

We grouped patients with *PPM1D* independently of other MDS/AML mutations given its unexpectedly high prevalence in our cohort and its typical strong association with chemotherapy when seen in MDS/AML which was absent in our cohort, suggesting an alternate driver in the TBD environment. We excluded patients with other more typical MDS/AML mutations from this group.

Telomere gene related

- Containing *TERT* *promotor* or *POT1* **with or without** a somatic variant from another group

TERTp alone

- Containing only *TERTp* with no other somatic variant

POT1 alone

- Containing only *POT1* with no other somatic variant

We grouped patients with mutations in telomere related genes separately due to their high prevalence in our cohort and our hypothesis that these were likely compensatory somatic reversion events.

MDS associated

- containing splicing factor genes, *RUNX1*, *SETBP1*, *ETV1*, *KRAS*, *STAG2*, *GATA2*, *TP53*, *KIT*, *PH6F* **with or without** a somatic variant from another group

Here we grouped mutations more typically associated with MDS/AML compared to age-related CHIP.

A subgroup of MDS/AML associated mutations was also assessed given the prevalence of splicing factor mutations in the cohort, and a possible association we noted with *TP53* and solid malignancy in our data set:

Splicing factors

- Containing *U2AF1*, *ZRSR2*, *SF3B1*, *SRSF2* **with or without** a somatic variant from another group

TP53

- containing *TP53* **with or without** a somatic variant from another group

4.2.2 Statistics

Univariable logistic regression was used to quantify the association of age, blood counts, and mutation groups on clinical outcomes including MDS, and survival. Multivariate analysis was performed to confirm whether predictors were independently associated. A p-value of <0.05 was considered for statistical significance. The Kaplan-Meier curves and the Cox Proportional Hazards models were used to estimate overall survival (OS) and cumulative incidence of MDS and compare the survival time distributions among mutation groups; time 0 was when patients were first assessed for somatic mutations. The log rank p-values value were computed to compare the survival time distributions among different sub-groups. All analysis was performed using R package “ggplot2”. Significant values and graphs are highlighted in red.

4.3 Results

4.3.1 Association with somatic mutation groups and age

Patients with somatic mutations were significantly older than those without somatic mutations and age distribution was different between the mutation groups. Patients with *TERTp*, *PPM1D*, and MDS/AML associated mutations (including splicing factors) were significantly older than non-mutated patients while those with CH associated, *POT1*, and *TP53* were not (Figure 19).

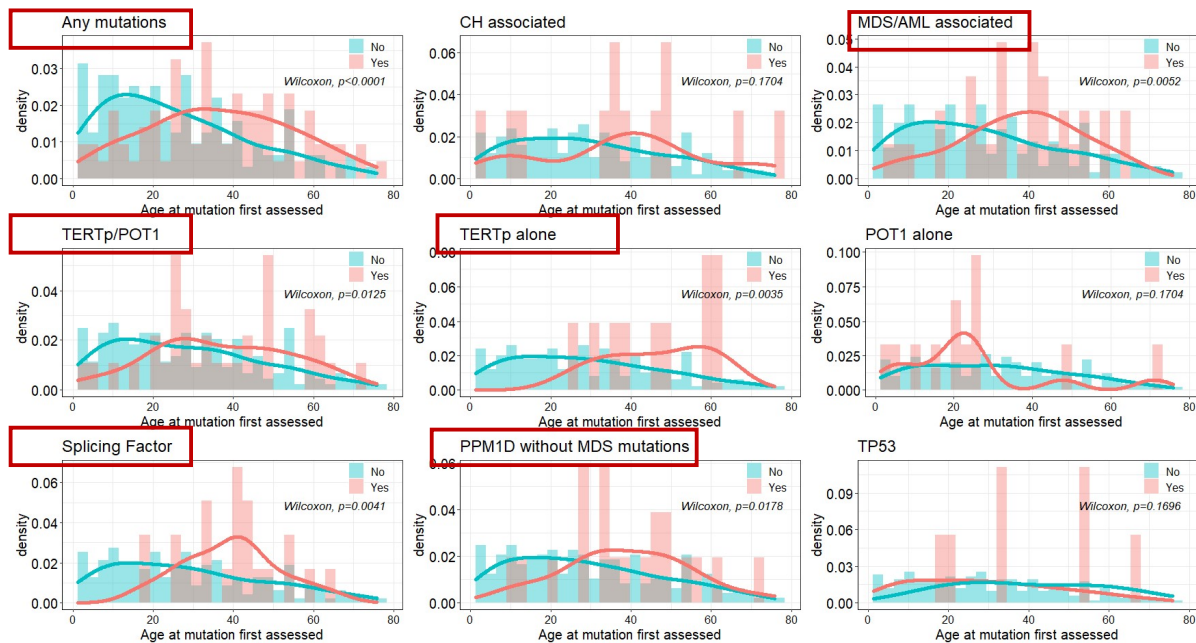


Figure 19. Age distribution of mutation groups as a continuous variable displayed as a histogram.

4.3.2 Association with somatic mutation groups and blood counts

We assessed whether somatic variants were associated with improved (adaptive) or worsening blood counts. *POT1* was associated with a lower haemoglobin ($p=0.0397$), however, no other mutation groups were associated with a lower haemoglobin.

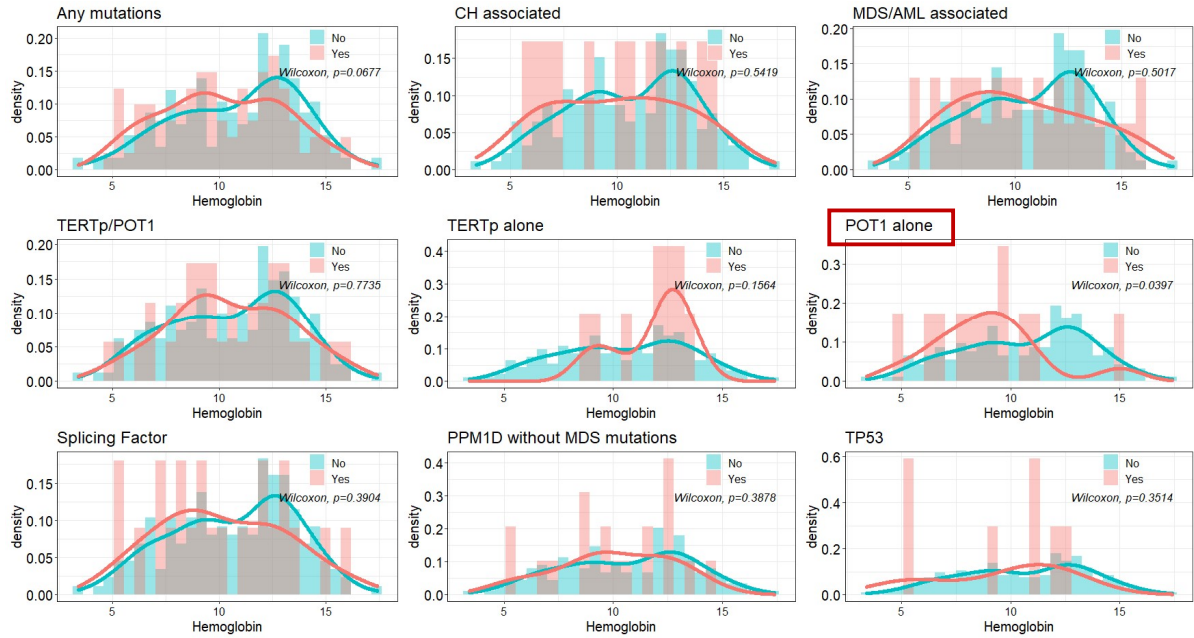


Figure 20. Haemoglobin distribution of mutation groups as a continuous variable

The presence of a somatic mutation was associated with lower platelet count ($p=0.01$); when mutations were grouped, this seen in CH mutations ($p=0.0451$) and *POT1* ($p=0.0031$).

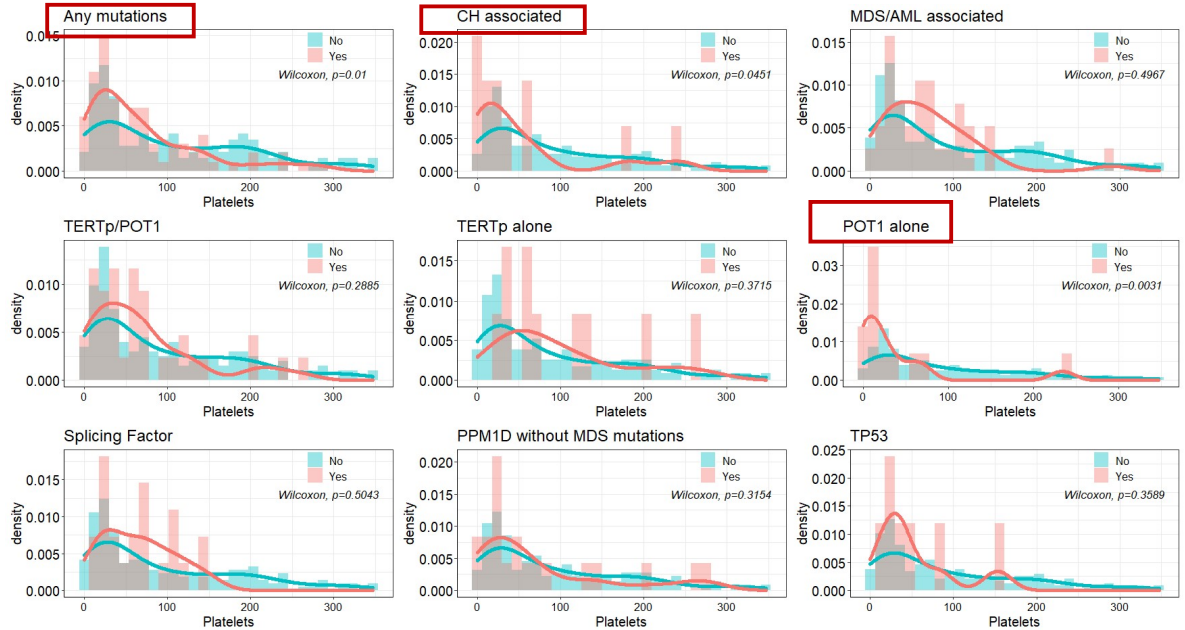


Figure 21. Platelet distribution of mutation groups as a continuous variable

The presence of somatic mutations was not associated with neutrophils. Lymphocyte count was associated with the presence of somatic mutations but not a specific group.

4.3.3 Predictors of MDS/AML using logistic regression

Using univariate logistic regression from the time of first patient assessment, the presence of a somatic mutation was predictive of MDS or AML (HR 2.94; $p=0.034$). Specifically, the presence of MDS/AML associated somatic mutations, which included splicing factors, was associated with the development of MDS/AML (HR 7.599; $p<0.001$). The presence of mutations in other groups, including CH associated, *TERTp*, *POT1*, *PPM1D*, and *TP53* was not associated with AML/MDS (Table 8).

Age was also associated with the development of MDS/AML, as was anaemia with haemoglobin $<12\text{g/dL}$ (HR 10.465; $p=0.023$).

Table 8. Univariate logistic regression of development of MDS/AML from baseline data

	HR (95% CI)	P-value	Concordance
Mutations	2.940 (1.086-7.957)	0.034	0.6279
CH associated	0.000 (0.000-Inf)	0.997	0.5298
MDS/AML associated	7.599 (2.921-19.769)	<0.0001	0.6946
TERTp/POT1	0.292 (0.039-2.205)	0.233	0.5681
TERTp alone	0.000 (0.000-Inf)	0.998	0.5218
POT1 alone	0.000 (0.000-Inf)	0.997	0.5335
Splicing factor	7.233 (2.739-19.104)	<0.0001	0.6707
PPM1D	0.515 (0.068-3.886)	0.52	0.5253
TP53	2.349 (0.308-17.924)	0.41	0.5202
Age	1.031 (1.006-1.057)	0.016	0.7061
Thrombocytopenia<70K	1.055 (0.401-2.775)	0.913	0.5002
Thrombocytopenia<100K	2.621 (0.753-9.126)	0.13	0.5932
Haemoglobin<12	10.465 (1.387-78.939)	0.023	0.669

Using multivariate logistic regression from the time of first patient assessment, MDS/AML associated mutations (HR 4.2; $p=0.03$), and age (HR 1.057; $p=0.002$) remained significant predictors of MDS/AML development (Table 9).

Table 9. Multivariate logistic regression of development of MDS/AML from baseline data

Concordance=0.9048		
	HR (95% CI)	P-value
CH associated	0.000 (0.000-Inf)	0.999
MDS/AML associated	4.207 (1.145-15.457)	0.03
TERTp alone	0.000 (0.000-Inf)	0.999
POT1 alone	0.000 (0.000-Inf)	0.999
PPM1D (no MDS mutations)	0.626 (0.069-5.661)	0.676
TP53	1.869 (0.184-18.993)	0.597
Age at mutations	1.057 (1.021-1.094)	0.002
Haemoglobin	0.751 (0.571-0.988)	0.041
Platelets	1.001 (0.987-1.015)	0.883
Neutrophils	0.598 (0.295-1.214)	0.155
ALC	1.061 (0.600-1.876)	0.838

4.3.4 Causes of death in the cohort

The commonest cause of death in the cohort was pulmonary complications, followed by infection and transplant related mortality. Five patients died directly due to malignancy.

Table 10. Causes of death in the TBD cohort

	Number of patients
Pulmonary complication	14
Infection	11
Transplant related complication	11
Malignancy related	5
Hepatopulmonary syndrome	3
Bleeding	2
Liver failure	2
Anaphylaxis	1
Cardiac failure	1
Post-surgery thrombosis	1
Cause of death unknown	18
Total	69

4.3.5 Predictors of survival using logistic regression

Using univariate analysis, multiple significant predictors of poor overall survival were present. In terms of presenting clinical features, the presence of non-AA cytopenia (HR 0.097; $p=0.023$) and being in the familial screen cohort (HR 0.058; $p=0.006$) showed a protective effect, while presence of MDS/AML was predictive of poorer overall survival (HR 2.558; $p=0.037$). Many mutation groups were predictive, including MDS/AML associated (HR 2.154; $p=0.008$), splicing factors (HR 2.161; $p=0.02$), *TERTp/POT1* (HR 2.154; $p=0.004$), *TERTp* ($p=0.04$), and TP53 (HR 7.55; $p<0.0001$) (Table 10). Blood counts were also predictive including Hb $<12\text{g/dL}$ (HR 1.8; $p=0.039$) and platelets $<70 \times 10^9/\text{L}$ (HR 1.8; $p=0.03$) (Table 10).

Table 11. Univariate logistic regression for development of death from baseline data

	HR (95% CI)	P-value	Concordance
Diagnosis			
Coats Plus	2.317 (0.306-17.557)	0.416	
Cytopenia (non AA)	0.097 (0.013-0.728)	0.023	
DC/HH	0.850 (0.479-1.511)	0.58	
Familial screen	0.058 (0.008-0.436)	0.006	0.6925
IPF	4.529 (0.567-36.176)	0.154	
MAA/SAA	0.430 (0.145-1.274)	0.128	
MDS/AML	2.558 (1.056-6.196)	0.037	
Mutations	2.420 (1.493-3.922)	<0.0001	0.5916
CH associated	0.903 (0.319-2.555)	0.848	0.511
MDS/AML associated	2.173 (1.220-3.870)	0.008	0.5473
TERTp/POT1	2.154 (1.272-3.646)	0.004	0.5507
TERTp alone	2.425 (1.042-5.642)	0.04	0.5234
POT1 alone	1.911 (0.869-4.203)	0.107	0.5136
Splicing factor	2.161 (1.129-4.135)	0.02	0.5424
PPM1D (no MDS mutations)	0.636 (0.273-1.483)	0.295	0.5272
TP53	7.556 (2.908-19.631)	<0.0001	0.5427
Age at mutations	1.010 (0.996-1.023)	0.169	0.5417
Thrombocytopenia<70K	1.804 (1.060-3.070)	0.03	0.5595
Thrombocytopenia<100K	1.511 (0.859-2.660)	0.152	0.5403
Haemoglobin<12	1.810 (1.031-3.178)	0.039	0.559
Haemoglobin	0.863 (0.793-0.939)	<0.0001	0.6319
Platelets	0.996 (0.992-0.999)	0.019	0.5989
Neutrophils	0.993 (0.801-1.230)	0.949	0.5466
ALC	0.745 (0.615-0.902)	0.003	0.6539

Using multivariate analysis only somatic mutations in *TERTp* and in *TP53* remained predictive of poorer overall survival. However, due to high numbers of variables possibly confounding results, we next used a step-wise model that was used to eliminate less significant variables for the multivariate analysis. This step-wise model showed that non-AA cytopenia (HR 0.116; p=0.04), and being part of the familial screen cohort (HR 0.105; p=0.105) remained protective, while presence of pulmonary fibrosis became (HR 14.439; p=0.02) and MDS/AML remained (HR 3.034; p=0.032) significant predictors of death. When looking at somatic mutations, MDS/AML associated mutations (HR 2.199; p=0.033), *TP53* (HR 15.805; <p=0.0001 and *TERTp* (HR 4.383; p=0.006) remained highly predictive of death (Table 11).

Table 12. Multivariate logistic regression for development of death from baseline data

	Full Model Concordance=0.767		Stepwise Model Concordance=0.7541	
	HR (95% CI)	P-value	HR (95% CI)	P-value
Diagnosis				
Coats Plus	1.270 (0.121-13.279)	0.842	0.930 (0.100-8.670)	0.949
Non-AA cytopenia	0.125 (0.014-1.094)	0.06	0.116 (0.015-0.907)	0.04
DC/HH	1.337 (0.496-3.608)	0.566	1.149 (0.565-2.337)	0.702
Familial screen	0.129 (0.014-1.202)	0.072	0.105 (0.012-0.907)	0.041
IPF	10.167 (0.872-118.592)	0.064	14.439 (1.482-140.670)	0.022
MAA/SAA	0.835 (0.253-2.761)	0.768	0.898 (0.282-2.858)	0.856
MDS/AML	2.624 (0.928-7.420)	0.069	3.034 (1.097-8.387)	0.032
CH associated	0.581 (0.130-2.589)	0.477		
MDS/AML associated	1.716 (0.761-3.871)	0.193	2.199 (1.066-4.535)	0.033
TERTp alone	4.043 (1.255-13.022)	0.019	4.383 (1.542-12.455)	0.006
POT1 alone	2.093 (0.841-5.210)	0.113	2.249 (0.963-5.256)	0.061
PPM1D (no MDS mutations)	0.804 (0.244-2.655)	0.72		
TP53	9.713 (2.255-41.831)	0.002	15.805 (5.344-46.742)	<0.0001
Age	1.011 (0.986-1.035)	0.393		
Haemoglobin	0.831 (0.730-0.946)	0.005	0.854 (0.755-0.966)	0.012
Platelets	1.005 (0.998-1.012)	0.156	1.004 (1.000-1.009)	0.049
Neutrophils	1.027 (0.750-1.405)	0.868		
ALC	0.891 (0.682-1.165)	0.4		

4.3.6 Association of somatic mutations with overall survival

Next, we used KM curves to assess the difference in overall survival between mutation groups. The presence of any somatic mutation was associated with decreased overall survival ($p=0.00021$) (Figure 22).

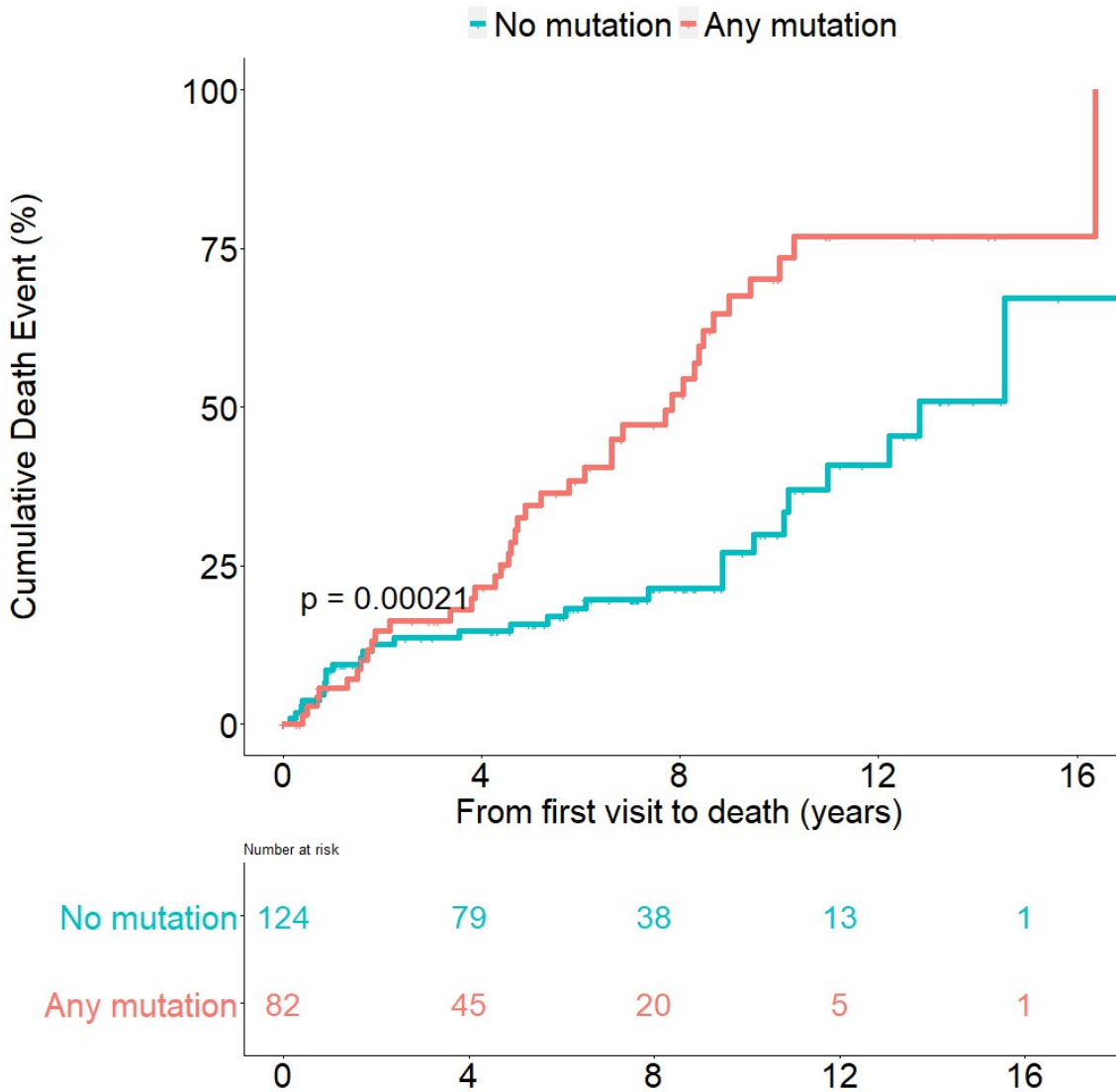


Figure 22. Overall survival comparing the presence and absence of any somatic mutation

When we assessed individual mutation groups, MDS/AML associated mutations ($p=0.0069$), including splicing factors ($p=0.0034$) and *TP53* ($p<0.0001$) had a significantly decreased overall survival compared to those without those mutations. The *TERTp* was also associated with decreased overall survival ($p=0.034$). Other mutation groups were not associated with decreased overall survival (Figure 23).

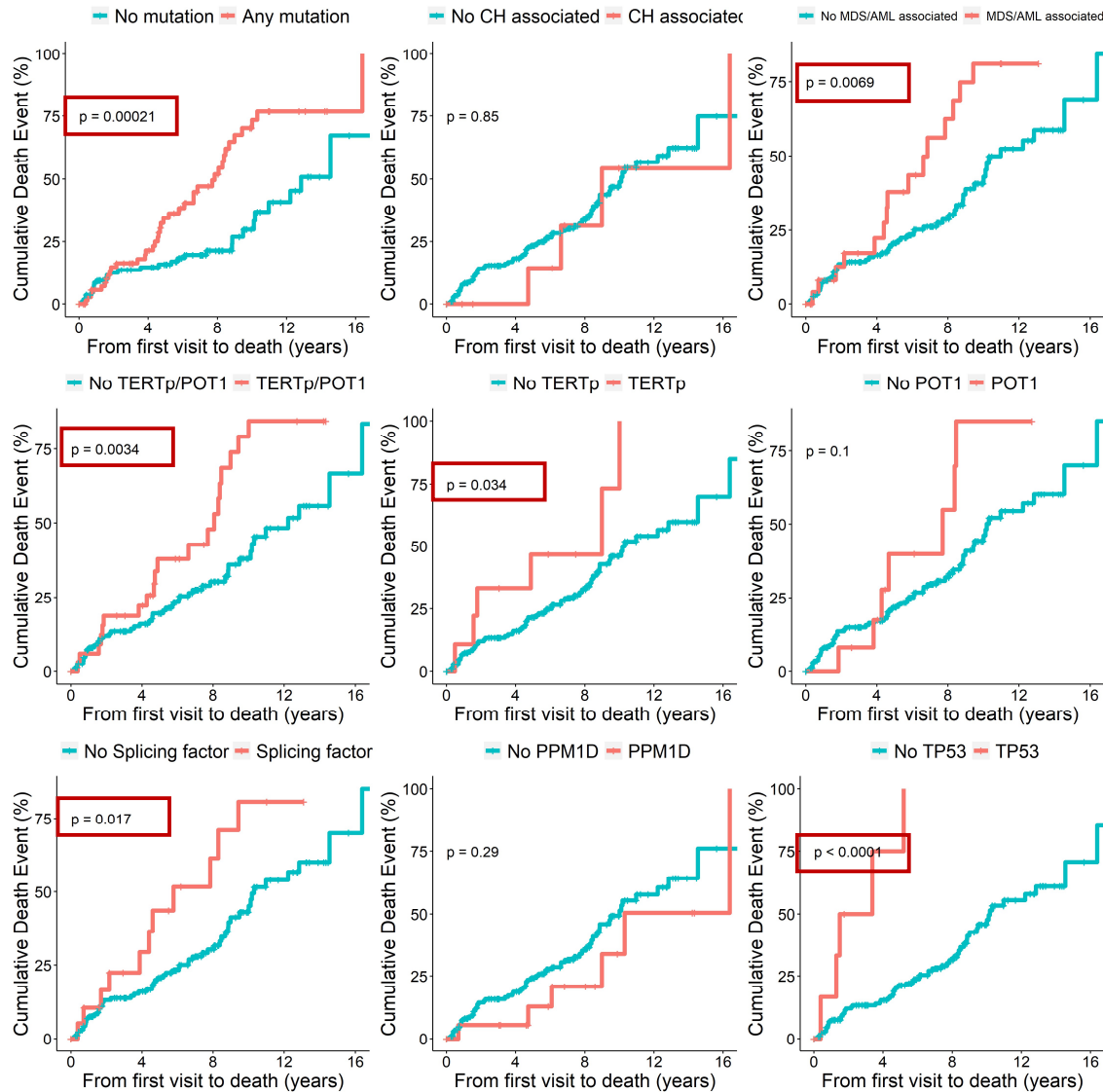


Figure 23. Overall survival comparing the presence and absence of any somatic mutation by somatic mutation grouping.

Given that age was an independent risk factor for MDS/AML, and the presence of a somatic mutation was significantly associated with increased age, we wished to ascertain whether somatic mutations remained predictive of overall survival in different age groups. When separated by age into <20 years, ≥ 20 years but <40 years, ≥ 40 years but less than 60 years, and >60 years, it appeared that in the ≥ 20 years but <40 years age bracket the effect of somatic mutations on survival was most pronounced (Figure 24).

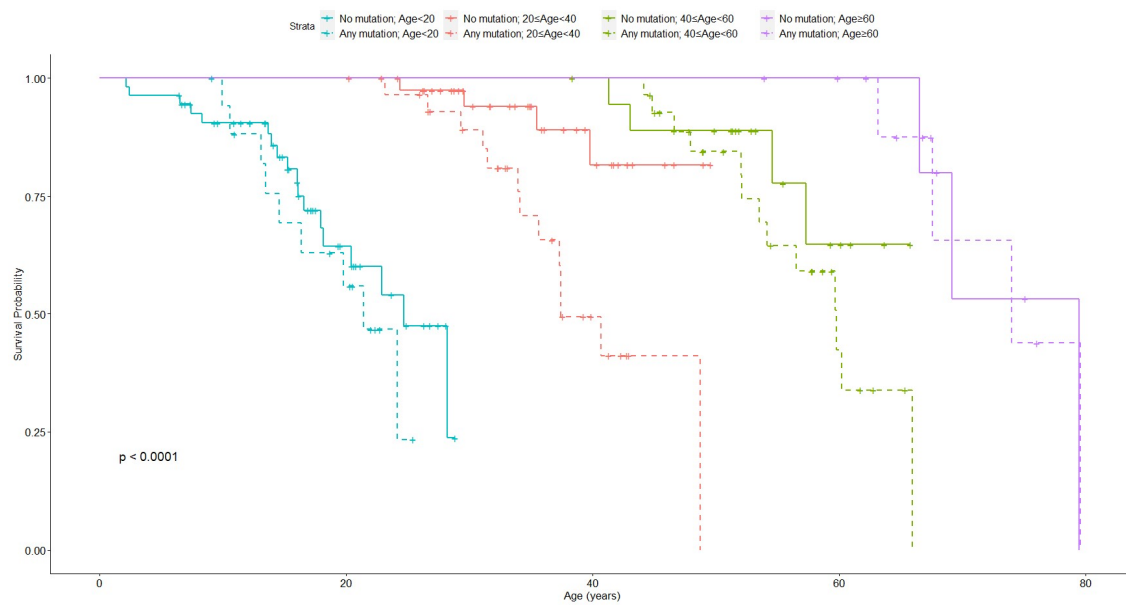


Figure 24. Overall survival of patients with or without mutations stratified by age. Dotted lines denote mutation. In blue is <20 years, in red ≥ 20 years but <40 years, in green ≥ 40 years but less than 60 years, and in purple ≥ 60 years.

4.3.7 Association of somatic mutations with overall survival in patients <20 years old

In patients <20 years old ($n=74$), the only mutations that predicted for worse overall survival were *POT1* ($p=0.023$) and *TP53* ($p=0.039$) (Figures 26 + 27). However, there were few patients overall who has somatic mutations in this age group. The *TERTp* group was excluded as no patients in this age group fell into the *TERTp* group.

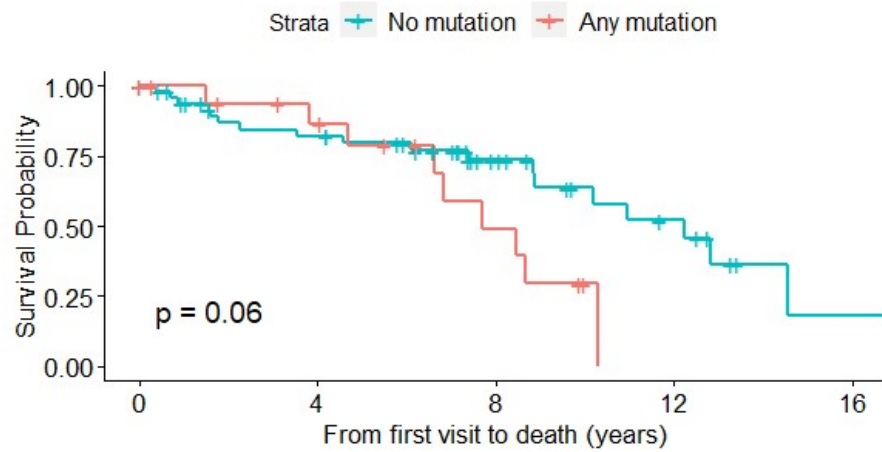


Figure 25. Overall survival of patients with or without mutations in patients <20 years. Blue lines denote no mutation while red denotes presence of mutation.

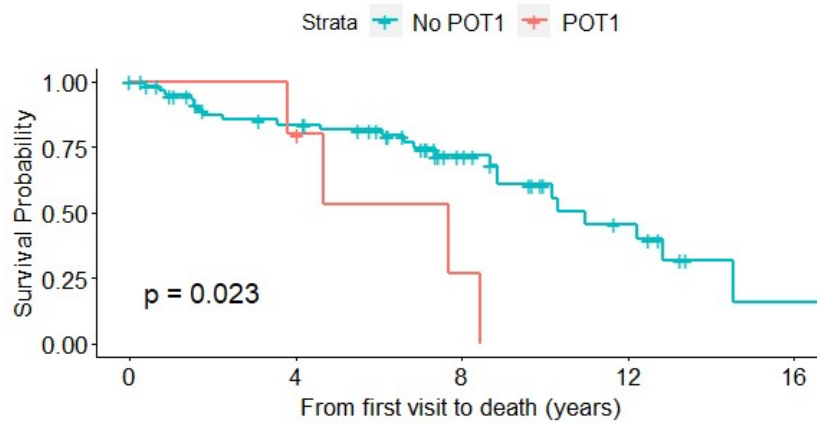


Figure 26. Overall survival of patients with *POT1* mutations compared to those without *POT1* mutations in patients <20 years. Blue lines denote no *POT1* mutation while red denotes presence of *POT1* mutation.

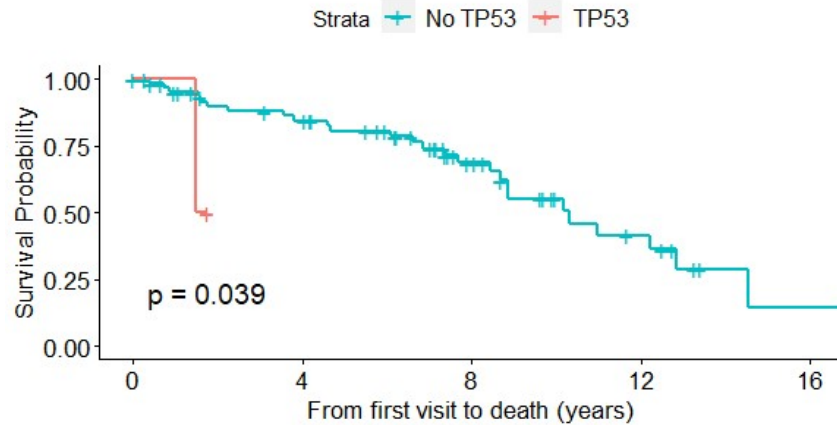


Figure 27. Overall survival of patients with TP53 mutations compared to those without TP53 mutations in patients <20 years. Blue lines denote no TP53 mutation while red denotes presence of TP53 mutation.

4.3.8 Association of somatic mutations with overall survival in patients ≥ 20 years but <40 years old

The strongest association between any somatic mutations and decreased survival was seen in the ≥ 20 years but <40 years age group (n=74). In this group having any somatic mutation strongly predicted lower overall survival (p=0.00089) (Figure 28). When broken down by mutation groups, splicing factor mutations (p=0.029), TP53 (p<0.0001) and TERTp (p=0.004) predicted poorer overall survival (Figures 29, 30, 31).

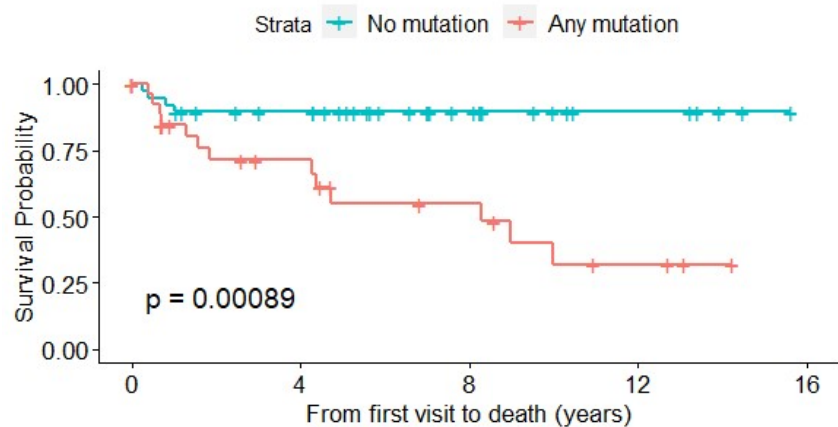


Figure 28. Overall survival of patients with or without mutations in patients ≥ 20 years but < 40 years. Blue lines denote no mutation while red denotes presence of mutation.

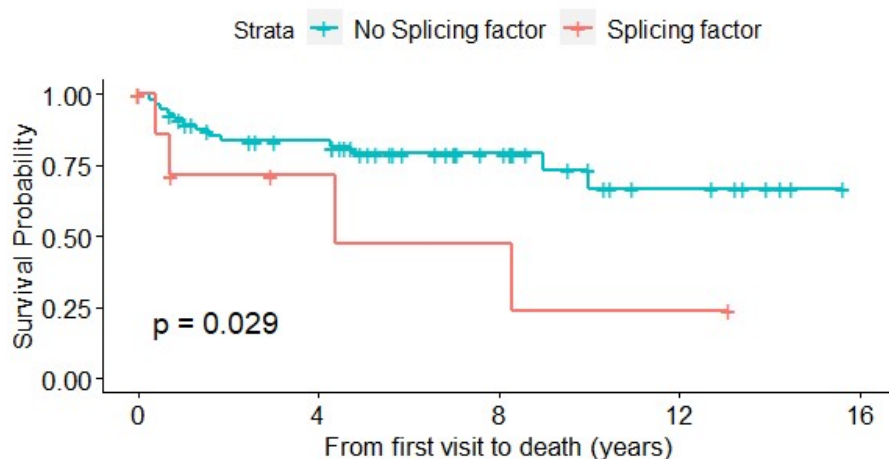


Figure 29. Overall survival of patients with splicing factor mutations compared to those without Splicing factor mutations in patients ≥ 20 years but < 40 . Blue lines denote no splicing factor mutation while red denotes presence of splicing factor mutation.

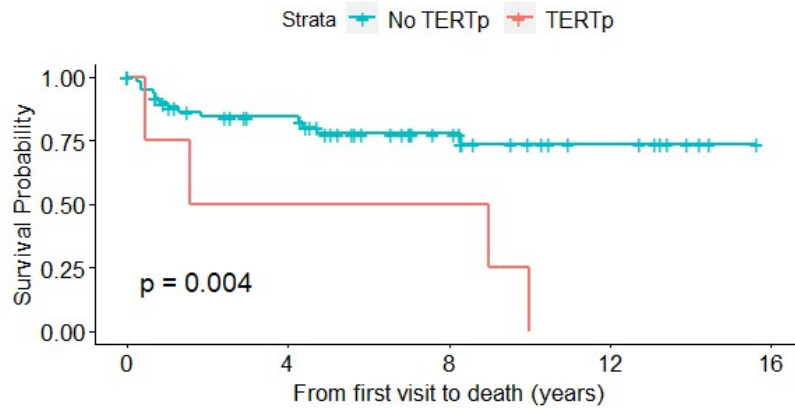


Figure 30. Overall survival of patients with *TERTp* mutations compared to those without *TERTp* mutations in patients ≥ 20 years but < 40 . Blue lines denote no *TERTp* mutation while red denotes presence of *TERTp* mutation.

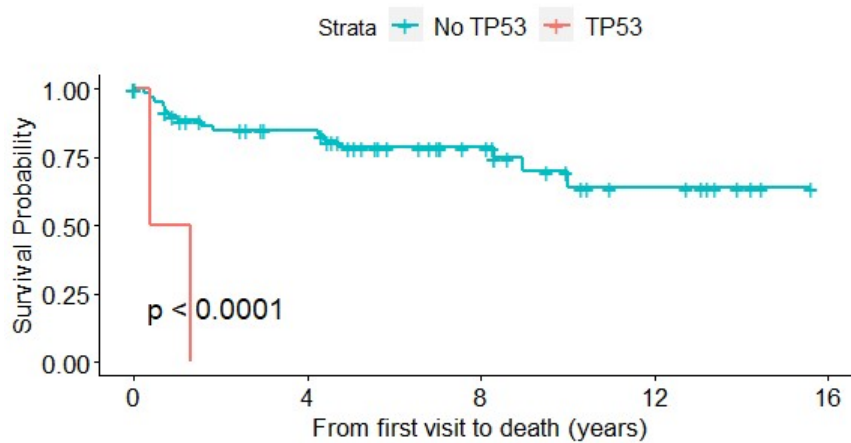


Figure 31. Overall survival of patients with *TP53* mutations compared to those without *TP53* mutations in patients ≥ 20 years but < 40 . Blue lines denote no *TP53* mutation while red denotes presence of *TP53* mutation.

4.3.9 Association of somatic mutations with overall survival in patients ≥ 40 years but less than 60 years old

In patients ≥ 40 years but less than 60 years ($n=45$) the presence of a somatic mutation was not significantly associated with survival. However, when grouped, the presence of MDS/AML associated mutations ($p=0.017$), PPM1D ($p=0.04$) or *TP53* ($p=0.011$) remained predictive of decreased survival (Figures 32, 33, 34).

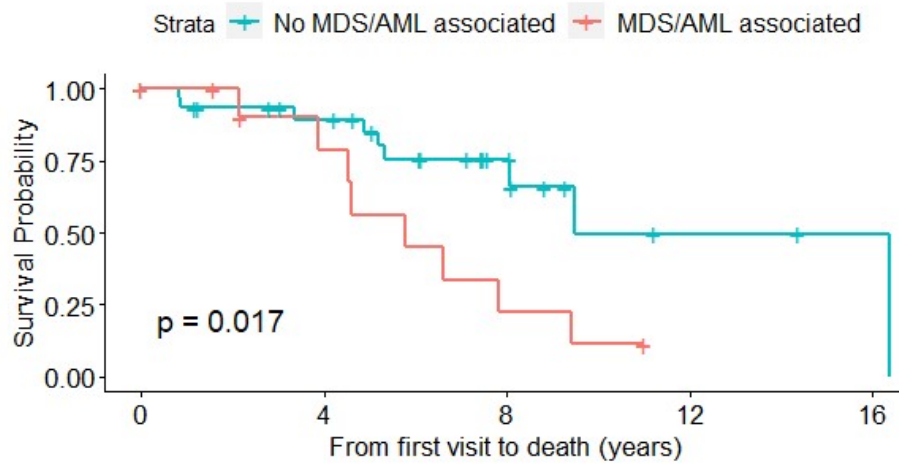


Figure 32. Overall survival of patients with MDS/AML associated mutations compared to those without MDS/AML associated mutations in patients ≥ 40 years but less than 60 years. Blue lines denote no MDS/AML associated mutation while red denotes presence of MDS/AML associated mutation.

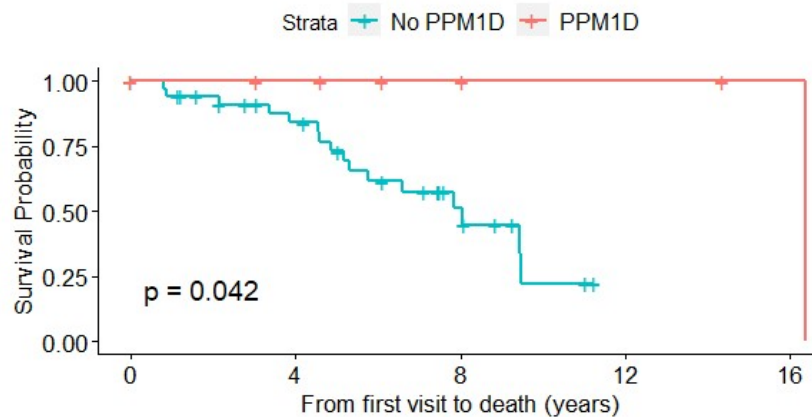


Figure 33. Overall survival of patients with PPM1D mutations compared to those without PPM1D mutations in patients ≥ 40 years but less than 60 years. Blue lines denote no PPM1D mutation while red denotes presence of PPM1D mutation.

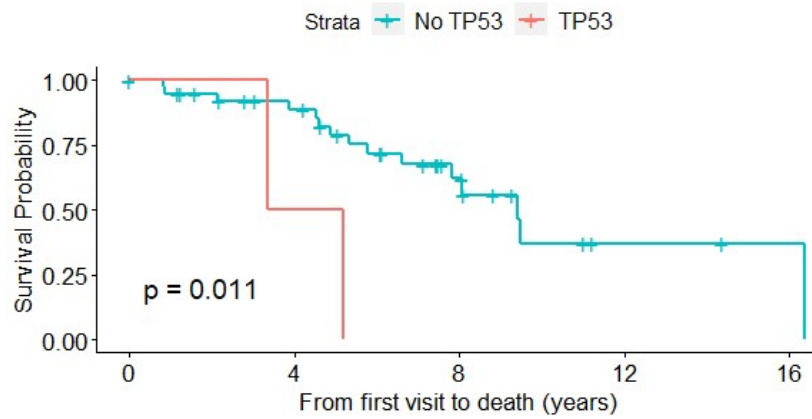


Figure 34. Overall survival of patients with TP53 mutations compared to those without TP53 mutations in patients ≥ 40 years but less than 60 years. Blue lines denote no TP53 mutation while red denotes presence of TP53 mutation.

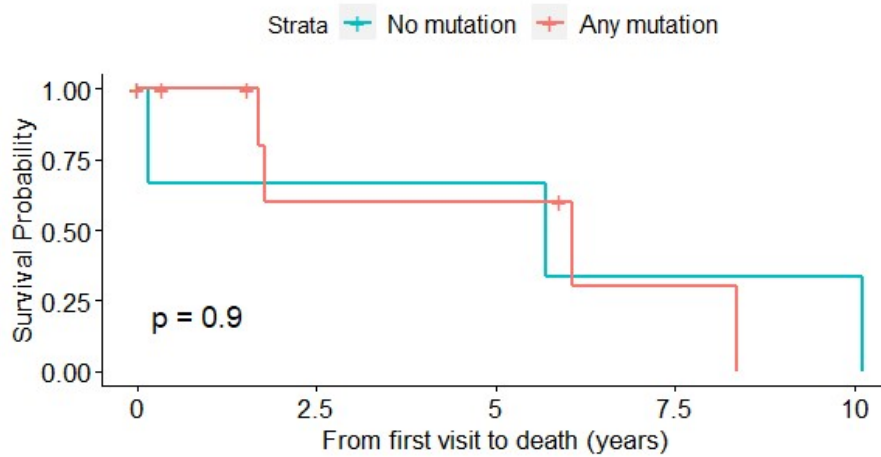


Figure 35. OS of patients with any somatic mutation compared to those without mutations in patients > 60 years. Blue lines denote no mutation; red denotes presence of mutation.

In patients > 60 ($n=15$), overall survival was not impacted by the presence of a somatic mutation (Figure 35). Curves were not generated for each mutation group in patients > 60 due to inadequate patient numbers

4.3.10 Predictors of solid malignancy using logistic regression

Given *TP53* was associated with poor overall survival but was not a predictor of MDS/AML, we looked to see whether *TP53* was associated with the development of solid malignancy. *TP53* was strongly associated with the development of solid malignancy in both univariate (HR 23.03; p=0.0001) and multivariate (HR 25.678; p=0.002) analysis (Table 12 and 13).

Table 13. *Univariate logistic regression for development of solid malignancy*

	OR (95% CI)	P-value	Concordance
Mutations	4.521 (1.388-14.723)	0.0123	0.6775
CH associated	1.182 (0.142-9.827)	0.8771	0.5048
MDS/AML associated	3.074 (0.975-9.689)	0.0552	0.5967
TERTp/POT1	0.719 (0.155-3.336)	0.674	0.5214
TERTp alone	0 (0-Inf)	0.9904	0.5259
POT1 alone	1.182 (0.142-9.827)	0.8771	0.5048
Splicing factor	3.33 (0.965-11.491)	0.0569	0.5841
PPM1D (no MDS mutations)	1.496 (0.312-7.159)	0.6143	0.52
TP53	23.03 (4.577-115.871)	0.0001	0.6256
Age	1.044 (1.014-1.075)	0.0035	0.7617

Table 14. *Multivariable logistic regression for development of solid malignancy*

Concordance=0.8933

	OR (95% CI)	P-value
Intercept	0.005 (0-0.12)	0.001
CH associated	0.659 (0.039-11.092)	0.772
MDS/AML associated	2.626 (0.489-14.11)	0.26
TERTp alone	0 (0-Inf)	0.994
POT1 alone	1.777 (0.126-25.158)	0.671
PPM1D (no MDS mutations)	2.162 (0.294-15.93)	0.449
TP53	25.678 (3.206-205.669)	0.002
Age	1.055 (1.011-1.101)	0.013

4.3.11 Predictors of lung and liver disease using logistic regression

The *TERTp* was significantly associated with liver fibrosis by both univariate (HR 61.33; p<0.0001) and multivariate (HR 123.3; p<0.0001) analysis. No other mutation groups were associated with liver fibrosis (Table 14 and 15).

While multiple mutations groups associated with pulmonary fibrosis on univariate analysis, none remained significant using multivariate analysis (Table 16 and 17).

Table 15. *Univariable logistic regression for development of liver fibrosis*

	OR (95% CI)	P-value	Concordance
Mutations	10.549 (2.981-37.332)	0.0003	0.7503
CH associated	2.185 (0.438-10.9)	0.3404	0.5258
MDS/AML associated	2.147 (0.717-6.425)	0.172	0.5578
TERTp/POT1	24.9 (8.178-75.812)	<0.0001	0.8212
TERTp alone	61.333 (11.71-321.251)	<0.0001	0.6946
POT1 alone	0.837 (0.102-6.846)	0.8685	0.5046
Splicing factor	2.333 (0.704-7.737)	0.1659	0.5516
PPM1D (no MDS mutations)	1.037 (0.222-4.836)	0.9631	0.5016
TP53	1.579 (0.18-13.818)	0.6798	0.5089
Age	1.032 (1.007-1.057)	0.013	0.6957

Table 16. *Multivariable logistic regression for development of liver fibrosis*

	Concordance=0.925	
	OR (95% CI)	P-value
Intercept	0.005 (0-0.174)	0.003
CH associated	2.57 (0.191-34.637)	0.477
MDS/AML associated	3.377 (0.655-17.409)	0.146
TERTp alone	123.303 (10.331-1471.579)	1E-04
POT1 alone	0 (0-Inf)	0.994
PPM1D (no MDS mutations)	1.039 (0.091-11.912)	0.975
TP53	0.916 (0.038-21.948)	0.957
Age	1.026 (0.977-1.077)	0.301

Table 17. *Univariable logistic regression for development of lung fibrosis*

	OR (95% CI)	P-value	Concordance
Mutations	3.284 (1.672-6.452)	0.0006	0.644
CH associated	4.478 (1.301-15.411)	0.0174	0.548
MDS/AML associated	2.667 (1.176-6.049)	0.0189	0.5707
TERTp/POT1	4.014 (1.844-8.737)	0.0005	0.6133
TERTp alone	4.583 (1.178-17.828)	0.028	0.5405
POT1 alone	1.129 (0.293-4.351)	0.8603	0.5034
Splicing factor	3.288 (1.319-8.198)	0.0106	0.5684
PPM1D (no MDS mutations)	1.952 (0.73-5.218)	0.1825	0.5333

TP53	2.625 (0.566-12.171)	0.2175	0.5193
Age	1.058 (1.037-1.08)	<0.0001	0.7672

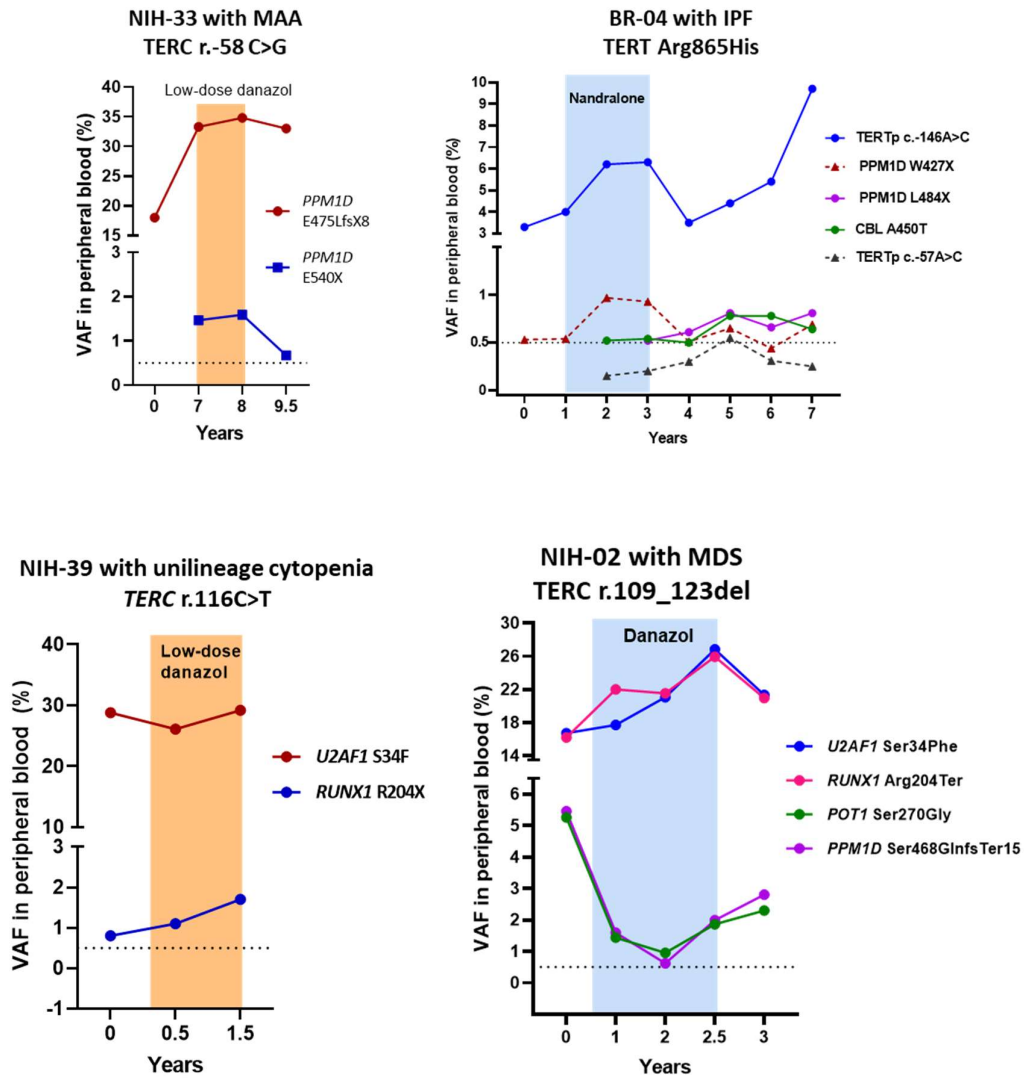
Table 18. *Multivariable logistic regression for development of lung fibrosis*

Concordance=0.8414		
	OR (95% CI)	P-value
Intercept	0.022 (0.003-0.18)	4E-04
CH associated	2.662 (0.491-14.434)	0.256
MDS/AML associated	1.463 (0.476-4.493)	0.507
TERTp alone	2.591 (0.49-13.704)	0.263
POT1 alone	2.328 (0.396-13.671)	0.349
PPM1D (no MDS mutations)	1.171 (0.321-4.277)	0.811
TP53	0.933 (0.126-6.895)	0.946
Age	1.072 (1.04-1.105)	<0.0001

4.3.12 Effect of androgens on somatic mutations

There was no clear association seen between androgen therapy and change in clone sizes. In 5 patients treated with androgens who had somatic mutations, two of whom had MDS, there was no obvious increase in mutational burden following the administration of therapy. Patients received either low (200-400mg) or high (800mg) dose danazol or IM nandrolone 5mg/kg every 2 weeks. Most mutations, including *U2AF1*, remained at stable VAF during the administration of androgen therapy (Figure 36).

E



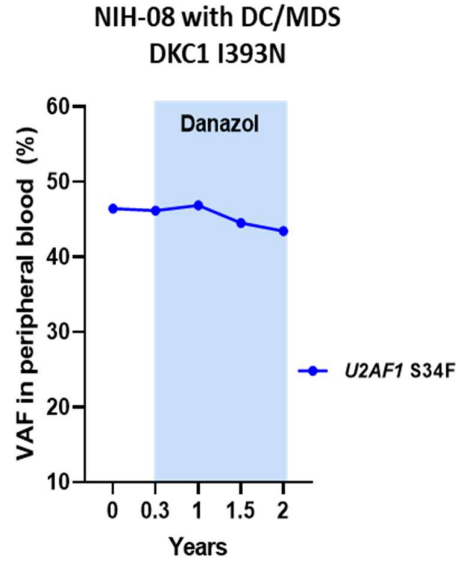


Figure 36. *Effect of androgen therapy on clonal trajectories.*

4.4 Conclusions

Somatic mutations were highly predictive of both the development of myeloid malignancy and lower overall survival in our cohort. When broken down by mutation type, only the MDS/AML associated mutations, which included splicing factors with *U2AF1* being the predominant gene mutated, were predictive of myeloid malignancy. However, even though they were not associated with MDS/AML, *TP53* and *TERTp* mutations were both also associated with poorer overall survival. *TP53* had a strong association with solid malignancy, while *TERTp* had a strong association with liver fibrosis, suggesting that causes other than myeloid neoplasia may be contributing to death in patients with these mutations.

Also important to consider is the strong association of older age with the presence of somatic mutations. In healthy individuals, CHIP is uncommon in those <60 years while the median age of patients with somatic mutations in our TBD cohort was only 38^{61,63}. Most patients with TBD die young due to organ dysfunction, cancer, or severe BMF. The strongest association between the presence of somatic mutations and death in our cohort was seen between the ages of 20 and 40. Possibly, this is because the presence of a somatic mutation is an indirect marker of disease like malignancy or organ dysfunction, that is likely to develop in that timeframe. We know that

TBD patients can exhibit a wide clinical phenotype in terms of degree of organ dysfunction and cancer risk, even within families carrying the same germline mutation¹⁵. In the case of *TERTp*, it may be that patients who develop *TERTp* mutations have inherently more severe disease, i.e. shorter telomeres or lower telomerase expression, that their body attempts to compensate. In our cohort, the commonest causes of death were not malignancy related but due to other organ dysfunction or infection, highlighting that somatic mutations are likely a marker of disease severity.

The association between the MDS/AML associated mutations, in particular *U2AF1*, and the presence or development of myeloid malignancy was unsurprising and in keeping with their predictive value in the setting of healthy individuals, and those with CCUS^{61,64,66}. However, it remains unclear why *U2AF1* is being selected over other mutations. When CHIP occurs in healthy individuals, splicing factor mutations are rare, and the clonal landscape is dominated by *DNMT3A*, *TET2*, and *ASXL1*, all of which were infrequent in our TBD cohort⁶¹. It is likely the recurrent *U2AF1* mutations in our cohort represent a compensatory mechanism, as it seen in other IBMFS, or a response to a specific environmental stimulus. However, in this case it appears maladaptive given its association with myeloid malignancy and worse survival. However, when looking at individual patients, there were some who had *U2AF1* clones over a number of years and did not develop MDS or AML. Therefore, it may be that *U2AF1* alone does not drive malignancy but represents a “first hit” in the development of cancer in TBD patients.

5. Chapter 5: Single cell analysis of U2AF1 and other somatic variants in TBD patients

5.1 Introduction

While bulk DNA sequencing techniques such as NGS or ECS can assess the genome of a cell population (such as peripheral blood or bone marrow), single cell DNA sequencing can assess individual cells¹⁰². Therefore, this technique allows us to determine whether clonal populations are co-occurring or separate in the same cell, and to further elucidate clonal trajectories.

Given the predominance of *U2AF1* in our patient cohort, and its clinical association with the development of myeloid neoplasia as the dominant mutation seen in the splicing factor mutation group, we wished to closely examine the clonal dynamics of TBD patients with *U2AF1* to see if it acts as a founder or driver of subsequent malignant transformation. Two patients were selected; both had large *U2AF1* clones and at least one other clone, one having developed subsequent myeloid neoplasia while the other was free of haematologic malignancy.

5.2 Methods

Two patients with >1 somatic mutation (one being *U2AF1*) detected on bulk sequencing selected for single cell analysis. Single cell DNA sequencing was performed using the Mission Bio Tapestri Platform.

5.2.1 Mission Bio Tapestri protocol overview

The assay was performed according to manufacturing instructions (see Appendix C for further details on full protocol). Briefly, fresh peripheral blood or bone marrow samples were collected in EDTA tubes, ACK lysed, and pelleted by centrifugation. Samples were stored in 10% DMSO at -80oC long-term. For the assay, fresh or frozen samples were processed or thawed respectively, cells were pelleted by centrifugation, counted with the Cellometer Auto 2000 (Nexcelom), and diluted to a concentration of 25,000 cells/μL. Cells (1×10^6) were stained with TotalSeq-D Heme Oncology Cocktail oligo-tagged antibodies (BioLegend) and individually partitioned into sub-nanoliter droplets. Barcoding Beads and PCR reagents were introduced using the Mission Bio Tapestri Instrument and DNA Cartridge. Cell lysis, protease digestion,

cell barcoding and targeted amplification using multiplexed PCR occurred within the droplets performed per Manufacturer's protocol (Appendix C). Droplets were then disrupted, and barcoded DNA extracted for Library Amplification. The Protein library was separated from the DNA library by biotinylated oligo pull-out. Protein and DNA libraries were indexed and amplified separately. Final libraries were purified and sequenced on Illumina Sequencer instrument. Targeted scDNA was performed with a customized panel that covered 120 amplicons, including the specific CH, myeloid malignancy, and telomere related mutations found in our patients on bulk sequencing.

5.2.2 Mission Bio Tapestri single cell analysis

Multi-omics data from targeted scDNA and protein were integrated for analysis using the Tapestri Pipeline V2 tool (MissionBio). Integrated data (Loom and h5 files) were then analysed with Tapestri Insights v2.2 (MissionBio) and the Python-based Mosaic package in a Jupyter notebook. Different cell subpopulations were classified according to their immunophenotypes identified via uniform manifold approximation and projection (UMAP) dimension reduction and k-means clustering. UMAP is algorithm that shows how single cells cluster in a multidimensional plot based on expression of 45 proteins. Clustering was used to identify different cell subpopulations classified according to their immunophenotype. The following proteins were assessed on the TotalSeq Heme Oncology Cocktail from BioLegend:

HEME ONCOLOGY PROTEIN PANEL

CD1c	CD11c	CD34	CD62P	CD141
CD2	CD13	CD38	CD64	CD163
CD3	CD14	CD44	CD69	CD303
CD4	CD16	CD45	CD71	CD304
CD5	CD19	CD45RA	CD83	FcεR1α
CD7	CD22	CD45RO	CD90	HLA-DR
CD8	CD25	CD49d	CD117	IgG1 control
CD10	CD30	CD56	CD123	IgG2a control
CD11b	CD33	CD62L	CD138	IgG2b control

Figure 37. Proteins used in MissionBio for cell identification¹⁰³

5.3 Results

5.3.1 Patient 1 with *U2AF1* and other mutations who developed MDS-EB2

Patient 1:

A 54 year old male with a pathogenic *TERT* variant (*TERT* p.Ile587Thr). Clinical characteristics of TBD included MDS-MLD diagnosed in 2010 with abnormal karyotype of +1q21 and trisomy 8 (46,XY,dup(1)(q21q32)[5]/47,idem,+8[6]) with subsequent evolution to MDS-EB2 with monosomy 7 occurring 5.5 years after initial MDS diagnosis. Patient received alemtuzumab in 2010 for treatment of HLA-DR+ MDS (prior to diagnosis of TBD) and received danazol therapy from 2011 to 2013. No evidence of PF or liver cirrhosis. Bulk sequencing revealed the following somatic variants occurring at various disease timepoints:

Gene	Amino acid change
<i>U2AF1</i>	p.Gln157Arg
<i>ASXL1</i>	p.Met837LeufsTer2
<i>GATA2</i>	p.Lys390del
<i>SETBP1</i>	p.Asp868Asn

<i>KRAS</i>	p.Gly12Val
-------------	------------

Clonal trajectory of each mutation and VAF over time are shown below based on results from bulk ECS.

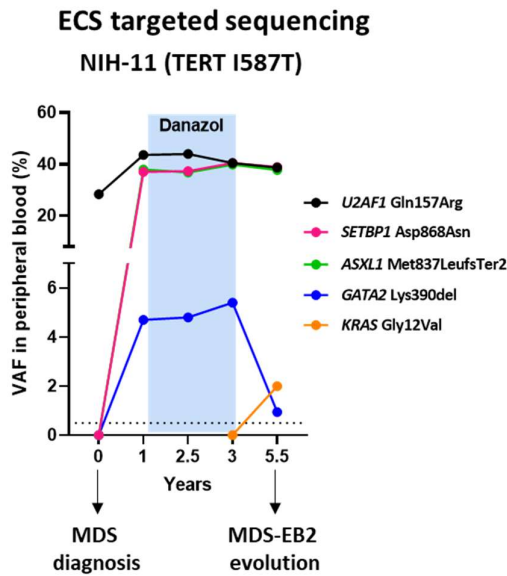


Figure 38. Clonal trajectory of each mutation and VAF for patient 1 (y-axis) over time (x-axis)

At MDS diagnosis, a *U2AF1* clone was present and remained at stable VAF over time. One year later, *SETBP1*, *GATA2*, *ASXL1*, clones appeared, also remaining stable until the development of MDS-EB2 when the *GATA2* clone decreased. At the time of MDS-EB2 progression, a new *KRAS* clone was detected.

Single cell DNA analysis was performed on BM sample from time of evolution to MDS-EB2. A total of 5570 single cells were analysed and showed 4 distinct clonal populations in addition to the wild-type cells. These subclones were: *U2AF1* alone, *U2AF1/KRAS/GATA2*, *U2AF1/SETBP1*, and *U2AF1/SETBP1/ASXL1*. The majority of cells were wild-type (3961cells), followed by *U2AF1/SETBP1* (604 cells), *U2AF1/SETBP1/ASXL1* (554 cells), *U2AF1/KRAS/GATA2* (258 cells), and *U2AF1* alone (193 cells) (Figure 39).

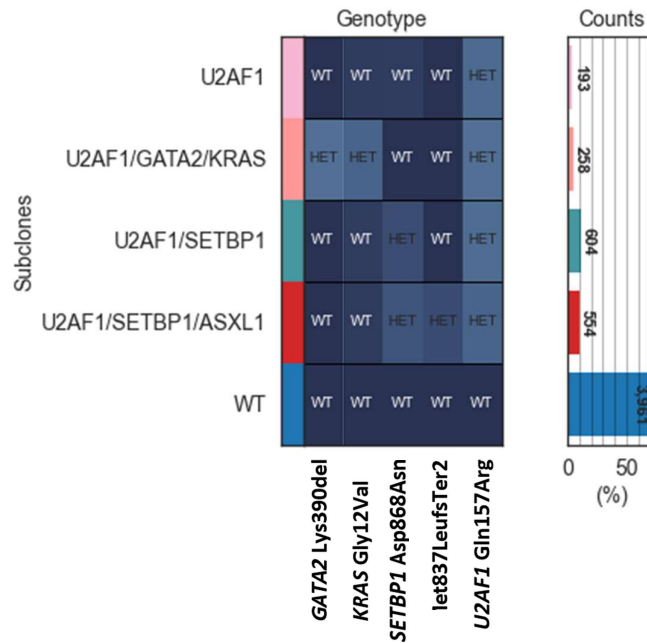


Figure 39. Single cell DNA clones and cell counts for patient 1. Genotype is showing 4 subclones. Wild type or mutant variants per clone are shown. Counts show the number of single cells containing each subclone. WT; wild-type, HET; heterozygous.

A total of 71% cells were wild type, while 29% were mutated. Of the 4 subclones, *U2AF1/SETBP1* occurred in 11% of cells, *U2AF1/SETBP1/ASXL1* in 10% of cells, *U2AF1/KRAS/GATA2* in 4.5% of cells, and *U2AF1* alone 3.5%. In a single cell, each mutation could be wildtype (absent), heterozygous (1 allele mutated), or homozygous (both alleles mutated). All subclones in this patient contained heterozygous variants.

Next, cells were subclassified into cell type according to immunophenotype using protein coding (Figure 40). As this was a BM sample, mature granulocytes were absent due to the Ficoll separation technique used.

The following cell types were identified:

Haematopoietic stem cells:

CD34+/CD117+

CD34-/CD117+

B-cells: CD19+/CD1c+/HLA-DR+

T-cells:

- 1. CD3+/CD8+
- 2. CD3+/CD4+/CD45RO+

NK-cells:CD56+/CD16+/CD3-

Monocytes: CD14+/CD16-/CD11b+/CD64+/HLA-DR+

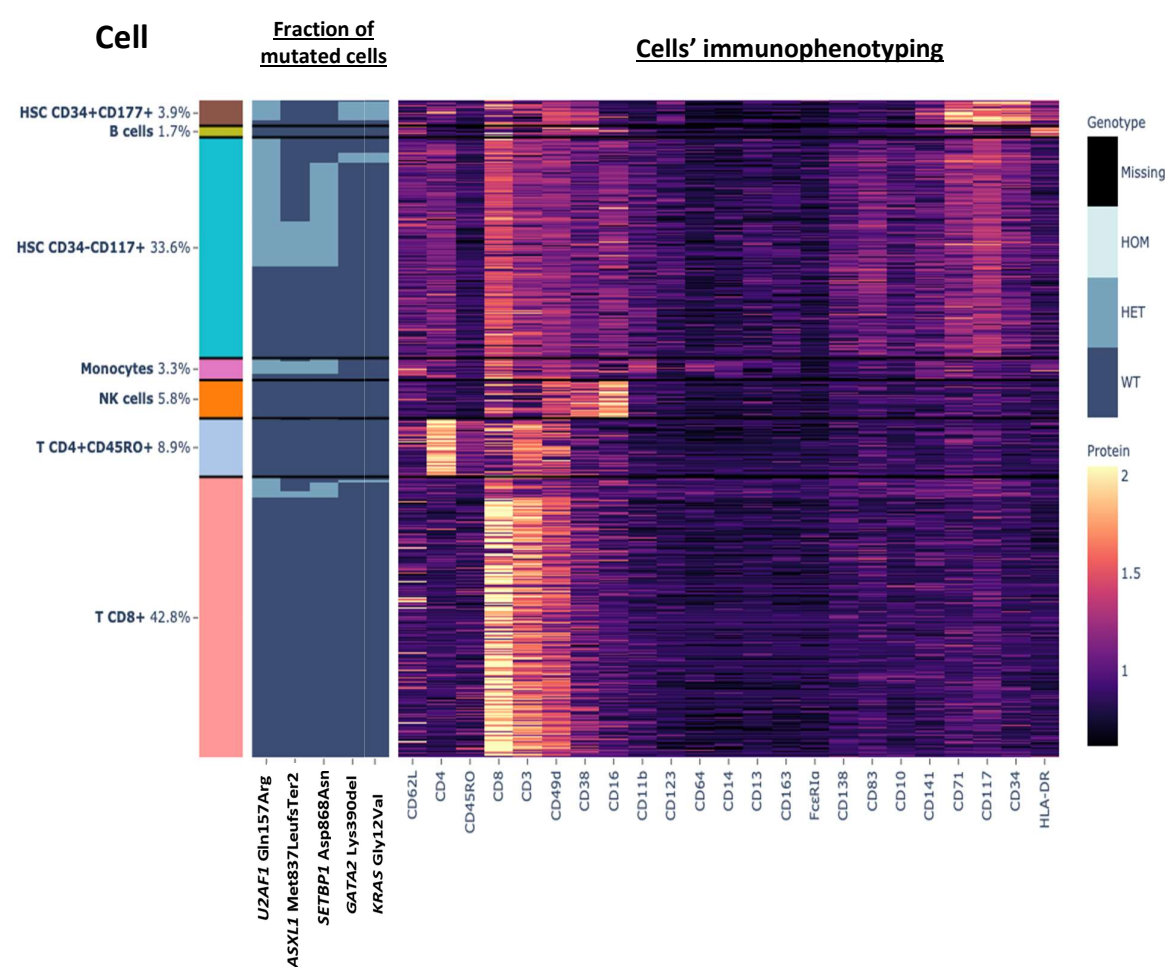


Figure 40. Single cells according to cell type using protein coding (immunophenotype) for patient 1

UMAP clustering (described in methods) was used to display cell protein data. Genotypes of the cells were superimposed over the protein UMAP to determine their presence in each cell type.

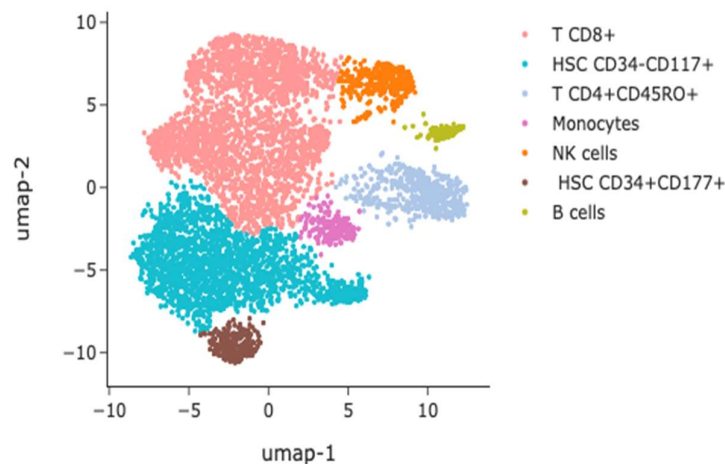


Figure 41. Cell populations classified by UMAP clustering for patient 1

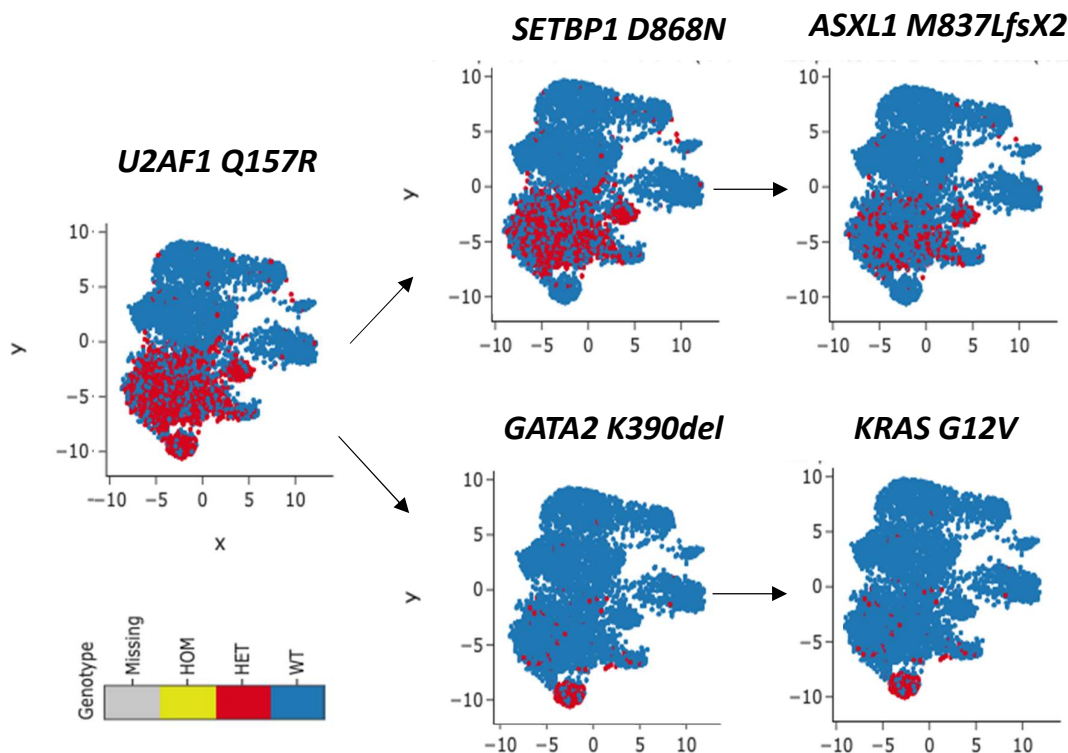


Figure 42. Cell populations with DNA genotype superimposed over protein UMAP clustering for patient 1. Denotations were: blue: wild type (mutation absent), red: Heterozygous, yellow: homozygous, grey: missing data

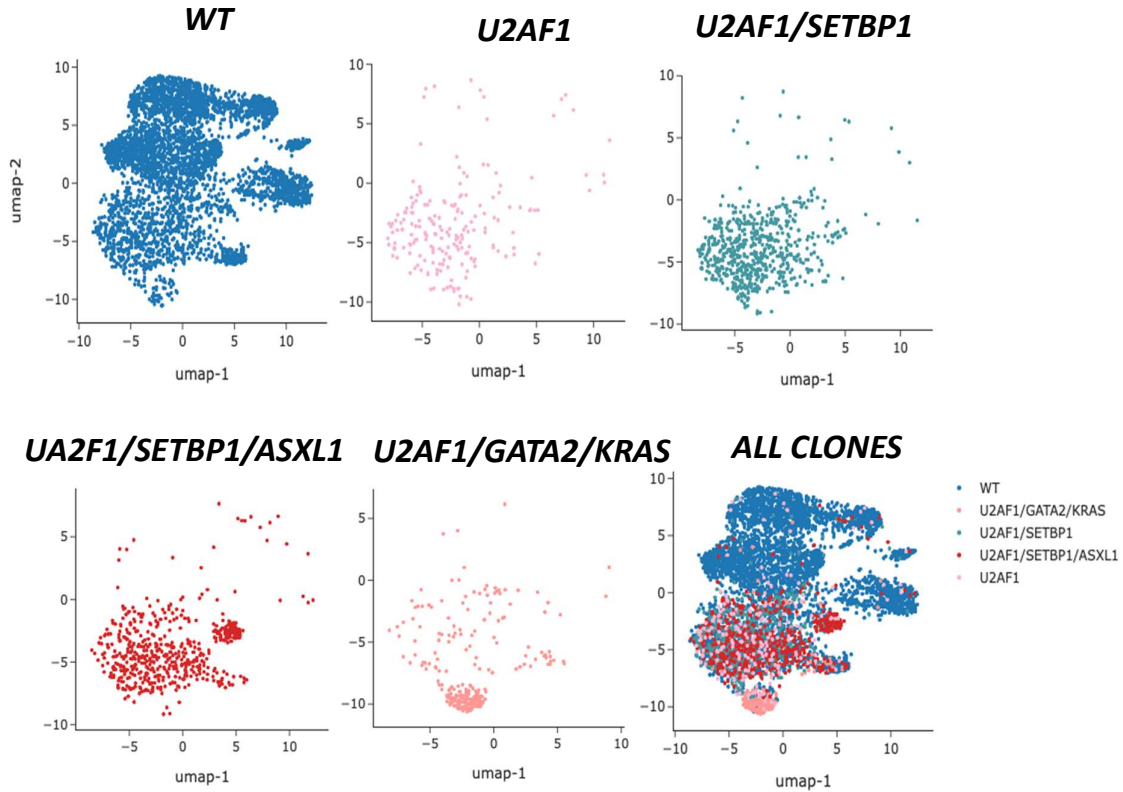


Figure 43. Mutant subclones detected by single cell superimposed over protein UMAP clustering for patient 1. Denotations were: blue: wild type (mutation absent), red: U2AF1/SETBP1/ASXL1, pink: U2AF1/GATA2/KRAS, green: U2AF1/SETBP1, salmon: U2AF1 alone

U2AF1 was enriched in stem cells, progenitors, and monocytes, while low or absent in lymphoid cells. Both *ASXL1* and *SETBP1* were also enriched in progenitors and monocytes but absent in stem cells. *GATA2* and *KRAS* were present in stem cells and absent in more mature cells, signalling that they may be driving leukaemic transformation.

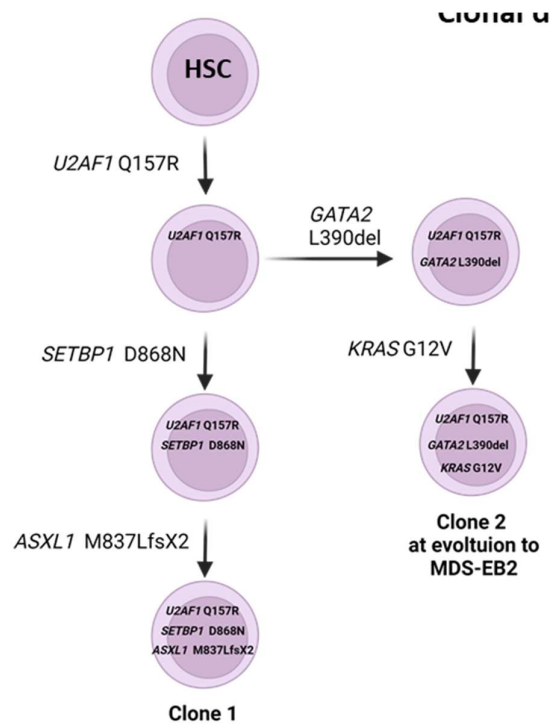


Figure 44. Clonal dynamics by single cell analysis for patient 1

Overall results were suggestive that *U2AF1* acted as a founder clone in this patient as it was present in all subsequent subclones, and the likely leukemic driver (*KRAS*) was absent in wild-type cells. The *U2AF1/GATA2/KRAS* being enriched in immature stem cells while *SETBP1* and *ASXL1* were not suggests that the former was the clone that ultimately drove leukemic evolution in this patient.

5.3.2 Patient 2 with *U2AF1* and other MDS mutations who did not develop myeloid malignancy

A 60-year-old female with a pathogenic *TERC* variant (*TERC* n.114_115del). Patient had bone marrow failure with counts meeting criteria for MAA. Patient also had reduced diffusion capacity on PFTs and evidence of fatty liver disease. Patient did not have a diagnosis of myeloid malignancy. Patient was treated with danazol therapy from 2012-2014 without haematological response.

Bulk sequencing revealed the following somatic variants occurring at various disease timepoints:

Gene	Amino Acid change
<i>ASXL1</i>	p.Glu635ArgfsTer15
<i>ETV6</i>	p.Gln143AlafsTer10
<i>U2AF1</i>	p.Ser34Tyr
<i>PHF6</i>	p.Arg275Gln

Clonal trajectory of each mutation and VAF over time are shown below based on results from bulk ECS.

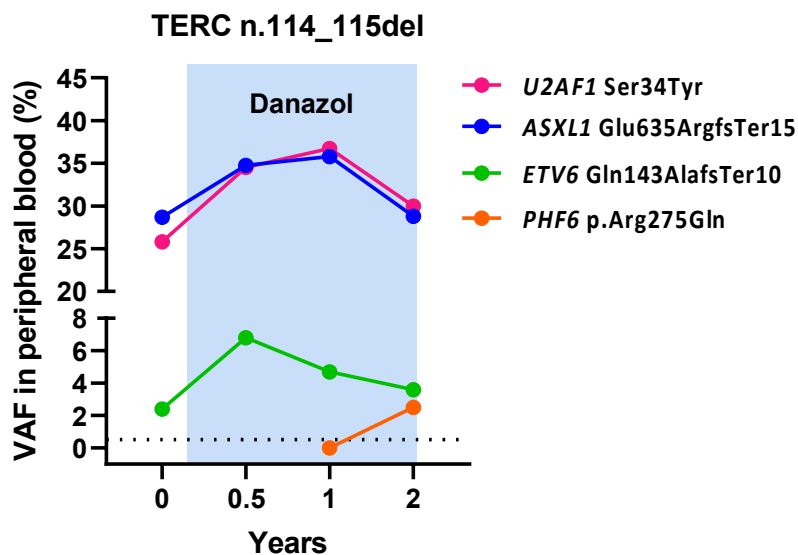


Figure 45. Clonal trajectory of each mutation and VAF for Patient 2 (y-axis) over time (x-axis)

U2AF1, *ASXL1*, and *ETV6* were present on the initial patient sample and remained stable over a period of two years. *PHF6* was acquired 1 year after the first sample was sequence and expanded on subsequent sample. This patient did not develop a myeloid malignancy.

Single cell DNA analysis was performed on BM sample 2 years after mutations were first assessed and detected. Of note, the amplicon for *ETV6* failed and as such this variant could not be assessed in this sample. A total of 3910 single cells were analysed and showed 3 distinct clonal populations in addition to the wild-type cells. These subclones were: *U2AF1* alone, *U2AF1/ASXL1*, and *U2AF1/PHF6/ASXL1*. The majority of cells were *U2AF1/ASXL1* (2111

cells), followed by wild-type (980 cells), *U2AF1* alone (594 cells), and *U2AF1/PHF6/ASXL1* (225 cells) (Figure 46)

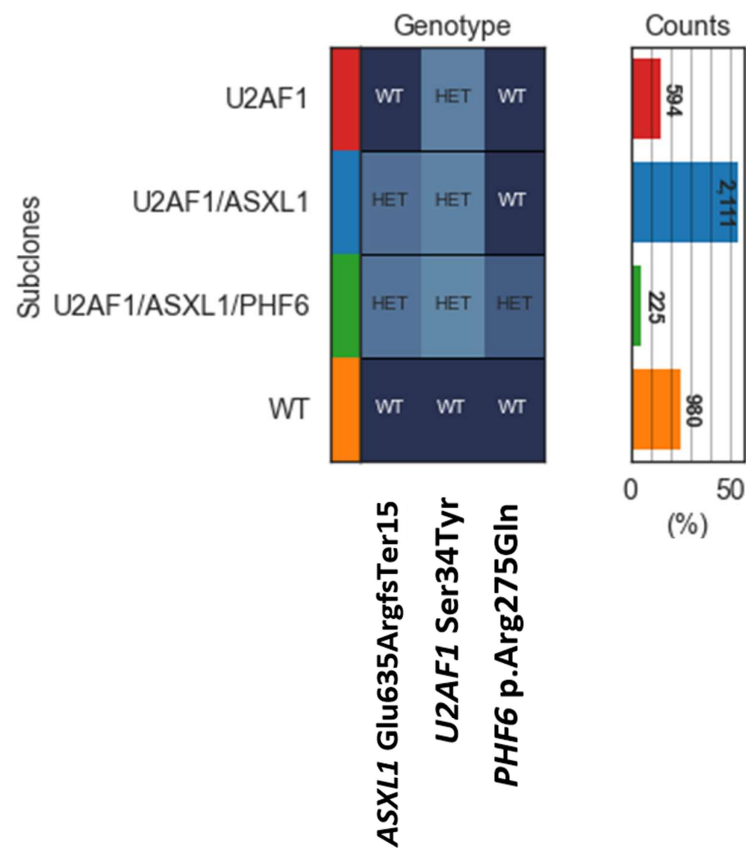


Figure 46. Single cell DNA clones and cell counts. Genotype is showing 3 subclones. Wild type or mutant variants per clone are shown. Counts show the number of single cells containing each subclone. WT; wild-type, HET; heterozygous.

A total of 25% cells were wild type, while 75% were mutated. Of the 3 subclones, *U2AF1/ASXL1* occurred in 54% of cells, *U2AF1* alone in 15% of cells, and *U2AF1/PHF6/ASXL1* in 6% of cells.

Again, cells were subclassified into cell type according to immunophenotype using protein coding.

The following cell types were identified (Figure 47):

Haematopoietic Stem Cells:

1. CD34+/CD117-
2. CD34-/CD117+

T-cells:

3. CD3+/CD8+
4. CD3+/CD4+

NK-cells: CD56+/CD16+/CD3-

Monocytes: CD14+/CD16+/CD11b+/CD64+/HLA-DR+

Neutrophils: CD16+/CD11b+/CD10+

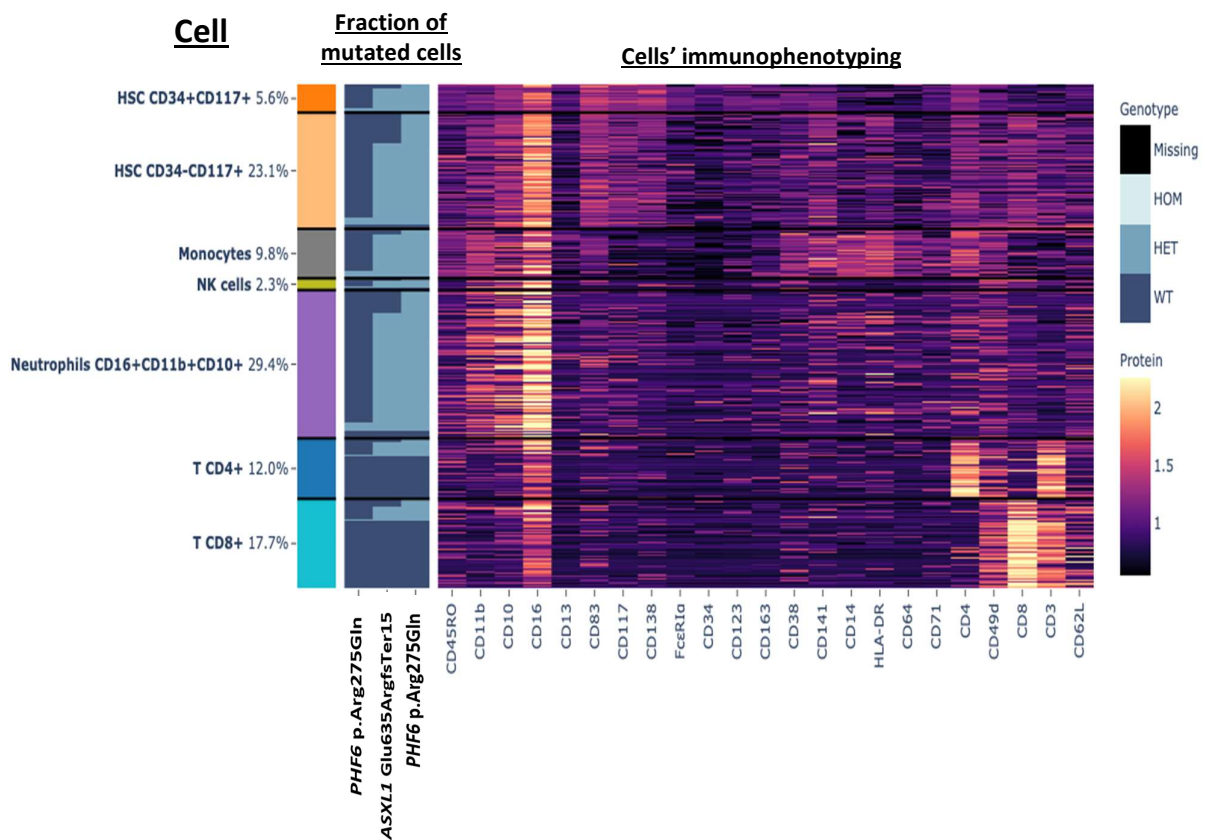


Figure 47. Single cells according to cell type using protein coding (immunophenotype)

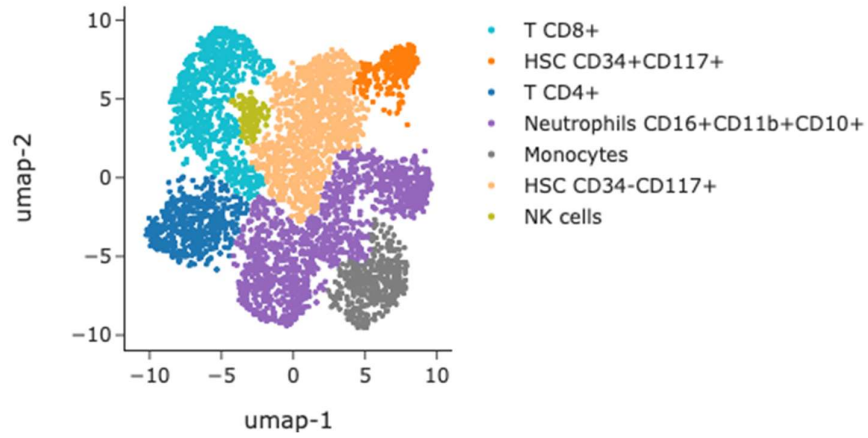


Figure 48. Cell populations classified by UMAP clustering

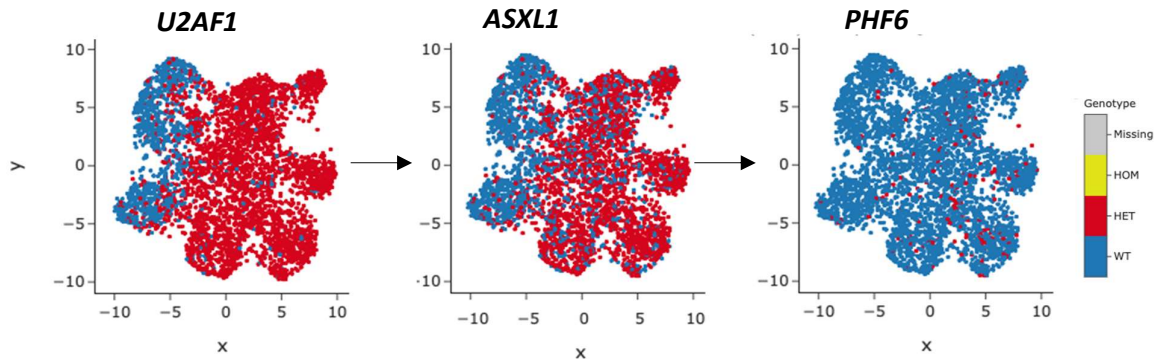


Figure 49. Cell populations with DNA genotype superimposed over protein UMAP clustering. Denotations were: blue: wild type (mutation absent), red: Heterozygous, yellow: homozygous, grey: missing data

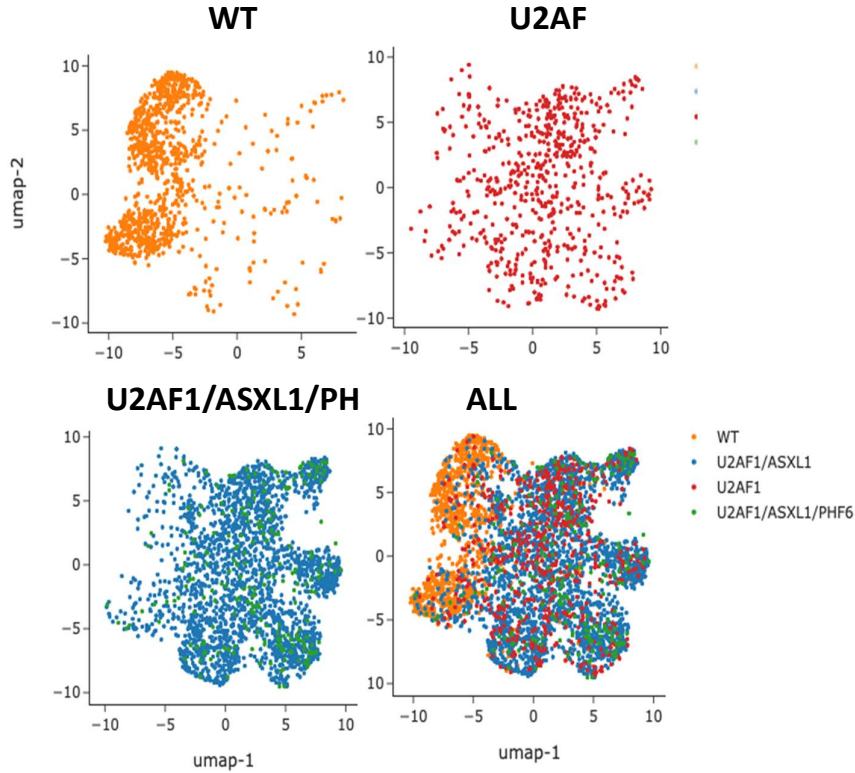


Figure 50. Mutant subclones detected by single cell superimposed over protein UMAP clustering. Denotations were: orange: wild type (mutation absent), red: *U2AF1* alone, green: *U2AF2/ASXL1/PHF6*, blue: *U2AF1/ASXL1*

U2AF1 and *ASXL1* were heavily enriched in most cell except T-cell populations while *PHF6* followed the same pattern in lower numbers of cells. This was suggestive that founding clonal event likely occurred in an early marrow progenitor cell.

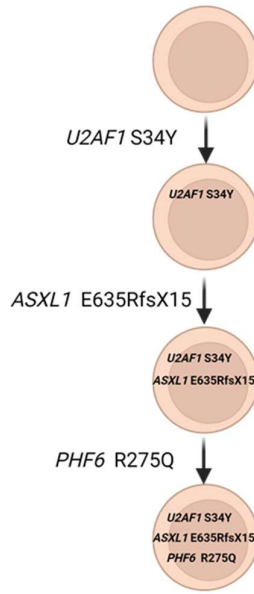


Figure 51. *Clonal dynamics by single cell analysis*

Since *U2AF1* was present in all subclones and *ASXL1* was present in all subclones containing *PHF6*, overall results were suggestive that *U2AF1* acted as a founder clone and that *ASXL1* and *PHF6* were subsequently linearly acquired in subclones.

5.4 Conclusions

Our data are suggestive that *U2AF1* acts as a founder clone with subsequent subclones developing within the *U2AF1* mutated cells rather than the wild-type ones. In our patient who ultimately evolved to MDS-EB2, this was associated with the development of a new subclone arising from the *U2AF1* mutated cells, that contained *KRAS* and *GATA2*. This stepwise progression has previously been demonstrated in de-novo MDS, where splicing factor mutations, including *U2AF1*, were shown to occur as early events in MDS pathogenesis while other mutations such as those in the *RAS* family or *FLT3* tend to be late events that drive high-risk evolution^{104,105}. It is hypothesized that the genetic mutations that occur early in MDS promote self-renewal and thus have a proliferative advantage. However, these cells may be more likely to accumulate subsequent mutations due to increased proliferation or direct downstream effects of the mutation on cell function¹⁰⁵.

Specific clones are known to be driven by environment; for instance, age related CHIP, such as *DNMT3A* is likely to be positively selected in an inflammatory setting^{106,107}. As these single cell data confirmed *U2AF1* as the founding clone in these two patients, it is likely that *U2AF1* is positively selected in TBD patients, possibly as a compensation to ongoing DDR activation, and once it develops can act as a driver for the development of subsequent MDS clones.

6. Final conclusions and discussion

6.1 Clinical and biological implications of the work

Telomere biology disorders are rare multisystem disorders due to accelerated telomere shortening that result in early mortality and an increased risk of cancer. Traditionally considered one of the classical IBMFS, TBD are increasingly being recognised as a heterogenous group of diseases where BMF may not predominate or may even be absent. Given the rarity of these disorders, research including large patient numbers is relatively scarce. Our work includes one of the largest cohorts of TBD patients reported; as the majority were on prospective clinical protocols and seen at our institutions, we were able to collect deep clinical phenotype data that may be lacking in registry studies. Our cohort was heterogenous, including both children and adults with a wide range of TBD related germline variants and coming from two different countries.

Clinical phenotype has recently been associated with genotype, with those having variants with autosomal recessive inheritance tending towards more severe disease and higher cancer risk than those with autosomal dominant disease. Our work builds upon this and identifies specific germline variants associated with a high risk of progression to haematological malignancy; *DKC1*, which is AR, but also *TERC* which is AD. Conversely, patients with *TINF2* and biallelic *RTEL1* had a very low risk of progression to MDS/AML. This has important implications for genetic counselling of patients and also for cancer screening needs assessment.

In examining our cohort, we noted that *U2AF1* was the predominant mutation seen in our TBD-MDS patients, prompting us to wonder whether it may be acting as a driver of haematological malignancy. Once we sequenced the entire cohort, it became clear that *U2AF1* was one of the commonest mutations seen in the entire cohort, though not exclusive to patients who had developed MDS. By using logistic regression, we found that the presence of mutations in the MDS/AML associated group, which were predominantly splicing factor mutations / *U2AF1*, did predict for the development of MDS/AML in patients >20 years old. No other mutation groups were associated with MDS/AML development, including *PPM1D* and *TP53*. Using DNA single cell analysis, we found that *U2AF1* acts as a founder clone, preceding the

development of subsequent clones, with these subsequent clones developing exclusively in the *U2AF1* mutated cells.

U2AF1 is an RNA spliceosome gene; it, and other RNA splicing factors, are recurrently mutated in MDS with *U2AF1* occurring in 10-15% of non-TBD MDS cases⁹³. *U2AF1* mutations may result abnormal splicing of genes involved in functionally important pathways such including cell cycle progression and RNA processing, however, it remains unclear how they directly contribute to the pathophysiology of malignancy⁹⁷. One of the genes previously shown to be affected downstream due to mutated *U2AF1* in MDS is ATR, a key components of the DDR, which may explain its predominance in the short-telomere environment⁹⁷. Therefore, it is possible that *U2AF1* arises as a direct consequence of the activated DDR that is known to occur in TBD, is ultimately maladaptive, and results in increased cancer susceptibility. Our lab is planning functional studies including CRISPR/Cas9 gene editing and RNA sequencing to further elucidate the effect that *U2AF1* has on cells with TBD-related germline variants.

PPM1D is a negative regulator of *TP53* that controls cellular responses to environmental stress, by promoting termination of the DDR pathway⁹⁰. It has been strongly associated with exposure to chemotherapy in patients with solid cancers and is enriched in therapy related MDS⁹⁰⁻⁹². In our cohort, *PPM1D* did not associate with the development of MDS/AML and did not influence survival. It was also not associated chemotherapy, suggesting an alternative driver of its predominance in TBD. Given the chronic activation of the DDR induced by telomere shortening that drives cellular senescence, it is possible that *PPM1D* arises as an attempt to compensate BMF driven by TP53 activation. However, *PPM1D* was not associated with difference in blood counts or survival compared to those without *PPM1D*, suggesting that its cell effect may be neutral.

Somatic reversion events in *POT1* and *TERTp* were also common in our cohort. POT1 variants were almost exclusively seen in patients with germline *TINF2*, who tend to have a severe clinical phenotype with DC and marrow failure. Somatic *POT1* mutations have been associated with telomere elongation in other malignancies but this was not seen in our cohort. As *POT1* is a telomere related gene in the shelterin complex, as is *TINF2*, it is possible that these mutations represent indirect somatic reversion to attempt to offset telomere shortening⁴². However, in our

patients *POT1* did not associate with improved blood counts. Further work is underway in our lab to assess whether *POT1* clones had any effect on telomere length or cell fitness over time.

6.2 Implications for future research

6.2.1 Cancer screening

Current practice for many seeing TBD patients is to perform annual bone marrow examinations due to the increased haematological cancer risk, though this is physician and patient dependent. This practice is onerous for patients as the procedure is invasive and can be painful. Currently the only marker that may guide clinicians on marrow frequency is the presence of a cytogenetic abnormality; guidelines recommend marrows every 3-4 months once a cytogenetic abnormality is detected for 2-3 marrows then annual thereafter⁴⁷. These recommendations are based on FA guidelines where strong data exist for the predictive value of cytogenetic abnormalities in developing haematological malignancy⁸⁴. In our study we identified two cohorts of patients at high risk of haematological malignancy based on their germline variant, those with *TERC* or *DKC1*. Conversely, patients with *TINF2* had no haematological malignancy develop and are likely much lower risk. We also identified that *U2AF1* and other mutations in the MDS/AML associated group, as assessed in the peripheral blood, act as strong biomarkers for the detection of patients with MDS/AML or those with a high risk of later developing it. While *TP53* was not associated with haematological malignancy, it was strongly associated with solid malignancy.

Therefore, we identified a particular cohort of patients where continued marrow surveillance should be implemented; those with *TERC* and *DKC1*. Patients with splicing factor mutations and other MDS/AML associated mutations should also have close surveillance. Patients with *TINF2* appear to have low rates of haematological cancer and less frequent bone marrow surveillance may be warranted. Based on our data, patients with *TP53* detected in marrow or peripheral blood should undergo screening for solid malignancy.

6.2.2 Androgen therapy

Androgens are used for treatment of BMF in TBD and have also been shown to elongate telomeres in patients with TBD⁵². The *TERTp* gene contains estrogen-response elements and

administration of high estrogen concentrations increases TERT transcription. Telomerase activity in haematopoietic stem cells is upregulated by androgens or estrogens but is inhibited by aromatase inhibitors (which block androgen to estrogen conversion) or tamoxifen (an estrogen antagonist)⁴⁸. However, there has been theoretical concern that androgen administration, by upregulating telomerase expression, may increase risk of clonal evolution to cancer, though there have been no clinical reports of such an effect to date. In our cohort of androgen treated patients who has serial samples performed, we did not observe that androgen therapy influenced clonal dynamics. However, larger numbers will be needed to confirm this.

6.2.3 Spliceosome inhibitors

Spliceosome inhibitors are current being assessed in standard MDS as a therapeutic option. These agents cause an excess in spliceosome dysfunction, based on the hypothesis that there is an amount of splicing dysfunction that is tolerated within a cell, but further alteration of splicing means the cell can no longer compensate resulting in selective cell death of the mutated cells¹⁰⁸. Given the predominance of splicing factors, in particular *U2AF1*, in our cohort of TBD patients, these agents may be efficacious in TBD-associated MDS/AML. Additionally, depending on the agents' safety and tolerability, they could be studied to see if eliminating the splicing factor clones prior to the development of MDS/AML prevents cancer development. A number of clinical trials are ongoing but to date no spliceosome inhibitors have been approved¹⁰⁹.

6.3.4 Gene therapy

As somatic reversion can occur in patients with IBMFS partially or completely rescuing haematopoiesis, it is possible that gene correction in the haemopoietic stem cell may restore haematopoietic function. Gene therapy has been successfully applied in some monogenic haematological disorders in which a single gene is involved, especially in the haemoglobinopathies and in haemophilia^{110,111}. However, in the IBMFS, though they are monogenic for each patient, there are a myriad of genes affected in each disorder, and multiple gene lesions are observed (point mutations, frameshifts, deletions, splicing variants), which complicates gene therapy⁴³.

To date, FA has seen the most success with two ongoing gene therapy clinical trials using direct lentiviral transfer, one in Spain and one in the USA. To date, the safety profile remains favourable but follow up remains short¹¹². Both trials target only one FANC gene, *FANCA*, limiting the treatments utility to only those with that specific mutation. The causative variant is found in *FANCA* for most FA patients (approximately 70%)¹¹³. However, in TBD the germline variants are more heterogenous with no one variant making up the majority of cases, meaning different there would be a need for different gene therapy protocols for each variant. In one recent study, direct correction of the *Tert* gene in mice using adeno-associated virus gene therapy rescued mice from severe BMF and lead to improved overall survival¹¹⁴. Additionally, direct gene editing of *TINF2* in ex vivo haematopoietic stem cells was able to restore telomere length¹¹⁵.

POT1 or *TERTp*, as examples of indirect somatic reversion, may represent possible targets for gene therapy. As the expression of genes affected in telomere maintenance is tightly regulated and overexpression may be associated with cancer development, there remains concern that gene therapy in TBD may result in unintended cancers. We did not observe increased cancer risk in patients with somatic *POT1* or *TERTp* mutations, however, we also did not observe any positive clinical consequences such as improved blood counts or survival. However, the VAF found in most patients was small and it is possible that increased correction of these genes may lead to functional effects. Given the multiple available approaches to explore gene therapy for TBD patients, through direct or indirect methods, it is hopeful that this may be an option for patients in the future.

6.3 Final conclusions

Our work delineated for the first time the clonal landscape of TBD and identified 4 recurrent somatic mutations: *U2AF1*, *PPM1D*, *TERTp*, and *POT1*. *U2AF1* and *TERTp* had significant clinical implications identified in this research; *U2AF1* was a marker of myeloid neoplasia and worse survival, and *TERTp* was predictive of lower OS. TP53 was also identified as a biomarker of solid malignancy. While the clinical implications of *POT1* and *PPM1D* are less clear, they act as molecular markers of TBD and further study will aid our understanding of these disorders. Based on our data, we believe that clonal selection in TBD is intrinsically driven by the impact of their germline mutation to correct cell fitness. Work is ongoing in our lab to better elucidate the functional effects of these somatic mutations.

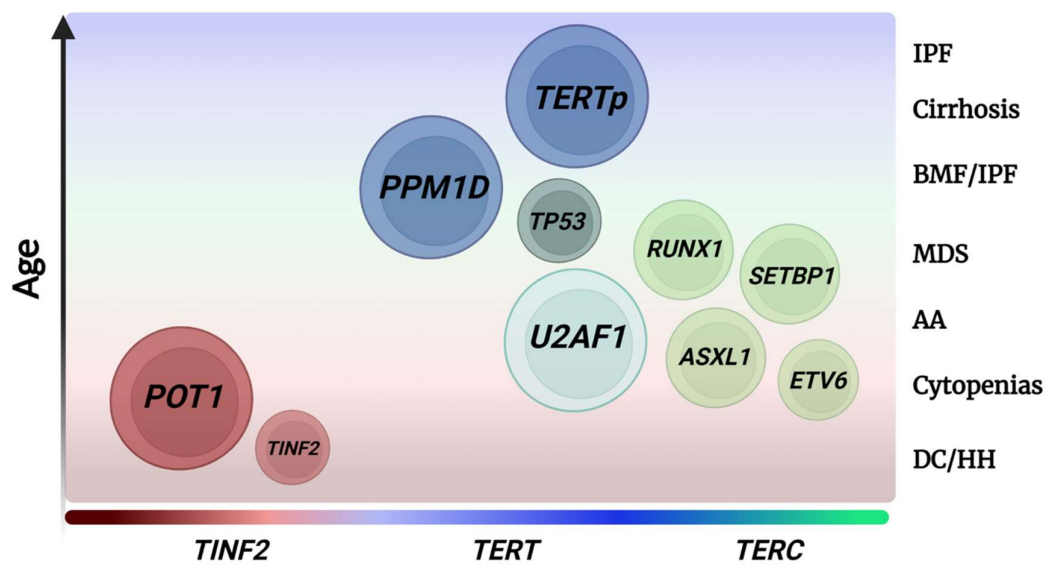


Figure 52. *Our conclusions regarding the factors influencing the clonal landscape of TBD*

References

1. Groarke EM, Young NS, Calvo KR. Distinguishing constitutional from acquired bone marrow failure in the hematology clinic. *Best Practice & Research Clinical Haematology* 2021;34(2):101275. DOI: <https://doi.org/10.1016/j.beha.2021.101275>.
2. Blackburn EH, Gall JG. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *Journal of Molecular Biology* 1978;120(1):33-53. DOI: [https://doi.org/10.1016/0022-2836\(78\)90294-2](https://doi.org/10.1016/0022-2836(78)90294-2).
3. Szostak JW, Blackburn EH. Cloning yeast telomeres on linear plasmid vectors. *Cell* 1982;29(1):245-255. DOI: [https://doi.org/10.1016/0092-8674\(82\)90109-X](https://doi.org/10.1016/0092-8674(82)90109-X).
4. Greider CW, Blackburn EH. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* 1987;51(6):887-98. (In eng). DOI: 10.1016/0092-8674(87)90576-9.
5. Gutierrez-Rodrigues F, Sahoo SS, Wlodarski MW, Young NS. Somatic mosaicism in inherited bone marrow failure syndromes. *Best Practice & Research Clinical Haematology* 2021;34(2):101279. DOI: <https://doi.org/10.1016/j.beha.2021.101279>.
6. Mustjoki S, Young NS. Somatic Mutations in “Benign” Disease. *New England Journal of Medicine* 2021;384(21):2039-2052. DOI: 10.1056/NEJMra2101920.
7. Cooper JN, Young NS. Clonality in context: hematopoietic clones in their marrow environment. *Blood* 2017;130(22):2363-2372. DOI: 10.1182/blood-2017-07-794362.
8. Dokal I. Dyskeratosis congenita in all its forms. *British Journal of Haematology* 2000;110(4):768-779. (Review). DOI: 10.1046/j.1365-2141.2000.02109.x.
9. Walne AJ, Dokal I. Dyskeratosis Congenita: A historical perspective. *Mechanisms of Ageing and Development* 2008;129(1):48-59. DOI: <https://doi.org/10.1016/j.mad.2007.10.006>.
10. Bryan HG, Nixon RK. Dyskeratosis Congenita and Familial Pancytopenia. *JAMA* 1965;192(3):203-208. DOI: 10.1001/jama.1965.03080160023005.
11. Heiss NS, Knight SW, Vulliamy TJ, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genetics* 1998;19(1):32-38. DOI: 10.1038/ng0598-32.
12. Vulliamy TJ, Knight SW, Mason PJ, Dokal I. Very Short Telomeres in the Peripheral Blood of Patients with X-Linked and Autosomal Dyskeratosis Congenita. *Blood Cells, Molecules, and Diseases* 2001;27(2):353-357. DOI: <https://doi.org/10.1006/bcmd.2001.0389>.
13. Vulliamy T, Marrone A, Goldman F, et al. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* 2001;413(6854):432-435. (Article). DOI: 10.1038/35096585.
14. Yamaguchi H, Calado RT, Ly H, et al. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *New England Journal of Medicine* 2005;352(14):1413-1424. (Article). DOI: 10.1056/NEJMoa042980.
15. Niewisch MR, Giri N, McReynolds LJ, et al. Disease progression and clinical outcomes in telomere biology disorders. *Blood* 2022;139(12):1807-1819. DOI: 10.1182/blood.2021013523.
16. Jyonouchi S, Forbes L, Ruchelli E, Sullivan KE. Dyskeratosis congenita: a combined immunodeficiency with broad clinical spectrum – a single-center pediatric experience. *Pediatric Allergy and Immunology* 2011;22(3):313-319. (<https://doi.org/10.1111/j.1399-3038.2010.01136.x>). DOI: <https://doi.org/10.1111/j.1399-3038.2010.01136.x>.
17. Wagner CL, Hanumanthu VS, Talbot CC, Jr., et al. Short telomere syndromes cause a primary T cell immunodeficiency. *The Journal of Clinical Investigation* 2018;128(12):5222-5234. DOI: 10.1172/JCI120216.
18. Shankar RK, Giri N, Lodish MB, et al. Bone mineral density in patients with inherited bone marrow failure syndromes. *Pediatric Research* 2017;82(3):458-464. DOI: 10.1038/pr.2017.117.

19. Himes RW, Chiou EH, Queliza K, et al. Gastrointestinal Hemorrhage: A Manifestation of the Telomere Biology Disorders. *J Pediatr* 2021;230:55-61.e4. (In eng). DOI: 10.1016/j.jpeds.2020.09.038.
20. Jonassaint NL, Guo N, Califano JA, Montgomery EA, Armanios M. The gastrointestinal manifestations of telomere-mediated disease. *Aging Cell* 2013;12(2):319-323. (<https://doi.org/10.1111/accel.12041>). DOI: <https://doi.org/10.1111/accel.12041>.
21. Higgs C, Crow YJ, Adams DM, et al. Understanding the evolving phenotype of vascular complications in telomere biology disorders. *Angiogenesis* 2019;22(1):95-102. (In eng). DOI: 10.1007/s10456-018-9640-7.
22. Armanios MY, Chen JJ, Cogan JD, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 2007;356(13):1317-26. (In eng). DOI: 10.1056/NEJMoa066157.
23. Cronkhite JT, Xing C, Raghu G, et al. Telomere shortening in familial and sporadic pulmonary fibrosis. *Am J Respir Crit Care Med* 2008;178(7):729-737. (In eng). DOI: 10.1164/rccm.200804-550OC.
24. Giri N, Ravichandran S, Wang Y, et al. Prognostic significance of pulmonary function tests in dyskeratosis congenita, a telomere biology disorder. *ERJ Open Res* 2019;5(4):00209-2019. (In eng). DOI: 10.1183/23120541.00209-2019.
25. Kapuria D, Ben-Yakov G, Ortolano R, et al. The Spectrum of Hepatic Involvement in Patients With Telomere Disease. *Hepatology* 2019;69(6):2579-2585. (In eng). DOI: 10.1002/hep.30578.
26. Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica* 2018;103(1):30-39. (In eng). DOI: 10.3324/haematol.2017.178111.
27. Schratz KE, Haley L, Danoff SK, et al. Cancer spectrum and outcomes in the Mendelian short telomere syndromes. *Blood* 2020;135(22):1946-1956. DOI: 10.1182/blood.2019003264.
28. Maschek H, Kaloutsi V, Rodriguez-Kaiser M, et al. Hypoplastic myelodysplastic syndrome: incidence, morphology, cytogenetics, and prognosis. *Annals of Hematology* 1993;66(3):117-122. DOI: 10.1007/BF01697619.
29. Rode L, Nordestgaard BG, Bojesen SE. Long telomeres and cancer risk among 95 568 individuals from the general population. *International Journal of Epidemiology* 2016;45(5):1634-1643. DOI: 10.1093/ije/dyw179.
30. Horn S, Figl A, Rachakonda PS, et al. *TERT* Promoter Mutations in Familial and Sporadic Melanoma. *Science* 2013;339(6122):959-961. DOI: doi:10.1126/science.1230062.
31. Collaboration TTM. Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. *JAMA Oncology* 2017;3(5):636-651. DOI: 10.1001/jamaoncol.2016.5945.
32. Ramsay AJ, Quesada V, Foronda M, et al. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nature Genetics* 2013;45(5):526-530. DOI: 10.1038/ng.2584.
33. Robles-Espinoza CD, Harland M, Ramsay AJ, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nature Genetics* 2014;46(5):478-481. DOI: 10.1038/ng.2947.
34. Swiggers SJ, Kuijpers MA, de Cort MJ, Beverloo HB, Zijlmans JM. Critically short telomeres in acute myeloid leukemia with loss or gain of parts of chromosomes. *Genes Chromosomes Cancer* 2006;45(3):247-56. (In eng). DOI: 10.1002/gcc.20286.
35. Meeker AK, Hicks JL, Platz EA, et al. Telomere shortening is an early somatic DNA alteration in human prostate tumorigenesis. *Cancer Res* 2002;62(22):6405-9. (In eng).
36. Maciejowski J, de Lange T. Telomeres in cancer: tumour suppression and genome instability. *Nature Reviews Molecular Cell Biology* 2017;18(3):175-186. DOI: 10.1038/nrm.2016.171.
37. Trybek T, Kowalik A, Góźdz S, Kowalska A. Telomeres and telomerase in oncogenesis (Review). *Oncol Lett* 2020;20(2):1015-1027. DOI: 10.3892/ol.2020.11659.

38. Barthel FP, Wei W, Tang M, et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nature Genetics* 2017;49(3):349-357. DOI: 10.1038/ng.3781.
39. Yuan X, Larsson C, Xu D. Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. *Oncogene* 2019;38(34):6172-6183. DOI: 10.1038/s41388-019-0872-9.
40. McMaster ML, Sun C, Landi MT, et al. Germline mutations in Protection of Telomeres 1 in two families with Hodgkin lymphoma. *British Journal of Haematology* 2018;181(3):372-377. (<https://doi.org/10.1111/bjh.15203>). DOI: <https://doi.org/10.1111/bjh.15203>.
41. Speedy HE, Kinnersley B, Chubb D, et al. Germ line mutations in shelterin complex genes are associated with familial chronic lymphocytic leukemia. *Blood* 2016;128(19):2319-2326. DOI: 10.1182/blood-2016-01-695692.
42. Gong Y, Stock AJ, Liu Y. The enigma of excessively long telomeres in cancer: lessons learned from rare human POT1 variants. *Current Opinion in Genetics & Development* 2020;60:48-55. DOI: <https://doi.org/10.1016/j.gde.2020.02.002>.
43. Groarke EM, Calado RT, Liu JM. Cell senescence and malignant transformation in the inherited bone marrow failure syndromes: Overlapping pathophysiology with therapeutic implications. *Seminars in Hematology* 2022. DOI: <https://doi.org/10.1053/j.seminhematol.2022.01.003>.
44. Calado RT, Cooper JN, Padilla-Nash HM, et al. Short telomeres result in chromosomal instability in hematopoietic cells and precede malignant evolution in human aplastic anemia. *Leukemia* 2012;26(4):700-707. DOI: 10.1038/leu.2011.272.
45. Artandi SE, Chang S, Lee S-L, et al. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* 2000;406(6796):641-645. DOI: 10.1038/35020592.
46. Riboni R, Casati A, Nardo T, et al. Telomeric fusions in cultured human fibroblasts as a source of genomic instability. *Cancer Genetics and Cytogenetics* 1997;95(2):130-136. DOI: [https://doi.org/10.1016/S0165-4608\(96\)00248-8](https://doi.org/10.1016/S0165-4608(96)00248-8).
47. Groarke EMG, N; Young, N.S. Medical Management of Bone Marrow Failure in Telomere Biology Disorders. In: Sharon Savage KS, Suneet Aggarwal, ed. *Telomere Biology Disorders: Diagnosis and Management Guidelines: Team Telomere*; 2022.
48. Calado RT, Yewdell WT, Wilkerson KL, et al. Sex hormones, acting on the TERT gene, increase telomerase activity in human primary hematopoietic cells. *Blood* 2009;114(11):2236-2243. DOI: <https://doi.org/10.1182/blood-2008-09-178871>.
49. Calado RT, Young NS. Telomere diseases. *N Engl J Med* 2009;361(24):2353-65. (In eng). DOI: 10.1056/NEJMra0903373.
50. Khincha PP, Wentzensen IM, Giri N, Alter BP, Savage SA. Response to androgen therapy in patients with dyskeratosis congenita. *British Journal of Haematology* 2014;165(3):349-357. DOI: <https://doi.org/10.1111/bjh.12748>.
51. Zlateska B, Ciccolini A, Dror Y. Treatment of dyskeratosis congenita-associated pulmonary fibrosis with danazol. *Pediatr Pulmonol* 2015;50(12):E48-51. (In eng). DOI: 10.1002/ppul.23235.
52. Townsley DM, Dumitriu B, Liu D, et al. Danazol Treatment for Telomere Diseases. *New England Journal of Medicine* 2016;374(20):1922-1931. DOI: 10.1056/NEJMoa1515319.
53. Khincha PP, Bertuch AA, Gadalla SM, Giri N, Alter BP, Savage SA. Similar telomere attrition rates in androgen-treated and untreated patients with dyskeratosis congenita. *Blood Adv* 2018;2(11):1243-1249. (In eng). DOI: 10.1182/bloodadvances.2018016964.
54. Kirschner M, Vieri M, Kricheldorf K, et al. Androgen derivatives improve blood counts and elongate telomere length in adult cryptic dyskeratosis congenita. *Br J Haematol* 2021;193(3):669-673. (In eng). DOI: 10.1111/bjh.16997.

55. Gadalla SM, Sales-Bonfim C, Carreras J, et al. Outcomes of Allogeneic Hematopoietic Cell Transplantation in Patients with Dyskeratosis Congenita. *Biology of Blood and Marrow Transplantation* 2013;19(8):1238-1243. DOI: <https://doi.org/10.1016/j.bbmt.2013.05.021>.
56. Barbaro P, Vede A. Survival after Hematopoietic Stem Cell Transplant in Patients with Dyskeratosis Congenita: Systematic Review of the Literature. *Biol Blood Marrow Transplant* 2016;22(7):1152-1158. (In eng). DOI: 10.1016/j.bbmt.2016.03.001.
57. Ayas M, Nassar A, Hamidieh AA, et al. Reduced intensity conditioning is effective for hematopoietic SCT in dyskeratosis congenita-related BM failure. *Bone Marrow Transplant* 2013;48(9):1168-72. (In eng). DOI: 10.1038/bmt.2013.35.
58. Nelson AS, Marsh RA, Myers KC, et al. A Reduced-Intensity Conditioning Regimen for Patients with Dyskeratosis Congenita Undergoing Hematopoietic Stem Cell Transplantation. *Biol Blood Marrow Transplant* 2016;22(5):884-8. (In eng). DOI: 10.1016/j.bbmt.2016.01.026.
59. Courtwright AM, Lamattina AM, Takahashi M, et al. Shorter telomere length following lung transplantation is associated with clinically significant leukopenia and decreased chronic lung allograft dysfunction-free survival. *ERJ Open Res* 2020;6(2) (In eng). DOI: 10.1183/23120541.00003-2020.
60. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* 2017;130(6):742-752. DOI: 10.1182/blood-2017-02-769869.
61. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* 2014;371(26):2488-2498. DOI: 10.1056/NEJMoa1408617.
62. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *New England Journal of Medicine* 2017;377(2):111-121. DOI: 10.1056/NEJMoa1701719.
63. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014;371(26):2477-87. (In eng). DOI: 10.1056/NEJMoa1409405.
64. Malcovati L, Galli A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood* 2017;129(25):3371-3378. (In eng). DOI: 10.1182/blood-2017-01-763425.
65. Groarke EM, Young NS. Aging and Hematopoiesis. *Clin Geriatr Med* 2019;35(3):285-293. (In eng). DOI: 10.1016/j.cger.2019.03.001.
66. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* 2018;559(7714):400-404. DOI: 10.1038/s41586-018-0317-6.
67. Estey EH. Acute myeloid leukemia: 2021 update on risk-stratification and management. *Am J Hematol* 2020;95(11):1368-1398. (In eng). DOI: 10.1002/ajh.25975.
68. Bersanelli M, Travaglino E, Meggendorfer M, et al. Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in Myelodysplastic Syndromes. *Journal of Clinical Oncology* 2021;39(11):1223-1233. DOI: 10.1200/JCO.20.01659.
69. Yoshizato T, Dumitriu B, Hosokawa K, et al. Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *New England Journal of Medicine* 2015;373(1):35-47. DOI: 10.1056/NEJMoa1414799.
70. Groarke EM, Patel BA, Shalhoub R, et al. Strong Predictors of Chromosomal Aberrations and Myeloid Neoplasia Following Immunosuppression Therapy for Severe Aplastic Anemia: A Retrospective Cohort Study. *Blood* 2021;138(Supplement 1):601-601. DOI: 10.1182/blood-2021-146070.
71. Nicoletti E, Rao G, Bueren JA, et al. Mosaicism in Fanconi anemia: concise review and evaluation of published cases with focus on clinical course of blood count normalization. *Annals of Hematology* 2020;99(5):913-924. DOI: 10.1007/s00277-020-03954-2.

72. Gross M, Hanenberg H, Lobitz S, et al. Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction. *Cytogenetic and Genome Research* 2002;98(2-3):126-135. DOI: 10.1159/000069805.
73. Dror Y, Durie P, Ginzberg H, et al. Clonal evolution in marrows of patients with Shwachman-Diamond syndrome: A prospective 5-year follow-up study. *Experimental Hematology* 2002;30(7):659-669. DOI: [https://doi.org/10.1016/S0301-472X\(02\)00815-9](https://doi.org/10.1016/S0301-472X(02)00815-9).
74. Pressato B, Valli R, Marletta C, et al. Deletion of chromosome 20 in bone marrow of patients with Shwachman-Diamond syndrome, loss of the EIF6 gene and benign prognosis. *British Journal of Haematology* 2012;157(4):503-505. (<https://doi.org/10.1111/j.1365-2141.2012.09033.x>). DOI: <https://doi.org/10.1111/j.1365-2141.2012.09033.x>.
75. Kennedy AL, Myers KC, Bowman J, et al. Distinct genetic pathways define pre-malignant versus compensatory clonal hematopoiesis in Shwachman-Diamond syndrome. *Nature Communications* 2021;12(1):1334. DOI: 10.1038/s41467-021-21588-4.
76. Wong JC, Bryant V, Lamprecht T, et al. Germline SAMD9 and SAMD9L mutations are associated with extensive genetic evolution and diverse hematologic outcomes. *JCI Insight* 2018;3(14):e121086. (In eng). DOI: 10.1172/jci.insight.121086.
77. Sahoo SS, Pastor VB, Goodings C, et al. Clinical evolution, genetic landscape and trajectories of clonal hematopoiesis in SAMD9/SAMD9L syndromes. *Nat Med* 2021;27(10):1806-1817. (In eng). DOI: 10.1038/s41591-021-01511-6.
78. Gutierrez-Rodrigues F, Donaires FS, Pinto A, et al. Pathogenic TERT promoter variants in telomere diseases. *Genetics in Medicine* 2019;21(7):1594-1602. DOI: 10.1038/s41436-018-0385-x.
79. Maryoung L, Yue Y, Young A, et al. Somatic mutations in telomerase promoter counterbalance germline loss-of-function mutations. *The Journal of clinical investigation* 2017;127(3):982-986. (In eng). DOI: 10.1172/JCI91161.
80. Schratz KE, Gaysinskaya V, Cosner ZL, et al. Somatic reversion impacts myelodysplastic syndromes and acute myeloid leukemia evolution in the short telomere disorders. *J Clin Invest* 2021;131(18) (In eng). DOI: 10.1172/jci147598.
81. Alter BP, Giri N, Savage SA, et al. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *British Journal of Haematology* 2010;150(2):179-188. (<https://doi.org/10.1111/j.1365-2141.2010.08212.x>). DOI: <https://doi.org/10.1111/j.1365-2141.2010.08212.x>.
82. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016;127(20):2391-2405. DOI: 10.1182/blood-2016-03-643544.
83. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2018;198(5):e44-e68. DOI: 10.1164/rccm.201807-1255ST.
84. Meyer S, Neitzel H, Tönnies H. Chromosomal aberrations associated with clonal evolution and leukemic transformation in fanconi anemia: clinical and biological implications. *Anemia* 2012;2012:349837-349837. (In eng). DOI: 10.1155/2012/349837.
85. Marie Sebert SGTL, Alix Rousseau, Olivier Bluteau, Rathana Kim, Raouf Benabelali, Flore Sicre de Fontbrune, Loïc Maillard, Carèle Fedronie, Nadia Vasquez, Yves Bertrand, Jean-Hugues Dalle, Francois Sigaux, Sylvie Chevret, Gerard Socie, Hugues de Thé, Christophe Antoniewski, Dominique Bluteau, Regis Peffault De Latour, Jean Soulier. Clonal Hematopoiesis Driven By MDM4 Amplification Defines a Canonical Route Towards Secondary MDS/AML in Fanconi Anemia Patients. *American Society of Hematology. Atlanta2021*.
86. Gregory JJ, Jr., Wagner JE, Verlander PC, et al. Somatic mosaicism in Fanconi anemia: evidence of genotypic reversion in lymphohematopoietic stem cells. *Proc Natl Acad Sci U S A* 2001;98(5):2532-2537. (In eng). DOI: 10.1073/pnas.051609898.

87. Kirschner M, Maurer A, Wlodarski MW, et al. Recurrent somatic mutations are rare in patients with cryptic dyskeratosis congenita. *Leukemia* 2018;32(8):1762-1767. DOI: 10.1038/s41375-018-0125-x.
88. ArcherDx. Archer® VariantPlex® HS/HGC Protocol for Illumina®.
(https://f.hubspotusercontent40.net/hubfs/4445440/Protocols/LA560.1%20Protocol,%20VariantPlex%C2%AE-HS_HGC%20Reagents%20for%20Illumina%C2%AE_pdf.pdf).
89. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. *Science (New York, NY)* 2019;366(6465):eaan4673. (In eng). DOI: 10.1126/science.aan4673.
90. Kleiblova P, Shaltiel IA, Benada J, et al. Gain-of-function mutations of PPM1D/Wip1 impair the p53-dependent G1 checkpoint. *J Cell Biol* 2013;201(4):511-521. (In eng). DOI: 10.1083/jcb.201210031.
91. Swisher EM, Harrell MI, Norquist BM, et al. Somatic Mosaic Mutations in PPM1D and TP53 in the Blood of Women With Ovarian Carcinoma. *JAMA oncology* 2016;2(3):370-372. (In eng). DOI: 10.1001/jamaoncol.2015.6053.
92. Lindsley RC, Saber W, Mar BG, et al. Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *The New England journal of medicine* 2017;376(6):536-547. (In eng). DOI: 10.1056/NEJMoa1611604.
93. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011;478(7367):64-69. DOI: 10.1038/nature10496.
94. Damm F, Kosmider O, Gelsi-Boyer V, et al. Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* 2012;119(14):3211-3218. DOI: 10.1182/blood-2011-12-400994.
95. Ogawa S. Genetics of MDS. *Blood* 2019;133(10):1049-1059. DOI: 10.1182/blood-2018-10-844621.
96. Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* 2012;119(15):3578-3584. DOI: 10.1182/blood-2011-12-399337.
97. Przychodzen B, Jerez A, Guinta K, et al. Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* 2013;122(6):999-1006. DOI: 10.1182/blood-2013-01-480970.
98. Esfahani MS, Lee LJ, Jeon YJ, et al. Functional significance of U2AF1 S34F mutations in lung adenocarcinomas. *Nat Commun* 2019;10(1):5712. (In eng). DOI: 10.1038/s41467-019-13392-y.
99. Desai P, Mencia-Trinchant N, Savenkov O, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nature Medicine* 2018;24(7):1015-1023. DOI: 10.1038/s41591-018-0081-z.
100. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* 2015;126(1):9-16. DOI: 10.1182/blood-2015-03-631747.
101. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nature Reviews Cancer* 2017;17(1):5-19. DOI: 10.1038/nrc.2016.112.
102. Wen L, Tang F. Recent advances in single-cell sequencing technologies. *Precision Clinical Medicine* 2022;5(1). DOI: 10.1093/pcmedi/pbac002.
103. MissionBio. TotalSeq™-D Heme Oncology Cocktail from BioLegend.
(https://missionbio.com/products/panels/totalseq-d-heme-oncology/?gclid=EAlaIqobChMlw73y05XJ9wIVdMyzCh15BQpEEAAYASAAEgLDRfD_BwE).
104. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013;122(22):3616-3627. DOI: 10.1182/blood-2013-08-518886.

105. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer* 2017;17(1):5-19. (In eng). DOI: 10.1038/nrc.2016.112.
106. Cobo I, Tanaka T, Glass CK, Yeang C. Clonal hematopoiesis driven by DNMT3A and TET2 mutations: role in monocyte and macrophage biology and atherosclerotic cardiovascular disease. *Curr Opin Hematol* 2022;29(1):1-7. (In eng). DOI: 10.1097/moh.0000000000000688.
107. Avagyan S, Henninger JE, Mannherz WP, et al. Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis. *Science* 2021;374(6568):768-772. DOI: doi:10.1126/science.aba9304.
108. Brunner AM, Steensma DP. Targeting Aberrant Splicing in Myelodysplastic Syndromes: Biologic Rationale and Clinical Opportunity. *Hematology/oncology clinics of North America* 2020;34(2):379-391. (In eng). DOI: 10.1016/j.hoc.2019.10.003.
109. Steensma DP, Wermke M, Klimek VM, et al. Phase I First-in-Human Dose Escalation Study of the oral SF3B1 modulator H3B-8800 in myeloid neoplasms. *Leukemia* 2021;35(12):3542-3550. DOI: 10.1038/s41375-021-01328-9.
110. Nathwani AC. Gene therapy for hemophilia. *Hematology* 2019;2019(1):1-8. DOI: 10.1182/hematology.2019000007.
111. Kanter J, Falcon C. Gene therapy for sickle cell disease: where we are now? *Hematology* 2021;2021(1):174-180. DOI: 10.1182/hematology.2021000250.
112. Czechowicz A, Sevilla J, Agarwal R, et al. Gene Therapy for Fanconi Anemia [Group A]: Interim Results of RP-L102 Clinical Trials. *Blood* 2021;138(Supplement 1):3968-3968. DOI: 10.1182/blood-2021-147071.
113. Shafqat S, Tariq E, Parnes AD, Dasouki MJ, Ahmed SO, Hashmi SK. Role of gene therapy in Fanconi anemia: A systematic and literature review with future directions. *Hematology/Oncology and Stem Cell Therapy* 2021;14(4):290-301. DOI: <https://doi.org/10.1016/j.hemonc.2021.02.001>.
114. Bär C, Povedano JM, Serrano R, et al. Telomerase gene therapy rescues telomere length, bone marrow aplasia, and survival in mice with aplastic anemia. *Blood* 2016;127(14):1770-1779. DOI: 10.1182/blood-2015-08-667485.
115. Choo S, Lorbeer FK, Regalado SG, et al. Editing TINF2 as a therapeutic approach to restore telomere length in dyskeratosis congenita. *Blood* 2022. DOI: <https://doi.org/10.1182/blood.2021013750>.
116. Bernard E, Tuechler H, Greenberg PL, et. al. Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. *NEJM Evid.* 2022 Jul;1(7):EVIDoa2200008. doi: 10.1056/EVIDoa2200008. Epub 2022 Jun 12

Appendix A:

Description of clinical protocols

Danazol for Genetic Bone Marrow and Lung Disorders

NCT01441037

Study Description (clinicaltrials.gov):

Objectives:

- To study the safety and effectiveness of Danazol for bone marrow and lung disorders caused by damaged genes.

Eligibility:

- Individuals at least 2 years of age who have cytopenia or lung fibrosis due to telomere biology disorder (TL length <1st percentile).

Design:

- Patients screened with a physical exam and medical history. Testing: blood and urine tests, CT chest, and PFTs, 6-minute walking test and bone marrow biopsy.
- Danazol 800mg twice a day for two years.
- Regular study visits at 6, 12, and 24 months, with blood tests, CT chest, PFTs, and a 6-minute walking test. A bone marrow sample collected at the 12-month visit.
- Study follow up of 2 years.

Primary endpoint:

- Number of patients with attenuated telomere attrition at 6 months

Sponsor: National Heart, Lung, and Blood Institute, National Institutes of Health

Principal Investigator: Neal S. Young

Study Status: Completed, has results

- Townsley DM, Dumitriu B, Liu D, Biancotto A, Weinstein B, Chen C, Hardy N, Mihalek AD, Lingala S, Kim YJ, Yao J, Jones E, Gochuico BR, Heller T, Wu CO, Calado RT, Scheinberg P, Young NS. Danazol Treatment for Telomere Diseases. N Engl J Med. 2016 May 19;374(20):1922-31. doi: 10.1056/NEJMoa1515319. PMID: 27192671)

Low-dose danazol for the treatment of telomere-related diseases

NCT03312400

Study Description (clinicaltrials.gov):

Objectives:

- To study the safety and effectiveness of Danazol for bone marrow and lung disorders caused by damaged genes.

Eligibility:

- Individuals at least 3 years of age who have cytopenias, lung fibrosis, or liver cirrhosis and a pathogenic or likely pathogenetic mutation in a telomere related gene and telomere length <10th percentile.

Design:

- Patients screened with a physical exam and medical history. Testing: blood and urine tests, CT chest, and PFTs, 6-minute walking test and bone marrow biopsy.
- Randomised to Danazol 200mg twice a day or 400mg twice daily for 6 months; then crossed over.
- Regular study visits at 6, 12, and 18 months, with blood tests, CT chest, PFTs, and a 6-minute walking test.
- Study follow up of 2 years.

Primary endpoint:

- Number of patients with attenuated telomere attrition at 6 months

Sponsor: National Heart, Lung, and Blood Institute, National Institutes of Health

Principal Investigator: Emma M. Groarke

Study Status: Actively enrolling

Natural History of Acquired and Inherited Bone Marrow Failure Syndromes

NCT05012111

Study Description (clinicaltrials.gov):

Objectives:

- To follow patients diagnosed with acquired or inherited bone marrow failure disease and study the long-term effects of the disease and its treatments on organ function

Eligibility:

- Individuals at least 3 years of age who have a diagnosis of acquired or inherited bone marrow failure

Design:

- Patients screened with a physical exam and medical history.
- Patients followed at minimum every 3 years to assess for outcomes
- Study follow up of 20 years.

Sponsor: National Heart, Lung, and Blood Institute, National Institutes of Health

Principal Investigator: Emma M. Groarke

Study Status: Actively enrolling

Cancer in Inherited Bone Marrow Failure Syndromes

NCT00027274

Study Description (clinicaltrials.gov):

Objectives:

- To follow patients diagnosed inherited bone marrow failure disease and study the long-term risk of hematologic or solid malignancy

Eligibility:

- Individuals who have a diagnosis of inherited bone marrow failure

Design:

- Natural history study, with questionnaires, clinical evaluations, clinical and research laboratory test, review of medical records, cancer surveillance

Sponsor: National Cancer Institute, National Institutes of Health

Principal Investigator: Neelam Giri

Study Status: Actively enrolling

Nandrolone Decanoate in the Treatment of Telomeropathies

NCT02055456

Study Description (clinicaltrials.gov):

Objectives:

- To study the safety and effectiveness of Danazol for bone marrow and lung disorders caused by damaged genes.

Eligibility:

- Individuals at least 2 years of age who have cytopenias, lung fibrosis, or liver cirrhosis and telomere length <1st percentile.

Design:

- Administered Nandrolone 5mg/kg IM every 2 weeks
- Study follow up of 2 years.

Primary endpoint:

- Number of patients with attenuated telomere attrition

Sponsor: Ribeirao Preto School of Medicine at University of Sao Paulo

Principal Investigator: Rodrigo T. Calado

Study Status: Completed, results pending

Appendix B:

Protocol for Archer VariantPlex -HS/HGC Protocol for Illumina

DNA Library Preparation

Reagents:

- Archer VariantPlex reagents in lyophilized format (DNA fragmentation solution, End Repair solution).
- Gene specific primers (GSPs) that target panel specific regions of interest during PCR amplification
- Archer Molecular Barcode Adapters (MBCs) adapters that tag each unique molecule with a barcode and common region prior to amplification.

Step 1: DNA fragmentation.

1. Purified DNA samples were diluted in 10mM Tris-HCL pH 8.0 to a final volume of 50uL and combined with (Archer prepared) DNA fragmentation solution reaction tubes stored on ice.
2. Samples were transferred to a thermal cycler and incubated as follows:
 - 1 minute at 4°C
 - 12 minutes at 37°C
 - 20 minutes at 72°C
3. Samples were then spun down and placed on ice

Step 2: End repair.

1. 50uL of DNA fragmentation reaction was added to (Archer prepared) End Repair solution. This mixture was dissolved, mixed and spun down.
2. Mixture was transferred to a thermal cycler for 30 minutes at 25°C.

Cleanup after end repair

1. AMPure XP beads (Beckman Coulter) were resuspended by vortexing
2. 2.5X volume (125µL) of AMPure XP beads were added to each End Repair reaction.
3. Mixture was vortexed and incubated for 5 minutes at room temperature
4. Tubes were spun and placed on magnet for 4 minutes to allow formation of a bead pellet against the tube wall
5. A pipette was used to remove and discard the supernatant without disturbing the bead pellet.
6. The beads were washed two times with 70% ethanol (200uL) while on the magnet.
7. After the final wash, the beads were re-suspended in 20uL of 10mM Tris-HCl pH 8.0 to elute DNA

Step 3: Ligation Step 1

1. Purified DNA (20uL) from Step 2: End repair is added to Ligation Step 1 reaction tubes (ArcherDx)
2. Mixture was dissolved, spun, and placed on ice.
3. Mixture was transferred to a thermal cycler for 15 minutes at 37°C

Cleanup after ligation 1

4. AMPure XP beads (ArcherDx) were resuspended by vortexing
5. 2.5X volume (125µL) of AMPure XP beads were added to each Ligation Step 1 reaction.
6. Mixture was vortexed and incubated for 5 minutes at room temperature
7. Tubes were spun and placed on magnet for 4 minutes to allow formation of a bead pellet against the tube wall
8. A pipette was used to remove and discard the supernatant without disturbing the bead pellet.
9. The beads were washed two times with 70% ethanol (200uL) while on the magnet.
10. After the final wash, the beads were re-suspended in 42uL of 10mM Tris-HCl pH 8.0 to elute DNA

Step 4: MBC (molecular barcode) Adapter Incorporation

1. Adapter tubes were labelled with the sample index tag letter and number
2. DNA (40uL) from Step 3: Ligation Step 1 reaction was pipetted into unique labelled MBC reaction tubes.
3. Mixture was dissolved, spun, and placed on ice.

Step 5: Ligation Step 2

1. The entire volume of each DNA sample from Step 4: MBC adapter incorporation was added to Ligation Step 2 reaction tubes (ArcherDx)
2. Mixture was dissolved, spun, and placed on ice.
3. Mixture was transferred to a thermal cycler for 5 minutes at 22°C

Cleanup after ligation 2

1. Prepare ligation clean-up beads
 - a. 50uL of Ligation Clean-up Beads (ArcherDx) were pipetted into a 0.2mL PCR tube.
 - b. Tube was placed on magnet until pellet formed allowing removal of supernatant
 - c. 50uL of Ligation Cleanup Buffer (ArcherDx) was pipetted into each tube to resuspend beads
2. Ligation cleanup procedure
 - a. Ligation Step 2 reaction (50uL) was transferred into the tubes with Ligation Cleanup beads
 - b. Mixture was vortexed and incubated for 5 mins x 2
 - c. Tubes placed on magnet and supernatant discarded
 - d. Beads were washed with 200uL Ligation Cleanup Buffer (ArcherDx) x 2 and supernatant discarded

- e. Beads were washed with 200uL Ultra-Pure Water (ArcherDx) and supernatant discarded
- f. Beads were resuspended in 5nM NaOH
- g. Mixture was transferred to a thermal cycler for 10 minutes at 75°C

Step 6: First PCR

1. Entire eluate from Step 5: Ligation step 2 and Gene specific primers mix 1 (ArcherDx) were pipetted into PCR tubes
2. Mixture was spun, and placed on ice
3. Mixture was transferred to a thermal cycler for PCR protocol

Step	Temperature (C)	Time	Cycles
1	95	3 min	1
2	95	30 sec	1
3	62	30 sec	1
4	72	3 min	1
5	4	Hold	1

Clean up after first PCR

1. 32uL of beads were added to each PCR1 reaction
2. Mixture was dissolved, spun, and incubated at room temperature for 5 minutes
3. Tubes were placed on magnets until beads formed pellet
4. A pipette was used to remove and discard the supernatant without disturbing the bead pellet.
5. The beads were washed two times with 70% ethanol (200uL) while on the magnet.
6. Supernatant was discarded
7. DNA eluted by resuspending beads in 10mM Tris-HCl pH 8.0
8. Tubes were placed on the magnet and the full volume of eluate transferred to a new 0.2mL PCR tube

Step 7: Second PCR

1. Second PCR tubes were placed on ice and labelled from 1-8 as shown:



2. This incorporates the index tag for sample level tracking. P7 index sequence was recorded that was used for each sample as below:

Index 1 (P7) sequence table:

Sample Number	Illumina Index 1 P7/i7 Sequence
1	TAAGGCGA
2	CGTACTAG
3	AGGCAGAA
4	TCCTGAGC
5	GGACTCCT
6	TAGGCATG
7	CTCTCTAC
8	CAGAGAGG

3. The entire volume of purified first PCR eluate was transferred into each second PCR tube.
4. Mixture was dissolved, spun, and placed on ice.
5. Mixture was transferred to a thermal cycler for PCR protocol

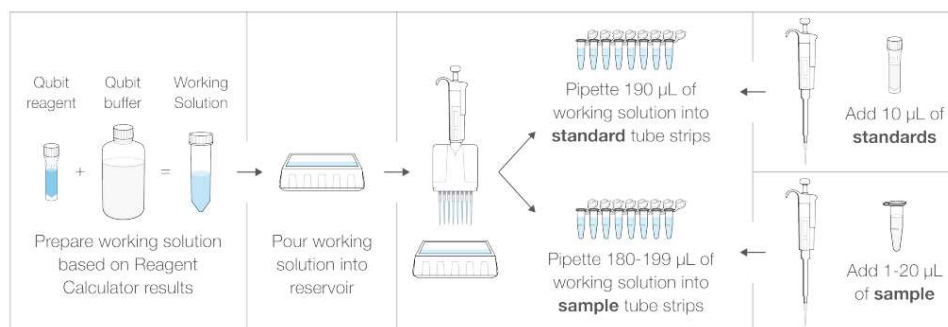
Step	Temperature (C)	Time	Cycles
1	95	3 min	1
2	95	30 sec	1
3	62	30 sec	1
4	72	3 min	1
5	4	Hold	1

Clean up after first PCR

6. 32uL of beads were added to each PCR2 reaction
7. Mixture was dissolved, spun, and incubated at room temperature for 5 minutes
8. Tubes were placed on magnets until beads formed pellet
9. A pipette was used to remove and discard the supernatant without disturbing the bead pellet.
10. The beads were washed two times with 70% ethanol (200uL) while on the magnet.
11. Supernatant was discarded
12. DNA eluted by resuspending beads in 20uL of 10mM Tris-HCl pH 8.0
13. Tubes were placed on the magnet for 2 minutes
14. 18uL of the purified solution was transferred to a new 0.2mL PCR tube

DNA quantification:

Double stranded DNA was quantified using Qubit (Thermofisher) assay as follows:



Vortex all tubes for 2-3 seconds, check for bubbles, then incubate for 2 minutes (15 minutes for protein) at room temperature

Qubit reagent, Qubit buffer, and Working solution were added to reservoir. Working solution volume was based on sample number per run. Working solution (180-199uL depending on sample volume) was pipetted into sample tube strips. The Qubit Normalization calculator was used to determine sample and buffer volume. Libraries were normalized based on DNA quantification with pooled equal amounts of libraries in a tube (5uL). Final library (pool of all individual libraries) was at concentration 20-40ng/uL and total volume >40uL.

Libraries were sent to the NIH sequencing core and sequenced on Illumina NovaSeq 6000 paired-end 150-8-8-150 at average coverage of 2000x which corresponds to 500x of de-duplicated reads.

Appendix C:

Protocol for Tapestri Single-Cell DNA + Protein Sequencing (MissionBio)

Full protocol available from MissionBio website: https://missionbio.com/wp-content/uploads/2021/02/Tapestri-Single-Cell-DNA-Protein-Sequencing-V2-User-Guide-PN_3360A.pdf accessed on 5/3/2022. Protocol description adapted from full protocol.

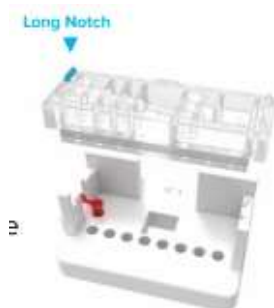
Prepare Cell Suspension

Thaw cells

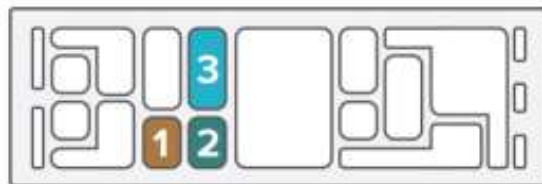
1. Patient BM cells were removed from -80°C freezer and thawed to 37°C in a water bath and 1mL of thawing media was added to the frozen vial
2. BM cells were transferred to 15mL conical tube; vial was washed with 1 mL thawing media.
3. 2mL of thawing media was added to the 15mL tube every few seconds until 12mL total volume reached
4. Mixture was centrifuged at 400g for 5 minutes at room temperature
5. The supernatant was aspirated without disturbing the cell pellet
6. The cell pellet was resuspended in thawing media and further 10mL thawing media added
7. Mixture was centrifuged at 400g for 5 minutes at room temperature
8. The supernatant was aspirated without disturbing the cell pellet
9. Cells were resuspended in 1mL of Cell Staining Buffer (CSB) (BioLegend 420201)
10. Mixture was centrifuged at 400g for 5 minutes at room temperature
11. The supernatant was aspirated and cells resuspended in 250 uL of CSB
12. Cells were quantified using automated cell counter and cells were diluted to 25,000 cells/uL suspension using CSB

Encapsulate cells

1. Opened Tapestri DNA cartridge
2. Mounted Base plate onto Tapestri instrument with recovery PCR tube
3. DNA cartridge placed onto the base plate (see below)



4. Lysis mix was prepared by adding 7.3uL of reverse primer pool into 92.7uL of lysis buffer, vortexed, and centrifuged
5. Lysis mix was pipetted (90uL) into reservoir 1, cell suspension (35uL) was pipetted into reservoir 2 and encapsulation oil pipetted into reservoir 3 of the DNA cartridge.



6. The Tapestri DNA gasket was applied to the top of the cartridge and the Tapestri machine run with program “Encapsulation” for 5 minutes.
7. Once “Encapsulation” program was complete, the gasket was removed
8. The emulsion sample tube was transferred to a 96-well plate holder
9. A gel loading tip was used to remove 100uL of oil from the bottom of the sample.

Lyse and Digest Cells

1. Samples were run in the thermal cycler for 60 minutes at 50°C then 10 minutes at 80°C

Barcode Cells

1. The DNA cartridge was primed for barcoding
 - a. 8 0.2mL PCR tubes were loaded into the Tapestri Base Plate
 - b. Tapestri DNA cartridge was mounted to base plate
 - c. Electrode solution was pipetted into reservoir 4 (200uL) and reservoir 5 (55uL) of the cartridge
 - d. The DNA gasket was applied and cartridge was run in the Tapestri machine in the “Priming” program for 20 minutes.
2. Loading the DNA cartridge
 - a. The encapsulated cell lysate emulsion from the thermal cycler was pipetted into reservoir 6 of the DNA cartridge
 - b. Barcoding beads were vortexed at high speed and pipetted (250uL) into reservoir 8 and pipetted (1.25mL) into reservoir 9.
 - c. The DNA gasket was applied and the “Barcoding” program run on the Tapestri machine for 35 minutes. Next, tubes were automatically exposed to light for 8 minutes inside the Tapestri machine
 - d. Once finished, the Base Plate and cartridge were removed with the 8 tubes containing the barcoded emulsion.
 - e. Barcoded DNA was visually inspected for quality
 - f. 90uL of Barcoding Oil was removed from each tube (bottom) and samples were put back onto the Base Plate

Targeted PCR Amplification and breaking emulsions

1. Samples were then PCR enriched by the Targeted PCR protocol as follows:

			2. Targeted PCR				
Amplicon Number			20 – 100	100 – 200	200 – 300	> 300	
Panel			Custom*	AML V2	THP/CLL	Myeloid	
Step	Ramp	Temperature	Time	Time	Time	Time	Cycle
1	4 °C/s	98 °C	6 min	6 min	6 min	6 min	
2	1 °C/s	95 °C	30 sec	30 sec	30 sec	30 sec	11
3		72 °C	10 sec	10 sec	10 sec	10 sec	
4		61 °C	3 min	4.5 min	6 min	9 min	
5		72 °C	20 sec	20 sec	20 sec	20 sec	
6	1 °C/s	95 °C	30 sec	30 sec	30 sec	30 sec	13
7		72 °C	10 sec	10 sec	10 sec	10 sec	
8		48 °C	3 min	4.5 min	6 min	9 min	
9		72 °C	20 sec	20 sec	20 sec	20 sec	
10	4 °C/s	72 °C	2 min	2 min	2 min	2 min	
11		4 °C	HOLD	HOLD	HOLD	HOLD	

Table 4. Thermal cycling programs for Targeted PCR.

2. Samples were removed from the thermocycler and emulsions were broken by adding 10uL of Extraction Agent to each sample tube, Vortexing for 20 seconds and incubating at room temperature for 3 minutes. 45uL of nuclease-free water was added to each tube which were then vortexed to separate the oil and aqueous layers. 42uL of each tube's aqueous layer was pipetted into an Eppendorf tube; tubes 1-4 in 1 Eppendorf and tubes 5-8 in the other

Cleanup PCR I products

1. Digestion Mix (MissionBio) was prepared by combining 20uL DNA Cleanup Buffer (MissionBio) and 12uL Cleanup Enzyme (MissionBio). Mixture was split into two tubes and run on thermal cycler for 60 minutes at 37°C
2. AMPure XP reagent (MissionBio) Library clean up
 - a. Sample tubes were recombined into two 1.5mL Eppendorf tubes
 - b. 140uL of AMPure XP reagent was added to each 200uL sample tube
 - c. Mixture was vortexed for 10 seconds and incubated at room temperature for 5 minutes
 - d. Mixture was placed on magnet and 340uL of supernatant was removed from each tube into new 1.5mL Eppendorf tubes and set aside
3. DNA Library cleanup I
 - a. The bead pellets were washed with 800uL of 80% ethanol for two wash cycles
 - b. Residual ethanol was removed and bead pellets were incubated at room air for 406 minutes
 - c. Magnet was removed
 - d. 60uL of nuclease-free water was added to each tube
 - e. Mixture was spun for 10 seconds and incubated at room temperature for 2 minutes.
 - f. Tubes were again placed on the magnet

- g. 50uL of purified PCR product was transferred from each tube into a new tube
- h. 1uL of purified PCR product was quantified using Qubit (as described earlier)

Protein Library Cleanup I

1. Streptavidin Beads were prepared
 - a. Streptavidin Beads (MissionBio) (100uL) were transferred to an Eppendorf tube and placed on a magnet
 - b. The supernatant was removed and discarded
 - c. The beads were washed using Wash buffer for 2 cycles
 - d. Beads were resuspended with 690uL of Wash Buffer and set aside
2. Antibody tags were isolated
 - a. 2uL Biotin Oligo (MissionBio) was added to two tubes of the supernatant from Cleanup PCR (section 2.d), vortexed, and incubated at 96°C for 5 minutes. Samples were then transferred to ice and incubated for 5 minutes.
 - b. 342 uL of Streptavidin Beads were added to each tube containing Biotin Oligo. Mixture was incubated for 20 minutes on a shaker at room temperature.
 - c. Mixture was placed on a magnet, supernatant was removed, and beads were washed with 1mL Wash Buffer.
 - d. Samples were removed from magnet and washed with 1mL nuclease-free water.
 - e. Samples were placed on magnet, supernatant removed, and beads resuspended with 25uL nuclease-free water.
 - f. Mixture was combined to 0.2mL PCR tube

PCR Target library

1. 2 0.2mL PCR tubes were labelled with the numbers of Index Primer
2. PCR reactions were set as follows:

	DNA Library	Protein Library
Reagent	Volume [μL]	
Library Mix V2	25	25
V2 Index Primer (●)	10	-
Protein Index Primer (●)	-	10
Targeted DNA PCR product	15	-
Resuspended Streptavidin Beads containing Antibody Tags		15
Total Volume	50	50

Source: https://missionbio.com/wp-content/uploads/2021/02/Tapestri-Single-Cell-DNA-Protein-Sequencing-V2-User-Guide-PN_3360A.pdf accessed on 4/5/2022

3. Samples were transferred to the thermal cycler and PCR protocol was run as follows (10 cycles for DNA and 20 for protein):
 - a. Step 1: 95°C for 3 minutes
 - b. Step 2: 98°C for 20 seconds
 - c. Step 3: 62°C for 20 seconds
 - d. Step 4: 72°C for 45 seconds
 - e. Step 5: 72°C for 2 minutes
 - f. Step 6: 4°C

DNA Library cleanup II

1. 50uL of nuclease-free water was added to sample tube
2. 69uL AMPureXP reagent (MissionBio) was added and mixture was vortexed, incubated at room temperature for 5 minutes, and put on magnet.
3. The bead pellets were washed with 200uL of 80% ethanol for two wash cycles
4. Residual ethanol was removed and bead pellets were incubated at room air for 2 minutes
5. Magnet was removed
6. 100uL of nuclease-free water was added to each tube
7. Mixture was spun for 5 seconds and incubated at room temperature for 2 minutes.
8. Tubes were again placed on the magnet
9. 100uL of purified PCR product was transferred from each tube into a new tube
10. 72uL of AMPureXP reagent (MissionBio) was added. Mixture was incubated for 5 minutes at room temperature and placed on magnet
11. The bead pellets were washed with 200uL of 80% ethanol for two wash cycles
12. Residual ethanol was removed and beads were incubated at room air for 2 minutes.
13. Tubes removed from magnet
14. Nuclease-free water (15uL) was added and tube was vortexed and incubated at room air for 2 minutes
15. Mixture was placed on magnets and 12uL purified PCR product place into a new 0.2mL tube.

Protein Library Cleanup II

1. Tube was placed on magnet to allow Streptavidin Beads (MissionBio) to separate
2. 50uL of supernatant was transferred to new PCR tube
3. AMPure XP reagent (MissionBio) (45uL) was added to sample; sample was vortexed for 10 seconds and incubated at room air for 5 minutes
4. Sample placed on magnet, supernatant removed and discarded.
5. Beads were washed using 200uL 80% ethanol for 2 wash cycles
6. Residual ethanol was removed and beads incubated at room temperature for 2 minutes.
7. Tube was removed from magnet and 17uL of nuclease-free water was added. Tube was then vortexed, and incubated at room temperature for 2 minutes.
8. Tubed placed on magnet and 15uL of purified PCR product was transferred to new tube

Quantification of the DNA and Protein Libraries:

Libraries were quantified using Qubit. In a new 0.2 mL PCR tube we normalized the DNA and protein libraries 5 nM in H₂O. We next combined the DNA and protein libraries at ratio of 1.3:1). Final library (pool of all individual libraries) was at concentration 1-2ng/uL.

Sequencing:

Libraries were sent to the NIH sequencing core and sequenced on Illumina NovaSeq 6000 paired-end 150-8-8-150 DNA libraries had average coverage of 70x per amplicon and protein libraries 200x per amplicon.