Leukemia Research Reports 19 (2023) 100362

Contents lists available at ScienceDirect

Leukemia Research Reports

journal homepage: www.elsevier.com/locate/lrr

Contribution of rare and common coding variants to haematological malignancies in the UK biobank

ARTICLE INFO

Keywords
Genetics
Hematologic neoplasms
Mutation
Whole exome sequencing

Haematological malignancies are a group of diverse conditions that originate from cells of the bone marrow and the lymphatic system [1]. There are three major groups of haematological malignancies: leukaemia, lymphoma, and plasma cell neoplasms [1]. The exact cause of most cases of haematological malignancies is not known although acquired factors like DNA-damaging agents and infections have been implicated [1,2]. Haematological malignancies have been linked to rare inherited syndromes and rare germline mutations in Mendelian disorders (*ANKRD26, CEBPA, DDX41, ELANE, ETV6, GATA2, HAX1, RUNX1, SAMD9, SAMD9L, SRP72*, and *LSD1*) [2]. Familial aggregation has been reported for several haematological malignancies [2]. Using aggregated whole exome sequencing (WES) data from UK biobank provided by two public portals (https://azphewas.com/ and https://app.genebass.org/) we have summarized the genetic contribution to haematological malignancies in UK biobank [3,4].

Wang et al reported the relationships between rare protein-coding variants and 17,361 binary phenotypes using WES data from 269,171 UK Biobank participants (https://azphewas.com/) [3], and Karczewski et al determined gene-based association investigating 4529 phenotypes in 394,841 UK biobank exomes (https://app.genebass.org/) [4]. AstraZeneca portals includes both variant-level and gene-level association tests (https://azphewas.com/) [3]. Gene-level association tests used collapsing analyses [3]. The proportion of cases with a qualifying variant was compared with the proportion of controls with a qualifying variant in each gene [3]. Twelve different sets of qualifying variant filters (models) were analyzed [3]. Genebass also includes single-variant tests and gene collapsing tests with gene-based burden (mean), SKAT (variance), and SKAT-O (hybrid variance/mean) tests (https://app.genebass.org/) [4]. Only candidate genes with p-values $< 2.5 \times 10^{-6}$ and candidate variants with p-values $< 5 \times 10^{-8}$ are shown in Table 1. To be comparable we only show the union of the ICD-10 codes (International Classification of Diseases 10th Revision) [3,4]. In Table 1 the genes with genome-wide significant results are shown with p-values for the best significant model. The involvement of rare variation using gene collapsing analysis was observed for 16 malignant phenotypes studied. Totally, thirty-one genes were linked to hematologic malignancies using gene collapsing analysis. Involvement of twenty-two specific coding variants in eighteen genes were observed for 14 malignant phenotypes. Many of the genes and variants are known to be involved in haematological malignancies as somatic mutations [5]. However, the data from UK biobank shows that germline inheritance of rare variants and specific coding variants are risk factors for hematologic malignancies (https://azphewas.com/ and https://app.genebass.org/) [3,4]. This is in line with two recent reports of patients with hematologic malignancies [6,7]. Some of the genes and variants are shared between different malignancies and others are unique (Table 1). There were associations to two syndromic genes associated with hematologic malignancies [2]: *ETV6* (myelodysplastic syndrome) and *RUNX1* (myeloid leukaemia). There were association between *CHEK2* and myeloid leukaemia and *ATM* and lymphoid leukaemia (Table 1) in agreement with previous studies [8]. Aberrations in *CHEK2* and *ATM* genes are known to increase the risk for all types of malignancies including hematologic malignancies [6]. Rare *POT1* (Protection Of Telomeres 1) mutations were associated with multiple myeloma and malignant plasma cell neoplasms and chronic lymphocytic leukaemia (Table 1). Mutations in *POT1* are known to be associated with malignant melanoma and glioma (https://www.genecards.org/). The data from UK biobank reported by Wang et al [3] and Karczewski et al [4] may give the molecular explanation for previously observed familial aggregation of haematological malignancies [2].

There were several mutations found in genes [*DNMT3A, TET2, ASXL1, JAK2, SF3B1, SRSF2*] associated with clonal haematopoiesis of indeterminate potential (CHIP) [9,10] (Table 1), suggesting that inherited germline mutations may predispose individuals for hematologic malignancies. However, a recent report by Ariste et al suggests these mutations should be interpreted with caution especially when they occur in elder individuals [10]. They could represent somatic alterations and not germline variants [10]. Determination of these mutations in other cell types, such as skin fibroblasts, could be assessed to avoid misclassification of variants in cancer-predisposition genes as inherited [10]. Thus, this will be an important research task to determine whether these mutations are somatic or germline.

In conclusion, rare variation and specific coding variants are risk factors for haematological malignancies. These findings provide the

https://doi.org/10.1016/j.lrr.2023.100362
Received 27 October 2022; Received in revised form 29 November 2022; Accepted 9 January 2023
Available online 10 January 2023

Table 1
Results of gene collapsing analysis of rare variants and individual coding variants for haematological malignancies according to ICD-10* codes (https://azphewas.com/ and https://app.genebass.org/).[3,4] Union was used to define phenotypes for https://azphewas.com. Only candidate genes with p-values $< 2.5 \times 10^{-6}$ and candidate variants with p-values $< 5 \times 10^{-8}$ are shown.

	Genebass LOF	Genebass missense/LC	Genebass variant level	Astra Zeneca gene collapsing	Astra Zeneca variant level
C81 Hodgkin's disease	NS	*ELF3* 7.09e-9	*ELF3* Asn104Ser 3.51e-8 *XRRA1* Gly772Asp 4.66e-8	NS	NS
C82 Follicular [nodular] non-Hodgkin's lymphoma	NS	NS	*HLA-DRB5* Asp66Ala 1.26e-8 *HLA-DRB5* Lys14ThrfsTer27 3.23e-8	NS	*CRYBG3* Thr1114Thr 3.71e-8
C83 Diffuse non-Hodgkin's lymphoma	NS	*CCND1* 1.22e-6	NS	NS	*TAF6* 7-100113586-C-T 8.24e-9
C84 Peripheral and cutaneous T-cell lymphomas	NS	NS	NS	NS	NS
C85 Other and unspecified types of non-Hodgkin's lymphoma	NS	*MYD88* 1.68e-8 *SCAMP1* 1.29e-6	*MYD88* Ter147ArgextTer8 3.8e-9	*TET2* 5.24e-7	*NOTCH4* Leu15_Leu16dup 4.32e-8
C86 Other specified types of T|NK-cell lymphoma	ND	ND	ND	NS	NS
C88 Malignant immunoproliferative diseases	NS	*OR1J1* 1.54e-6 *IWS1* 1.77e-6	NS	NS	NS
C90 Multiple myeloma and malignant plasma cell neoplasms	NS	*MKRN2* 1.76e-7	NS	*POT1* 2.80e-7 *TET2* 2.42e-6	NS
C91 Lymphoid leukaemia	*NOTCH1* 1e-18 *CHD2* 3.25e-12	*MYD88* 1.6e-14 *SF3B1* 9.43e-12 *IGLL5* 1.71e-12 *IGHV4*-39 1.53e-8 *GIP* 6.92e-7	*MYD88* Ter147ArgextTer8 1.5e-16 *IGLL5* Glu15Asp 3.47e-14 *NOTCH1* Pro2514ArgfsTer4 5.52e-14 *SF3B1* Lys700Glu 1.5e-11	*NOTCH1* 6.32e-15 *IGLL5* 3.79e-14 *CHD2* 1.50e-13 *SF3B1* 9.93e-9 *ATM* 4.34e-7 *POT1* 4.57e-7 *CHEK2* 5.41e-7 *KMT2D* 1.48e-6	NS
C91.0 Acute lymphoblastic leukaemia	ND	ND	ND	NS	NS
C91.1 Chronic lymphocytic leukaemia	ND	ND	ND	*IGLL5* 3.08e-16 *NOTCH1* 2.08e-15 *CHD2* 9.60e-14 *SF3B1* 1.34e-10 *POT1* 2.21e-7 *KMT2D* 6.08e-7	*BCL2L11* Ile95Ile 3.32e-8
C91.4 Hairy-cell leukaemia	ND	ND	ND	NS	NS
C92 Myeloid leukaemia	*TET2* 2.45e-20 *EZH2* 1.53e-8 *CHEK2* 5.78e-7	*JAK2* 6.29e-12 *DNMT3A* 2.58e-7	*IDH2* Arg140Gln 3.33e-21 *JAK2* Val617Phe 2.54e-16 *SRSF2* Pro95_Arg102del 1.45e-8	*SRSF2* 1.43e-58 *TET2* 5.83e-41 *ASXL1* 5.97e-35 *DDX41* 1.76e-13 *RUNX1* 2.49e-11 *CHEK2* 1.65e-10 *DNMT3A* 8.74e-10 *EZH2* 2.35e-8	*SRSF2* Pro95His 7.13e-42 *JAK2* Val617Phe 1.09e-24
C92.1 Chronic myeloid leukaemia	ND	ND	ND	*TET2* 4.80e-13 *SRSF2* 7.22e-13	NS
C93 Monocytic leukaemia	ND	ND	ND	*TET2* 9.58e-37 *SRSF2* 5.77e-26 *ASXL1* 1.11e-9 *ALX4* 2.06e-6	*SRSF2* Pro95His 4.57e-10
C94 Other leukaemias of specified cell type	ND	ND	ND	*TET2* 1.43e-8	*JAK2* Val617Phe 5.76e-34
C95 Leukaemia of unspecified cell type	ND	ND	ND	NS	NS
C96 Other and unspecified malignant neoplasms of lymphoid| haematopoietic and related tissue	ND	ND	ND	*TET2* 5.29e-11 *HIST1H4G* 1.77e-7	*JAK2* Val617Phe 2.05e-31
D45 Polycythaemia vera	ND	ND	ND	*JAK2* 7.71e-13 *TET2* 5.97e-9	*JAK2* Val617Phe 6.79E-255 *HFE* Cys282Tyr 1.50E-22 *JAK2* His163His 5.33E-17 *INSL4* Arg63Arg 1.50E-14 *JAK2* Leu830Leu 2.14E-11 *HIST1H1T* Gln178Lys 2.77E-08

(*continued on next page*)

Table 1 (*continued*)

	Genebass LOF	Genebass missense/LC	Genebass variant level	Astra Zeneca gene collapsing	Astra Zeneca variant level
					HIST1H4C Leu91Leu 3.06E-08
D46 Myelodysplastic syndromes (MDS)	ND	ND	ND	*SRSF2* 4.98e-34 *TET2* 3.94e-32 *ASXL1* 1.43e-25 *SF3B1* 8.45e-15 *DDX41* 1.97e-13 *ETV6* 2.50e-7 *EZH2* 7.80e-7 *ZNF528* 1.35e-6	*SRSF2* Pro95His 4.63e-26 *JAK2* Val617Phe 2.91e-25
D47 Other neoplasms of uncertain or unknown behaviour of lymphoid| haematopoietic and related tissue	ND	ND	ND	*TET2* 1.61e-28 *SRSF2* 2.38e-17 *ASXL1* 2.36e-16 *CALR* 3.58e-12 *JAK2* 4.23e-10 *SF3B1* 1.97e-7 *DNMT3A* 3.06e-7	*JAK2* Val617Phe 1.83e-272 *SRSF2* Pro95His 9.72e-12 *JAK2* His163His 6.47e-9 *JAK2* Leu830Leu 3.49e-8
D47.2 Monoclonal gammopathy	ND	ND	ND	NS	NS
D47.3 Essential (haemorrhagic) thrombocythaemia	ND	ND	ND	*TET2* 5.41e-13 *CALR* 1.52e-10 *ZIC5* 9.14e-7	*JAK2* Val617Phe 4.35e-93

*ICD-10=International Classification of Diseases 10th Revision; NS=no genome wide significant gene; ND=not determined; Significance threshold was p-values < 0.05/20,000 genes = 2.5×10^{-6} commonly used for WES studies and $p < 5 \times 10^{-8}$ for variant level. For Genebass portal (https://app.genebass.org) the highest p-value for SKAT, SKAT-O or burden test is shown.[4] For Astra Zeneca portal (https://azphewas.com) the highest p-value of the 12 tested models is shown.[3]
LoF: High-confidence Loss of function variants indicated by LOFTEE.[4] Missense/LC: Missense variants are grouped with in-frame insertions and deletions, as well as low-confidence LoF variants filtered out by LOFTEE.[4] The latter have a frequency spectrum consistent with missense variation and affect a set of amino acids in a similar way.[4]

molecular fundament for the previously reported familial aggregation of haematological malignancies.

Declaration of Competing Interest

None

Acknowledgments

This work was supported by a grant awarded to Dr Bengt Zöller by ALF-funding from Region Skåne, Sparbanken Skåne, and by the Swedish Research Council.

References

[1] D Rodriguez-Abreu, A Bordoni, E. Zucca, Epidemiology of hematological malignancies, Ann. Oncol. 18 (2007) I3–I8.
[2] A Sud, S Chattopadhyay, H Thomsen, et al., Analysis of 153 115 patients with hematological malignancies refines the spectrum of familial risk, Blood 134 (2019) 960–969.
[3] Q Wang, RS Dhindsa, K Carss, et al., Rare variant contribution to human disease in 281,104 UK Biobank exomes, Nature 597 (2021) 527–532.
[4] K.J. Karczewski, M. Solomonson, Chao, et al., Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes, Cell Genomics 2 (2022), 100168.
[5] J Taylor, W Xiao, O Abdel-Wahab, Diagnosis and classification of hematologic malignancies on the basis of genetics, Blood 130 (2017) 410–423.
[6] ES Martin, A Ferrer, AA Mangaonkar, et al., Spectrum of hematological malignancies, clonal evolution and outcomes in 144 mayo clinic patients with germline predisposition syndromes, Am. J. Hematol. 96 (2021) 1450–1460.
[7] D Singhal, CN Hahn, S Feurstein, et al., Targeted gene panels identify a high frequency of pathogenic germline variants in patients diagnosed with a hematological malignancy and at least one other independent cancer, Leukemia 35 (2021) 3245–3256.
[8] RJ Stubbins, S Korotev, LA. Godley, Germline CHEK2 and ATM variants in myeloid and other hematopoietic malignancies, Curr. Hematol. Malig. Rep. 17 (2022) 94–104.
[9] M Bou Zerdan, L Nasr, L Saba, et al., A synopsis clonal hematopoiesis of indeterminate potential in hematology, Cancers (Basel) 14 (2022) 3663.
[10] O Ariste, P de la Grange, RA. Veitia, Recurrent missense variants in clonal hematopoiesis-related genes present in the general population, Clin. Genet. 10 (2022), https://doi.org/10.1111/cge.14259. NovOnline ahead of print.

Bengt Zöller[a,*], Eric Manderstedt[a], Christina Lind-Halldén[b], Christer Halldén[b]
[a] *Center for Primary Health Care Research, Department of Clinical Sciences, Lund University and Region Skåne, Malmö, Sweden*
[b] *Department of Environmental Science and Bioscience, Kristianstad University, Kristianstad, Sweden*

[*] Corresponding author at: Center for Primary Health Care Research, CRC, Building 28, Floor 11, Jan Waldenströms gata 35, Skåne University Hospital, S-205 02 Malmö, Sweden.
E-mail address: bengt.zoller@med.lu.se (B. Zöller).