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

Concordance for clonal hematopoiesis is limited in elderly twins

Tracking no: BLD-2019-001807R1

Margarete Fabre (Wellcome Sanger Institute, United Kingdom) Thomas McKerrell (Wellcome Trust Sanger Institute, United Kingdom) Maximillian Zwiebel (German Consortium for Translational Cancer Research, Germany) MS Vijayabaskar (Wellcome Trust Sanger Institute, United Kingdom) Naomi Park (Sanger institute, United Kingdom) Philippa Wells (King's College London, United Kingdom) Roland Rad (Technical University of Munich, Germany) Panagiotis Deloukas (Queen Mary University of London, United Kingdom) Kerrin Small (King's College London, United Kingdom) Claire Steves (King's College London, United Kingdom) George Vassiliou (University of Cambridge & Wellcome Sanger Institute, United Kingdom)

Abstract:
While acquisition of leukemia-associated somatic mutations by one or more hematopoietic stem cells (HSCs) is inevitable with advancing age, its consequences are highly variable, ranging from clinically silent clonal hematopoiesis (CH) to leukemic progression. To investigate the influence of heritable factors on CH, we performed deep targeted sequencing of blood DNA from 52 monozygotic (MZ) and 27 dizygotic (DZ) twin pairs (aged 70-99 years). Using this highly sensitive approach, we identified CH (Variant Allele Fraction (VAF) ≥0.5%) in 62% of individuals. We did not observe higher concordance for CH within MZ twin pairs as compared to that within DZ twin pairs, or to that expected by chance. However, we did identify two MZ pairs in which both twins harbored identical rare somatic mutations, suggesting a shared cell of origin. Finally, in three MZ twin pairs harboring mutations in the same driver genes, serial blood samples taken 4-5 years apart showed substantial twin-to-twin variability in clonal trajectories. Our findings propose that the inherited genome does not exert a dominant influence on the behavior of adult CH and provide evidence that CH mutations may be acquired *in utero*.

Conflict of interest: COI declared - see note

COI notes: G.S.V. is a consultant for Kymab Ltd and OxStem Ltd, and receives a research grant from Celgene.

Preprint server: No;

Author contributions and disclosures: G.S.V., T.M. and P.D. conceived and designed the study; G.S.V. supervised the study; M.A.F., T.M. and M.Z. performed the bulk of bioinformatic and statistical analyses with help from M.S.V; N.P. performed sequencing validation experiments; P.M.W. performed heritability modeling; K.S. and C.S. guided the choice of twin samples and advised on analysis; G.S.V., M.A.F., T.M. and M.Z. wrote the manuscript with input from all co-authors.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Public repository (EGAD00001005055)

Clinical trial registration information (if any):

Concordance for clonal hematopoiesis is limited in elderly twins

Margarete A Fabre[1-3]*, Thomas McKerrell[1,2]*, Maximillian Zwiebel[1,4]*, MS Vijayabaskar[1,2], Naomi Park[1], Philippa M Wells[5], Roland Rad[4], Panagiotis Deloukas[6], Kerrin Small[5], Claire J Steves[5], and George S Vassiliou[1-3†].

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom.
2. Wellcome-MRC Cambridge Stem Cell Institute, Hills Road, Cambridge, CB2 0XY, United Kingdom.
3. Department of Haematology, Cambridge University Hospitals NHS Trust, Cambridge, United Kingdom.
4. German Consortium for Translational Cancer Research (DKTK), Partnering Site, Munich, 81377 Munich, Germany.
5. Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, SE1 7EH, United Kingdom.
6. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, United Kingdom.

* These authors contributed equally to this work

† Correspondence: gsv20@sanger.ac.uk

Short title for the running head: **Clonal hematopoiesis in twins**

Abstract word count: 186
Text word count: 1236
Figure count: 2
Reference count: 25

Scientific category: Hematopoiesis and Stem Cells

Key Points

- Limited concordance and differing trajectories of clonal hematopoiesis in identical twins emphasize the importance of non-heritable factors

- Identification of elderly monozygotic twins with identical driver mutations suggests a common cellular origin *in utero*

Abstract

While acquisition of leukemia-associated somatic mutations by one or more hematopoietic stem cells (HSCs) is inevitable with advancing age, its consequences are highly variable, ranging from clinically silent clonal hematopoiesis (CH) to leukemic progression. To investigate the influence of heritable factors on CH, we performed deep targeted sequencing of blood DNA from 52 monozygotic (MZ) and 27 dizygotic (DZ) twin pairs (aged 70-99 years). Using this highly sensitive approach, we identified CH (Variant Allele Fraction (VAF) ≥0.5%) in 62% of individuals. We did not observe higher concordance for CH within MZ twin pairs as compared to that within DZ twin pairs, or to that expected by chance. However, we did identify two MZ pairs in which both twins harbored identical rare somatic mutations, suggesting a shared cell of origin. Finally, in three MZ twin pairs harboring mutations in the same driver genes, serial blood samples taken 4-5 years apart showed substantial twin-to-twin variability in clonal trajectories. Our findings propose that the inherited genome does not exert a dominant influence on the behavior of adult CH and provide evidence that CH mutations may be acquired *in utero*.

Introduction

Clonal hematopoiesis (CH), the disproportionate expansion of blood cell clones harboring leukemia-associated somatic mutations, becomes more prevalent with advancing age and is the precursor of many hematological malignancies[1-5]. Acquisition of such mutations in one or more hematopoietic stem cells (HSCs) is inevitable by the age of 50-60 years[6], yet the consequences of mutation acquisition are highly variable between individuals. A number of small studies tracking clonal size longitudinally suggest that clones in different individuals with similar or even identical mutations behave differently over time[5-7]. Indeed, this could also be inferred by the fact that clinically silent CH is common, whilst hematological cancers are rare[5]. Importantly, acquisition of additional driver mutations is not always necessary for malignant progression; for example, the *JAK2*-V617F mutation[8] is the sole identifiable driver in many cases of myeloproliferative neoplasms (MPN), and *SF3B1* mutations[9,10] are often the only driver in Myelodysplastic Syndromes (MDS). The factors that allow mutant clones to expand in some individuals, and those restraining them in others, are not understood.

There is some evidence that the inherited genome might play a role in this process. For example, recent studies reported (i) heritable genetic variants associated with increased risk of developing MPNs[11,12], (ii) familial clustering of CH driven by *TET2* mutations[13], and (iii) increased prevalence of CH among relatives of individuals with myeloid, but not lymphoid, malignancies[14]. Moreover, a number of germline variants have emerged as important

determinants of hematological phenotypes in the general population and it is plausible that these exert epistatic effects on CH evolution[15]. In order to investigate whether the inherited genome influences CH development, we performed deep targeted sequencing on blood DNA from 52 monozygotic (MZ) and 27 dizygotic (DZ) twin pairs, and analyzed patterns of twin-to-twin concordance for CH.

Methods

We studied blood DNA from 158 individuals from the TwinsUK cohort, comprising 52 MZ and 27 DZ twin pairs with no history of hematological malignancy, aged 70-99 years (150 females, 8 males)[16]. Samples were obtained with informed consent and appropriate ethics committee approval (REC reference EC04/015). Target enrichment for 41 genes implicated in CH and myeloid malignancies (Agilent SureSelect, ELID 0735431, Supplemental Table 1) was performed successfully for 154 samples. Libraries were sequenced on Illumina HiSeq 2000 and variant-calling was performed as we described previously[5]. Briefly, somatic single nucleotide variants (SNV) and small indels were called using Shearwater (v.1.21.5), an algorithm designed to detect subclonal mutations in deep sequencing experiments[17]. Two additional variant-calling algorithms were applied to complement this approach: CaVEMan (v.1.11.2) for SNVs, and Pindel (v.2.2) for indels. Finally, allele counts at recurrent mutation hotspots were verified using an in-house script (*https://github.com/cancerit/alleleCount*). Driver mutations were defined according to evidence for functional relevance in CH and hematological malignancy (Supplemental Table 2). Methodological validation of our approach is outlined in Supplemental Figure 1, and Supplemental Tables 3 and 4.

Statistical analyses were performed in R (version 3.4.0). Fisher's Exact Test was used to assess twin concordance for CH. Null distributions of CH within the MZ and DZ groups were generated using random sample permutation (1000 iterations). The openMX R package was used for maximum likelihood modeling of genetic and environmental contributions to CH[18].

Results and Discussion

Mutational landscape in the cohort

Using deep sequencing (mean 1650X) and sensitive variant-calling, we identified CH (VAF ≥0.5%) in 62% of individuals (95/154; Figure 1A), with larger clones (VAF ≥2%) present in 41 individuals (Figure 1B). Somatic driver mutations were identified in 16 of the 41 genes sequenced, with mutations in the epigenetic regulators *DNMT3A* and *TET2* predominant (Figure 1C-E; Supplemental Table 5). Almost one third of individuals (48/154) harbored multiple distinct mutations (Figure 1F), often in the same gene and at different VAFs (Supplemental Figure 2), suggesting the presence of multiple clones or subclones.

CH in twin pairs

Comparing age-matched samples from MZ and DZ groups (age range 70-80 years; n=33 MZ pairs and 24 DZ pairs), the overall prevalence of CH was very similar (59% and 54% respectively; p=0.70). We did not observe significantly higher concordance for CH within MZ twin pairs as compared to DZ pairs (p=0.59, Figure 2A). Furthermore, using random

permutation to model the null distribution, we found no difference in the observed distributions of CH among either MZ or DZ twins as compared to those expected by chance (p=1 for MZ; p=0.86 for DZ; Figure 2A). Excess twin concordance was also not observed when CH positivity was defined by (i) mutation in DNMT3A, (ii) mutation in TET2, and (iii) mutation in any gene with VAF > 2% (Supplemental Figure 3).

In 8 individuals (4 MZ twin pairs), serial blood samples were taken 4-5 years apart. CH clones were identified in both twins in 3/4 of these pairs, and inter-twin variability in clonal size and trajectory was seen in all three (Figure 2B). This was the case even for clones harboring mutations in the same gene.

In summary, we find no evidence of high concordance for CH in elderly MZ twins. In addition, we observe disparity in clonal size and trajectory over time, even between MZ twins harboring mutations in the same driver genes. Whilst our cohort size is too small to precisely quantify genetic versus environmental contributions, maximum likelihood modeling provides support for a substantial influence of non-inherited factors upon CH emergence and behavior (Supplemental Table 6). Overwhelming evidence shows that mutation acquisition is widespread and inevitable in the aging hematopoietic system[6], as is the case in other normal tissues studied to date[19,20], indicating that this is not the rate-limiting step in CH development. By suggesting that the inherited genome does not play a dominant role, our study frames non-genetic events as important factors in CH emergence. Altered interactions of the HSC with its environment, associated with processes such as aging, senescence, inflammation and infection are plausible operators, some of which are supported by evidence from experimental models[21].

Despite the overall lack of concordance for CH, we did identify two MZ pairs in which both twins harbored identical nonsense mutations, namely *KDM6A* Q692X in one pair and *DNMT3A* R598X in the other (Figure 2C-D; Supplemental Table 5). KDM6A (=UTX) is a histone H3 lysine 27 demethylase which acts as a tumor suppressor in a number of different cancers, including 2-3% of myeloid malignancies[22]. There are no somatic mutation hotspots in *KDM6A* and the substitution identified here is not reported in either the COSMIC database (https://cancer.sanger.ac.uk/cosmic) or in several large, albeit less sensitive, CH studies (total number of participants >30,000)[1,2,4]. Although mutations in DNMT3A are generally more prevalent, the particular mutation detected here is not common. In this light, the likelihood that each member of these two twin pairs acquired the same mutation independently and by chance is extremely small. A more plausible explanation is that the somatic mutation occurred just once during embryogenesis, either prior to twinning or in an HSC whose progeny reached both twins through shared circulation in utero. While monozygotic twin sharing of somatic mutations has been demonstrated in other settings, including pediatric leukemia, this is the first description of possible acquisition of adult-type CH driver mutations *in utero*[23-25].

In conclusion, the lack of strong concordance for CH and the variable clonal trajectories between MZ twins, indicate that the inherited genome does not exert a profound influence on the emergence and behavior of CH in older adults. In addition, sharing of rare somatic mutations by MZ twins raises the possibility that mutations driving adult CH may sometimes be acquired *in utero*.

Acknowledgments

This work was funded by Wellcome Trust (WT098051). G.S.V. is funded by a Cancer Research UK Senior Cancer Fellowship (C22324/A23015) and work in his lab is also funded by the European Research Council, Kay Kendall Leukaemia Fund, Bloodwise, Leukaemia Lymphoma Society and Rising Tide Foundation for Clinical Cancer Research. M.A.F. is funded by a Wellcome Clinical PhD Fellowship. We thank the participants of the TwinsUK Registry. We thank Kirsten Grønbæk, Kaare Christensen, Jacob Werner Hansen and Hannes Ponstingl for helpful discussions.

Data Availability

Sequencing data are deposited at the European Genome-phenome Archive with accession number EGAD00001005055.

Author Contributions

G.S.V., T.M. and P.D. conceived and designed the study; G.S.V. supervised the study; M.A.F., T.M. and M.Z. performed the bulk of bioinformatic and statistical analyses with help from M.S.V; N.P. performed sequencing validation experiments; P.M.W. performed heritability modeling; K.S. and C.S. guided the choice of twin samples and advised on analysis; G.S.V., M.A.F., T.M. and M.Z. wrote the manuscript with input from all co-authors.

Disclosure of Conflicts of Interest

G.S.V. is a consultant for Kymab Ltd and OxStem Ltd, and receives a research grant from Celgene.

References

1. Genovese G, Kahler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-2487.
2. 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.
3. 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.
4. 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.
5. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature.* 2018;559(7714):400-404.
6. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat Commun.* 2016;7:12484.

7. McKerrell T, Park N, Chi J, et al. JAK2 V617F hematopoietic clones are present several years prior to MPN diagnosis and follow different expansion kinetics. *Blood Adv.* 2017;1(14):968-971.
8. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med.* 2013;369(25):2391-2405.
9. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* 2013;122(22):3616-3627; quiz 3699.
10. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia.* 2014;28(2):241-247.
11. Hinds DA, Barnholt KE, Mesa RA, et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood.* 2016;128(8):1121-1128.
12. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet.* 2009;41(4):446-449.
13. 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.
14. 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.
15. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 2016;167(5):1415-1429 e1419.
16. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol.* 2013;42(1):76-85.
17. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics.* 2014;30(9):1198-1204.
18. Neale MC, Hunter MD, Pritikin JN, et al. OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika.* 2016;81(2):535-549.
19. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science.* 2015;348(6237):880-886.
20. Yokoyama A, Kakiuchi N, Yoshizato T, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature.* 2019;565(7739):312-317.
21. Meisel M, Hinterleitner R, Pacis A, et al. Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature.* 2018;557(7706):580-584.
22. Gozdecka M, Meduri E, Mazan M, et al. UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. *Nat Genet.* 2018;50(6):883-894.
23. Greaves M, Hughes W. Cancer cell transmission via the placenta. *Evol Med Public Health.* 2018;2018(1):106-115.
24. Ju YS, Martincorena I, Gerstung M, et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature.* 2017;543(7647):714-718.
25. Lee-Six H, Obro NF, Shepherd MS, et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature.* 2018;561(7724):473-478.

Figure Legends

Figure 1. CH in individuals. (A) Grey bars depict the proportions of individuals *with* CH (dark grey = MZs, light grey = DZs), and white bars represent the proportions *without* CH, as identified in respective age groups. Absolute numbers of individuals in each proportion are shown within each bar. There were no DZ individuals above the age of 80 years. (B) Distribution of the maximum Variant Allele Frequency (VAF) per individual among those with CH. VAFs are divided into 0.5% bins. (C) In the main grid, each column represents one individual, and each row one gene. If a grid square is colored, a mutation was detected, and the specific color indicates the mutation type (see key). The right-hand plot shows the proportion of the cohort harboring a mutation in each gene. (D) and (E) Somatic variants identified in DNMT3A (D) and TET2 (E). Conserved / functional protein domains are colored red, and intervening domains grey. Each circle connected to the protein cartoon represents a mutation. Missense mutations are represented above and truncating mutations below the protein, with the color of the circle indicating specific mutation type (as per the key in (C)). (F) Distribution of the total number of mutations per individual.

Figure 2. CH in twin pairs. (A) Concordance for CH status in age-matched MZ (n=33) and DZ (n=24) twin pairs. Red bars represent the proportion of twin pairs in which *neither* has CH, blue bars show the proportion in which only *one* twin has CH, and green bars show the proportion in which *both* twins have CH. Observed (obs) proportions are those identified in the cohort, and expected (exp) are those generated by random sample permutation. There was no significant difference when comparing (i) the observed distributions between MZ and DZ twins (p=0.59), (ii) the observed vs expected distributions within in the MZ (p=1) or DZ (p=0.86) twin groups (Fisher's exact test). (B) Change in VAF over time in the 3 twin pairs in which both individuals had CH and serial samples were available. Each box surrounded by a solid line represents a twin pair. The fourth box surrounded by a dashed line is data from the third twin pair with the y-scale zoomed in to lower VAFs. In each box, change in VAF over time is represented by a solid line for one twin, and a dotted line for the other, with line color indicating which gene was mutated. Figures (C) and (D) focus on the variants identified in two MZ twin pairs in which both twins harbored identical nonsense mutations: KDM6A (NM_021140:c.C2074T:p.Q692X) (C), and DNMT3A (NM_175629:c.C1792T:p.R598X) (D). Each 'triplet' represents the three non-reference bases at each genomic position, centered on the identified variant position (denoted as 0). For each alternate allele, the VAF of each of the 154 individuals in the cohort is plotted. The horizontal dashed line represents the lower limit of sensitivity of variant-calling, with calls below this VAF considered error. MZ twins with identical mutations are plotted in red, all other individuals in black.