Supplemental Information

Article Title: Somatic HLA Mutations Expose the Role of Class I-Mediated Autoimmunity in Aplastic Anemia and its Clonal Complications

Authors: Daria V. Babushok[1,2], Jamie L. Duke[3], Hongbo M. Xie[4], Natasha Stanley[2], Jamie Atienza[2], Nieves Perdigones[2], Peter Nicholas[2], Deborah Ferriola[3], Yimei Li[5], Hugh Huang[2], Wenda Ye[2], Jennifer J.D. Morrissette[6], Jane Kearns[6], David L. Porter[1], Gregory M. Podsakoff[7], Laurence C. Eisenlohr[3,6], Jaclyn A. Biegel[8,9], Stella T. Chou[10], Dimitrios S. Monos[3], Monica Bessler[1,2], Timothy S. Olson[2, 11]

[1] Division of Hematology-Oncology, Department of Medicine, Hospital of the University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104
[2] Comprehensive Bone Marrow Failure Center, Department of Pediatrics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA 19104
[3] Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA 19104
[4] Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, 3535 Market Street, Philadelphia, PA 19104
[5] Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, 3501 Civic Center Boulevard, Philadelphia, PA 19104
[6] Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania, United States of America
[7] Office of Clinical and Translational Research, Children's Hospital of Philadelphia, 3501 Civic Center Boulevard, Philadelphia, PA 19104
[8] Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA 19104
[9] Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles and Keck School of Medicine, University of Southern California, 4650 W. Sunset Boulevard, Los Angeles, CA 90033
[10] Division of Hematology, Department of Pediatrics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA 19104
[11] Division of Oncology, Department of Pediatrics, Children's Hospital of Philadelphia, 3615 Civic Center Boulevard, Philadelphia, PA 19104

Corresponding Author:

Daria Babushok, M.D., Ph.D.

Division of Hematology-Oncology, Hospital of the University of Pennsylvania

3615 Civic Center Boulevard, ARC 302, Philadelphia PA 19104

Email: daria.babushok@uphs.upenn.edu

Phone: 215-426-9888, Fax: 267-426-9892

TABLE OF CONTENTS

SUPPLEMENTAL METHODS **4**

SUPPLEMENTAL DISCUSSION **10**

SUPPLEMENTAL TABLES **12**

Supplemental Table S1. Baseline clinical characteristics of patients with acquired aplastic anemia.. 12

Supplemental Table S2. HLA allele loss in patients with acquired aplastic anemia.. 13

Supplemental Table S3. Race- and ethnicity-stratified frequency of HLA risk alleles in 66 aplastic anemia patients 15

Supplemental Table S4. *HLA-B**14:02 and *HLA-B**40:02 are enriched in aAA patients compared to race-matched controls 15

Supplemental Table S5. Frequencies and association analysis of HLA class I alleles in aAA Patients (Caucasian, n=42) 16

Supplemental Table S6. Haplotypes Containing *HLA-B**14:02 Allele in USA Caucasian Populations 17

Supplemental Table S7. HLA A, B, C and DRB1 alleles in 66 patients with acquired aplastic anemia 18

Supplemental Table S8. Clinical characteristics of patients with aplastic anemia stratified by presence or absence of HLA risk alleles. 20

Supplemental Table S9. Clonal Hematopoiesis in 45 Aplastic Anemia Patients Analyzed by Comparative WES..... 21

Supplemental Table S10. Comprehensive analysis of clonal hematopoiesis in 45 patients with aplastic anemia 22

Supplemental Table S11. Longitudinal follow-up of clonal hematopoiesis 25

Supplemental Table S12. Hematologic malignancy targeted NGS gene pane..... 28

SUPPLEMENTAL FIGURES **29**

Supplemental Figure S1. Validation of putative somatic mutations detected by comparative Whole Exome Sequencing..... 34

SUPPLEMENTAL FIGURE S2. AGE DISTRIBUTION OF THE 66-PATIENT ACQUIRED APLASTIC ANEMIA COHORT.. 29

SUPPLEMENTAL FIGURE S3. CLONAL ARCHITECTURE ANALYSIS OF A PATIENT WITH THREE LOSS-OF-FUNCTION SOMATIC MUTATIONS IN THE *HLA-B**40:02 ALLELE DEMONSTRATING PRESENCE OF OTHER SOMATIC MUTATIONS WITHIN HLA CLONES. 30

SUPPLEMENTAL FIGURE S4. CLONAL ARCHITECTURE ANALYSIS OF A PATIENT WITH THREE 6P CN-LOH CLONES DEMONSTRATING SUBCLONAL SOMATIC MUTATIONS 32

SUPPLEMENTAL FIGURE S5. VALIDATION OF LOSS-OF-FUNCTION SOMATIC HLA MUTATIONS DETECTED BY TARGETED NEXT-GENERATION SEQUENCING..... 35

SUPPLEMENTAL REFERENCES 37

Methods

Patients and Study Oversight

The Penn-CHOP Bone Marrow Failure Syndrome (BMFS) cohort is an open prospective/retrospective cohort for the study of molecular mechanisms of BMFS, approved by the Institutional Review Boards of Children's Hospital of Philadelphia and of the University of Pennsylvania. All pediatric and adult patients evaluated for bone marrow failure were invited to participate. Written informed consent from all study participants or their legal guardians was obtained prior to study participation in accordance with the Declaration of Helsinki. The diagnosis of aAA was established according to the International Study of Agranulocytosis and Aplastic Anemia[1], and required exclusion of congenital BMFS such as Fanconi Anemia and Dyskeratosis Congenita, and other conditions mimicking aAA . The 66 aAA patients analyzed in this study included 16 aAA patients from the previously published WES cohort[2]. In all patients, the diagnosis of aAA was made prospectively prior to study enrollment. In accordance with the clinical practice at our center, all patients with a diagnosis of aAA, including those who achieved complete remission, continued hematology follow-up. Patients with an immunophenotypic PNH clone of 50% or greater, as measured by flow cytometry of granulocytes, were categorized as having hemolytic PNH, and were analyzed separately. The diagnosis of myelodysplastic syndrome (MDS) was established based on the 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia[3].

A separate cohort of MDS patients without pre-existing aAA were used as a comparator population for association analyses; these were adult patients who underwent high resolution HLA typing as a part of an unrelated donor transplant evaluation for a diagnosis of MDS at the Hospital of the University of Pennsylvania. The healthy control population was a combination of healthy volunteers and of patients' relatives, previously HLA typed by next-generation

sequencing as a part of a clinical evaluation to be a bone marrow donor. All patients and controls were enrolled with the approval of the Institutional Review Board. Race and ethnicity were self-reported. In accordance with the American Academy of Pediatrics Council on Child and Adolescent Health, pediatric-onset aAA was defined as a diagnosis of aAA under the age of 22[4].

Hematopathology, Cytogenetics, and SNP-A Analysis

Bone marrow histology was evaluated by a clinical hematopathologist in a blinded fashion, as patients were entered into the study only after completion of the diagnostic review. According to the department policy, all controversial cases were subject to a clinical consensus conference. Cytogenetic analysis and fluorescence in situ hybridization (FISH) were performed according to standard methods. Illumina Infinium SNP-A genotyping of bone marrow aspirate or peripheral blood DNA was performed at the CHOP Center for Applied Genomics using one of the Illumina beadchips (Quad610, Omni1 Quad, CytoSNP 850, Human Omni Express-24, Human Core, Human Omni Express-8-EXOME) according to the manufacturer's protocol, as previously described[5]. Arrays were analyzed for regions of acquired CN-LOH in GenomeStudio (Illumina, Inc., San Diego, CA), which allows direct visualization of B-Allele Frequency and log R ratio. The characterization of specific 6p CN-LOH breakpoints and analysis of missing haplotypes are presented elsewhere[6]. SNP-A data have been deposited into Gene Expression Omnibus (GEO) (accession GSE48483 and GSE48484).

WES

Comparative WES of paired bone marrow and skin fibroblast DNA was performed as previously described[2]. Briefly, WES of DNA, extracted using Qiagen DNeasy Blood & Tissue Kit (Qiagen Inc., Valencia, CA), was performed at the BGI@CHOP High Throughput Sequencing Center to 150X average depth on the Illumina HiSeq 2500 platform. Exome libraries were constructed

with Agilent SureSelect All Exon V4 + UTRs kit (Agilent Technologies, Santa Clara, CA). Somatic variant calling on bone marrow-skin biopsy pairs was performed with VarScan2[7], with parameters *–min-coverage 4,–min-var-freq 0.08,–p-value 0.05,–strand-filter 1–min-avg-qual 20*. Filtering and annotation of somatic mutations was performed using SNP & Variation Suite v8.0 (Golden Helix, Inc., Bozeman, MT). All mutations were manually curated in Integrative Genomics Viewer[8]. Somatic mutations identified by WES were validated with bi-directional Sanger sequencing of paired skin and bone marrow DNA, and were confirmed as somatic if they were present in the bone marrow and absent in the skin DNA (Figure S1). All chromosome coordinates were based on hg19 (NCBI build 37). WES for 16 patients was previously reported[2].

Targeted HLA NGS and NGS of malignancy-associated genes

Paired-end next-generation sequencing (2×251 bp) of the Human Leukocyte Antigen (HLA) region was performed on the Illumina MiSeq platform at >10,000X average depth as previously described[9]. Sequence alignment and HLA typing were performed using NGSengine (GenDx Utrecht, Netherlands), Target (Omixon, Budapest, Hungary) and Twin (Omixon) software in the CHOP CLIA-approved Immunogenetics Laboratory. Variant calling was performed using the GATK Haplotype Caller[10], using the sequence of each patient's HLA type as published in the IPD-IMGT/HLA Database[11] as the patient's reference sequence, with the following GATK parameters -stand_call_conf 20.0 -ploidy 5 -stand_emit_conf 10.0. All variants were manually curated in Integrative Genomics Viewer[8] and in the NGSengine software (GenDx). Eleven of thirteen identified HLA mutations were independently validated by WES of the patients' bone marrow DNA and/or targeted Sanger sequencing of single cell clones (Supplemental Figure S5). Two mutations that were not validated by WES were of lower allelic frequency and fell within regions with low depth of coverage in WES (read depths of 25 and 42 reads). Somatic status of mutations was confirmed by WES of paired skin fibroblast DNA for 5 of 6 patients, as permitted by skin sample availability.

Targeted sequencing of 67 hematologic malignancy-associated genes (Table S12) was performed by the Penn Clinical CytoGenomics Laboratory by enriching targeted regions using the Illumina Truseq Amplicon Assay, followed by paired-end, multiplex sequencing on the Illumina MiSeq platform (Illumina, Inc., San Diego, CA).

Sensitivity of HLA mutation detection

The median read depths for *HLA-A*, *-B* and *-C* gene sequencing in 66 aAA patients were 20,370 (range 7,401-81,268), 19,991 (range 11,523-84,784) and 16,388 (range 7,412-101,653), respectively. Twenty five patients with hemolytic PNH were also analyzed by deep HLA targeted sequencing. The NGS read depth for the 25 PNH patients was similar to aAA patients, with median depths of 19,281 (11,843-3126), 18,638 (13,410-30,102) and 15,190 (range 10,662-27,144) for the *HLA-A*, *-B* and *-C* gene, respectively. Because normal controls were a combination of patients' relatives and healthy volunteers, who were previously HLA-typed by NGS at the CHOP Immunogenetics Clinical laboratory, the sequencing depth for normal controls was lower, with median read depths of 3,578 (range 1,506-18,430), 3,861 (range 253-19,255) and 3,899 (range 193-22,546) for HLA-A, -B, and –C genes, respectively. To address whether HLA sequencing depth for normal controls was adequate to detect somatic mutations, we performed *in silico* simulation of mutation detection at lower depths by random downsampling of NGS reads. We then tested the sensitivity of our mutation detection pipeline to detect the 13 somatic mutations in aAA patients at lower read depths. Ten replicates at each read depth were analyzed. Mutations of >5% allele frequency were consistently detected at all read depths ≥150 reads. For smaller allele frequencies, detection rate was not linearly related to the read depth. Simulated detection of low frequency variants at read depths below 3,500 reads allowed for a 46-62% chance of detection of mutation of 3.5-3.9% allele frequency, and a 71%

chance of detection of mutation at 4.8% allele frequency. Mutations below 2% were not reliably detected.

Cell Sorting

Lymphoid- and myeloid-enriched cell fractions were obtained from peripheral blood by immunomagnetic selection for CD3 and CD19-positive cells using anti-CD3 and anti-CD19 microbeads (Miltenyi Biotec, San Diego, CA). The cells were separated using “LS” columns into the myeloid-enriched (CD3- and CD19-depleted) and lymphocyte (CD3+ and CD19+ selected) fractions, which were used for DNA extraction.

Colony Assays

Peripheral blood mononuclear cells from patients with multiple identified mutations were cultured in Methocult H4434 Classic Medium (Stem Cell Technologies, Vancouver, Canada) for 14 days at 37°C and 5% CO2. Individual colonies were picked into 100 µL of sterile water and boiled at 99°C for 10 minutes to prepare crude lysates suitable for subsequent PCR and Sanger sequencing (Figures S1 and S2). Primer sequences are available upon request.

HLA Allele Association and Other Statistical Analyses

HLA allele frequencies for ethnically-stratified populations were obtained from The Allele Frequency Net Database, a database and repository for immune gene frequencies in worldwide populations[12]. The frequency of aAA patients with HLA alleles of interest was analyzed for the largest ethnic group in our cohort (White, n=42 patients) and compared to the corresponding frequency in the USA NMDP European Caucasian population (n=1,242,890), with the number of individuals carrying the selected alleles in controls estimated at 2 * allele frequency[12]. The analysis was performed using Fisher’s exact test, with a two-tailed significance level of 0.05.

For clinical correlation analysis, clinical variables were defined as follows: aAA severity was defined based on the Camitta criteria[13], modified to include absolute reticulocytes < 60 cells/microliter as assessed by automated technologies[14]. PNH clone size was categorized based on established flow cytometric methods as none (<1%), subclinical (1-10%), moderate (10-50%), and large (>50%). Response to IST was assessed 6 months after therapy according to the previously published criteria[15], with complete response (CR) defined as meeting all three peripheral blood parameters of absolute neutrophil count >1000 cells/microliter, hemoglobin > 10g/dL, and platelet count > 100 cells/microliter, and partial response (PR) defined as no longer satisfying criteria for severe aAA but insufficient recovery to be classified as CR. Analysis of clinical categories was performed using Mann-Whitney U nonparametric test for continuous variables, and Fisher's exact test for categorical variables, with a two-tailed significance level of 0.05.

Supplemental Discussion

Due to linkage disequilibrium and heterogeneity of ethnic populations, HLA association studies require a very large study population, which is difficult to accrue for rare diseases such as aAA. In contrast, recurrent loss-of-function mutations that cause clonal expansion provide unambiguous genetic evidence of the pathogenicity of the targeted allele, and, thus, allow the identification of pathogenic alleles in a much smaller patient population. We chose not to include HLA alleles that were lost solely through 6p CN-LOH as risk alleles in our analysis because in all cases the region affected by 6p CN-LOH included multiple HLA loci, sometimes encompassing the whole MHC locus (Table S6), and thus did not permit more precise identification of specific pathogenic alleles. Supporting our findings, an association with *HLA-B**14 was previously noted in a German/Austrian cohort of pediatric aAA[16] and a North American study of 6p CN-LOH[6], and, for HLA-B*40, in a 32-patient Japanese cohort [17] as well as a Japanese study of 6p CN-LOH[18]. Although the less prevalent HLA alleles *HLA-A**33:03 and *HLA-A**68:01 were not enriched in our cohort, our study was not adequately powered to detect rare HLA allele associations, and our cohort did not have sufficient numbers of patients with ethnicities where these alleles are more common. An example of ethnicity-based differences in aAA is illustrated by a Japanese aAA cohort[18,19], where *HLA-B**14:02 allele, which is very rare in the Japanese population, was not found to be associated with aAA. In contrast, *HLA-B**14:02 was identified in our study and also in several association studies of Caucasian aAA and PNH patients[6,16,20].

We have shown that HLA loss is sufficient for clonal expansion, frequently leading to emergence of several independent clones with recurrent somatic loss of the same HLA allele. A unique property of HLA as a driver of clonal hematopoiesis is the requisite reliance on the differential immunogenicity between the missing and remaining HLA alleles. Thus, population and inter-individual differences in frequencies of HLA alleles and HLA haplotypes likely underlie

the variable rates of HLA loss and are likely to influence patterns of clonal evolution in aAA cohorts of different ethnicities[18,21]. An intriguing finding in light of the crucial role of HLA class I-autoimmunity in aAA is a prior report of a somatic missense mutation in the non-polymorphic short β chain of the HLA class I complex, β2-microglobulin, in a patient with aAA [22]. We did not observe *B2M* gene mutations in our aAA patient population. This likely reflects the need to preserve a general expression of HLA class I molecules to avoid destruction by Natural Killer cells[23], which would preclude somatic *B2M* loss (which causes a loss of all HLA class I allele expression) from being a general mechanism of immune escape in aAA.

The lack of somatic HLA loss in patients with classical hemolytic PNH is interesting, because of the close etiologic association of PNH and aAA. The unique, close relationship of PNH and aAA suggests that the loss of GPI-linked proteins confers a strong growth advantage within the aAA bone marrow environment, long hypothesized to occur via immune escape. We did not observe simultaneous somatic HLA loss in patients with classical hemolytic PNH with large PNH clones. To the best of our knowledge, PNH and HLA loss have never been detected within the same hematopoietic clone; however, we and others (e.g. Yoshizato et al.) did observe independent clones—either smaller PNH clones or clones with HLA loss—occurring in separate clones or in succession within the same patient. The most likely explanation for lack of HLA loss in classical PNH is that both the loss of PIGA and the loss of HLA act via the same pathway, meaning that having both PNH and HLA loss does not add any additional growth advantage over having just one of these.

Table S1. Baseline clinical characteristics of patients with acquired aplastic anemia

Baseline Characteristics	Total Patients (n=66)	HLA Risk Allele Present* (n=24)	HLA Risk Allele Absent (n=42)	P value
Median age at diagnosis, yrs (range)	11.9 (1.5-65.6)	12.8 (3.2-61.5)	11.1 (1.5-65.6)	0.342
Pediatric-onset, n (%)	46 (69.7%)	15 (62.5%)	31 (73.8%)	0.408
Adult-onset, n (%)	20 (30.3%)	9 (37.5%)	11 (26.2%)	
Severity of aAA at diagnosis, n (%)				
NSAA	12 (18.2%)	5 (20.8%)	7 (16.7%)	1.000
SAA or VSAA	44 (67.7%)	17 (70.8%)	27 (64.3%)	
n/a	10 (15.2%)	2 (8.3%)	8 (19.0%)	
Median disease duration at HLA NGS, yrs (range)	1.1 (0-30.4)	1.1 (0-30.4)	1.1 (0-13.0)	0.516
Disease status at HLA NGS, n (%)^				0.106
Diagnosis prior to therapy/stable NSAA	8/5 (20.0%)	3/3 (25%)	5/2 (17%)	
PR	19 (28.8%)	7 (29.2%)	12 (28.6%)	
CR	17 (25.8%)	4 (16.7%)	13 (31.0%)	
Refractory/Relapsed/post-aAA MDS	3/4/1 (12.1%)	1/4/1 (25.0%)	2/0/0 (4.8%)	
NE	9 (13.6%)	1 (4.2%)	8 (19.0%)	
Median follow-up, yrs (range)	1.8 (0-12.6)	1.8 (0.3-4.6)	1.8 (0-12.6)	0.682
Clinical follow-up unavailable, n (%)	12 (18.2%)	2 (8.0%)	10 (23.8%)	
Therapy prior to HLA NGS, n				
Therapy-naïve	12 (18%)	7 (29.2%)	5 (11.9%)	0.523
First-line IST with hATG and CsA	44 (66.7%)	16 (66.7%)	28 (66.7%)	
Second-Line Therapy	9 (13.6%)	7 (29.2%)	2 (4.8%)	
n/a	10 (15.2%)	1 (4.2%)	9 (21.4%)	
Patients with PNH granulocytes, n				
none (<1%)	41 (62.1%)	16 (66.7%)	25 (59.5%)	0.760
Subclinical (1-10%)	8 (12.1%)	3 (12.5%)	5 (11.9%)	
Moderate (10-50%)	7 (10.6%)	4 (16.7%)	3 (7.1%)	
n/a	10 (15.2%)	1 (4.2%)	9 (21.4%)	
WES available	45 (68.1%)	17 (70.8%)	28 (66.7%)	0.789

Yrs, years; *P-value pertains to the comparison of the 24 patients who carry at least one of the 4 HLA risk alleles (*HLA-A**33:03, *HLA-A**68:01, *HLA-B**14:02, or *HLA-B**40:02) to the 42 patients who carry none of these HLA alleles. N, number of patients. NSAA, non-severe aAA, SAA, severe aAA, VSAA, very severe aAA. IST, immunosuppressive therapy. hATG, horse antithymocyte globulin. CsA, cyclosporine A. ^ Disease status at HLA NGS; for composite categories, such as relapse/refractory/MDS, the numbers of patients in each sub-category are listed on the same line, separated by "/".

Supplemental Table S2. HLA allele loss in patients with acquired aplastic anemia.

Mechanism of HLA Loss	Patient ID	6p CN-LOH clones (n)	6p CN-LOH Region	Mutations (n)	Mutated Allele	Type of Mutation	Exon	Mutation (coding)	Mutation (protein)	Allele Frequency, %	HLA Class I Alleles Lost through 6p CN-LOH	HLA Class I Alleles Retained
Mutational Inactivation	29.01	n/a		5	B*40:02	Start	1	c.1A>G	p.M1V	7.0	n/a	A*02:01, B*44:02 C*03:05, C*05:01
					B*40:02	Nonsense	1	c.19C>T	p.R7*	12.5		
					B*40:02	Nonsense	4	c.742C>T	p.Q248*	3.9		
					B*40:02	Nonsense	4	c.862G>T	p.E288*	4.8		
					B*40:02	Frameshift del	4	c.870delG	p.L290fs	1.6		
	435.01	n/a		3	B*40:02	Frameshift del	1	c.19delC	p.R7*	17.3	n/a	A*02:01, A*11:01 B*40:01 C*02:02,C*03:04
					B*40:02	Frameshift ins	1	c.60insGA	p.T20fs	19.2		
					B*40:02	Frameshift del	2	c.286delAG	p.Q96fs	5.1		
	506.01	n/a		1	B*14:02	Frameshift del	4	c.890delAT	p.R297fs	17.0	n/a	A*02:01 A*03:01 B*44:03 C*04:09, C*08:02
Mutational Inactivation and 6p CN-LOH	54.01	2	6pterp21.31 6pterp21.1	2	B*14:02	Nonsense	1	c.19C>T	p.R7*	6.6	A*33:01 B*14:02# C*08:02	A*24:02, B*44:03 C*04:01
					B*14:02	Frameshift del	4	c.880delC	p.L294fs	20.1		
	281.01	4	6pterp22.1 6pterp21.33 6pterp12.1	1	A*33:03	Nonsense	3	c.526C>G	p.Y142fs	15.1	A*33:03# B*44:03 C*07:06	A*24:02^ B*52:01 C*12:02
	505.01	2	6pterp21.31 6pterp12.1	1	A*68:01	Frameshift del	4	c.794delC	p.F265fs	3.5	^^	^^
6p CN-LOH	56.01	4	6pterp21.31 6pterp21.1 6pterp12.1 6pterp11.1	0	n/a						A*03:01, B*14:02, C*08:02	A*03:01, B*07:02, C*07:02
	471.01	4	6pterp21.32 6pterp21.31 6pterp21.2 6pterp12.1	0	n/a						A*11:01 B*35:01 C*04:01	A*01:01 B*18:01 C*07:01
	284.01	2	6pterp21.32 6pterp11.1	0	n/a						A*02:01 B*35:03 C*12:03	A*26:01, B*18:01, C*12:03
	348.01	1	6pterp21.2	0	n/a						A*29:02 B*13:02 C*06:02	A*02:05, B*51:01, C*16:01
	390.01	1	6WC	0	n/a						^^	^^

CN-LOH, copy number-neutral loss of heterozygosity; del, deletion; ins, insertion; WC, whole chromosome; A, *HLA-A*; B, *HLA-B*; #, allele lost through both mutational inactivation and 6p CN-LOH; ^^, insufficient clone size to determine minor and major alleles; Allele frequency, determined as a fraction of NGS reads containing the mutation; note, that a 50% allele frequency would correspond to 100% of diploid cells carrying the mutation. A*24:02^, a new HLA allele most closely related to A*24:02. Please see detailed table of all HLA alleles in Supplemental Table S3.

Supplemental Table S3. Race- and ethnicity-stratified frequency of HLA risk alleles in 66 aplastic anemia patients.

Race/Ethnicity	Aplastic Anemia Patients (n)	*HLA-A**33:03 (aAA patients, n)	*HLA-A**68:01 (aAA patients, n)	*HLA-B**14:02 (aAA patients, n)	*HLA-B**40:02 (aAA patients, n)
White	42	0	2	12	4
Black	11	2	0	1	0
Hispanic	5	1	0	0	2
Indian	2	1	0	0	0
Asian	1	0	0	0	0
Middle Eastern	1	0	0	1	0
Other	1	0	0	0	0
Unknown	3	0	1	0	0
Total	66	4	3	14	6

Supplemental Table S4. *HLA-B14:02 and *HLA*B*-40:02 are enriched in aAA patients compared to race-matched controls**

Population	n	*HLA-B*14:02*			*HLA-B*40:02*		
		Individuals with Allele, n (%)	Individuals Without Allele, n (%)	P-value, aAA vs control/ MDS vs control	Individuals with Allele, n (%)	Individuals Without Allele, n (%)	P-value, aAA vs control/ MDS vs control
aAA, Caucasian	42	12 (28.6%)	30 (71.4%)		4 (9.5%)	38 (90.5%)	
MDS, Caucasian	163	13 (8.0%)	150 (92.0%)		4 (2.5%)	159 (97.5%)	
USA NMDP European Caucasian	1,242,890	71,093 (5.7%)	1,171,797 (94.3%)	**0.000**/0.2333	31,321 (2.5%)	1,211,569 (97.5%)	**0.021**/1.000
USA European American Population 2	1,245	57 (4.6%)	1188 (95.4%)	**0.000**/0.081	35 (2.8%)	1210 (97.2)	**0.035**/1.000
USA Caucasian Population 4	1,070	72 (6.7%)	998 (93.3%)	**0.000**/0.511	22 (2.1%)	1048 (97.9%)	**0.014**/0.768
USA Caucasian Bethesda	307	26 (8.5%)	281 (91.5%)	**0.001**/1.000	7 (2.3%)	300 (97.7%)	**0.032**/1.000
USA Caucasian Population 2	265	18 (6.8%)	247 (93.2%)	**0.000**/0.702	8 (3.0%)	257 (97.0%)	0.066/1.000
Philadelphia Caucasian	141	6 (4.3%)	135 (95.7%)	**0.000**/0.143	4 (2.8%)	137 (97.2%)	0.241/1.000

P-values correspond to the comparison of frequencies of HLA alleles in aAA and MDS patients as compared to independent race-matched controls populations. P-values <0.05 are shown in bold.

Supplemental Table S5. Frequencies and association analysis of HLA class I alleles in aAA Patients (Caucasian, n=42)

HLA-A				HLA-B				HLA-C			
HLA-A Allele	**Frequency (%)**	**Patients with Allele (n)**	**P-value**	**HLA-B Allele**	**Frequency (%)**	**Patients with Allele (n)**	**P-value**	**HLA-C Allele**	**Frequency (%)**	**Patients with Allele (n)**	**P-value**
02:01	50.0	21	0.537	07:02	19.0	8	0.380	02:02	11.9	5	0.410
01:01	40.5	17	0.325	08:01	31.0	13	0.203	03:03	16.7	7	0.208
03:01	33.3	14	0.491	13:02	7.1	3	0.453	03:04	9.5	4	0.394
11:01	14.3	6	0.636	**14:02**	**28.6**	**12**	**0.000**	04:01	21.4	9	1.000
24:02	7.1	3	0.101	15:01	9.5	4	0.813	05:01	9.5	4	0.165
32:01	7.1	3	1.000	18:01	7.1	3	1.000	06:02	9.5	4	0.165
33:01	**7.1**	**3**	**0.031**	35:01	11.9	5	0.807	07:01	35.7	15	0.621
29:02	4.8	2		40:01	9.5	4	1.000	07:02	19.0	8	0.231
68:01	4.8	2		**40:02**	**9.5**	**4**	**0.021**	**08:02**	**28.6**	**12**	**0.000**
26:01	4.8	2		44:02	11.9	5	0.325	12:03	7.1	3	0.795
30:01	4.8	2		44:03	9.5	4	1.000	16:01	4.8	2	
30:02	2.4	1		51:01	9.5	4	1.000	04:09	2.4	1	
02:09	2.4	1		55:01	7.1	3	0.205	07:04	2.4	1	
68:02	2.4	1		35:02	4.8	2		12:02	2.4	1	
				49:01	2.4	1		12:05	2.4	1	
				35:03	2.4	1		15:11	2.4	1	
				37:01	2.4	1		16:04	2.4	1	
				52:01	2.4	1		17:03	2.4	1	
				41:02	2.4	1					
				39:01	2.4	1					
				27:05	2.4	1					
				38:01	2.4	1					

P-values correspond to the comparison of frequencies of HLA alleles in aAA patients as compared to the published frequencies in race-matched controls (USA NMDP European Caucasian cohort, n=1,242,890[12,24,25]). HLA alleles with a statistically significant association with aAA are shown in bold. Association analysis was performed for all HLA alleles identified in more than 2 patients within the cohort.

Supplemental Table S6. Haplotypes Containing *HLA-B14:02 Allele in USA Caucasian Populations**

Population	Sample size (n)	Haplotype	Frequency (%)
USA Caucasian pop 2	265	A*02:01-B*14:02	1
USA San Francisco Caucasian	220	A*02:01-B*14:02-C*08:02	1
USA San Francisco Caucasian	220	A*33:01-B*14:02-C*08:02	1.4
USA Italy Ancestry	273	A*33:01-B*14:02-C*08:02-DRB1*01:02	1.1
USA NMDP European Caucasian	1,242,890	A*33:01-B*14:02-C*08:02-DRB1*01:02-DQB1*05:01	0.44
USA NMDP European Caucasian	1,242,890	A*33:01-B*14:02-C*08:02-DRB1*03:01-DQB1*02:01	0.09
USA NMDP European Caucasian	1,242,890	A*68:02-B*14:02-C*08:02-DRB1*13:03-DQB1*03:01	0.27
USA Caucasian pop 2	265	B*14:02-C*08:02	3.2

Haplotype information represents all *HLA-B**14:02 containing haplotypes in the United States Caucasian individuals in The Allele Frequency Net Database[12,24,25].

Supplemental Table S7. HLA A, B, C and DRB1 alleles in 66 patients with acquired aplastic anemia

Patient ID	HLA-A (1)	HLA-A (2)	HLA-B (1)	HLA-B (2)	HLA-C (1)	HLA-C (2)	HLA-DRB1 (1)	HLA-DRB1 (2)
1.01	03:01:01	33:01:01	14:02:01	44:03:01	08:02:01	16:01:01	01:01:01	11:01:01
3.01	11:01:01	11:01:01	15:11:01	40:01:02	03:03:01	07:02:01	04:80	12:02:01
5.01	01:01:01	02:02:01	08:01:01	15:03:01	02:10:01	07:01:01	03:01:01	04:05:01
19.01	03:01:01	33:03:01	53:01:01	53:01:01	04:01:01	04:01:01	08:04:01	11:02:01
20.01	02:01:01	03:01:01	40:02:01	51:01:01	02:02:02	03:03:01	09:01:02	13:01:01
29.01	02:01:01	02:01:01	**40:02:01**	44:02:01	03:05:01	05:01:01	01:01:01	13:01:01
41.01	01:01:01	02:01:01	07:02:01	50:01:01	06:02:01	07:02:01	04:01:01	16:02:01
45.01	01:01:01	03:01:01	07:02:01	08:01:01	07:01:01	07:02:01	01:01:01	03:01:01
54.01*	24:02:01	**33:01:01**	**14:02:01**	44:03:01	04:01:01	**08:02:01**	07:01:01	**11:01:01**
56.01*	03:01:01	03:01:01	07:02:01	**14:02:01**	07:02:01	**08:02**	03:01	**04:04**
65.01	02:01:01	02:01:01	14:02:01	51:01:01	02:02:02	08:02:01	04:03:01	15:01:01
79.01	02:01:01	02:01:01	15:01:01	40:01:02	03:03:01	03:04:01	01:01:01	14:54:01
102.01	02:01:01	03:01:01	14:02:01	35:01:01	04:01:01	08:02:01	01:01:01	01:02:01
160.01	11:01:01	24:02:01	07:02:01	35:01:01	04:01:01	07:02:01	11:01:01	15:01:01
164.01	02:01:01	11:01:01	35:01:01	44:03:01	04:01:01	16:01:01	04:07:01	13:02:01
168.01	02:01:01	03:01:01	15:01:01	35:01:01	03:03:01	04:01:01	01:01:01	14:54:01
170.01	01:01:01	03:01:01	15:01:01	55:01:01	03:03:01	03:03:01	04:01:01	04:07:01
173.01	30:01:01	33:01:01	18:01:01	42:01:01	02:02:02	17:01:01	07:01:01	11:02:01
180.01	01:01:01	03:01:01	08:01:01	18:01:01	07:01:01	12:05	04:01:01	11:01:01
260.01	02:01:01	29:02:01	44:02:01	49:01:01	07:01:01	07:04:01	11:01:01	13:01:01
263.01	02:05:01	03:01:01	58:01:01	58:02:01	06:02:01	07:18	07:01:01	11:02:01
281.01*	24:02:01	**33:03:01**	**44:03:01**	52:01:01	**07:06:01**	12:02	**07:01**	13:01
284.01*	26:01:01	**02:01:01**	18:01:01	**35:03:01**	12:03:01	12:03	13:02	**04:08**
286.01	01:01:01	03:01:01	14:02:01	37:01:01	06:02:01	08:02:01	01:01:01	01:01:01
326.01	01:01:01	11:01:01	08:01:01	44:02:01	05:01:01	07:01:01	03:01:01	12:01:01
332.01	02:05:01	33:03:01	18:01:01	58:01:01	07:18	12:03:01	03:01:01	11:04:01
337.01	01:01:01	01:01:01	08:01:01	52:01:01	07:01:01	12:02:02	03:01:01	15:02:01
348.01*	02:05:01	**29:02:01**	51:01:01	**13:02:01**	16:01:01	**06:02:01**	13:01	**07:01**
352.01	30:02:01	68:01:01	57:03:01	58:02:01	06:02:01	08:02:01	12:01:01	13:03:01
356.01	02:01:01	33:01:01	14:02:01	41:02:01	08:02:01	17:03	01:02:01	13:03:01
362.01	01:01:01	01:01:01	08:01:01	35:02:01	04:01:01	07:01:01	03:01:01	11:04:01
364.01	02:01:01	11:01:01	14:02:01	44:03:01	04:01:01	08:02:01	13:02:01	16:02:01

376.01	01:01:01	02:01:01	08:01:01	44:02:01	05:01:01	07:01:01	03:01:01	12:01:01
378.01	01:01:01	01:01:01	08:01:01	18:01:01	07:01:01	07:01:01	03:01:01	11:04:01
383.01	02:01:01	11:01:01	40:02:01	55:01:01	02:02:02	03:03:01	04:08:01	14:54:01
385.01	23:01:01	34:02:01	15:03:01	53:01:01	02:10	03:03:04	07:01:01	11:01:02
387.01	02:01:01	36:01:01	47:01:01	53:01:01	04:01:01	07:18	08:06	11:01:02
390.01*^	32:01:01	68:01:01	14:02:01	51:01:01	08:02:01	14:02	11:04	15:01
408.01	03:01:01	24:02:01	07:02:01	44:02:01	07:02:01	07:04:01	11:04:01	15:01:01
409.01	03:01:01	23:01:01	07:02:01	08:01:01	07:01:01	07:02:01	03:01:01	15:01:01
410.01	01:01:01	02:01:01	07:02:01	08:01:01	07:01:01	07:02:01	04:01:01	15:01:01
427.01	02:01:01	03:01:01	07:02:01	44:02:01	05:01:01	07:02:01	04:01:01	07:01:01
429.01	01:01:01	03:01:01	08:01:01	14:02:01	07:01:01	08:02:01	03:01:01	13:02:01
434.01	01:01:01	01:01:01	08:01:01	08:01:01	07:01:01	07:01:01	03:01:01	11:01:01
435.01	02:01:01	11:01:01	40:01:02	**40:02:01**	02:02:02	03:04:01	04:04:01	08:01:01G
439.01	03:01:01	26:01:01	14:02:01	55:01:01	01:02:01	08:02:01	01:02:01	03:01:01
447.01	02:01:01	32:01:01	13:02:01	40:02:01	02:02:02	06:02:01	11:04:01	13:01:01
450.01	02:01:01	26:01:01	07:02:01	44:02:01	05:01:01	07:02:01	04:01:01	15:01:01
466.01	01:01:01	24:02:01	08:01:01	35:08:01	04:01:01	07:01:01	03:01:01	08:03:02
471.01*	01:01:01	**11:01:01**	18:01:01	**35:01:01**	07:01:01	**04:01:01**	04:01	**14:54**
474.01	30:01:01	68:02:01	13:02:01	55:01:03	03:03:01	06:02:01	03:01:01	13:02:01
482.01	01:01:01	03:01:01	08:01:01	40:01:02	03:04:01	07:01:01	04:04:01	11:01:01
483.01	01:01:01	02:01:01	08:01:01	13:02:01	06:02:01	07:01:01	04:01:01	07:01:01
486.01	01:01:01	32:01:01	07:02:01	08:01:01	07:01:01	07:02:01	13:02:01	15:01:01
487.01	02:01:01	03:01:01	15:01:01	40:01:02	03:03:01	03:04:01	04:01:01	09:01:02
488.01	03:01:01	03:01:01	07:02:01	15:03:01	02:10	07:02:01	04:04:01	07:01:01
503.01	02:01:01	11:01:01	40:02:01	50:01:01	03:04:01	04:01:01	08:02:01	13:01:01
505.01*^	29:02:01	68:01:01	07:02:01	14:02:01	07:02:01	08:02:01	13:02	15:01
506.01	02:01:01	03:01:01	**14:02:01**	44:03:01	04:09:01	08:02:01	07:01:01	13:02:01
510.01	02:09:01	30:01:01	35:02:01	39:01:01	04:01:01	12:03:01	15:01:01	15:01:01
519.01	01:01:01	11:01:01	15:02:01	15:17:01	07:01:02	08:01:01	12:02:01	13:02:01
524.01	01:01:01	11:01:01	14:02:01	27:05:02	08:02:01	15:11:01	13:02:01	15:01:01
484.01	01:01:01:01	02:01:01:01	08:01:01	44:02:01:01	05:01:01:01	07:01:01:01	04:01	15:01
533.01	01:01:01:01	02:01:01:01	44:03:01:01	44:03:01:01	02:02:02:01	16:01:01:01	07:01	08:04
539.01	02:01:01:01	30:02:01:01	38:01:01	51:01:01:01	12:03:01:01	16:04:01	4[&]	11[&]
542.01	30:01:01	33:03:01	42:01:01	53:01:01	04:01:01:01	17:01:01:01	08:04	13:03

Patients with somatic HLA loss are indicated by bold formatting of Patient ID, with 6p CN-LOH indicated by an asterisk (*). In patients with 6p CN-LOH, the alleles lost due to loss-of-heterozygosity are shown in bold with shading. ^, missing alleles were not able to be resolved due to the small clone size of 6p CN-LOH. &, HLA type obtained by serological equivalent.

Supplemental Table S8. Clinical characteristics of patients with aplastic anemia stratified by presence or absence of HLA risk alleles.

Clinical Characteristic	Total Patients (n=66)	HLA Risk Allele Present* (n=24)	No HLA Risk Allele (n=42)	P value
First-line therapy, patients, n				0.222
MRD AlloBMT	3	1	2	
hATG and CsA	47	18	29	
CsA monotherapy	1	0	1	
None	5	4	1	
n/a	10	1	9	
Median time to IST, months	1.0 (<1-82.8)	1.2 (<1-50.0)	1.0 (<1-82.8)	0.849

MRD-AlloBMT, matched related donor allogeneic bone marrow transplant; hATG, horse anti-thymocyte globulin; CsA, cyclosporine A; IST, immunosuppressive therapy.

Supplemental Table S9. Clonal Hematopoiesis in 45 Aplastic Anemia Patients Analyzed by Comparative WES

Baseline Characteristics	WES Cohort (n=45)	HLA Risk Allele* Present (n=17)	HLA Risk Allele Absent (n=28)	P-value
Median age at diagnosis, years (range)	11.6 (1.5-65.6)	11.9 (3.2-61.5)	10.1 (1.5-65.6)	0.465
Pediatric-onset, patients, n	34 (75.6%)	12 (70.6%)	22 (78.6%)	0.722
Adult-onset, patients, n	11 (24.4%)	5 (29.4%)	6 (21.4%)	
Median disease duration at WES, yrs (range)	1.1 (0.0-30.4)	1.1 (0.1-30.4)	1.1 (0.0-13.0)	0.992
Median age at WES, yrs (range)	14.3 (2.6-67.2)	14.1 (4.3-61.8)	15.0 (2.6-67.2)	0.529
Median follow-up, yrs (range)	1.6 (0.1-5.2)	1.69 (0.4-4.1)	1.6 (0.1-5.2)	0.719
Clonal hematopoiesis, any modality, patients, n (%)	32 (71.1%)	14 (82.4%)	18 (64.3%)	0.311
Somatic HLA loss, n (%)	9 (20.0%)	7 (41.2%)	2 (7.1%)	**0.009**
PNH clone, patients, n (%)	15 (33.3%)	7 (41.2%)	8 (28.6%)	0.517
Non-HLA, non-PIGA somatic mutations, patients, n (%)	23 (51.1%)	9 (52.9%)	14 (50.0%)	1.000
Non-PNH somatic events recurrent in aAA, patients, n (%)	16 (35.6%)	11 (64.7%)	5 (17.9%)	**0.004**
Prognostically adverse somatic mutations (*ASXL1, DNMT3A, RUNX1*), patients, n (%)	6 (13.3%)	5 (29.4%)	1 (3.6%)	**0.022**

P-value pertains to the comparison of the 17 patients who carry at least one of the 4 HLA risk alleles* (*HLA-A**33:03, *HLA-A**68:01, *HLA-B**14:02, or *HLA-B**40:02) to the 28 patients who carry none of the HLA risk alleles.

Supplemental Table S10. Comprehensive analysis of clonal hematopoiesis in 45 patients with aplastic anemia

Study ID	HLA Risk Allele	Age at Dx, yrs		Duration of AA at WES, yrs	Karyotype	Acquired CN-LOH	HLA Loss	PNH Granulocytes >1%	MDS-associated mutations	Non-PIGA, non-HLA Somatic Mutations, n	Nonsynonymous Coding Mutations, gene (mutation, clone size)	Mutations in Untranslated Regulatory Regions, gene (mutation type, clone size)	Lineage
001.01	B*14:02	3.2		5.6	der(5)t(1;5)(q11;q11.2)	none	No	none	none	2	*SH3KBP1* (R252T, 28%)	*KCNMB4* (3'UTR, 4%)	Both
439.01	B*14:02	3.9		1.1	+der(15)(p13→q10::?)(c)	none	No	none	none	0	none	none	n/a
281.01	A*33:03	4.3		0.1	normal	6p	Yes	none	none	0	none	none	n/a
056.01	B*14:02	6.0		8.0	normal	6p	Yes	none	none	0	none	none	n/a
505.01	A*68:01, B*14:02	7.9		0.1	normal	6p	Yes	none	none	1	none	*ACSL4* (3'UTR, 8%)	n/a
447.01	B*40:02	8.0		2.1	normal	none	No	none	none	0	none	none	n/a
054.01	B*14:02	10.5		11.1	normal	6p	Yes	none	none	0	none	none	M
506.01	B*14:02	11.6		1.0	normal	none	Yes	none	none	0	none	none	n/a
435.01	B*40:02	11.9		1.9	normal	none	Yes	none	none	4	*BCL9* (G195D, 8%)	*COPA* (3'UTR, 9%), *ABHD10* (3'UTR, 15%), *RGMB* (3'UTR, 12%)	Both
332.01	A*33:03	12.8		1.1	normal	none	No	none	none	0	none	none	n/a
503.01	B*40:02	19.6		0.1	t(1;13)(q24;q14)	none	No	Yes	none	0	none	none	n/a
020.01	B*40:02	21.1		5.6	normal	5q	No	Yes	none	5	*CAMK2G*(T306M, 12%), *BPTF* (D312H, 13%), *WDR18* (A156T, 16%)	*PDE10A* (3'UTR, 12%), *SLC2A13* (3'UTR, 8%)	n/a
429.01	B*14:02	24.7		30.4	normal	17q	No	Yes→ none	*SUZ12, RUNX1, PHF6, ASXL1*	15	*IL22RA1* (R318W, 33%), *CD58* (E65K, 38%), *CRB1* (G97*, 43%), *GRM3* (Y434*, 38%), *KCNU1* (R1002*,41%), *KCNJ8*(G304V, 33%), *SUZ12* (D605N, 81%), *RUNX1* (S295*, 35%), *TAF7L* (D60N, 35%), *PHF6* (splice, 41%),*H3F3C* (Q6P, 38%), *ASXL1* (G646fs, 23%)	*CCDC6* (3'UTR, 41%), *PPM1A* (3'UTR, 38%), *KAT5/RNASEH2C* (3'UTR, 37%)	n/a
364.01	B*14:02	31.3		4.9	Monosomy 7	none	No	Yes	*ASXL1, SETBP1*	13	*SLC22A6* (R336C, 11%), *STAT5B* (N642H, 23%), *WFDC12* (G27C, 16%), *DNAJC24*(c.251delAGA, 17%), *ASXL1* (E635fs, 7%), *SETBP1* (G870S, 9%)	*OSBPL11* (3'UTR, 20%), *SENP6*(3'UTR, 16%), *MET*(3'UTR, 16%), *NFIB*(3'UTR, 15%), *ZNF747*(3'UTR, 14%), *BLCAP*(3'UTR, 16%), *CLCN5* (3'UTR, 15%)	M

356.01	B*14:02	53.5		0.4	normal	none	No	Yes→ none	*ASXL1*, *PHF6*	9	*SYNPO2*(E709K, 39%), *IL20RA*(E473K, 49%), *FOS*(G97R, 43%), *NINL*(R667C, 37%), *SEPT3*(R165H, 34%), *ASXL1*(C687fs, 40%), *PHF6* (G64fs, 2%)	*LMTK2* (5'UTR, 27%), *ATP2B3* (3'UTR, 18%)	Both
524.01	B*14:02	61.2		0.5	Monosomy 7	none	No	Yes→ none	*SRSF2*, *ASXL1*	18	*HSD3B2*(P153A,27%), *NOVA1* (A273T, 25%), *H2AFY2*(E288del, 24%), *SRSF2*(P95H, 23%), *SHANK3*(A281T,23%), *WWC3*(Q250H, 22%), *HAS2*(A64S, 22%), *GRIN2A*(C1412S, 19%), *SOX6*(P707S, 17%), *SCRIB*(P964A, 9%), *MYH3*(I304V, 8%), *MAMDC4* (G1071R, 12%), *ASXL1* (G646fs, 3%)	*AEBP1*(3'UTR, 29%), *VWC2*(3'UTR, 26%), *HNF4A*(3'UTR, 23%), *MORC2-AS1* (NCExonic, 18%), *CAMKK1*(3'UTR, 9%), *DPP6*(3'UTR, 8%)	Both
390.01	A*68:01 , B*14:02	61.5		0.3	del(13) (q12q14)	6WC	Yes	Yes	*ASXL1*, *DNMT3A*, *RUNX1*	3	*ASXL1*(V751fs, 44.7%), *DNMT3A* (splice, 43.4%), *RUNX1* (R204Q, 41.8%)	none	n/a
362.01		1.5		1.1	normal	none	No	none	none	0	none	none	n/a
387.01		2.6		0.0	+3, der(3;15) (q10;q10)	none	No	none	none	0	none	none	n/a
450.01		3.2		1.1	normal	none	No	none	none	0	none	none	n/a
474.01		3.3		1.1	normal	none	No	none	none	2	*CCDC180* (E766K, 21%)	*SLC17A9* (3'UTR, 17%)	Both
041.01		3.6		12.6	normal	none	No	none	none	2	*ANK2* (S2942N, 26%), *FAM104A*(N121I, 26%)	none	Both
519.01		4.2		5.1	normal	none	No	none	none	1	GANAB (R19H, 8%)		n/a
003.01		4.8		1.0	normal	none	No	none	none	0	none	none	n/a
263.01		5.3		0.6	normal	none	No	Yes	none	1	none	*CYP4F24P* (NCExonic, 12%)	n/a
471.01		6.3		10.2	normal	6p	Yes	none	none	4	*TOP2B* (T1562I, 8%), *FBN2* (R2267H, 12%), *RSPH9* (W220R, 11%), *PRR14* (V68M, 20%)	none	Both
510.01		7.6		1.1	normal	none	No	none	none	0	none	none	n/a
005.01		7.7		5.4	normal	none	No	none	*BCOR*	1	*BCOR* (splice, 45%)	none	n/a
484.01		8.3		1.1	normal	none	No	none	none	0	none	none	n/a
487.01		8.8		1.2	normal	none	No	none	none	0	none	none	n/a
533.01		9.1		1.0	normal	none	No	none	none	3	*R3HDM1* (D908N, 7%), *RNF139* (H503P, 8%)	*NCOR2* (5'UTR, 9%)	n/a
488.01		11.2		13.0	normal	none	No	none	none	0	none	none	n/a
466.01		12.2		1.2	normal	none	No	none	none	0	none	none	n/a

542.01		13.9		0.3	normal	none	No	Yes	none	1	none	*FPGS* (3'UTR, 7%)	Both
045.01		14.2		6.0	normal	none	No	Yes	none	0	none	none	n/a
376.01		15.2		0.5	normal	none	No	none	none	0	none	none	n/a
385.01		18.5		0.0	normal	none	No	Yes	none	3	*EMX2* (A172V, 9%), *DSG1* (M208V, 9%)	*LARP* (3'UTR 9%)	n/a
434.01		19.2		1.1	normal	none	No	none	none	0	none	none	Both
173.01		20.2		0.2	normal	none	No	none	none	0	none	none	n/a
483.01		41.1		0.6	normal	none	No	Yes→ none	none	5	*IQCE* (R183TW, 26%), *SUPT20H* (G590V, 24%)	*CXCR2* (3'UTR, 26%), *C19orf44* (3'UTR, 29%), *GABRA3* (3'UTR, 28%)	Both
284.01		46.9		2.7	normal	6p	Yes	none	none	5	*EXO1* (G274W, 11%), *PXDNL* (W992*, 14%), *LRRC37B* (I926T, 15%), *ZNFX1* (S262C, 13%)	*GRIN1* (3'UTR, 11%)	Both
482.01		50.7		1.8	normal	none	No	Yes	*BCOR*	18	*ST6GALNAC1* (E139K, 14%), *CDH4* (A813S, 12%), *PI16* (S4F, 11%), *USP24* (R1102Q, 10%), *BCOR* (E812*, 10%), *CCDC40*(R431Q, 10%), *ACTA2*(P266S, 10%), *EML4*(R306C, 9%), *PNLIPRP1*(I373N, 9%), *SHANK2*(A770D, 8%), *OR1S2*(S31P, 8%), *TTN* (R4683C, 9%), *TTN* (K11777R, 12%), *MROH2A* (I859V, 11%)	*TMEM26*(3'UTR, 12%), *CALML4*(3'UTR, 10%), *KLHL15*(3'UTR, 9%), *GART* (5'UTR, 10%)	M
168.01		53.5		5.5	normal	none	No	Yes	none	1	*ADAM21* (P40L, 14%)	none	n/a
486.01		61.2		1.1	normal	none	No	none	*ASXL1*	9	*IGF1R* (K453R, 31%), *MMD* (L145fs, 23%), *ITFG3* (V157M, 10%), *FAXC* (E257fs, 8%),*ASXL1* (Q733*, 6%)	*GPR123*(3'UTR, 19%), *RTL1* (5'UTR, 14%), *FBRS* (3'UTR, 18%), *AVPR2* (UTR3, 14%)	Both
408.01		65.6		1.6	normal	none	No	Yes	none	0	none	none	Both

A, HLA-A; B, HLA-B; yrs, years; PNH granulocytes >1%, greater than 1% of granulocytes with PNH immunophenotype as measured by flow-cytometry; Nonsyn. coding, nonsynonymous coding somatic mutations; UTR, somatic mutations in untranslated regulatory regions; Lineage, mutations present in myeloid (M), Lymphoid (L), or both myeloid and lymphoid lineages (Both). Clone size was calculated as a percentage of mutant reads.

Supplemental Table S11. Longitudinal follow-up of clonal hematopoiesis

Study ID	HLA Risk Allele	Age at AA Dx, yrs	Duration of Disease at WES, yrs	Cytogene-tics at Dx	Interim Cytogenetics (if different)	Cytogenetics at follow-up	Acquired CN-LOH at Dx	Interim CN-LOH (if different)	Acquired CN-LOH at follow-up	PNH Granulocytes (>1%) at dx	PNH Granulocytes (>1%), interim	PNH Granulocytes (>1%) follow-up	Malignancy-associated mutations	Serial NGS
1.01	B*14:02	3.2	5.6	**normal**		**46,XX,der(5)t(1;5)(q11;q11.2)[3]/46,XX[17]**	n/a		none	None	None	None	none	WES
439.01	B*14:02	3.9	1.1	Constitutional 47, XX, +der(15)(p13->q10)		Constitutional 47, XX, +der(15)(p13->q10)	none		none	None	None	None	none	WES
281.01	A*33:03	4.3	0.1	normal		normal	6pterp22.1 6pterp21.33 6pterp12.1		n/a	None	None	None	none	
56.01	B*14:02	6.0	8.0	n/a		normal	n/a		6pterp21.31 6pterp21.1 6pterp12.1 6pterp11.1	n/a	None		none	
505.01	A*68:01, B*14:02	7.9	0.1	normal		normal	6pterp21.31 6pterp12.1		n/a	n/a	n/a	None	none	
447.01	B*40:02	8.0	2.1	normal		normal	none		none	None	n/a	n/a	none	
54.01	B*14:02	10.5	11.1	normal		normal	n/a		6pterp21.31 6pterp21.1	None	None	None	none	Targeted NGS
506.01	B*14:02	11.6	1.0	normal		normal	n/a		none	None	n/a	n/a	none	
435.01	B*40:02	11.9	1.9	normal		normal	none		none	None	None	None	none	WES
332.01	A*33:03	12.8	1.1	normal		normal	none		none	None	None	None	none	
503.01	B*40:02	19.6	0.1	**46, XY, t(1;13)(q24;q14)[13]/46,XY[4]**		**normal**	none		none	Yes (Moderate, 14%)	n/a	n/a	none	
20.01	B*40:02	21.1	5.6	normal		normal	**none**		**5q15qter**	**None**	**Yes (Subclinical, 8%)**	**Yes (Moderate, 16%)**	none	
429.01	B*14:02	24.7	30.4	n/a		normal	n/a		17q11.2qter	**Yes (hemolytic PNH, clone not quantified)**	**n/a**	**None**	SUZ12, RUNX1, PHF6, ASXL1	
364.01	B*14:02	31.3	4.9	**normal**		**45,XX,-7[2] ; FISH positive 28/100 cells**	n/a		none	**n/a**	**None**	**Yes, Subclinical (4.6%)**	**ASXL1, SETBP1**	Targeted NGS
356.01	B*14:02	53.5	0.4	normal		normal	none		none	**Yes (Subclinical, 2-9%)**	**Yes, Subclinical (1.6%)**	**None**	ASXL1, PHF6	Targeted NGS

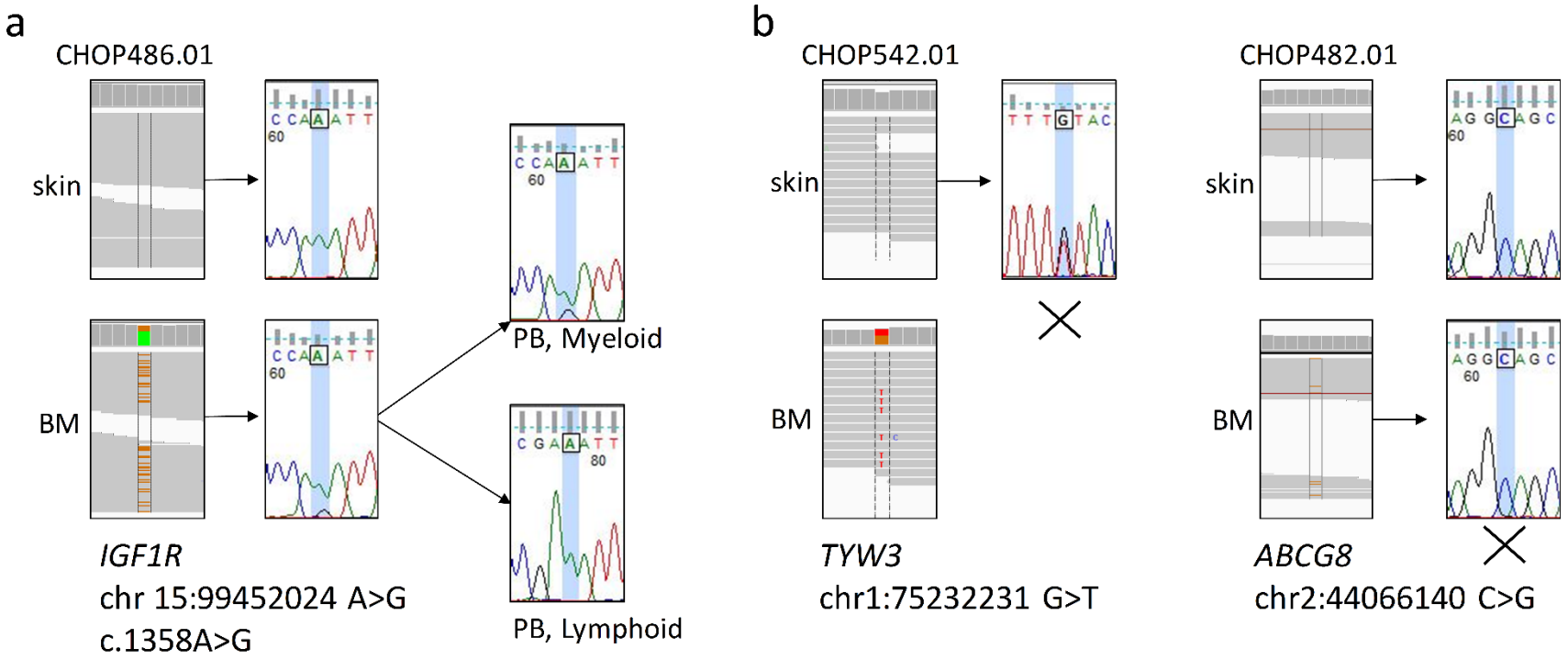
524.01	B*14:02	61.2	0.5	**normal**	**45,XX,-7[5]/46,XX[16]**	**normal**	none		none	**Yes (Subclinical, 1%)**	**None**	**None**	SRSF2, ASXL1	Targeted NGS
390.01	A*68:01, B*14:02	61.5	0.3	**normal**	**46,XX, del(13)(q12q14) [6]/46,XX[12]**	**normal**	**none**	**6WC (~15-20%) →(~5%)**	**none**	**Yes (Moderate, 26%)**	**Yes (Subclinical, 5%)**	**Yes, Subclinical (1.3%)**	**ASXL1, DNMT3A, RUNX1**	Targeted NGS
362.01		1.5	1.1	normal			none		none	None	None	None	none	
387.01		2.6	0.0	**normal**		**46,XY,+3,der(3;15)(q10;p10),-15[12]/46,XY[8]**	none		none	n/a	None	None	none	
450.01		3.2	1.1	normal		normal	none		none	None	None	None	none	WES
474.01		3.3	1.1	normal		normal	none		none	None	None	None	none	
41.01		3.6	12.6	normal		normal	n/a		none	n/a	None	None	none	WES
519.01		4.2	5.1	n/a		normal	n/a		none	n/a	n/a	None	none	
3.01		4.8	1.0	normal		normal	none		none	None	None	None	none	
263.01		5.3	0.6	normal		normal	none		none	Yes (Subclinical, 3.4%)	n/a	n/a	none	
471.01		6.3	10.2	n/a		normal	n/a		6pterp21.32 6pterp21.31 6pterp21.2 6pterp12.1	None	n/a	n/a	none	
510.01		7.6	1.1	normal		normal	none		none	None	None	None	none	
5.01		7.7	5.4	normal		normal	none		none	None	None	None	BCOR	
484.01		8.3	1.1	normal		normal	none		none	None	None	None	none	
487.01		8.8	1.2	normal		normal	none		none	None	None	None	none	
533.01		9.1	1.0	normal		normal	none		none	None	n/a	n/a	none	
488.01		11.2	13.0	normal		normal	n/a		none	n/a	None	None	none	
466.01		12.2	1.2	normal		normal	none		none	n/a	None	None	none	
542.01		13.9	0.3	normal		normal	n/a		none	Yes (Moderate, 25%)	Yes (Moderate, 31%)	n/a	none	
45.01		14.2	6.0	n/a		normal	n/a		none	**n/a**	**None**	**Subclinical (1.9%)**	none	Targeted NGS
376.01		15.2	0.5	n/a		normal	n/a		none	n/a	n/a	None	none	
385.01		18.5	0.0	normal		normal	none		none	Yes*	n/a	n/a	none	
434.01		19.2	1.1	normal		normal	none		none	None	None	None	none	WES
173.01		20.2	0.2	normal		normal	n/a		none	None	None	None	none	

483.01		41.1	0.6	normal		normal	n/a		none	**None**	**Yes, (Subclinical, 6%)**	**None**	none	Targeted NGS
284.01		46.9	2.7	n/a		normal	**6W CN-LOH**		**6pterp21.32 6pterp11.1**	None	None	None	none	Targeted NGS
482.01		50.7	1.8	n/a		normal	n/a		none	Yes, (Subclinical, 2.5%)	Yes (Subclinical, 5%)	Yes (Moderate, 12%)	BCOR	WES, Targeted NGS
168.01		53.5	5.5	normal		normal	n/a		none	n/a	Yes (Moderate, 10-12%)	Yes (Moderate, 10-12%)	none	
486.01		61.2	1.1	normal		normal	n/a		none	None	None	None	ASXL1	Targeted NGS
408.01		65.6	1.6	normal		normal	n/a		none	**n/a**	**None**	**Yes (Subclinical, 1.65%)**	none	Targeted NGS

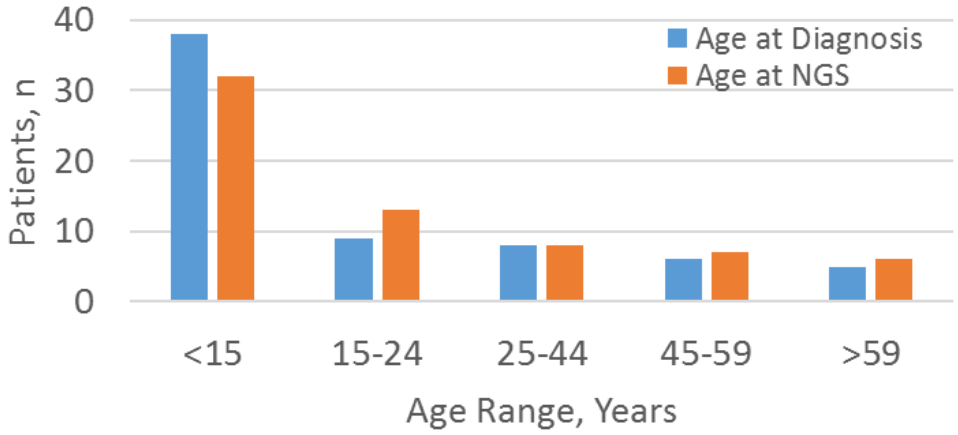
A, HLA-A; B, HLA-B; yrs, years; PNH granulocytes >1%, greater than 1% of granulocytes with PNH immunophenotype as measured by flow-cytometry; Nonsyn. coding, nonsynonymous coding somatic mutations; UTR, somatic mutations in untranslated regulatory regions; Lineage, mutations present in myeloid (M), Lymphoid (L), or both myeloid and lymphoid lineages (Both).

Supplemental Table S12. Hematologic malignancy targeted sequencing gene panel

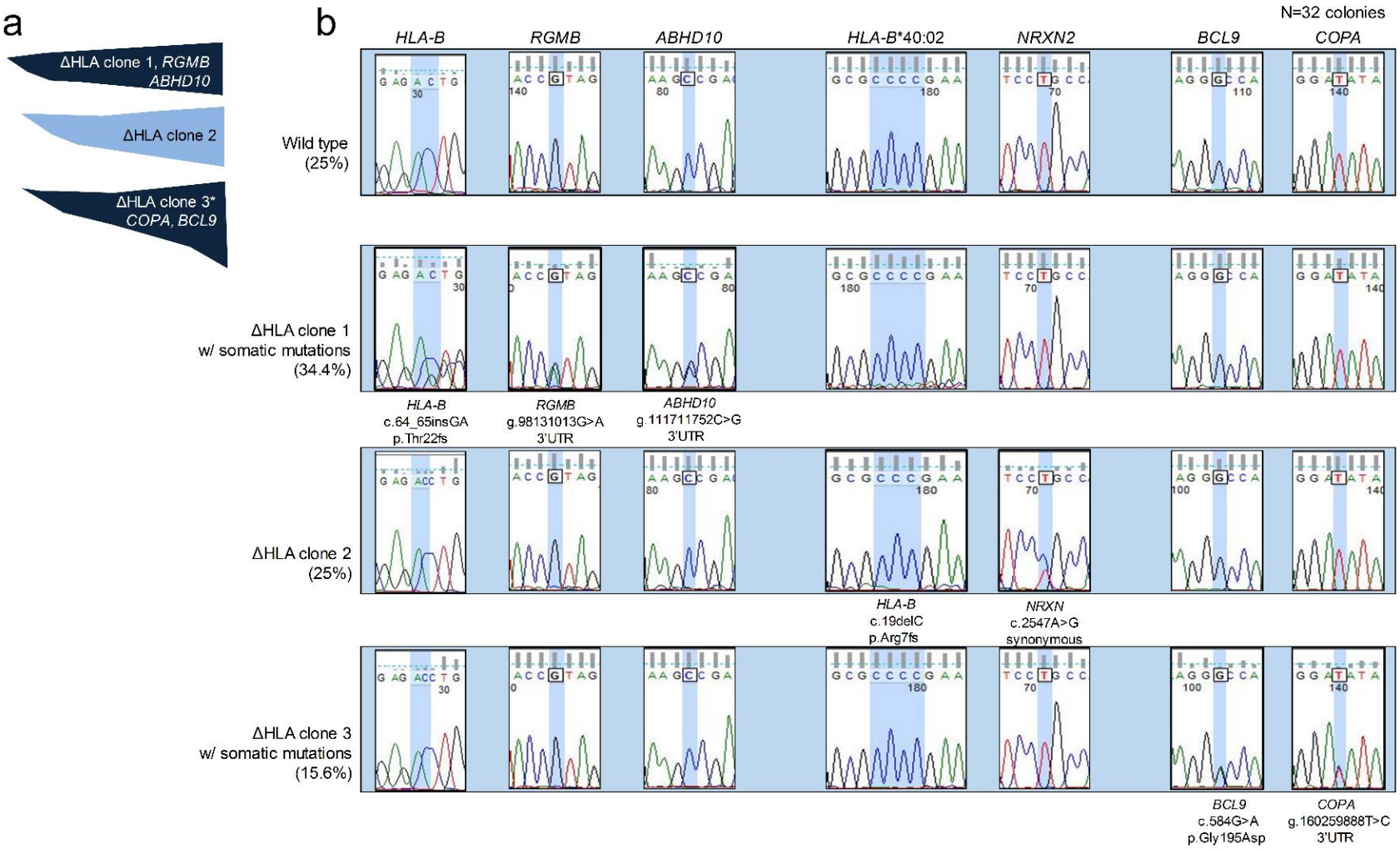
ABL1	*CSF1R*	*GNAS*	*MAPK1*	*NRAS*	*SETBP1*	*TPMT*
ASXL1	*CSF3R*	*HNRNPK*	*MIR142*	*PDGFRA*	*SF1*	*U2AF1*
ATM	*DDX3X*	*IDH1*	*MPL*	*PHF6*	*SF3A1*	*U2AF2*
BCOR	*DNMT3A*	*IDH2*	*MYC*	*POT1*	*SF3B1*	*WT1*
BCORL1	*ETV6*	*IL7R*	*MYCN*	*PRPF40B*	*SMC1A*	*XPO1*
BIRC3	*EZH2*	*JAK2*	*MYD88*	*PTEN*	*SRSF2*	*ZMYM3*
BRAF	*FAM5C*	*KIT*	*NF1*	*PTPN11*	*STAG2*	*ZRSR2*
CALR	*FBXW7*	*KLHL6*	*NOTCH1*	*RAD21*	*TBL1XR1*	
CBL	*FLT3*	*KRAS*	*NOTCH2*	*RIT1*	*TET2*	
CDKN2A	*GATA2*	*MAP2K1*	*NPM1*	*RUNX1*	*TP53*	



Supplemental Figure S1. Validation of putative somatic mutations detected by comparative Whole Exome Sequencing. (a) IGV screenshots of WES of constitutional DNA (skin) and bone marrow (BM) showing a c.1358A>G mutation in *IGF1R* gene present in the bone marrow but not in the paired constitutional DNA, which was confirmed by the corresponding Sanger sequencing chromatographs. The mutation was detected in the immunomagnetically sorted myeloid cell fraction of peripheral blood (PB, Myeloid) but was not detected in the lymphocyte fraction (PB, Lymphoid). (b) An example of two putative somatic mutations identified by WES that failed orthogonal validation by Sanger sequencing. The left-hand panel shows an IGV screenshot of a candidate somatic mutation in *TYW3* gene, which failed verification by Sanger sequencing, due to the variant being constitutional. The right-hand panel shows an IGV screenshot of a low-frequency putative mutation in the *ABCG8* gene, which failed verification by Sanger sequencing due to either being below the level of detection or due to being artefactual.

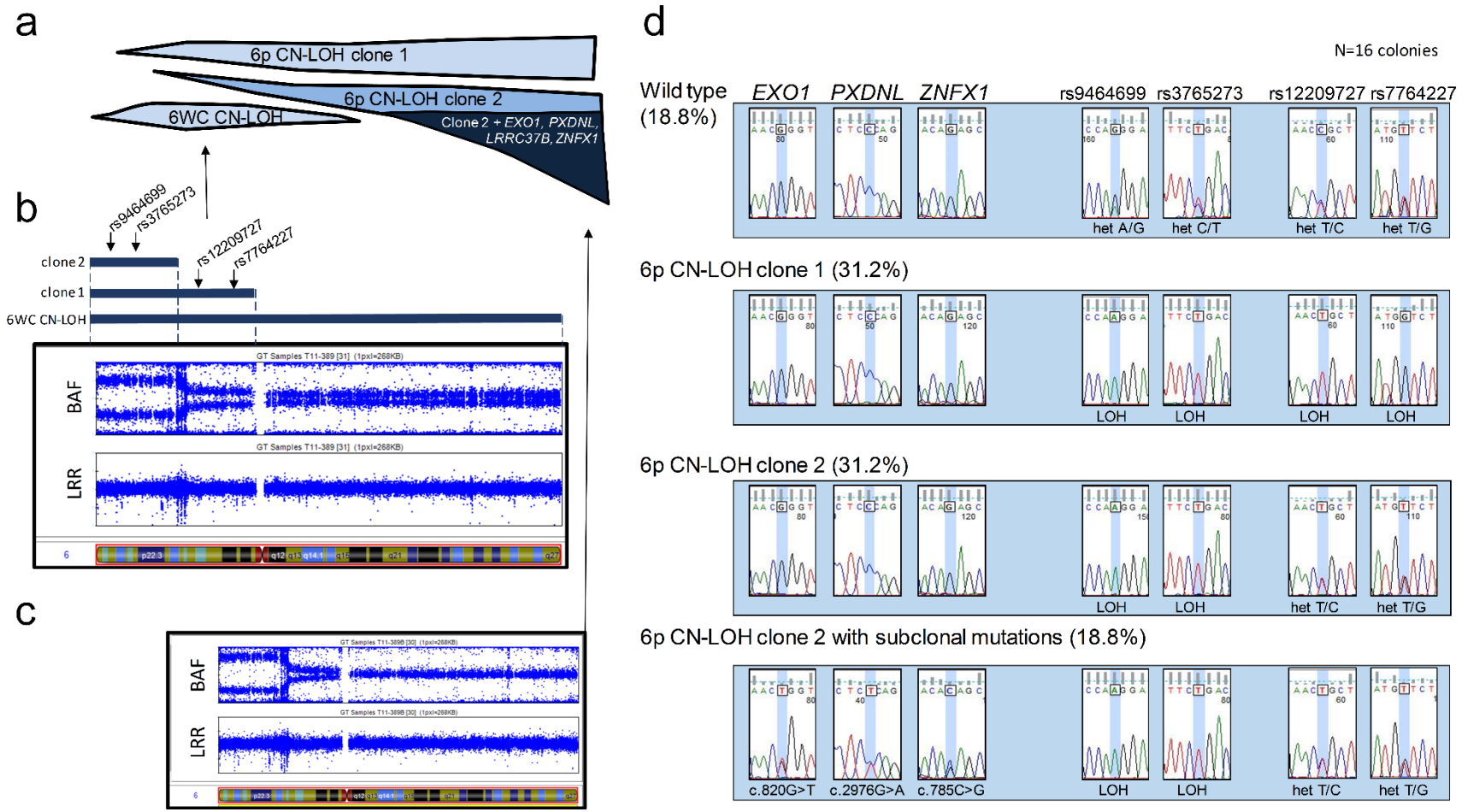


Supplemental Figure S2. Age distribution of 66 patients with acquired aplastic anemia. Shown is a histogram of the patients' age at the time of aplastic anemia diagnosis (blue), and at the time of HLA next-generation sequencing analysis (orange).



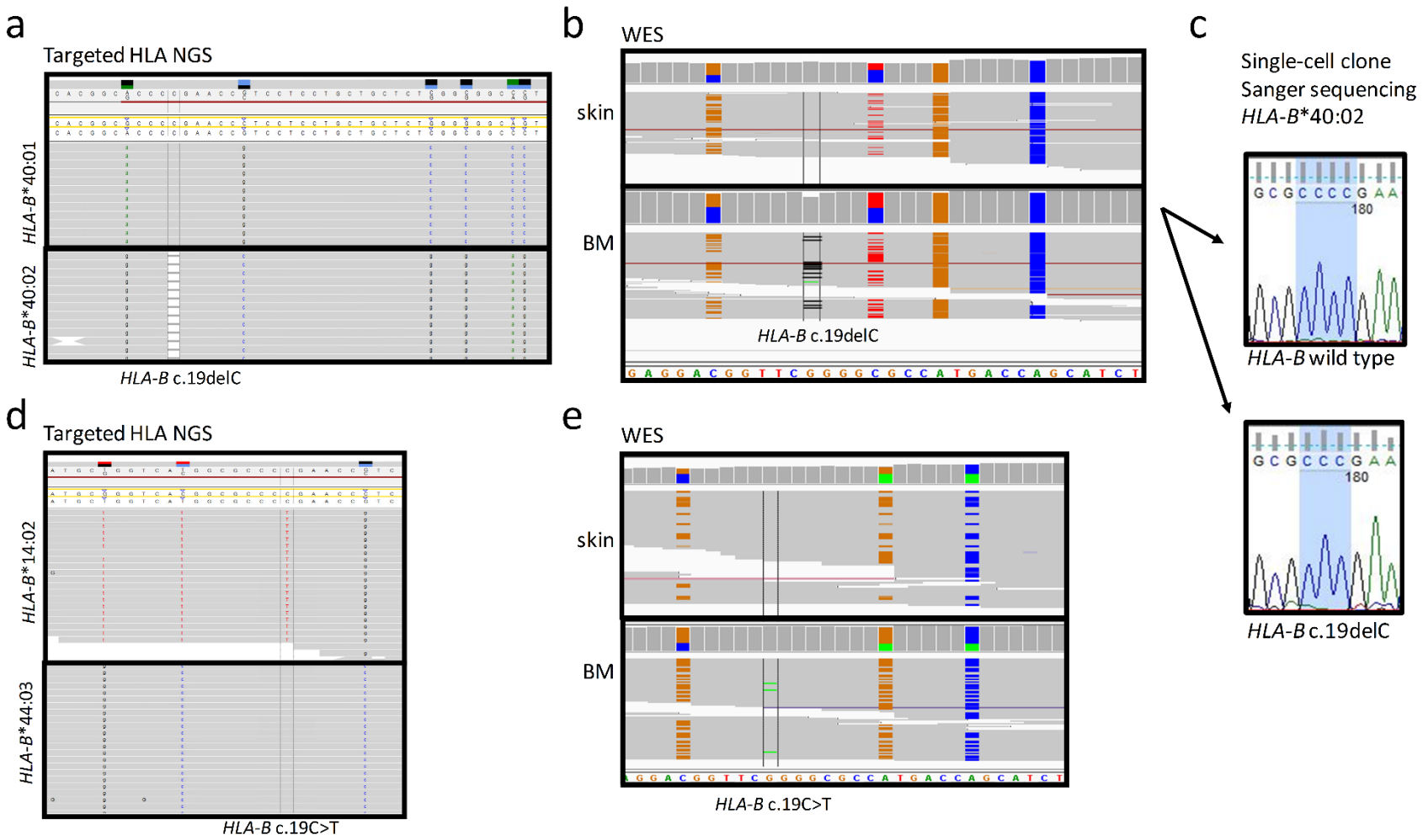
Supplemental Figure S3. Clonal architecture analysis of a patient with three loss-of-function somatic mutations in the *HLA-B40:02 allele demonstrating presence of other somatic mutations within HLA clones.** Targeted NGS of HLA class I

alleles in bone marrow DNA of patient CHOP435.01 identified three loss-of-function mutations *HLA-B**40:02. Using comparative WES we identified multiple non-HLA somatic mutations. Single cell clone genotyping with Sanger sequencing confirmed three independent hematopoietic clones carrying independent loss-of-function mutations in *HLA-B**40:02, with two clones also carrying other non-synonymous or regulatory region mutations. (a) A schematic depicting the patient's hematopoietic clonal architecture. (b) Sanger sequencing chromatographs of 4 representative colonies demonstrating the mutational profile of the three mutant clones, as well as the corresponding chromatographs of wild type cells. HLA mutation p.Q96fs was not amenable to PCR-based genotyping of colony DNA; its presence in clone 3 was inferred by comparing somatic allele frequencies from targeted NGS and WES genotyping.



Supplemental Figure S4. Clonal architecture analysis of a patient with three 6p CN-LOH clones demonstrating subclonal somatic mutations. SNP-A genotyping of bone marrow DNA of patient CHOP284.01 at the time of aAA diagnosis revealed 3

acquired CN-LOH clones: one involved the whole chromosome 6 (6WC) in one clone, and two involved different parts of chromosome arm 6p (6pterp11.1 and 6pterp21.32). Comparative WES of bone marrow and skin DNA identified 4 nonsynonymous coding somatic mutations in *EXO1*, *PXDNL*, *LRRC376* and *ZNFX1* genes. SNP-A analysis of the patient's bone marrow DNA after the patient achieved partial hematopoietic recovery revealed presence of the two 6p CN-LOH clones, and disappearance of the 6WC CN-LOH clone. Sanger sequencing of single cell clones showed that the 4 somatic mutations were subclonal to the 6pterp21.32 CN-LOH clone. (a) A schematic depicting the patient's hematopoietic clonal architecture. (b) SNP-A genotyping of the patient's bone marrow is shown as a pair of scatter plots. The top plot shows B-allele Frequency (BAF, a relative frequency of the minor allele) on the Y-axis, and the chromosomal location on the X-axis. The bottom plot shows Log R Ratio (LRR, a measure of normalized total signal intensity for both alleles) on the Y-axis, and the chromosomal location on the X-axis. In a region with acquired CN-LOH, the copy number (indicated by the LRR) remains constant, while there is a decreased frequency of the heterozygous alleles (indicated by the change of BAF plot). The regions of CN-LOH affected in each of the three clones are depicted by the diagram above the SNP-A plot. The four heterozygous SNP locations used for genotyping CN-LOH clones are indicated above the diagram by their SNP ID numbers. (c) SNP-A genotyping at the time of partial hematopoietic recovery shows disappearance of the 6WC CN-LOH clone. (d) Sanger sequencing chromatographs of 4 representative colonies demonstrating the mutational profile of the two different 6p CN-LOH clones, one of which developed other subclonal somatic mutations, as well as the corresponding chromatographs of wild type cells.



Supplemental Supplemental Figure S5. Validation of inactivating somatic HLA mutations detected by targeted next-generation sequencing. (a) An NGSengine (GeneDx, Utrecht, Netherlands) screenshot of targeted HLA NGS in the bone marrow of

patient CHOP435.01 showing the *HLA-B* c.19delC mutation occurring *in cis* to the *HLA-B**40:02 allele. Allele-defining polymorphisms are shown by nucleotide letters surrounding the c. 19delC mutation. (b) An IGV screenshot showing the *HLA-B* c.19delC mutation detected by WES in the patient's bone marrow (BM), and absent in the patient's skin DNA. (c) A chromatograph of *HLA-B**40:02 allele-specific Sanger sequencing of single-cell clones showing an *HLA-B* c.19delC mutation-positive clone, as well as an example of a wild type clone. (d) An NGSengine (GeneDx, Utrecht, Netherlands) screenshot of targeted HLA NGS in the bone marrow of patient CHOP54.01 showing the *HLA-B* c.19C>T mutation occurring *in cis* to the *HLA-B**14:02 allele. (b) An IGV screenshot showing the *HLA-B* c.19C>T mutation detected by WES in the patient's bone marrow (BM), and absent in the patient's skin DNA.

Supplemental References

1. International Agranulocytosis and Aplastic Anemia Study. Incidence of aplastic anemia: the relevance of diagnostic criteria. . *Blood*. 1987;70(6):1718-1721.

2. Babushok DV, Perdigones N, Perin JC, et al. Emergence of clonal hematopoiesis in the majority of patients with acquired aplastic anemia. *Cancer Genet*. 2015;208(4):115-128.

3. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.

4. American Academy of Pediatrics Council on Child and Adolescent Health: Age limits of pediatrics. *Pediatrics*. 1988;81(5):736.

5. Babushok DV, Xie HM, Roth JJ, et al. Single nucleotide polymorphism array analysis of bone marrow failure patients reveals characteristic patterns of genetic changes. *British journal of haematology*. 2013.

6. Betensky M, Babushok D, Roth JJ, et al. Clonal evolution and clinical significance of copy number neutral loss of heterozygosity of chromosome arm 6p in acquired aplastic anemia. *Cancer Genet*. 2016;209(1-2):1-10.

7. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576.

8. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-192.

9. Duke JL, Lind C, Mackiewicz K, et al. Determining performance characteristics of an NGS-based HLA typing method for clinical applications. *HLA*. 2016;87(3):141-152.

10. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.

11. Robinson J, Soormally AR, Hayhurst JD, Marsh SG. The IPD-IMGT/HLA Database - New developments in reporting HLA variation. *Hum Immunol*. 2016;77(3):233-237.

12. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res*. 2011;39(Database issue):D913-919.

13. Camitta BM. Pathogenesis and treatment of aplastic anemia. *Rinsho Ketsueki*. 1984;25(4):459-469.

14. Rovo A, Tichelli A, Dufour C. Diagnosis of acquired aplastic anemia. *Bone marrow transplantation*. 2013;48(2):162-167.

15. Scheinberg P, Wu CO, Nunez O, Boss C, Sloand EM, Young NS. Treatment of severe aplastic anemia with a combination of horse antithymocyte globulin and cyclosporine, with or without sirolimus: a prospective randomized study. *Haematologica*. 2009;94(3):348-354.

16. Fuhrer M, Durner J, Brunnler G, et al. HLA association is different in children and adults with severe acquired aplastic anemia. *Pediatr Blood Cancer*. 2007;48(2):186-191.

17. Shichishima T, Noji H, Ikeda K, Akutsu K, Maruyama Y. The frequency of HLA class I alleles in Japanese patients with bone marrow failure. *Haematologica*. 2006;91(6):856-857.

18. Katagiri T, Sato-Otsubo A, Kashiwase K, et al. Frequent loss of HLA alleles associated with copy number-neutral 6pLOH in acquired aplastic anemia. *Blood*. 2011;118(25):6601-6609.

19. Zaimoku Y, Takamatsu H, Hosomichi K, et al. Identification of an HLA class I allele closely involved in the auto-antigen presentation in acquired aplastic anemia. *Blood*. 2017.

20. Lombardi ML, Terrazzano G, Cosentini E, et al. Paroxysmal nocturnal hemoglobinuria: significant association with specific HLA-A, -B, -C, and -DR alleles in an Italian population. *Hum Immunol*. 2008;69(3):202-206.

21. Kulasekararaj AG, Jiang J, Smith AE, et al. Somatic mutations identify a subgroup of aplastic anemia patients who progress to myelodysplastic syndrome. *Blood*. 2014;124(17):2698-2704.

22. Lane AA, Odejide O, Kopp N, et al. Low frequency clonal mutations recoverable by deep sequencing in patients with aplastic anemia. *Leukemia*. 2013;27(4):968-971.

23. Brooks AG, Boyington JC, Sun PD. Natural killer cell recognition of HLA class I molecules. *Rev Immunogenet*. 2000;2(3):433-448.

24. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res*. 2015;43(Database issue):D784-788.

25. Middleton D, Menchaca L, Rood H, Komerofsky R. New allele frequency database: http://www.allelefrequencies.net. *Tissue Antigens*. 2003;61(5):403-407.