Clonal hematopoiesis in elderly twins

concordance, discordance and mortality

Hansen, Jakob Werner; Pedersen, Dorthe Almind; Larsen, Lisbeth Aagaard; Husby, Simon; Clemmensen, Signe Bedsted; Hjelmborg, Jacob; Favero, Francesco; Weischenfeldt, Joachim; Christensen, Kaare; Grønbæk, Kirsten

Published in:
Blood

DOI:
10.1182/blood.2019001793

Publication date:
2020

Document version:
Accepted manuscript

Citation for pulished version (APA):
Hansen, J. W., Pedersen, D. A., Larsen, L. A., Husby, S., Clemmensen, S. B., Hjelmborg, J., Favero, F., Weischenfeldt, J., Christensen, K., & Grønbæk, K. (2020). Clonal hematopoiesis in elderly twins: concordance, discordance and mortality. *Blood*, *135*(4), 261-268. https://doi.org/10.1182/blood.2019001793

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Clonal hematopoiesis in elderly twins: concordance, discordance and mortality

Tracking no: BLD-2019-001793R1

Jakob Werner Hansen (Department of Hematology, Rigshospitalet and Biotech Research and Innovation Centre (BRIC), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark and The Danish Stem Cell Center (Danstem), University of Copenhagen, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, Denmark) Dorthe Almind Pedersen (The Danish Twin Registry, University of Southern Denmark, Odense, Denmark and Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark) Lisbeth Aagaard Larsen (The Danish Twin Registry, University of Southern Denmark, Odense, Denmark and Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark) Simon Husby (Rigshospitalet, Denmark) Signe Bedsted Clemmensen (The Danish Twin Registry, University of Southern Denmark, Odense, Denmark and Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark) Jacob Hjelmborg (The Danish Twin Registry, University of Southern Denmark, Odense, Denmark and Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark) Francesco Favero (Biotech Research & Innovation Center (BRIC), Denmark) Joachim Weischenfeldt (RIgshospitalet, Denmark) Kaare Christensen (University of Southern Denmark, Denmark) Kirsten Grønbæk (Department of Hematology, Rigshospitalet and Biotech Research and Innovation Centre (BRIC), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark and The Danish Stem Cell Center (Danstem), University of Copenhagen, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, Denmark)

Abstract:
Clonal hematopoiesis (CH) of indeterminate potential (CHIP) is defined by mutations in myeloid cancer-associated genes with a variant allele frequency of at least 2%. Recent studies have suggested a possible genetic predisposition to CH. To further explore this phenomenon, we conducted a population-based study of 594 twins from 299 pairs aged 73-94 years, all with more than 20 years follow-up. We sequenced DNA from peripheral blood with a customized 21 genes panel at a median coverage of 6179X. The casewise concordance rates for mutations were calculated to assess genetic predisposition. Mutations were identified in 214 (36%) of the twins. Whereas 20 twin pairs had mutations within the same genes, the exact same mutation was only observed in two twin pairs. No significant difference in casewise concordance between monozygotic and dizygotic twins were found for any specific gene, subgroup or CHIP mutations overall and no significant heritability could be detected. In pairs discordant for CHIP mutations, we tested if the affected twin died before the unaffected twin, as a direct measurement of the association of having CH when controlling for familial factors. A total of 127 twin pairs were discordant for carrying a mutation, and in 61 (48%) cases the affected twin died first, p=0.72. Overall, we did not find a genetic predisposition to CHIP mutations in this twin study. The previously described negative association of CHIP mutations on survival, could not be confirmed in a direct comparison among twin pairs that were discordant for CHIP mutations.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: J.W.H, K.C and K.G conceived the study and wrote the manuscript. J.W.H, S.H, F.F and J.W. did the sequencing work and analyzed the sequencing data. D.A.P, L.A.L, S.B.C and J.H linked the data and performed the twin and survival analyses. All authors contributed to the manuscript and interpretation of the data, and approved the final version of the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Publication-related data will be available by contacting the corresponding author by email.

Clinical trial registration information (if any):

CLONAL HEMATOPOIESIS IN ELDERLY TWINS: CONCORDANCE, DISCORDANCE AND MORTALITY

Running title: Clonal hematopoiesis in elderly twins

Jakob Werner Hansen[1,2,3], Dorthe Almind Pedersen[4,5], Lisbeth Aagaard Larsen[4,5], Simon Husby[1,2], Signe Bedsted Clemmensen[4,5], Jacob Hjelmborg[4,5], Francesco Favero[2,6], Joachim Weischenfeldt[2,6], Kaare Christensen[4,5], Kirsten Grønbæk[1,2,3]

1) Department of Hematology, Rigshospitalet, Copenhagen, Denmark

2) Biotech Research and Innovation Centre (BRIC), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

3) The Danish Stem Cell Center (Danstem), University of Copenhagen, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

4) The Danish Twin Registry, University of Southern Denmark, Odense, Denmark

5) Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense

6) Finsen Laboratory, Rigshospitalet, Copenhagen, Denmark;

Key Points

1) **A genetic predisposition to clonal hematopoiesis is not detectable in this cohort of 594 elderly twins**
2) **CHIP mutations are not associated with mortality in a direct comparison among discordant twin pairs**

Corresponding author: Professor Kirsten Grønbæk, Kirsten.groenbaek@regionh.dk , BRIC, Ole Maaløesvej 5, building 2, 3. Floor, Section 3733. 2200 Copenhagen, Denmark. Phone +4535456060

Abstract:

Clonal hematopoiesis (CH) of indeterminate potential (CHIP) is defined by mutations in myeloid cancer-associated genes with a variant allele frequency of at least 2%. Recent studies have suggested a possible genetic predisposition to CH. To further explore this phenomenon, we conducted a population-based study of 594 twins from 299 pairs aged 73-94 years, all with more than 20 years follow-up. We sequenced DNA from peripheral blood with a customized 21 genes panel at a median coverage of 6179X. The casewise concordance rates for mutations were calculated to assess genetic predisposition. Mutations were identified in 214 (36%) of the twins. Whereas 20 twin pairs had mutations within the same genes, the exact same mutation was only observed in two twin pairs. No significant difference in casewise concordance between monozygotic and dizygotic twins were found for any specific gene, subgroup or CHIP mutations overall and no significant heritability could be detected. In pairs discordant for CHIP mutations, we tested if the affected twin died before the unaffected twin, as a direct measurement of the association of having CH when controlling for familial factors. A total of 127 twin pairs were discordant for carrying a mutation, and in 61 (48%) cases the affected twin died first, p=0.72.

Overall, we did not find a genetic predisposition to CHIP mutations in this twin study. The previously described negative association of CHIP mutations on survival, could not be confirmed in a direct comparison among twin pairs that were discordant for CHIP mutations.

Introduction

More than twenty years ago, studies of X-chromosome inactivation in elderly females suggested that selected hematopoietic clones expand in the aging bone marrow[1], and our studies of aging twins proposed that this aberrant hematopoietic stem cell kinetics is not random[2]. The accumulation of clonal stem cells has previously been associated with myeloid skewing, lymphopenia and progressing anemia in the elderly[3].

In recent years, comprehensive exome and targeted sequencing studies of large cohorts revealed that the hematopoietic stem cells of healthy elderly individuals accumulate somatic mutations in the same genes that are drivers of myeloid cancer, thus giving rise to clonal hematopoiesis (CH)[4–6]. Subsequent studies have confirmed and extended these findings[7–10], and several terms have been proposed to describe the condition including age related clonal hematopoiesis (ARCH) that signify CH of any acquired clonal event[1,11]; and clonal hematopoiesis of undetermined significance (CHIP), which has been defined to require a variant allele frequency of driver genes of minimum 2%[9]. Ultra-deep sequencing studies have revealed that very small clones may be ubiquitously present at young age[12], however the prevalence of CHIP increases dramatically after middle age[4,5,8,10].

The most frequent aberrations in CHIP are loss-of-function mutations in the epigenetic modifiers *DNMT3A*, *TET2* and *ASXL1*, so-called DTA mutations.

Stem cells with mutations in *DNMT3A* had already been shown to precede the development of acute myeloid leukemia (AML)[13], when analyses of the large cohorts revealed that CHIP carriers have a significantly increased (10-fold) risk of hematological cancers, including, but not restricted to, AML and myelodysplastic syndrome (MDS)[4,5]. This was particularly the case in patients with CHIP mutations of a variant allele frequency (VAF) > 10%, where the risk of hematological cancer was increased by a factor 50[5]. Further studies have shown that CHIP mutations frequently persist in

patients in complete remission after treatment for AML[14,15], but only the non-DTA mutations are associated with an increased risk of relapse[15].

Twin studies from the Scandinavia have shown an increased familial risk of cancer, and in chronic myeloid neoplasms, Andersen et al[16] demonstrated a casewise concordance (cc) rate of 15% among monozygotic (MZ) twins compared to 0% among dizygotic (DZ) twins p = 0.016, suggesting a genetic predisposition. In addition, it has been known for decades that some germline aberrancies predispose to hematological cancer such as AML and MDS. In some of these genes, driver mutations occur both in the germline and as somatic events; e.g. in *RUNX1, GATA2* and *ETV6*. When patients with predisposing germline variants are diagnosed with overt AML or MDS, co-occurring somatic mutations, which are involved in leukemogenesis, are often present[17].

In adults, hematological cancers are rarely caused by germline aberrancies, although recent sequencing studies have identified germline mutations that predispose to hereditary myeloid cancers after middle age[18]. While the link between somatic CHIP mutations and hematological cancer is evident, it is unknown why some aging individuals develop CHIP while others do not. One Icelandic study suggested that besides age, smoking and psychiatric disease is associated with CH. They performed a genome wide association study, and identified an 8 base pair (bp) deletion in *TERT* intron 3 (RS 34002450)[19] with an increased risk of CH with an odds ratio of 1.37 and an allele frequency of 40.6% in Iceland. Meanwhile, Buscarlet et al described a genetic predisposition to *TET2* mutations within sib-ships with a recurrence risk of 0.09 and 0.13 corresponding to a recurrence risk ratio of 2.2 and 2.7 in age groups above 54 and 64, respectively. However, no association was observed for *DNMT3A* mutations[20]. Lastly Frick et al. recently investigated CHIP in related donors to patients undergoing allogeneic stem cell transplantation. They identified a

higher prevalence of CHIP in related donors to patients with a myeloid malignancy compared to a lymphoid malignancy 19.2% vs 6.3%, p<0.001, also suggesting a possible familial predisposition[21].

However, these studies were not designed to distinguish whether CHIP is resulting from a common germline pre-disposition, and whether such an aberration also predisposes to hematological cancer. In addition, it is not clear whether the reported aggregation of *TET2* mutations in elderly siblings is merely driven by a genetic predisposition or by common environmental factors. To clarify whether CHIP mutations arise on a common genetic basis or are a result of shared environmental exposures we conducted a study of 299 elderly twin pairs who were prospectively followed for more than 20 years.

Methods

Cohort

In the present study, twin pairs were identified from the Longitudinal Study of Aging Danish Twins (LSADT) in which 689 twins donated blood in 1997[22], 299 intact same sex twin pairs were included in the present study. Twin zygosity was classified using questionnaire-based zygosity assessment, which is found to be correct in 95% of the cases[23]. Information about smoking (pack-years) was obtained from the LSADT questionnaire in 1997. The LSADT study was approved by the Regional Committees on Health Research Ethics for Southern Denmark (S-VF-20040241) and the present study with (S-20170053). Research was conducted in accordance with the Declaration of Helsinki.

Sequencing and bioinformatic analyses

A targeted sequencing approach was applied, using an Illumina TruSeq Custom Amplicon panel (Illumina, San Diego, CA, USA) covering more than 95% of mutations commonly associated with CH[4–6,24] (Supplementary Material). DNA was quantified using the Qubit fluorometer (Life Technologies, CA, USA), and 100–200 ng of DNA was used for library preparation. To optimize

variant calling and identification of low-level mutations, unique molecular identifiers (UMI's), consisting of 6 random index nucleotides, were added to each sample before amplification. Libraries were pooled and sequenced on the Illumina NextSeq platform (Illumina, San Diego, CA, USA). We used a 300 cycles mid output kit as specified by the manufacturer.

Variant calling

In brief, raw sequence reads and corresponding UMI's were aligned to the human reference genome (hg19 build) with the BWA-mem algorithm[25]. Identification of possible variants was performed using two variant callers; Freebayes v. 1.1.0[26] and VarDict v.1.5.1.[27]. We investigated CHIP mutations, as defined by Steensma et al. [9], with a VAF of 2% or above. Common variants, defined by a population frequency of above 1% in large public variant databases (ExAC, TOPMED, 1000Genomes) were excluded, and furthermore, we removed rare variants with a VAF of 40-60%, which were concordant in related twins, so these should not bias the casewise concordance rates. Additional details are provided in supplemental material.

Vital status and diagnoses

Date of birth, vital status and date of death was retrieved from the Danish Civil Registration System[28]. To investigate the role of CHIP mutations for the risk of subsequent hematological cancer or cytopenia, we extracted diagnoses from The Danish National Patient Register, which essentially contains all discharges from Danish hospitals since 1977[29]. From 1994 and onwards diagnoses were classified according to the International Classification of Diseases (ICD-10). In the present study, we included primary and secondary diagnoses and all patient types (in-patient, out-patient,

emergency room patient) from 1997-2014. We defined two groups of ICD-10 codes representing hematological cancers or cytopenia, respectively (supplementary material).

Statistical analyses

Concordance and heritability

The classic twin-study methodology is based on monozygotic twins having identical genotypes, whereas dizygotic twins share, on average, half of their genetic variants like biologic full siblings. A greater phenotypic similarity in monozygotic than in dizygotic twins is expected if there is a substantial genetic component in the etiology of the phenotype. We assessed the similarity of monozygotic and dizygotic twins using casewise concordance rates and tetrachoric correlations. The casewise concordance rate is defined as the probability that a twin is affected given the co-twin is affected and corresponds to recurrence rates in siblings. Assuming that there is a normally distributed liability to develop a complex phenotype we used standard quantitative biometric models to estimate the tetrachoric correlations and the relative contribution of genetic and environmental factors in CHIP occurrence[30]

ADCE-Model: The phenotypic variance can then be separated into four variance components: variance attributable to additive genetic effects (A), genetic dominance (D), shared environment (C), and nonshared (individual-specific) environment (E). Only nonshared environments contribute to dissimilarity within MZ twin pairs because of their presumed genetic identity, whereas the effects of additive genetic factors and genetic dominance may also contribute to dissimilarity within DZ pairs, who share, on average, half of the additive and one- quarter of the dominant genetic factors[30].

Time to event analyses

Follow-up were initiated at the date of blood donation. Cox proportional hazard model with time since blood sample was used as the underlying time scale to compute hazard ratios (HRs) of time to

death for individuals with CHIP mutations compared with individuals without CHIP mutations. In these analyses, the follow-up was terminated at the date of death or 1 August 2018, whichever came first.

To assess the risk of developing hematological cancer or cytopenia, competing risks regression analyses were performed treating death as a competing risk using the STATA command *stcrreg*, which is based on Fine and Gray's proportional sub-hazards model. Follow-up were terminated at date of first diagnosis of the specific diagnosis group, death or 1 March 2014, whichever came first.

In addition, since some studies indicate that bias can be introduced using time since sampling as timescale[31,32], all time to event analyses were also performed with age as the underlying time scale.

The proportional hazard assumption was assessed based on Schoenfeld residuals. If the proportional hazard assumption was violated for the covariates, we performed Cox regression with strata on covariates to assess if results changed noteworthy.

Furthermore, in twins discordant for mutation, we tested if the affected twin died before the unaffected twin using binomial probability test.

In addition, all analyses were performed for different classifications of mutations (*DNMT3A*, *TET2* and DTA mutations overall) and based on VAF.

Results

In this study, we included a total of 598 elderly twins from 299 pairs, which were followed prospectively for 20 years. Four twins were excluded from the study due to poor sequencing quality, and their corresponding twin were removed from the concordance analyses, which then included a total of 295 twin pairs. The median age at inclusion were 77 years (range 73-94), 33% of the twins were male, while 256 (43%) and 338 (57%) were MZ and DZ twins, respectively (Table 1). By targeted next generation sequencing of DNA from peripheral blood drawn at study entry, we identified a total of 286 mutations in 15 known, recurrent CHIP genes, distributed among 214 (36%) twins (Figure 1). Thus, some twins carried more than one mutation, with a maximum of 3 mutations being detected in one individual (Supplementary Figure 1).

The proportion of individuals with a CHIP mutation was increasing with age ranging from 29% at the age of 75 years, to more than 64% in the group above 85 years (Supplementary Figure 2). The most commonly mutated gene was *DNMT3A* observed in 104 twins, followed by *TET2*, which was mutated in 81 twins. The higher number of mutations in *DNMT3A* compared to *TET2* (17.2% and 10.1%, respectively) were found in the age group 73-79 years. However, in individuals between 80-94 years this trend was reversed, with *TET2* mutations being most abundant i.e. 21.4% mutations in *TET2* vs 18.2% in *DNMT3A* (Supplementary Figure 3a and b). In total, DTA mutations accounted for 79.4% of all the mutations observed in this cohort. We identified 96 *TET2* mutations: 36 nonsense-, 35 frameshift-, 19 missense-, 5 splice site mutations and one deletion. None of the missense mutations had a VAF above 40%. The 115 mutations in *DNMT3A* were 61 missense-, 20 frameshift-, 16 nonsense- and 14 splice site mutations, as well as 4 in-frame deletions. Six of the missense mutations were at the hot spot position R882. Twins with CHIP mutations were older than the non-mutated twins (p =<0.001), and CHIP mutations were equally distributed among MZ and DZ twins (Table 1). The median VAF in the entire cohort was 4.4% (range = 2-68%), with two

individuals having a VAF above 50%. Based on VAF of the largest clone found in each twin, they were grouped as having below 10% vs 10% and higher. Of the 214 mutated cases, 63 (29%) had a VAF of 10% or above, and the frequency increased with age (Supplementary Figure 4).

Co-occurrence of mutations in twin pairs

Whereas 20 twin pairs had mutations within the same genes, the exact same mutations (involving the exact same base exchange) were only observed among two twin pairs. One twin pair, which was DZ, both had an *SRSF2* mutation at position c.284C>A causing P95H, with a VAF of 4.5% and 31%, respectively, without any other co-occurring mutations. The second twin pair, which were MZ, both had a 5 bp deletion c.912_916delCTGGT in *DNMT3A* giving rise to a frameshift at p.Trp305fs with a VAF of 11.7% and 26.8%, respectively. Both of these twins each had an additional *DNMT3A* mutation; the first twin had a missense mutation c.2096G>T causing an amino acid change at p.Gly699Val with a VAF of 7.3%, and the second twin had a 4 bp deletion c.2167_2168insGTAG giving rise to a frameshift p.Leu723fs with a VAF of 2%. The common 5 bp deletion with VAF among the MZ twin pair is intriguing, as it could indicate mosaicism due to a deletion occurring at a very early developmental stage, however it is beyond the scope of the current study to further analyze this issue.

Concordance and heritability

We estimated the casewise concordance rates for carrying a mutation in any of the 21 examined genes and found that the concordance rate for MZ twins was 0.40 [0.32;0.49] and for DZ twins 0.40 [0.32;0.49], ccMZ = ccDZ, p=1. Since *TET2* mutations had previously been reported to co-occur in families, while *DNMT3A* did not[20], we further tested the concordance rates for *TET2* and *DNMT3A* separately. For *TET2* we observed a concordance rate for MZ and DZ of 0.20 [0.10;0.34] and 0.20

[0.10;0.34], respectively, ccMZ = ccDZ, p=1. Similar results were obtained for *DNMT3A* 0.25 [0.12;0.43] and 0.22 [0.11;0.40], for MZ and DZ twins, respectively, ccMZ = ccDZ, p=0.86. When examining all DTA mutations together, we also found similar concordance rates for MZ- and DZ twins; 0.33 [0.24;0.43] and 0.33 [0.24;0.43], respectively, ccMZ = ccDZ, p=1. Lastly, we examined concordance rates of the groups having maximum VAF<10% vs the group with no mutation. Again, we did not observe any significant difference between MZ or DZ twins using these cut-offs (Table 2). To evaluate the environmental effects, we applied the polygenic ADCE model and estimated what was governing the presence of CHIP mutations, and subsequently the presence of *TET2* and *DNMT3A* mutations. The biometrical models revealed that 89% (95% CI: 71-100%) of the variation in the liability to have any CHIP mutations was accounted for by environmental factors not shared by co-twins, while 0% (95% CI: NA) was accounted for by genetic factors and 11% (95% CI: 0-29%) was explained by common environmental factors. For *TET2* and *DNMT3A* the corresponding numbers were 84% (95% CI:58-100%) and 83% (95% CI: 48-100%) for environmental factors not shared by co-twins, while the genetic factors and common environmental effects were so small that they could not be reliably estimated with the present sample size. Hence a genetic predisposition to CH is not detectable in this cohort of elderly twins neither overall or for specific mutations or subgroups.

Mortality

At the end of follow-up (1 August 2018), 96% of the study population were deceased. In a cox regression analysis, not controlling for co-variates, CHIP mutation carriers had a HR [95%CI] of 1.31 [1.10;1.56], $p < 0.001$. When adjusted for age, sex and tobacco consumption in a Cox

regression analysis, CHIP was borderline significantly associated with increased mortality: HR 1.17 [0.97;1.40], p=0.096. (Table 3a). When we used age as the underlying time scale in a Cox regression analysis the impact of CHIP mutations was HR 1.13 [0.94;1.36] p=0.192 after adjusting for sex and tobacco consumption (Table 3b). We also tested whether the number of mutations grouped as 1 mutation or 2-3 mutations had any impact on survival, however both had a HR at 1.17, which was non-significant. We next grouped the twins according to whether they had "DTA mutations" only or "other mutations" (i.e. other mutation +/- DTA mutation). When adjusted for age, sex and smoking, the negative effect of "other mutations" was not significant HR 1.01 [0.65;1.54], p=0.98 (Table 4a). In contrast, DTA mutations retained negative impact on overall survival when adjusted for age, sex and tobacco consumption; HR 1.21 [1.01;1.45], p=0.035. Using age as the underlying time scale DTA mutations were borderline significant HR 1.18 [0.98;1.41], p=0.077, after adjusting for sex and tobacco consumption (Table 4b).

Lastly, we investigated the impact of clone size on overall mortality. Subjects with mutations were categorized according to VAF<10% or ≥10%, respectively, however, no difference in overall mortality could be seen according to these cut-offs (Supplementary Table 1a). Using age as the underlying timescale did not change the results notably (Supplementary Table 1b).

Mortality in twins discordant for clonal hematopoiesis

In a direct comparison, we tested the association of CH with mortality after controlling for familial factors, which is a unique possibility in this twin study where 96% of the cohort has deceased. In discordant pairs, we tested if the affected twin died before the unaffected twin. A total of 127 twin pairs were discordant for carrying a mutation, and in 61 (48%) cases the affected twin died first, p=0.72. Furthermore, we tested whether certain subtypes of CHIP mutations were directly linked to

mortality by comparing discordant pairs with DTA- vs non-mutated (n=104), and "other mutations" vs non-mutated (n=23). No significant difference was observed in either group, 50 (48%) DTA-mutated twins and 11 (48%) twins with "other mutations" died before their non-mutated related twin, p=0.77 and p=1.00, respectively. Thus, these data indicate that CHIP mutations are unlikely to affect mortality in the healthy elderly, when controlling for familial factors.

Risk of hematological cancer and unspecified cytopenia

A total of 27 twins developed a hematological cancer during the follow-up, of those 8 could be categorized as myeloid neoplasms and the remaining 19 were classified as B- or T-cell malignancies. For confidentiality reasons we do not report on groups with less than five individuals affected, thus the specific numbers within each hematological cancer cannot be listed. First, we tested the risk of developing a hematological cancer and its association with the presence of CHIP mutations. After adjusting for sex, age and tobacco consumption we did not find any association between CHIP mutations and the risk of developing a hematological cancer, HR [95%CI] 1.80 [0.87;3.75], p=0.15 (Supplementary Table 2). However, in the group with unspecified cytopenia (n=74), when we adjusted for age, sex and tobacco consumption, we found a borderline significant association HR [95%CI] 1.59 [0.99;2.54], p=0.052, however CHIP mutations were the only variable with a potential association to cytopenia (Supplementary Table 3).

Discussion

Here, we report the hitherto first analyses of CHIP in twins. Our primary aim was to uncover the role of genetic and familial factors for the development of CHIP. Since this twin cohort ~~was~~is population-based and followed prospectively for 20 years, we were also able to report data on the association to hematological disease and to overall mortality. Importantly, a direct comparison of discordant twins allowed us to estimate the association between CH and mortality, controlling for familial factors.

We identified CHIP mutations in 36% of the twins, which is in line with what is reported in other studies of this age group[8,10], and, as expected, we mainly identified DTA mutations. It is still not clear why DTA genes are so frequently mutated during aging, and it has been speculated whether this may be caused by a genetic predisposition[19,20]. However, we did not find any indication of a genetic predisposition to CH in the current study. This is in contrast to what was suggested by two previous studies, which detected an aggregation of *TET2* mutations in siblings[20], and observed a possible germline defect which may associate with CH with or without driver mutations[19]. Our data were cleaned for polymorphisms and rare variants with a VAF of 40-60%, so these data did not bias the results on genetic predisposition. Using this strategy, we report only four mutations with a VAF between 40-60%, which were two truncating *TET2* mutations and two *JAK2* V617F mutations.

We next tested different aspects of CH such as the impact of individual mutations and allele frequency, but none of these pointed to a genetic predisposition, in particular we could not confirm a predisposition to *TET2* mutations as previously suggested[20]. Jaiswal et al have shown a significant impact of CHIP mutations on overall mortality, with an HR of 1.4 on all-cause mortality after adjusting for both age and sex[5]. By contrast, we did not find CHIP mutations have any significant impact on overall mortality in the elderly population in this study where age was used as the

underlying time scale in the survival analysis. DTA mutations did retain a borderline significant impact using age as underlying time scale (p=0.077) in contrast to "other mutations" which did not have any impact on overall mortality, even though this group included patients with more than one mutation and mutations in DNA repair genes. Meanwhile, our findings are in line with the study from van den Akker et al, who applied a left truncated Cox proportional hazard model and showed that CHIP mutations were not associated with overall mortality in the elderly population[33]. In the current study we had the unique possibility to investigate the impact of CHIP when controlling for familial factors i.e. by comparing discordant twin pairs. Interestingly, we identified no significant difference between affected and unaffected discordant twins. Thus, based on the current data, CHIP mutations per se are not associated with increased mortality. However, it is important to acknowledge the differences in population age, when comparing studies, as our study population is older than those reported in most of the previous studies investigating the impact of CHIP on survival[4,5]. CHIP might have an impact on survival if present in younger individuals, but age-dependent incidents such as CHIP can lose their negative impact in the elderly.

In addition, we did not observe any association with the development of hematological cancers, however, there was a tendency (p=0.052) that twins with CHIP mutations had an increased risk of developing "unspecified cytopenia". It can be speculated whether this relates to the development of clonal cytopenia of undetermined significance (CCUS)[9], however the MDS diagnostic tools and criteria have also been refined and optimized significantly over the past two decades, so it is likely that MDS has been underdiagnosed in this study where cases are registered from 1997 and onwards. The association between CHIP and "unspecified cytopenia" needs to be validated in other cohorts, as the definition of "unspecified cytopenia" is not consistent.

In the current study we did not find any aggregation of hematological cancer or unspecified cytopenia within twin pairs. This said, it is becoming increasingly clear that some myeloid

neoplasms are caused by a genetic predisposition[34]. Some hereditary myeloid cancers present late in life, without any obvious predisposition syndromes and with variable penetrance[35]. However, we were not able to detect any genetic predisposition to CH or *TET2* mutations, specifically, in this population-based cohort of elderly twins.

Limitations

There were no complete blood counts performed at the time of blood collection in our study, so it was not possible to assess the impact and correlation with peripheral blood counts. However, this does not influence the main purpose with the study, which was to investigate the possible genetic predisposition to CH.

In this study we only examined elderly twins above the age of 73 years, where CH is a common phenomenon. By only investigating elderly twins we may theoretically have missed a genetic predisposition to CH appearing in younger individuals. However, investigating younger twins will require a larger cohort, as the proportion with CHIP mutations would be much lower.

In conclusion

This is hitherto the first twin study investigating a genetic predisposition to CH and the impact of CH on mortality in twins. In this cohort no germline predisposition to CHIP mutations could be detected. Furthermore, no difference in mortality was observed in a direct comparison among twin pairs that were discordant for CHIP mutations.

Contributions

J.W.H, K.C and K.G conceived the study and wrote the manuscript. J.W.H, S.H, F.F and J.W. did the sequencing work and analyzed the sequencing data. D.A.P, L.A.L, S.B.C and J.H linked the data and performed the twin and survival analyses. All authors contributed to the manuscript and interpretation of the data, and approved the final version of the manuscript.

Conflict of interest: The authors have no disclosures

Funding

K.G. and J.W.H. are funded by center grants from the Danish Cancer Society (Danish Research Center for Precision Medicine in Blood Cancer; grant 223-A13071-18-S68), and the Novo Nordisk Foundation (Novo Nordisk Foundation Center for Stem Cell Biology, DanStem; grant NNF17CC0027852) and the Greater Copenhagen Health Science Partners (Clinical Academic Group in Translational Hematology). This work and D.A.P, L.A.L, S.B.C, J.H and K.C. is supported by the Odense University Hospital AgeCare program (Academy of Geriatric Cancer Research). The Danish Aging Research Center is supported by a grant from the VELUX Foundation (grant number: Velux 31205). The Danish Twin Registry has been supported by grants from The National Program for Research Infrastructure 2007 (09-063256) from the Danish Agency for Science Technology and Innovation, the Velux Foundation and the US National Institute of Health (P01 AG08761)

Data sharing statement: Please contact the corresponding author for questions regarding original data

References

1. Busque L, Mio R, Mattioli J, et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*. 1996;88(1):59–65.

2. Christensen K, Kristiansen M, Hagen-Larsen H, et al. X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood*. 2000;95(7):2449–51.

3. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells*. 2018;36(9):1287–1294.

4. 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.

5. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* 2014;371(26):2488–2498.

6. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* 2014;20(12):1472–1478.

7. McKerrell T, Park N, Moreno T, et al. Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. *Cell Rep.* 2015;10(8):1239–1245.

8. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. 2018;559(7714):400–404.

9. 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.

10. Desai P, Mencia-Trinchant N, Savenkov O, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat. Med.* 2018;24(7):1015–1023.

11. Shlush LI. Age-related clonal hematopoiesis. *Blood*. 2018;131(5):496–504.

12. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* 2016;7:1–7.

13. Shlush LI, Zandi S, Mitchell A, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. 2014;506(7488):328–33.

14. Pløen GG, Nederby L, Guldberg P, et al. Persistence of DNMT3A mutations at long-term remission in adult patients with AML. *Br. J. Haematol.* 2014;167(4):478–486.

15. Jongen-Lavrencic M, Grob T, Hanekamp D, et al. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N. Engl. J. Med.* 2018;378(13):1189–1199.

16. Andersen MA, Bjerrum OW, Ranjan A, et al. Myeloproliferative Neoplasms in Danish Twins. *Acta Haematol.* 2018;139(3):195–198.

17. Churpek JE, Pyrtel K, Kanchi K, et al. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. *Blood*. 2015;126(22):2484–90.

18. Polprasert C, Schulze I, Sekeres MA, et al. Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell*. 2015;27(5):658–670.

19. 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.

20. Buscarlet M, Provost S, Zada YF, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*. 2017;130(6):753–762.

21. Frick M, Chan W, Arends CM, et al. Role of Donor Clonal Hematopoiesis in Allogeneic

Hematopoietic Stem-Cell Transplantation. *J. Clin. Oncol.* 2019;37(5):375–385.

22. Christensen K, Gaist D, Vaupel JW, McGue M. Genetic contribution to rate of change in functional abilities among Danish twins aged 75 years or more. *Am. J. Epidemiol.* 2002;155(2):132–9.

23. Christiansen L, Frederiksen H, Schousboe K, et al. Age- and sex-differences in the validity of questionnaire-based zygosity in twins. *Twin Res.* 2003;6(4):275–8.

24. Gibson CJ, Lindsley RC, Tchekmedyian V, et al. Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. *J Clin Oncol*. 2017;35:.

25. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.

26. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012;

27. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 2016;44(11):e108.

28. Pedersen CB, Gøtzsche H, Møller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Dan. Med. Bull.* 2006;53(4):441–9.

29. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand. J. Public Health*. 2011;39(7 Suppl):30–33.

30. MC N, Cardon L. Methodology for genetic studies of twins and families. Dordrecht: Kluwer Academic Publishers BV; 1992.

31. Thiébaut ACM, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic

cohort data: a simulation study. *Stat. Med.* 2004;23(24):3803–20.

32. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am. J. Epidemiol.* 1997;145(1):72–80.

33. van den Akker EB, Pitts SJ, Deelen J, et al. Uncompromised 10-year survival of oldest old carrying somatic mutations in DNMT3A and TET2. *Blood*. 2016;127(11):1512–5.

34. Churpek JE. Familial myelodysplastic syndrome/acute myeloid leukemia. *Best Pract. Res. Clin. Haematol.* 2017;30(4):287–289.

35. Lewinsohn M, Brown AL, Weinel LM, et al. Novel germ line DDX41 mutations define families with a lower age of MDS/AML onset and lymphoid malignancies. *Blood*. 2016;127(8):1017–1023.

Figure 1 Waterfall plot of the 214 twins with a total of 286 mutations.

Waterfall plot of the 214 (36%) cases with at least one mutation detected, accounting for a total of 286 mutations being 114 missenses, 76 frameshifts, 69 nonsenses, 21 splice sites and five inframe deletions.

Figure 2a and b Clone size depicted as a density plot and by individual mutation

We only reported mutations with an allele frequency of 2% and above, which defines clonal hematopoiesis of indeterminate potential (CHIP). The distribution of the allele frequencies is shown by density in figure 2a and the variant of each case shown in figure 2b. The median VAF in the entire cohort was 4.4%.

Table 1 Clinical characteristics

	CHIP-mutation negative		CHIP-mutation positive			Total	
	N	%	N	%	p-value	N	%
Total	380	63.97	214	36.03		594	100.00
Males	127	33.42	69	32.24	0.769	196	33.00
Age at blood collection							
73-75	72	18.95	29	13.55	<0.001	101	17.00
75-80	205	53.95	101	47.20		306	51.52
80-85	82	21.58	47	21.96		129	21.72
85-93	21	5.53	37	17.29		58	9.76
Zygosity							
MZ	155	40.79	101	47.20	0.130	256	43.10
DZ	225	59.21	113	52.80		338	56.90
Tobacco consumption (Pack years)							
0	146	38.42	79	36.92	0.815	225	37.88
1st tertile (1-8.55)	746	20.00	40	18.69		116	19.53
2nd tertile (8.55-25)	77	20.26	46	21.50		123	20.71
3rd tertile (25+)	65	17.11	43	20.09		108	18.18
Missing	16	4.21	6	2.80		22	3.70
Dead prior to August 2018	359	94.47	210	98.13	0.033	569	95.79

We included 299 twin pairs and sequenced peripheral blood DNA from all twins. Four twins had low quality DNA and were excluded from the study, why 594 twins were included in the study. In august 2018 96% of the cohort had deceased and follow up was available for all twins. Only 38% of the cohort was never smokers, the smokers were divided into tertiles according to tobacco consumption for survival analysis.

Table 2 Concordance rates according to any mutation, *DNMT3A*, *TET2*, "DTA mutations" and maximum clone size below 10%

	Monozygotic				Dizygotic				
	Number of pairs	Concordant pairs	Discordant pairs	casewise concordance rate	Number of pairs	Concordant pairs	Discordant pairs	casewise concordance rate	p-values for ccMZ=ccDZ
Variable	N	N	N	pc [95%CI]	N	N	N	pc [95%CI]	p
Any Mutation (Yes/No)	81	20	61	0.40 [0.32;0.49]	89	23	66	0.40 [0.32;0.49]	1
TET2 mutation (Yes/No)	35	4	31	0.20 [0.10;0.34]	38	4	34	0.20 [0.10;0.34]	1
DNMT3A mutation (Yes/No)	41	6	35	0.25 [0.12;0.43]	50	6	44	0.22 [0.11;0.40]	0.8
"Only TET2/DNMT3A/ASXL1" vs "No mutation"	65	14	51	0.33 [0.24;0.43]	63	11	52	0.33 [0.24;0.43]	1
"Max VAF < 10" vs "No mutation"	57	11	46	0.28 [0.25;0.32]	57	7	50	0.28 [0.25;0.32]	0.9

cc: casewise concordance

Casewise concordance rates were calculated to estimate a genetic predisposition. If the casewise concordance rates are higher in the monozygotic twins compared to monozygotic twins, it indicates a genetic predisposition to the variable. We did not find any difference in the casewise concordance rates between monozygotic or dizygotic in any of the tested variables.

Table 3a and b Cox regression analysis including mutational status, sex, age and tobacco consumption.

A

Variable	HR	95% CI	P
Mutation (yes/no)	**1.17**	**0.97-1.40**	**0.096**
Age (years)	**1.11**	**1.08-1.13**	**<0.001**
Sex (male)	**1.15**	**0.90-1.45**	**0.259**
Smoking (1st tertile)	**1.02**	**0.81-1.29**	**0.840**

Smoking (2nd tertile)	**1.22**	**0.96-1.55**	**0.100**
Smoking (3rd tertile)	**1.71**	**1.24-2.35**	**0.001**

B

Variable	**HR**	**95% CI**	**P**
Mutation (yes/no)	**1.13**	**0.94-1.36**	**0.192**
Sex (male)	**1.14**	**0.90-1.44**	**0.274**
Smoking (1st tertile)	**1.03**	**0.83-1.30**	**0.769**
Smoking (2nd tertile)	**1.26**	**0.99-1.59**	**0.052**
Smoking (3rd tertile)	**1.73**	**1.26-2.37**	**0.001**

In table **A** time since blood sample is used as the underlying time scale whereas age is used as the underlying the time scale in **B**, emphasizing the impact of age on overall survival in this elderly cohort. The borderline association of CHIP mutations (p = 0.096), was not confirmed when age was used as the underlying timescale in the Cox-regression analysis.

Table 4 A and B Cox regression analysis including mutational status groups as DTA mutation and "other mutation", sex, age and tobacco consumption.

A

Variable	**HR**	**95% CI**	**P**
Only DTA* mutation	**1.21**	**1.01-1.45**	**0.035**
Other mutation	**1.01**	**0.65-1.55**	**0.981**
Age (years)	**1.11**	**1.08-1.13**	**<0.001**
Sex (male)	**1.15**	**0.91-1.46**	**0.233**
Smoking (1st tertile)	**1.03**	**0.82-1.30**	**0.772**
Smoking (2nd tertile)	**1.22**	**0.96-1.55**	**0.108**
Smoking (3rd tertile)	**1.72**	**1.26-2.36**	**0.001**

B

Variable	**HR**	**95% CI**	**P**
Only DTA* mutation	**1.18**	**0.98-1.41**	**0.077**
Other mutation	**0.97**	**0.63-1.50**	**0.899**
Sex (male)	**1.15**	**0.91-1.44**	**0.247**

Smoking (1st tertile)	**1.04**	**0.83-1.31**	**0.715**
Smoking (2nd tertile)	**1.25**	**0.99-1.58**	**0.060**
Smoking (3rd tertile)	**1.75**	**1.28-2.38**	**<0.001**

*DTA=*DNMT3A,TET2 and ASXL1*

In table **A** time since blood sample is used as the underlying time scale whereas age is used as the underlying the time scale in **B**. When we divided the mutations into DTA and "other mutations" defined as non-DTA(*DNMT3A*, *TET2*, *ASXL1*) mutations, but with or without co-occurring DTA mutations, we found a significant association of DTA mutations after adjusting for age, sex and tobacco consumption. This association was retained as a borderline significant association when we used age as the underlying timescale in the Cox-regression.

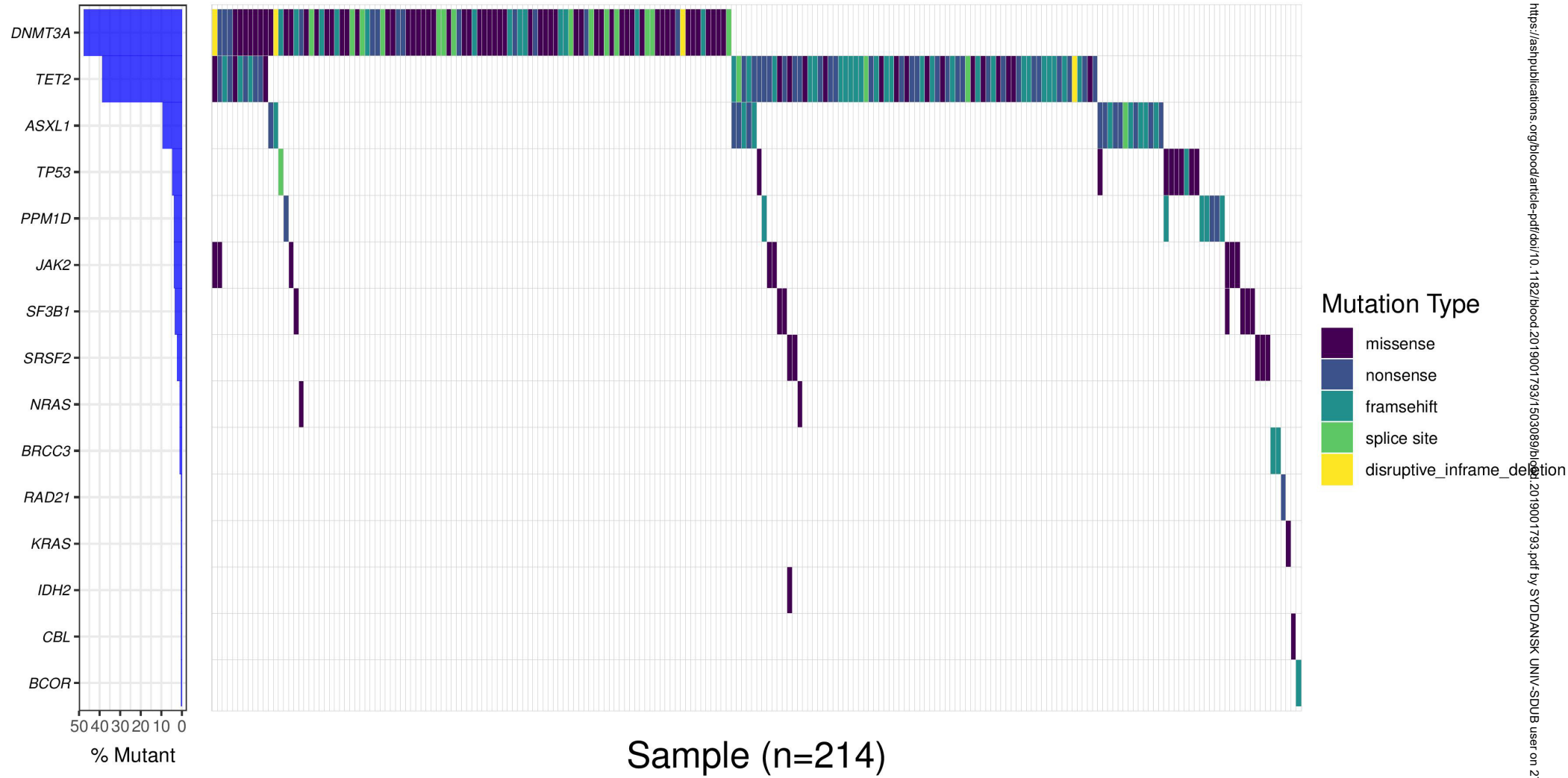
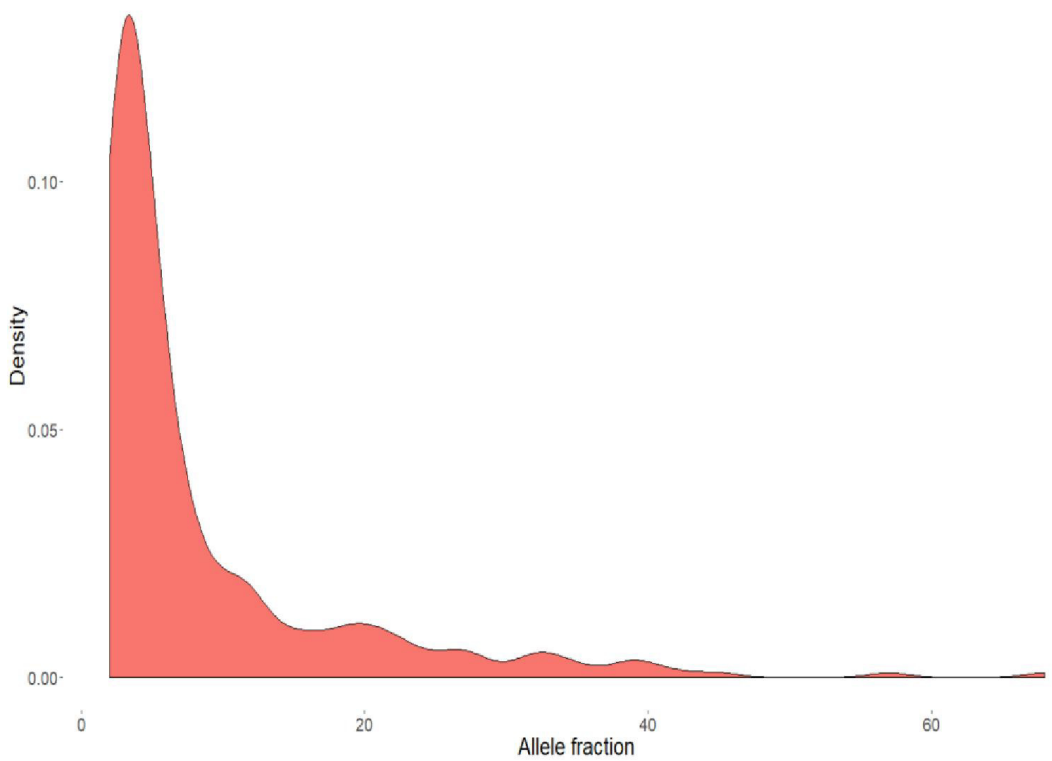
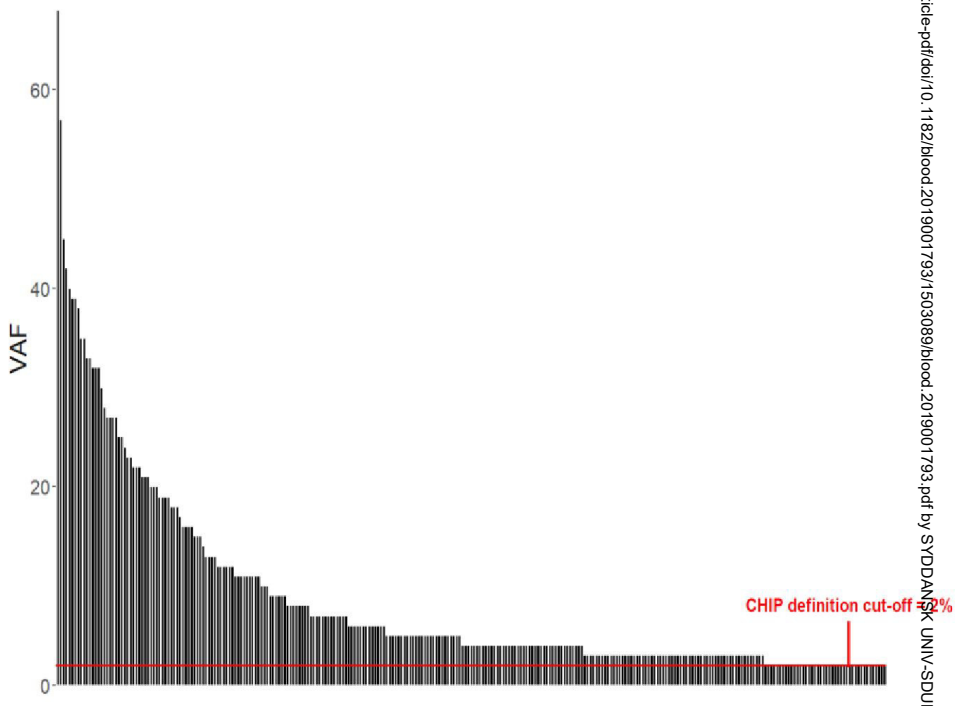


Figure 2



A



B