Clonal Hematopoiesis: genetic and phenotypic associations

Chao Tian, Pierre Fontanillas, Joyce Y Tung, David A Hinds, the 23andMe Research Team

[1]23andMe, Inc., Mountain View, CA.

Introduction

Age-related clonal hematopoiesis is a common condition that is associated with an increase in hematologic cancer and all-cause mortality. Studies have shown that 10% of individuals over the age of 65 years have recurrent somatic mutations and large-scale somatic events[1]. We analyzed SNP array data with custom probes for previously characterized single-nucleotide somatic mutations and were able to detect somatic point mutations at *DNMT3A* r882c/r882h, *IDH2* r140q, *JAK2* v617f and *SF3B1* k700e. SNP array data are used routinely to scan for large chromosomal anomalies genome-wide, and it is also useful for detecting large-scale chromosomal mosaicism where a relatively high frequency of mutated cells are present (>5-10%) [2,3]. We were able to use our custom SNP array to detect such large-scale mosaic structure alterations in the autosomes of 0.5% of study subjects who were unselected for hematologic phenotypes.

Methods

Study Cohort

The study cohort is composed of ~1.6 million research participants of 23andMe, Inc., who were all consented under an IRB-approved protocol[4] and were genotyped on a fully custom array from Illumina.

Single-nucleotide Somatic Mutations

The fluorescence intensity data imported from Illumina GenomeStudio software are transformed to intensity R = signal.a + signal.b and minor allelic burden signal.a/R, where signal.a and signal.b represent normalized signals from alleles A and B for a particular locus. The detectable single-nucleotide somatic locus has an unusual intensity distribution (**Fig 1**), with a smear of individuals showing evidence for non-reference allelic burden, but no clear clusters. Over 99% of study subjects are homozygous non-carriers (blue dots in **Fig1**) and somatic carriers are defined as those who have mutation burdens at least 10 standard deviation (sd) away from the mean burden (red dots in **Fig 1**). We required that the carrier status is significantly associated (OR>3 and P value <1e-10) with old age (age >=65).

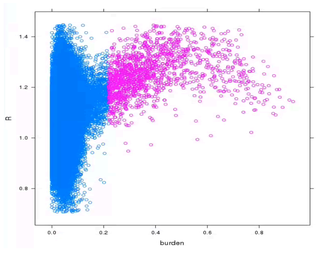

Figure 1. Intensity distribution of a typical somatic point mutation.

Somatic Structural Alterations

The log R ratio (LRR) of intensities and b-allelic frequency (BAF) are derived from the estimated genotype-specific cluster centers[3]. The data is then analyzed with the mosaic alteration detection (MAD) algorithm implemented in R genomic Alteration detection Analysis (R-GADA) software[5] with modifications.

Genetic and Phenotypic Associations

To understand risk factors that contributed to having a detectable somatic mutation, we performed a genome-wide and phenome-wide association scan that included age, sex and ancestry as covariates for the carrier status of the four detectable somatic mutations.

Mutation	cytoband	snp	position	pvalue	OR	gene
JAK2_v617f	9p24.1	rs12349785	5076613	1.20E-109	0.396	JAK2
JAK2_v617f	5p15.33	rs2853677	1287194	5.90E-39	1.712	TERT
JAK2_v617f	7q32.3	rs7803075	130742066	4.00E-18	0.686	klf14---[]---MKLN1
JAK2_v617f	4q24	rs199741557	105864530	2.70E-17	0.449	TET2
JAK2_v617f	3q25.33	rs201009932	160086719	1.00E-16	3.387	**[RP11-432B6.3,IFT80]**
JAK2_v617f	6p21.31	rs77136196	34247047	4.90E-09	1.729	**C6orf1--[]NUDT3**
JAK2_v617f	2q37.3	rs4675939	242448518	2.40E-08	0.762	**STK25[]--BOK**
JAK2_v617f	9q21.32	rs145543052	85632307	4.90E-08	1.727	**RASEF**
DNMT3A r882h/c	5p15.33	rs2853677	1287194	2.50E-23	1.318	**TERT**
	3q25.33	rs7356111	160077846	4.30E-07	1.152	**[RP11-432B6.3,IFT80]**
IDH2_r140q	2q37.3	rs62192032	242699491	1.10E-09	0.345	**D2HGDH**
SF3B1_k700e	6q21	rs1546723	109625879	7.60E-10	1.547	**CEP57L1---[]--CD164**

Table 1: Top associations from GWASes

Results

Single-nucleotide Somatic Mutations

The presence of the somatic single-nucleotide mutations was detectable at *DNMT3A* r882c/r882h, *IDH2* r140q, *JAK2* v617f, *SF3B1* k700e, and some other locations that under further verification. The frequency of carrying any of the four somatic mutations increased from <0.1% in young individuals (age<= 45) to 1.7% in old individuals (age>85) (**Fig 2**).

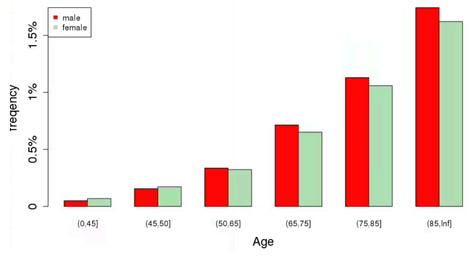

Figure 2. Frequency of the detectable somatic mutations at *DNMT3A* r882c/r882h, *IDH2* r140q, *JAK2* v617f and *SF3B1* k700e by age.

Somatic Structural Alterations

We detected 8350 (0.5%) large mosaic structural alterations in ~1.6 million individuals (~2093 deletions, ~1921 duplication, ~4336 uniparental disomy (UPD)), The frequency increased from 0.2% in young individual (age<=45) to over 2% in old individuals (age>85) (**Fig 3**). The distribution across genome is showed in **Fig.5**.

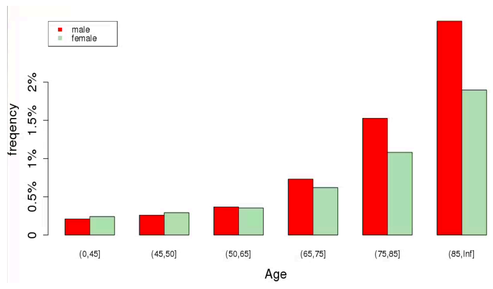

Figure 3. Frequency of the detectable somatic structural variations by age.

Genetic Associations

Multiple germline variants (**Table 1**) were associated with *JAK2* v617f clonal hematopoiesis, some of which have been described previously[6]. Novel associations were identified with *IFT80* (rs201009932, P=1.0e-16, OR=3.387), splice site variants in *HMGA1* (rs77136196, P=4.9e-9, OR=1.73), eQTL in *STK25* (rs4675939, P=2.4e-8, OR=0.762) and *RASEF* (rs145543052, P=4.9e-8, OR=1.727). *HMGA1* chromatin biding proteins are potent oncogenes and overexpressed in acute leukemia and have been shown to be drivers of clonal expansion in myeloid disease in humans[7]. *RASEF* is involved in cell-growth mechanisms and may play role as an oncoprotein[8]. *TERT* is also significantly associated with having *DNMT3A* r882c/r882h mutations. Variants in *D2HGDH* (rs76685039, P=6.3e-16, OR=1.3) are found to be significantly associated with having *IDH2 r140q* somatic mutations. *H2HGDH* was previously showed to regulate alpha-ketoglutarate levels and dioxygenase function by modulating *IDH2*. A region in 6q21 indexed by rs1832777 (P=2.7e-11, OR=1.5), previously reported to be associated with red blood traits, is significantly associated with having the *SF3B1* k700e mutations. These data indicate that the germline variants predispose individuals to clonal hematopoiesis, which may later cause an overt neoplasm.

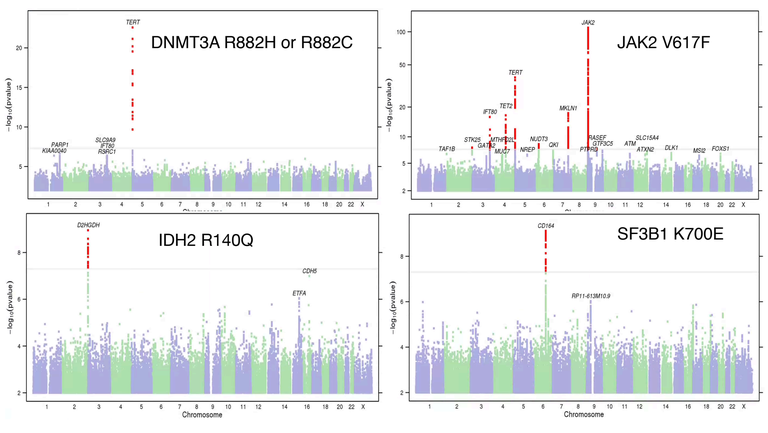

Figure 4. Manhattan plot of GWAS results.

Results (cont.)

We observed that the presence of somatic point mutations was often accompanied by large-scale chromosomal alterations (**Fig 6**), which is consistent with a previous report[6] that people who carry the *JAK2* v617f mutation are more likely to also have the UPD across chromosome 9p. People who have high burden of *DNMT3A* r882c/r882h mutations tend to have deletions across the DNMT3A gene location.

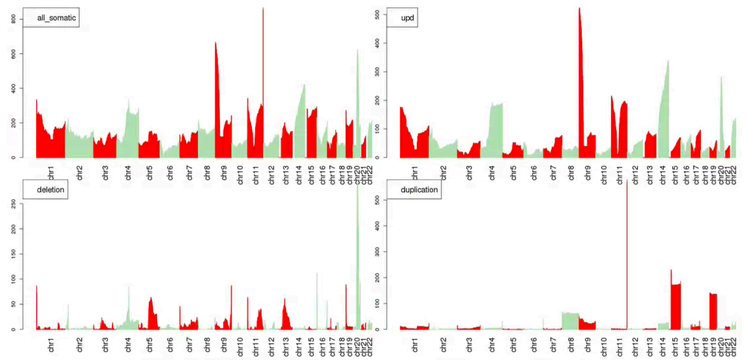

Figure 5. Genome distribution of somatic structural alterations (UPD, deletion and duplication).

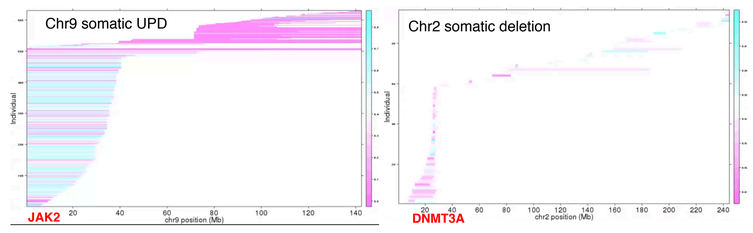

Figure 6. Co-localization of single-nucleotide somatic mutation and somatic structural alterations. The heat map intensity is showing for the mutated allelic burden at JAK2 v617f and DMNT3A r882c/r882h, separately.

Phenotype associations

The presence of somatic mutations was also analyzed for association with self-reported phenotypes, adjusting for age, gender and ancestry. The sample sizes for each individual mutation are still too small to find any significant phenotypic indications, except for *JAK2* v617f which has relatively larger number of carriers in our database. The *JAK2* v617f carriers are less likely to have high cholesterol (P = 4.3e-25, OR = 0.66) and seasonal allergies (P=3.68e-12, OR=0.64), and report higher instances of leukemia (P=1.16e-14, OR=4.15) and blood clots (P=3.8e-16, OR=3).

Discussion

Our work suggested important roles of specific germline mutations underlying the clonal hematopoiesis at several loci. We are extending the analysis to many other commonly mutated genes including *TET2*, *ASXL1* and *TP53*. Our algorithm for detecting somatic structural alterations is not perfect and still requires improvement. The numbers of somatic structural alterations are expected to be higher in the general populations. Also with the growth of our database, we will have higher sensitivity to detect any phenotypic indications of having certain somatic mutations, such as to determine the relationship between detectable mosaic abnormalities and cancers, which has been actively studied recently mostly via next-generation sequencing technology.

Acknowledgments

We thank 23andMe customers who consented to participate in research for enabling this study. We also thank employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References

1. 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.
2. Laurie CC, Laurie CA, Rice K, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012;44(6):642-50.
3. Jacobs KB, Yeager M, Zhou W, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012;44(6):651-8.
4. Eriksson N, Mountain J et al., Web-Based, participant-driven studies yield novel associations for common traits. Plos Genetics. 2010; 6(6).
5. González JR, Rodríguez-santiago B, Cáceres A, et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. BMC Bioinformatics. 2011;12:166.
6. 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):1121-8.
7. Resar L., Moliterno AR *et al.*, High Mobility Group A1/2 Chromatin Remodeling Proteins Associate with Polycythemia Vera Transformation to Acute Leukemia in Humans and a JAK2 V617F Transgenic mouse model. Blood 2016 128:1958.
8. Oshita H, Nishino R, Takano A, et al. RASEF is a novel diagnostic biomarker and a therapeutic target for lung cancer. Mol Cancer Res. 2013;11(8):937-51.