OXFORD

Human Molecular Genetics, 2022, Vol. 31, 14, 2386–2395

https://doi.org/10.1093/hmg/ddac034

Advance access publication date: 18 February 2022

General Article

Clonal hematopoiesis as a pitfall in germline variant interpretation in the context of Mendelian disorders

Theresa Brunet[1,2,*], Riccardo Berutti[1,2], Veronika Dill[3], Judith S. Hecker[3], Daniela Choukair[4], Stephanie Andres[5], Marcus Deschauer[6], Janine Diehl-Schmid[7,8], Martin Krenn[1,9], Gertrud Eckstein[2,10], Elisabeth Graf[1], Thomas Gasser[11,12], Tim M. Strom[1], Julia Hoefele[1], Katharina S. Götze[3], Thomas Meitinger[1] and Matias Wagner[1,2]

[1]Institute of Human Genetics, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, 81675 Munich, Germany
[2]Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg 85764, Germany
[3]Department of Medicine III, Technical University of Munich, Klinikum rechts der Isar, 81675 Munich, Germany
[4]Division of Paediatric Endocrinology and Diabetology, University Children's Hospital, 69120 Heidelberg, Germany
[5]Center of Human Genetics and Laboratory Diagnostics, 82152 Martinsried, Germany
[6]Department of Neurology, Technical University of Munich, School of Medicine, 81675 Munich, Germany
[7]Technical University of Munich, School of Medicine, Department of Psychiatry and Psychotherapy, 81675 Munich, Germany
[8]Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany
[9]Department of Neurology, Medical University of Vienna, 1090 Vienna, Austria
[10]Core Facility Genomics, Helmholtz Center Munich, Neuherberg 85764, Germany
[11]Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany
[12]German Center for Neurodegenerative Diseases (DZNE), 72076 Tübingen, Germany

*To whom correspondence should be addressed at: Institute of Human Genetics, Technical University Munich, Trogerstr. 32, 81675 Munich, Germany. Tel: +49 8941406381; Fax: +49 8941406382; Email: theresa.brunet@mri.tum.de

Abstract

Clonal hematopoiesis because of somatic mutations in hematopoietic stem/progenitor cells is an age-related phenomenon and commonly observed when sequencing blood DNA in elderly individuals. Several genes that are implicated in clonal hematopoiesis are also associated with Mendelian disorders when mutated in the germline, potentially leading to variant misinterpretation. We performed a literature search to identify genes associated with age-related clonal hematopoiesis followed by an OMIM query to identify the subset of genes in which germline variants are associated with Mendelian disorders. We retrospectively screened for diagnostic cases in which the presence of age-related clonal hematopoiesis confounded exome sequencing data interpretation. We found 58 genes in which somatic mutations are implicated in clonal hematopoiesis, while germline variants in the same genes are associated with Mendelian (mostly neurodevelopmental) disorders. Using five selected cases of individuals with suspected monogenic disorders, we illustrate how clonal hematopoiesis in either variant databases or exome sequencing datasets poses a pitfall, potentially leading to variant misclassification and erroneous conclusions regarding gene–disease associations.

Introduction

Clonal hematopoiesis (CH) is characterized by a genetically distinct subpopulation of mature blood cells derived from acquired mutations in a hematopoietic stem cell (HSC) conferring a selective growth advantage (1,2). CH increases with age with a prevalence of 10–20% in those older than age 70 years and >20% in those older than 90 years (3–5). Low-level mosaicism in circulating blood cells resulting from somatic mutations in hematopoietic stem/progenitor cells can be found in individuals with normal blood cell counts and without any signs of hematologic disorders (6,7). The clinical term 'clonal hematopoiesis of indeterminate potential' (CHIP) was introduced for individuals carrying somatic mutations with variant allele frequencies (VAFs) $\geq$ 2% in the peripheral blood without any evidence of hematologic malignancy (6). Nevertheless, CH is considered a premalignant state, and progression to myeloid neoplasms including myelodysplastic syndromes (MDSs) and acute myeloid leukemia (AML) in CHIP carriers occurs in 0.5–1% per year (8). Clonal expansion of hematopoietic cells is mainly driven by a relatively small set of somatic mutations in genes known to be implicated in hematologic malignancies, most frequently observed in *DNMT3A*, *TET2*, *ASXL1*, *JAK2* and *TP53*, but also by mutations not affecting known driver genes (9). To date, large-scale sequencing studies from peripheral blood have suggested >150 genes associated with CH (10). Indeed, the same variants that are associated with CH can typically be observed in a broad range of malignancies and are therefore listed in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (10). CH arising from somatic mutations in hematopoietic stem/progenitor cells may occur in peripheral blood and a VAF $\leq$ 0.35 has been proposed as a threshold to distinguish between germline and somatic events (7).

Germline variants in a large number of CH-associated genes have also been implicated in Mendelian

Received: May 18, 2021. **Revised:** January 30, 2022. **Accepted:** January 31, 2022
© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

disorders, mainly neurodevelopmental disorders (NDDs) (7). However, the scope of this genetic overlap between somatic mutations in CH and variants in Mendelian Disorders has not been systematically elucidated yet. This genetic overlap has been addressed for the two most significant CH genes, *DNMT3A* and *ASXL1* (11,12). While CH-associated somatic mutations in these genes are frequently observed in the elderly, *de novo* germline variants in *ASXL1* and *DNMT3A* are the cause of the developmental disorders 'Bohring-Opitz syndrome' [MIM: #605039], 'Tatton-Brown-Rahman syndrome' [MIM: #615879] and 'Heyn-Sproul-Jackson syndrome' [MIM: #618724], respectively (13,14).

When interpreting germline variants in the context of Mendelian disorders, CH may be encountered in two different scenarios. First, somatic mutations in CH-associated genes have been found to be present in reference population databases such as the Exome Aggregation Consortium (ExAC) database and the Genome Aggregation Database (gnomAD) (11,15). Second, the increased utilization of next-generation sequencing (NGS), especially in adult patients, leads to an increased detection of CH-associated mosaicism and the subsequent pitfall of variant misclassification.

Recently, variants in *TET2*, one of the most common genes associated with age-related CH, were found to be enriched in individuals with neurodegenerative disorders (16). However, it was later shown that loss-of-function variants in *TET2* most likely arose from CH (17–19).

In this study, we compiled a comprehensive list of CH-associated genes that are also linked to Mendelian disorders when mutated in the germline. Moreover, we report a series of five individuals in which the presence of CH—either in population databases or in the analyzed exome dataset—confounded variant analysis and interpretation. Finally, we propose different approaches to overcome these challenges.

Results

CH-associated genes implicated in Mendelian disorders

Our literature search identified 72 genes, which are currently reliably associated with age-related CH (6,7,20,21) (Supplementary Material, Table S1). Next, we performed an query of Online Mendelian Inheritance in Men (OMIM) and revised published literature resulting in a subset of 58 genes in which germline variants are associated with Mendelian disorders (Table 2) (22). Forty-three genes were associated with autosomal dominant, six with X-linked and nine with autosomal recessive disorders, respectively. Three genes (*LZTR1*, *MPL, STAT5B*) were associated with both an autosomal recessive and an autosomal dominant mode of inheritance. In addition, we identified five genes (*BRCC3*, *PTPRD*, *SF3B1*, *SRSF2*, *U2AF1*) previously proposed as novel candidate genes for NDDs (23). Overall, NDDs ($n = 30$) represented the largest disease group with the majority of genes encoding for components of the epigenetic machinery (24,25). Table 2 gives an overview of all compiled CH-associated genes and the subset of CH-associated genes in which variants are associated with Mendelian disorders. Of note, from the genes most commonly implicated in CH, *DNMT3A* and *ASXL1* have been associated with autosomal dominant NDDs in contrast to *TET2* which is related to autosomal recessive immunodeficiency with lymphoma (26). Supplementary Material, Table S1 provides additional information on the level of evidence as a CHIP gene as well as information on the frequency of occurrence of CH per gene.

Cases in which CHIP confounded ES interpretation

Here, we present five cases in which CHIP confounded exome sequencing (ES) data interpretation. Three individuals (individuals 1–3) harbor rare pathogenic loss-of-function or missense variants in *ASXL1, PPM1D* or *DNMT3A* with VAFs between 10% and 35% in peripheral blood (Supplementary Material, Fig. S2). Other tissues were only available from individual 1 as individuals 2 and 3 had already deceased. Sanger sequencing in individual 1 did not identify the respective variants in DNA from fingernails and buccal swap (Supplementary Material, Fig. S3) thus confirming CH in individual 1.

In addition, we present two individuals (individuals 4 and 5) with de novo variants in *KMT2D* and *DNMT3A* with VAFs from 50% to 51% in peripheral blood (Supplementary Material, Fig. S2), respectively, which were listed in gnomAD each, however as somatic mosaicism. In individual 5, the variants was also present in DNA from other tissues (fingernails and buccal swap) with VAFs of ~50% in line with a germline origin (Supplementary Material, Fig. S4). Individual 4 was lost for follow-up. Pedigrees of all individuals are presented in Supplementary Material, Fig. S5.

Case descriptions

Table 1 gives an overview of the variants identified in individuals 1–5, including VAFs and presence in gnomAD v2.1.1. The visualization of the variants is shown in Supplementary Material, Figures S2–S4. Complete blood counts (CBCs) for individuals 1–3 are summarized in Supplementary Material, Table S2.

Individual 1 is a 59-year-old male who presented with a neuromuscular disorder characterized by a 3-year history of progressive muscle weakness predominantly at the lower legs with elevated creatine kinase. ES was performed and identified the heterozygous missense variant c.2245C > T, p.(Arg749Cys) in *DNMT3A* (NM_175629.2) (Supplementary Material, Fig. S2A) which had previously been associated with Tatton-Brown-Rahman disease, an overgrowth syndrome with intellectual disability, and was therefore considered as 'pathogenic' (13,27). We compared the patient's phenotype with previously published features and observed a great discrepancy, hence questioning a causal relationship. Moreover, a VAF

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

Table 1. Variants identified in individuals 1–5

Individual	Age at genetic testing	Gene	Variant	VAF (tissue)	Method	Depth	gnomAD Link	Allele Count gnomAD v2.1.1	VAF gnomAD v2.1.1	pLI o/e	Z-score o/e
1	59 years	*DNMT3A* (NM_175629.2)	c.2245C > T, p.(Arg749Cys) chr2:g.25463248G > A	23% (peripheral blood)	Exome sequencing	124	https://gnomad.broadinstitute.org/variant/chr2-25463248-G-A	6	10–45%	0 1.26 (1–1.58)	3.45 0.59 (0.54–0.65)
1				Not detectable (fingernails)	Sanger sequencing	–					
2	85 years	*ASXL1* (NM_015338.5)	c.1773C > G, p.(Tyr591*) chr20:g.31022288C > G	36% (peripheral blood)	Exome sequencing	159	https://gnomad.broadinstitute.org/variant/chr20-31022288-C-G	1	<30%	0 0.59 (0.44–0.79)	0.64 0.94 (0.88–0.99)
3	80 years	*PPM1D* (NM_003620.3)	c.1654C > T, p.(Arg552*) chr17:g.58740749C > T	24% (peripheral blood)	Exome sequencing	206	https://gnomad.broadinstitute.org/variant/chr17-58740749-C-T	3	10–30%	0 0.75 (0.53–1.1)	2.7 0.59 (0.52–0.66)
4	3 months	*KMT2D* (NM_003482.3)	c.15257G > A, p.(Arg5086Gln) chr12:g.49420492C > T	50% (peripheral blood)	Exome sequencing	292	https://gnomad.broadinstitute.org/variant/chr12-49420492-C-T	1	17%	1 0.07 (0.04–0.1)	3.73 0.81 (0.79–0.84)
5	23 years	*DNMT3A* (NM_175629.2)	c.994G > A, p.(Gly332Arg) chr2:g.25470480C > T	51% (peripheral blood)	Exome sequencing	315	https://gnomad.broadinstitute.org/variant/chr2-25470480-C-T	3	10–35%	0 1.26 (1–1.58)	3.45 0.59 (0.54–0.65)
5				~50% (fingernails)	Sanger sequencing	–					
5				~50% (buccal swap)	Sanger sequencing	–					

VAF = variant allele frequency; pLI = probability of loss of function intolerance; o/e = observed/expected ratio.

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

Table 2. List of CH genes and their association with Mendelian Disorders

Gene	Linked to Mendelian disorder	Mendelian Disorder when mutated in germline	Inheritance
ABL2	–		
AFF3	+	KINSSHIP syndrome	AD
APC	+	APC-associated polyposis conditions	AD
AR	+	Androgen insensitivity	XLR
ARID2	+	Coffin-Siris syndrome 6	AD
ASXL1	+	Bohring-Opitz syndrome	AD
ATE1	–		
ATM	+	Ataxia-telangiectasia/{Breast cancer, susceptibility to}	AR/AD
BCOR	+	Microphthalmia, syndromic 2	XLD
BCORL1	+	Shukla-Vernon syndrome	XLR
BRAF	+	Noonan syndrome 7/LEOPARD syndrome 3/cardiofaciocutaneous syndrome	AD
BRCC3		Candidate for NDD	AD
CALR	–		
CBL	+	Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia	AD
CDKN1B	+	Multiple endocrine neoplasia, type IV	AD
CHEK2	+	{Breast cancer, susceptibility to}	AD
CREBBP	+	Menke-Hennekam syndrome 1/Rubinstein-Taybi syndrome 1	AD
CTCF	+	Mental retardation, autosomal dominant 21	AD
CUX1	+	Global developmental delay with or without impaired intellectual development	AD
DNM2	+	Centronuclear myopathy 1/Charcot–Marie–Tooth disease/Lethal congenital contracture syndrome 5	AD/AR
DNMT3A	+	Tatton-Brown-Rahman syndrome/Heyn-Sproul-Jackson syndrome	AD
DNMT3B	+	Immunodeficiency-centromeric instability-facial anomalies syndrome 1	AR
ERCC2	+	Xeroderma pigmentosum, group D	AR
ERF	+	Chitayat syndrome/Craniosynostosis 4	AD
EZH2	+	Weaver syndrome	AD
FOXP1	+	Mental retardation with language impairment and with or without autistic features	AD
GNAS	+	Disorders of GNAS inactivation	AD
GNB1	+	Mental retardation, autosomal dominant 42	AD
IDH1	–		
IDH2	+	D-2-hydroxyglutaric aciduria 2	AR
JAK2	+	Thrombocythemia 3	AD
KDM5C	+	Mental retardation, X-linked, syndromic, Claes-Jensen type	XLR
KDM6A	+	Kabuki syndrome 2	XLD
KMT2C	+	Kleefstra syndrome 2	AD
KMT2D	+	Kabuki syndrome 1	AD
KRAS	+	Noonan syndrome 3/Cardiofaciocutaneous syndrome 2	AD
LZTR1	+	Noonan syndrome 10/Noonan syndrome 2	AD/AR
MDM4	+	?Bone marrow failure syndrome 6	AD
MGA	–		
MKL1	+	?Immunodeficiency 66	AR
MPL	+	Thrombocythemia 2/Thrombocytopenia, congenital amegakaryocytic	AD/AR
MYCN	+	Feingold syndrome 1	AD
MYD88	+	Pyogenic bacterial infections, recurrent, due to of MYD88 deficiency	AR
MYO5A	+	Griscelli syndrome, type 1	AR
NF1	+	Neurofibromatosis, type 1	AD
NOTCH1	+	Adams-Oliver syndrome 5/Aortic valve disease 1	AD
NRAS	+	Noonan syndrome 6	AD
PABPC1	–		
PPARG	+	Lipodystrophy, familial partial, type 3	AD
PPM1D	+	Jansen de Vries syndrome	AD
PTPN11	+	Noonan syndrome 1/LEOPARD syndrome 1	AD
PTPRD		Candidate for NDD	AD
RAD21	+	Cornelia de Lange syndrome 4	AD
RET	+	Multiple endocrine neoplasia II	AD
RUNX1	+	Platelet disorder, familial, with associated myeloid malignancy	AD
SDHAF2	+	Paragangliomas 2	AD
SETD2	+	Luscan-Lumish syndrome	AD
SETDB1	+	*SETDB1*-associated Disorder	AD
SF3B1		Candidate for NDD	AD
SH2B3	–		

(Continued)

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

Table 2. Continued

Gene	Linked to Mendelian disorder	Mendelian Disorder when mutated in germline	Inheritance
SRSF2		Candidate for NDD	AD
STAG2	+	Holoprosencephaly 13, X-linked/Mullegama-Klein-Martinez syndrome	XL
STAT3	+	Autoimmune disease, multisystem, infantile-onset, 1/Hyper-IgE recurrent infection syndrome	AD
STAT5B	+	Growth hormone insensitivity with immunodeficiency/Growth hormone insensitivity with short stature and mild immune dysregulation	AR/AD
SUZ12	+	Imagawa-Matsumoto syndrome	AD
TET2	+	Immunodeficiency 75	AR
TMEM127	+	{Pheochromocytoma, susceptibility to}	AD
TP53	+	Li-Fraumeni syndrome	AD
TP63	+	TP63-related disorders	AD
U2AF1		Candidate for NDD	AD
ZRSR2	−		

AD = autosomal dominant; AR = autosomal recessive; XL = X-linked; XLD = X-linked dominant; XLR = X-linked recessive.

of 23% was observed and the variant was not present in DNA from fingernails. In gnomAD, the variant is listed in heterozygous state in six individuals (age group 70–75 years) with VAFs ranging from 10% to 45% (28). The substitution p.(Arg749Cys) had previously been established as a recurrent somatic hotspot in hematologic malignancies (29,30), as well as pathogenic germline variant associated with 'Tatton-Brown-Rahman syndrome' (12,13,27). Overall, we interpreted the variant in *DNMT3A* as a somatic event related to CHIP. Normal blood counts of individual 1 confirmed no evidence of hematological disease (Supplementary Material, Table S2).

Individual 2 was an 85-year-old male patient, who presented with frontotemporal dementia. ES identified a heterozygous nonsense variant [c.1773C > G, p.(Tyr591*)] in *ASXL1* (NM_015338.5) (Supplementary Material, Fig. S2B). Loss-of-function germline variants in *ASXL1* cause 'Bohring-Opitz syndrome', an infantile-onset malformation syndrome (14). The variant has been reported as somatic mutation in MDS and AML. (30–34) It is listed in gnomAD once in a heterozygous state with a VAF < 30% in a female in the age group of 65–70 years, which is also in line with CH (28). The observed VAF of 36% in our patient did not allow us to unambiguously discriminate between somatic or germline origin. We therefore reassessed the clinical features of the index patient and did not observe any phenotypic overlap with individuals affected by 'Bohring-Opitz syndrome', a condition for which reduced penetrance has not yet been described. CBC of individual 2 were not indicative of hematologic disease (Supplementary Material, Table S2). Taken together, we interpreted the patient's variant in *ASXL1* most likely as mutational event in the context of CHIP and as not disease-causative for the patient's neurological disorder.

Individual 3 is an already deceased male, who underwent genetic testing at the age of 80 years because of Parkinson's disease. ES did not reveal any disease-associated variants; however, a filter for rare variants identified the heterozygous nonsense variant c.1654C > T, p.(Arg552*) located in the last exon of *PPM1D* (NM_003620.3) with a VAF of 24% (Supplementary Material, Fig. S2C). The variant was found in three individuals in gnomAD (no age group assigned) with VAFs ranging from 10% to 30% (28). Germline truncating variants in the last and penultimate exons of *PPM1D* have been described as the cause of 'Jansen de Vries syndrome' [MIM: #617450], an infantile-onset NDD. Yet, somatic truncating mutations also affecting the C-terminal part of the protein have been implicated in CH (4,10,35). The variant present in individual 3 has already been reported as germline variant in one individual diagnosed with 'Jansen de Vries syndrome' (36). Individual 3 did not show any phenotypic overlap with this condition. In view of the VAF of 24%, we considered the variant in *PPM1D* as a somatic event related to CHIP (normal blood counts excluding an underlying hematological disease are shown in Supplementary Material, Table S2).

Individual 4 presented with persistent hypoglycemia on the third day of life. Hypoglycemia was attributed to hyperinsulinism and adrenal insufficiency most likely secondary to adrenocorticotropic hormone deficiency. No dysmorphic features were noted, and a comprehensive diagnostic workup regarding acquired etiologies was unremarkable. Subsequent trio ES in the first year of life identified the heterozygous de novo missense variant c.15257G > A, p.(Arg5086Gln) in *KMT2D* (NM_003482.3) (Supplementary Material, Fig. S2D). The variant, which has not been associated with Kabuki syndrome 1 [MIM: #147920] so far, was listed once in gnomAD (https://gnomad.broadinstitute.org/variant/chr12-49420492-C-T). Subsequently, the substitution was classified as a variant of uncertain significance (VUS) according to the American College of Medical Genetics and Genomics (ACMG) criteria (PS2, PP3, BS2) (37). We therefore reassessed the phenotypic features of the index patient. The individual did not show the typical structural anomalies reminiscent of Kabuki syndrome and did not exhibit any striking dysmorphic features. However, the recognition of dysmorphic features associated with

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

Kabuki syndrome in the neonatal period is known to be challenging in some cases (38,39). As Kabuki syndrome has recently been shown to be an important differential diagnosis of neonatal hyperinsulinism accounting for 1% of patients (38), variant reassessment was performed. When visualizing the variant listed in gnomAD using the integrative genomics viewer, it was present in only 6/35 reads (VAF of 17%), indicating CH (28). Of note, the missense variant c.15257G > A, p.(Arg5086Gln) is also listed once as somatic mutation in tumor tissue (cervix carcinoma) in the COSMIC database (30). As the variant in gnomAD is most likely due to CH, we applied the ACMG criterion PM2 [i.e. 'Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genome Project, or Exome Aggregation Consortium'], thus allowing a reclassification to 'likely pathogenic' (40).

Individual 5, a 23-year-old female, presented with autistic spectrum disorder, dysmorphic facial features, mild intellectual disability and schizophrenia. Disease-causing variants could not be identified by proband-only (single) ES in the context of the patient's clinical presentation. Subsequent trio ES from peripheral blood, however, revealed the heterozygous de novo missense variant c.994G > A, p.(Gly332Arg) in *DNMT3A* (NM_175629.2) which was later confirmed in DNA from fingernails and buccal swap with a VAF of ∼50% (Supplementary Material, Fig. S3). The variant was classified as 'likely pathogenic' (ACMG criteria PS2, PM1, PP3). The variant had not been prioritized in the preceded analysis as (1) the variant was listed in three individuals in gnomAD (https://gnomad.broadinstitute.org/variant/chr2-25470480-C-T), (2) the variant was reported as 'uncertain significance' in ClinVar (Accession: SCV001437701.1) and (3) the provided clinical information was not specific for a *DNMT3A*-associated disorder. (41) Germline loss-of-function variants in *DNMT3A* have been described as the cause of 'Tatton-Brown-Rahman syndrome' [MIM #615879], whereas gain-of-function variants in *DNMT3A* have been linked to microcephalic dwarfism ('Heyn-Sproul-Jackson syndrome'[MIM #618724]) and paraganglioma (13,42,43). In turn, somatic mutations in *DNMT3A* are a frequent cause of CH and variants affecting the p.Gly332 residue have been identified in individuals with CH as somatic mutation (4,9,44,45). Indeed, the variant identified in this patient is listed in gnomAD with VAFs ranging from 10 to 35% in line with CH (28).

Discussion

CH is an age-related phenomenon characterized by a genetically distinct subpopulation of mature blood cells derived from acquired mutations in HSC (2). CH-associated mutations occur in reference population database as well as in ES of adult patients. As the genetic spectrum of CH overlaps with germline variants of Mendelian disorders, CH poses a relevant pitfall in variant interpretation. In this study, we illustrate this phenomenon by reporting five individuals in which the presence of CH biased variant interpretation.

First, the presence of CH in reference databases may confound germline variant interpretation in patients with suspected Mendelian disorders. One key prerequisite for the interpretation of sequence variants is that reference databases such as gnomAD do not contain data from individuals with severe childhood-onset diseases (46). In two of our reported cases (individuals 4 and 5), the presence of the *KMT2D* and *DNMT3A* variant in gnomAD was therefore thought to be inconsistent with pathogenicity, thus leading to the misclassification of the variant as VUSs. In contrast, individuals 1–3 demonstrate how CH, especially in adult patients, may complicate variant interpretation. These three patients harbored pathogenic variants in known disease genes that would confer early onset NDDs, if present in the germline. The three cases illustrate how CH, if overlooked, may result in misdiagnoses.

A literature review identified several publications in which the presence of CH might have impaired genetic analyses: (1) Recently, variants in *TET2*, one of the most common genes associated with age-related CH, were found to be enriched in individuals with neurodegenerative disorders (16). However, it was later shown that loss-of-function variants in *TET2* most likely arose from CH (17,18). (2) Not considering the possibility of CHin reference population databases may lead to inaccurate penetrance estimates for autosomal dominant disorders. One example is a study in which the presence of pathogenic loss-of-function variants in *ASXL1* in the ExAC database was interpreted as evidence for incomplete penetrance, but it seems more likely that such variants are of somatic nature (11,47).

Furthermore, we propose that CH in reference population database may also confound constraint metrics. For example, truncating variants in the last and penultimate exons of *PPM1D* are commonly somatically mutated in CH, but are also the cause of Jansen-de-Vries syndrome when mutated in the germline. Several *PPM1D* loss-of-function variants clustering in the last exons are currently listed in gnomAD raising the questions whether those variants are CH resulting in a falsely low calculated probability of loss-of-function intolerance. In theory, CH could even complicate disease gene discovery as well as the understanding of underlying pathomechanisms.

To date, there is still a lack of systematic data evaluating the role of CH-related genes that are also linked to single gene disorders when mutated in the germline. With the aim to partly fill this gap, we provide a list of 58 genes associated with Mendelian disorders that have also been described in the context of CH. Of note, NDD genes ($n = 30$) represented the largest disease group with the majority of genes implicated in epigenetic regulation. We acknowledge that the extent of this list is far from complete as the number of novel Mendelian disease genes as well as CH genes is constantly increasing.

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

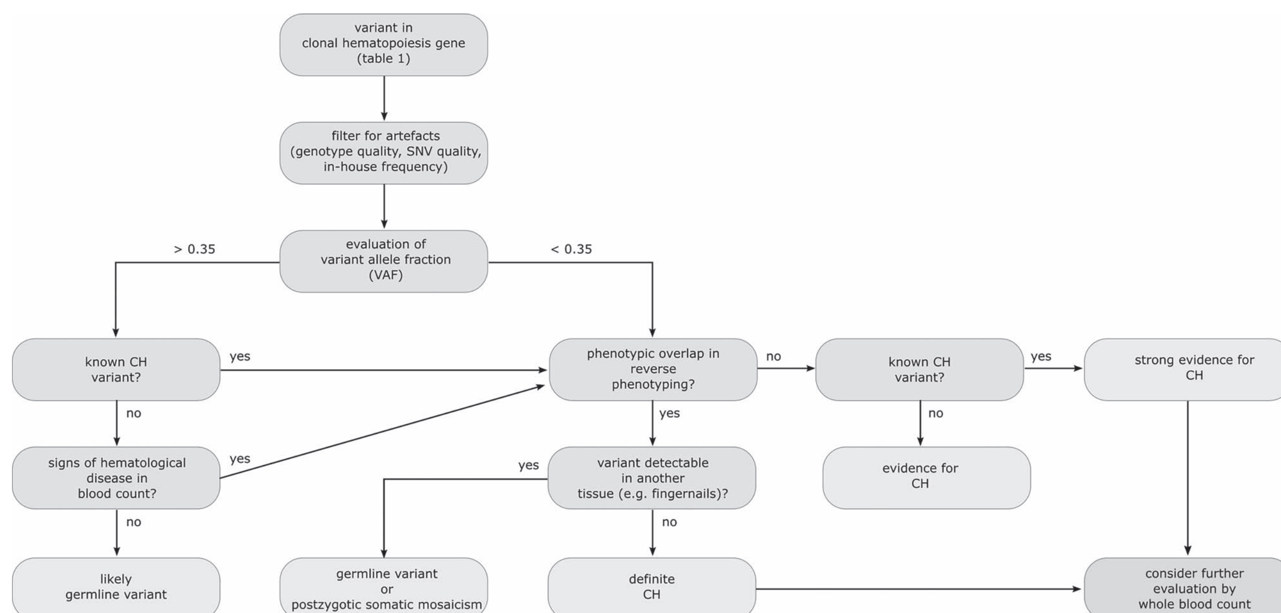

Figure 1. Flow chart for the distinction between germline and somatic variants in CH genes. After filtering of artifacts, the VAF gives the first indication of germline versus somatic origin. Variants with a VAF ≤ 35% as well as known somatic variants with a VAF > 35% are suggestive of somatic mosaicism and need to be further evaluated using 'reverse phenotyping' and databases such as COSMIC and ClinVar. Ultimately, testing DNA from another tissue enables the discrimination between somatic and germline origin.

Furthermore, we developed a workflow providing guidance in the interpretation of (germline) variants in CH-relevant genes (Fig. 1). In this context, it should be noted that increased age as well as a history for chemotherapy increases the risk for presence of CH. An important element in this workflow is the careful examination of the VAF. A VAF < 35% is considered as suggestive of CH but could also be indicative for a hematological malignancy. This cutoff is on the basis of the VAF distribution of confirmed inherited and de novo germline variants with most confirmed variants having a VAF > 35% (48). However, the quality metrics to detect mosaicism by ES depends on the read depth with decreasing positive predictive value and sensitivity associated with decreasing read depth. Thus, the suggestions presented here should be applied cautiously when interpreting variants with lower sequencing coverage. Another indicator for CH is the presence of the respective variant in a neoplastic tissue sample, for example in the COSMIC database.

In addition, we introduce reverse phenotyping as a powerful tool for the interpretation of potentially CH-associated somatic variants in adult patients, particularly in those cases with inconclusive VAFs. Ultimately, testing DNA from another tissue such as urine, mucosa or fingernails enables the confirmation of somatic versus germline origin as well as the discrimination between postzygotic mosaicism and CH. Furthermore, we suggest that CH-associated genes in general and somatic mutations in particular should be flagged in reference population databases as well as diagnostic sequencing centers to raise awareness of CH.

Finally, it is widely acknowledged that the detection of germline variants known as driver mutations in cancer tissue or CH supports the variant's pathogenicity. Therefore, we stress that the integration of CH and/or cancer NGS data into the analysis pipeline of germline variants has great potential.

Notably, CH is associated with an increased risk of developing myeloid malignancies. Moreover, the presence of somatic mutations at a high VAF in high-risk genes such as *ASXL1* might even be an indicator for an undiagnosed and treatable malignancy. This raises the question whether the detection of CH in clinical exome and genome sequencing should be reported as an incidental finding in individuals without any hematologic abnormalities. To date, recommendations are missing, necessitating prospective studies to provide guidance for the clinical management in such cases.

Overall, the presence of CH in either variant databases or ES datasets poses a pitfall, potentially leading to variant misclassification and erroneous conclusions regarding gene–disease associations.

Materials and Methods

Compilation of CH-associated genes

We compiled a list of recurrently mutated genes in CH and hematologic malignancies on the basis of a literature search followed by a systematic OMIM database query (https://www.omim.org/) and literature search using PubMed (https://pubmed.ncbi.nlm.nih.gov/) (both accessed on 1 October 2021) to identify the subset of genes in which germline variants are associated with monogenic disorders.

Exome sequencing

Exome sequencing (ES) was performed in agreement with the ethical standards of the responsible committee on

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

human experimentation (institutional and national) and with the Declaration of Helsinki as previously described, and was approved by the local ethics committee of the Technical University Munich (#5360/12S) (49). ES was performed using genomic DNA which was extracted from leukocytes or whole blood using a chemagic 360 Instrument (PerkinElmer, US). Sure Select Human All Exon 60 Mb V6 Kit (Agilent) was used for exome enrichment. Libraries were sequenced on an *Illumina NovaSeq6000* system (Illumina, San Diego, California, USA) and reads were aligned to the UCSC human reference assembly (hg19) with BWA v.0.7.5a. (50) On average > 97% of targeted regions were covered at least 20× and the median of the average coverages was 134× (interquartile range: 120–149) across all diagnostic samples (see Supplementary Material, Fig. S1). Single-nucleotide variants and small insertions and deletions were detected using both SAMtools v.0.1.19 and GATK 4.1. (51) Copy number variations were detected using ExomeDepth and Pindel (52,53). Variant prioritization was performed on the basis of an autosomal recessive (minor allele frequency (MAF) < 0.1%) and autosomal dominant (MAF < 0.01%) inheritance. Variants' pathogenicity was classified according to the ACMG guidelines (37). Variants with a VAF between 10% and 35% were considered somatic mosaicism.

Supplementary material

Supplementary Material is available at *HMG online*.

Acknowledgements

There are no financial conflicts of interest to disclose.

Conflict of Interest statement. None declared.

Authors' Contributions

Conceptualization: T.B., T.M., M.W. Data curation: T.B., M.W., R.B., D.C., S.A., M.D., J.DS., G.E., E.G., T.S., J.H., T.M. Formal analysis: T.B., R.B., M.W. Investigation: T.B., D.C., S.A., M.D., J.DS., J.H., M.W. Methodology: T.B., R.B., G.E., E.G. Software: R.B., T.S. Visualization: T.B., V.D., J.H., K.G., M.W. Writing—original draft: T.B., M.W. Writing—review and editing: V.D., J.H., R.B., M.D., J.DS, M.K., E.G., J.H., K.G., T.M.

References

1. Jan, M., Ebert, B.L. and Jaiswal, S. (2017) Clonal hematopoiesis. *Semin. Hematol.*, **54**, 43–50.
2. Sperling, A.S., Gibson, C.J. and Ebert, B.L. (2017) The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat. Rev. Cancer*, **17**, 5–19.
3. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A. *et al.* (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.*, **371**, 2488–2498.
4. Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M. *et al.* (2014) Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.*, **371**, 2477–2487.
5. Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A. *et al.* (2014) Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.*, **20**, 1472–1478.
6. Steensma, D.P., Bejar, R., Jaiswal, S., Lindsley, R.C., Sekeres, M.A., Hasserjian, R.P. and Ebert, B.L. (2015) Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*, **126**, 9–16.
7. Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S.H., Kiemeney, L., Veltman, J.A., Gilissen, C. and Hoischen, A. (2017) Ultra-sensitive sequencing identifies high prevalence of clonal Hematopoiesis-associated mutations throughout adult life. *Am. J. Hum. Genet.*, **101**, 50–64.
8. Silver, A.J. and Jaiswal, S. (2019) Clonal hematopoiesis: pre-cancer PLUS. *Adv. Cancer Res.*, **141**, 85–128.
9. Buscarlet, M., Provost, S., Zada, Y.F., Barhdadi, A., Bourgoin, V., Lepine, G., Mollica, L., Szuber, N., Dube, M.P. and Busque, L. (2017) DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*, **130**, 753–762.
10. Zink, F., Stacey, S.N., Norddahl, G.L., Frigge, M.L., Magnusson, O.T., Jonsdottir, I., Thorgeirsson, T.E., Sigurdsson, A., Gudjonsson, S.A., Gudmundsson, J. *et al.* (2017) Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*, **130**, 742–752.
11. Carlston, C.M., O'Donnell-Luria, A.H., Underhill, H.R., Cummings, B.B., Weisburd, B., Minikel, E.V., Birnbaum, D.P., Exome Aggregation, C., Tvrdik, T., MacArthur, D.G. *et al.* (2017) Pathogenic ASXL1 somatic variants in reference databases complicate germline variant interpretation for Bohring-Opitz syndrome. *Hum. Mutat.*, **38**, 517–523.
12. Shen, W., Heeley, J.M., Carlston, C.M., Acuna-Hidalgo, R., Nillesen, W.M., Dent, K.M., Douglas, G.V., Levine, K.L., Bayrak-Toydemir, P., Marcelis, C.L. *et al.* (2017) The spectrum of DNMT3A variants in Tatton-Brown-Rahman syndrome overlaps with that in hematologic malignancies. *Am. J. Med. Genet. A*, **173**, 3022–3028.
13. Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L. *et al.* (2014) Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.*, **46**, 385–388.
14. Hoischen, A., van Bon, B.W., Rodriguez-Santiago, B., Gilissen, C., Vissers, L.E., de Vries, P., Janssen, I., van Lier, B., Hastings, R., Smithson, S.F. *et al.* (2011) De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.*, **43**, 729–731.
15. Soussi, T., Leroy, B., Devir, M. and Rosenberg, S. (2019) High prevalence of cancer-associated TP53 variants in the gnomAD database: a word of caution concerning the use of variant filtering. *Hum. Mutat.*, **40**, 516–524.
16. Cochran, J.N., Geier, E.G., Bonham, L.W., Newberry, J.S., Amaral, M.D., Thompson, M.L., Lasseigne, B.N., Karydas, A.M., Roberson, E.D., Cooper, G.M. *et al.* (2020) Non-coding and loss-of-function coding variants in TET2 are associated with multiple neurodegenerative diseases. *Am. J. Hum. Genet.*, **106**, 632–645.
17. Holstege, H., Hulsman, M., van der Lee, S.J. and van den Akker, E.B. (2020) The role of age-related clonal Hematopoiesis in genetic sequencing studies. *Am. J. Hum. Genet.*, **107**, 575–576.
18. Cochran, J.N., Myers, R.M. and Yokoyama, J.S. (2020) Response to Holstege et al. *Am. J. Hum. Genet.*, **107**, 577–578.

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

19. Wagner, M., Lorenz, G., Volk, A.E., Brunet, T., Edbauer, D., Berutti, R., Zhao, C., Anderl-Straub, S., Bertram, L., Danek, A. *et al.* (2021) Clinico-genetic findings in 509 frontotemporal dementia patients. *Mol. Psychiatry*, **26**, 5824–5832.
20. Fuster, J.J. and Walsh, K. (2018) Somatic mutations and clonal Hematopoiesis: unexpected potential new drivers of age-related cardiovascular disease. *Circ. Res.*, **122**, 523–532.
21. Oriol, Pich, I.R.-S., Gonzalez-Perez, A. and Lopez-Bigas, N. (2020) Discovering the drivers of clonal hematopoiesis. *bioRxiv*, in press.
22. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. and McKusick, V.A. (2005) Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
23. Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y., Gallone, G., Lelieveld, S.H., Martin, H.C., McRae, J.F. *et al.* (2020) Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*, **586**, 757–762.
24. Bjornsson, H.T. (2015) The Mendelian disorders of the epigenetic machinery. *Genome Res.*, **25**, 1473–1481.
25. Hartmann, L. and Metzeler, K.H. (2019) Clonal hematopoiesis and preleukemia-genetics, biology, and clinical implications. *Genes Chromosomes Cancer*, **58**, 828–838.
26. Stremenova Spegarova, J., Lawless, D., Mohamad, S.M.B., Engelhardt, K.R., Doody, G., Shrimpton, J., Rensing-Ehl, A., Ehl, S., Rieux-Laucat, F., Cargo, C. *et al.* (2020) Germline TET2 loss of function causes childhood immunodeficiency and lymphoma. *Blood*, **136**, 1055–1066.
27. Tatton-Brown, K., Loveday, C., Yost, S., Clarke, M., Ramsay, E., Zachariou, A., Elliott, A., Wylie, H., Ardissone, A., Rittinger, O. *et al.* (2017) Mutations in epigenetic regulation genes are a major cause of overgrowth with intellectual disability. *Am. J. Hum. Genet.*, **100**, 725–736.
28. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
29. Brecqueville, M., Cervera, N., Gelsi-Boyer, V., Murati, A., Adelaide, J., Chaffanet, M., Rey, J., Vey, N., Mozziconacci, M.J. and Birnbaum, D. (2011) Rare mutations in DNMT3A in myeloproliferative neoplasms and myelodysplastic syndromes. *Blood Cancer J*, **1**, e18.
30. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E. *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–D947.
31. Chou, W.C., Huang, H.H., Hou, H.A., Chen, C.Y., Tang, J.L., Yao, M., Tsay, W., Ko, B.S., Wu, S.J., Huang, S.Y. *et al.* (2010) Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations. *Blood*, **116**, 4086–4094.
32. Pratcorona, M., Abbas, S., Sanders, M.A., Koenders, J.E., Kavelaars, F.G., Erpelinck-Verschueren, C.A., Zeilemakers, A., Lowenberg, B. and Valk, P.J. (2012) Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value. *Haematologica*, **97**, 388–392.
33. Lundberg, P., Karow, A., Nienhold, R., Looser, R., Hao-Shen, H., Nissen, I., Girsberger, S., Lehmann, T., Passweg, J., Stern, M. *et al.* (2014) Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood*, **123**, 2220–2228.
34. Bartels, S., Faisal, M., Busche, G., Schlue, J., Hasemeier, B., Schipper, E., Vogtmann, J., Westphal, L., Lehmann, U. and Kreipe, H. (2020) Mutations associated with age-related clonal hematopoiesis in PMF patients with rapid progression to myelofibrosis. *Leukemia*, **34**, 1364–1372.
35. Kahn, J.D., Miller, P.G., Silver, A.J., Sellar, R.S., Bhatt, S., Gibson, C., McConkey, M., Adams, D., Mar, B., Mertins, P. *et al.* (2018) PPM1D-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. *Blood*, **132**, 1095–1105.
36. Jansen, S., Geuer, S., Pfundt, R., Brough, R., Ghongane, P., Herkert, J.C., Marco, E.J., Willemsen, M.H., Kleefstra, T., Hannibal, M. *et al.* (2017) De novo truncating mutations in the last and penultimate exons of PPM1D cause an intellectual disability syndrome. *Am. J. Hum. Genet.*, **100**, 650–658.
37. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, **17**, 405–424.
38. Yap, K.L., Johnson, A.E.K., Fischer, D., Kandikatla, P., Deml, J., Nelakuditi, V., Halbach, S., Jeha, G.S., Burrage, L.C., Bodamer, O. *et al.* (2019) Congenital hyperinsulinism as the presenting feature of kabuki syndrome: clinical and molecular characterization of 9 affected individuals. *Genet Med*, **21**, 233–242.
39. Bogershausen, N., Gatinois, V., Riehmer, V., Kayserili, H., Becker, J., Thoenes, M., Simsek-Kiper, P.O., Barat-Houari, M., Elcioglu, N.H., Wieczorek, D. *et al.* (2016) Mutation update for kabuki syndrome genes KMT2D and KDM6A and further delineation of X-linked kabuki syndrome subtype 2. *Hum. Mutat.*, **37**, 847–864.
40. Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C. *et al.* (2010) Exome sequencing identifies MLL2 mutations as a cause of kabuki syndrome. *Nat. Genet.*, **42**, 790–793.
41. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
42. Heyn, P., Logan, C.V., Fluteau, A., Challis, R.C., Auchynnikava, T., Martin, C.A., Marsh, J.A., Taglini, F., Kilanowski, F., Parry, D.A. *et al.* (2019) Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.*, **51**, 96–105.
43. Remacha, L., Curras-Freixes, M., Torres-Ruiz, R., Schiavi, F., Torres-Perez, R., Calsina, B., Leton, R., Comino-Mendez, I., Roldan-Romero, J.M., Montero-Conde, C. *et al.* (2018) Gain-of-function mutations in DNMT3A in patients with paraganglioma. *Genet Med*, **20**, 1644–1651.
44. Young, A.L., Tong, R.S., Birmann, B.M. and Druley, T.E. (2019) Clonal hematopoiesis and risk of acute myeloid leukemia. *Haematologica*, **104**, 2410–2417.
45. Watson, C.J., Papula, A.L., Poon, G.Y.P., Wong, W.H., Young, A.L., Druley, T.E., Fisher, D.S. and Blundell, J.R. (2020) The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*, **367**, 1449–1454.
46. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A and Birnbaum, D.P. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**:434–443.
47. Ropers, H.H. and Wienker, T. (2015) Penetrance of pathogenic mutations in haploinsufficient genes for intellectual disability and related disorders. *Eur. J. Med. Genet.*, **58**, 715–718.

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024

48. Krupp, D.R., Barnard, R.A., Duffourd, Y., Evans, S.A., Mulqueen, R.M., Bernier, R., Rivière, J.B., Fombonne, E. and O'Roak, B.J. (2017) Exonic mosaic mutations contribute risk for autism Spectrum disorder. *Am. J. Hum. Genet.*, **101**, 369–390.
49. Brunet, T., Jech, R., Brugger, M., Kovacs, R., Alhaddad, B., Leszinski, G., Riedhammer, K.M., Westphal, D.S., Mahle, I., Mayerhanser, K. *et al.* (2021) De novo variants in neurodevelopmental disorders-experiences from a tertiary care center. *Clin. Genet.*, **100**, 14–28.
50. Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, **26**, 589–595.
51. Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.
52. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, **25**, 2865–2871.
53. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J. *et al.* (2012) A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, **28**, 2747–2754.

Downloaded from https://academic.oup.com/hmg/article/31/14/2386/6530529 by guest on 10 June 2024