Cell Reports, Volume *25*

Supplemental Information

Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis

Fernando G. Osorio, Axel Rosendahl Huber, Rurika Oka, Mark Verheul, Sachin H. Patel, Karlijn Hasaart, Lisanne de la Fonteijne, Ignacio Varela, Fernando D. Camargo, and Ruben van Boxtel

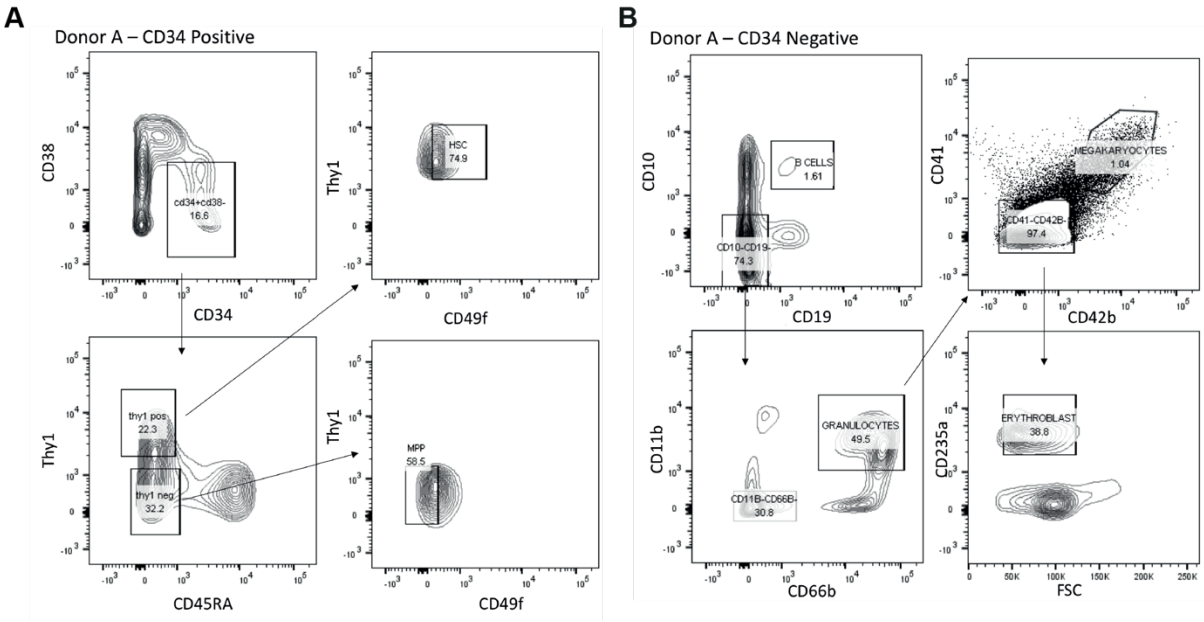

Figure S1. HSC and MPP isolation strategy by FACSs. Related to Figure 1A.
(**A**) Representative FACS plot for purifying HSCs and MPPs, starting with a CD34+ enriched bone marrow-derived cell suspension.
(**B**) Representative FACS plot for isolation of pre-B-cells, granulocytes, megakaryocyte progenitors and erythroblasts.

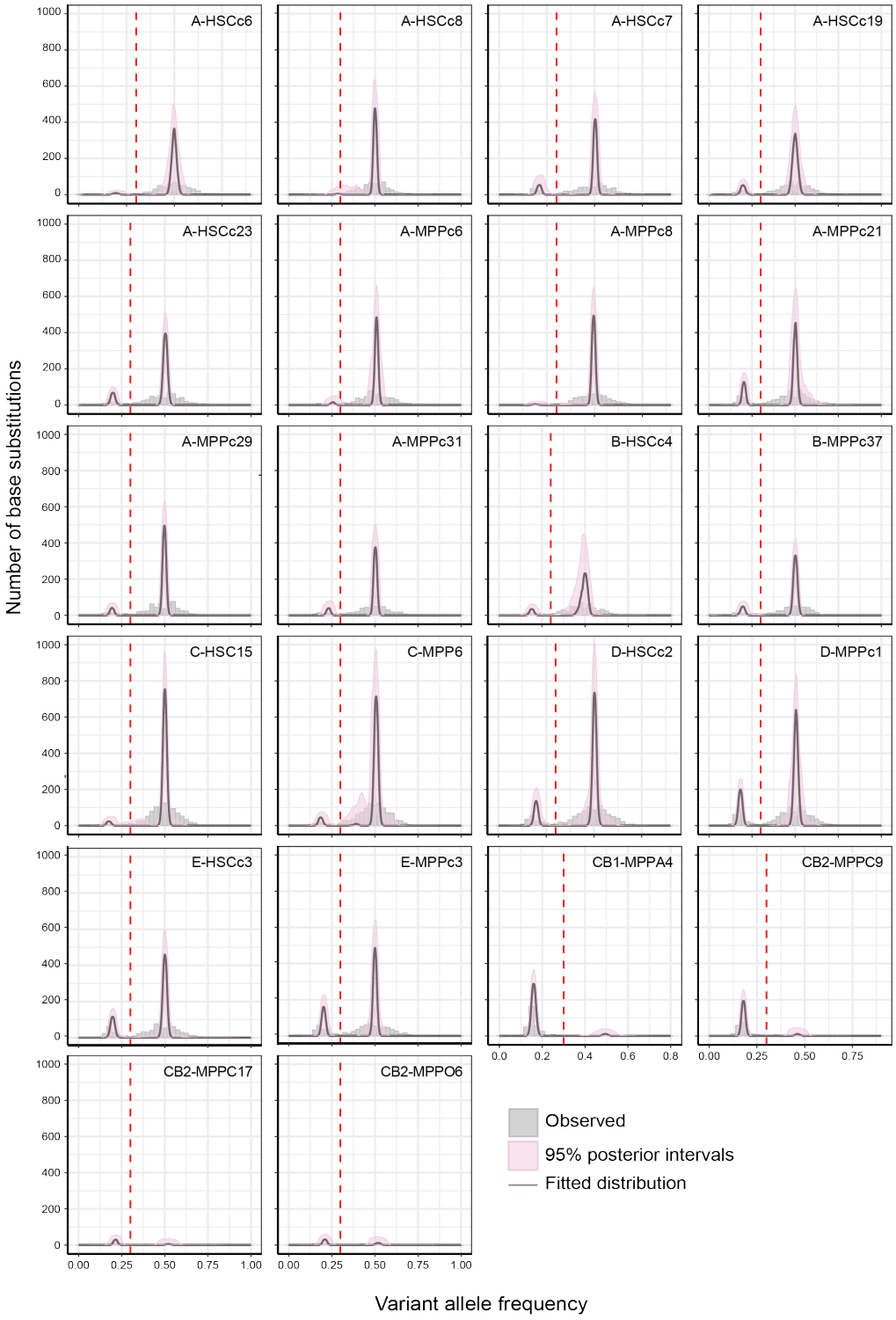

Figure S2. Variant allele fractions (VAF)of base substitutions in sequenced clones. Related to STAR methods section ‘ Mutation calling and filtering’.
Histograms of the variant allele frequency of all detected single base substitutions before the last filtering step (VAF > 0.3). Clonal heterozygous mutations peak at VAF = 0.5. A threshold of VAF 0.3 was used to obtain mutations that were clonal in the organoid culture and thus present in the original sorted HSPC. Mutations aquired during or after clonal culture have lower VAFs and are therefore excluded. Dirichlet modeling was used to determine the clonality of each culture, and a peak at VAF = 0.5 indicates mutations aquired during life. Shaded area represents the 95% posterior confidence intervals for the fitted distribution. (pink area). In most samples, two clusters of mutations can be identified.

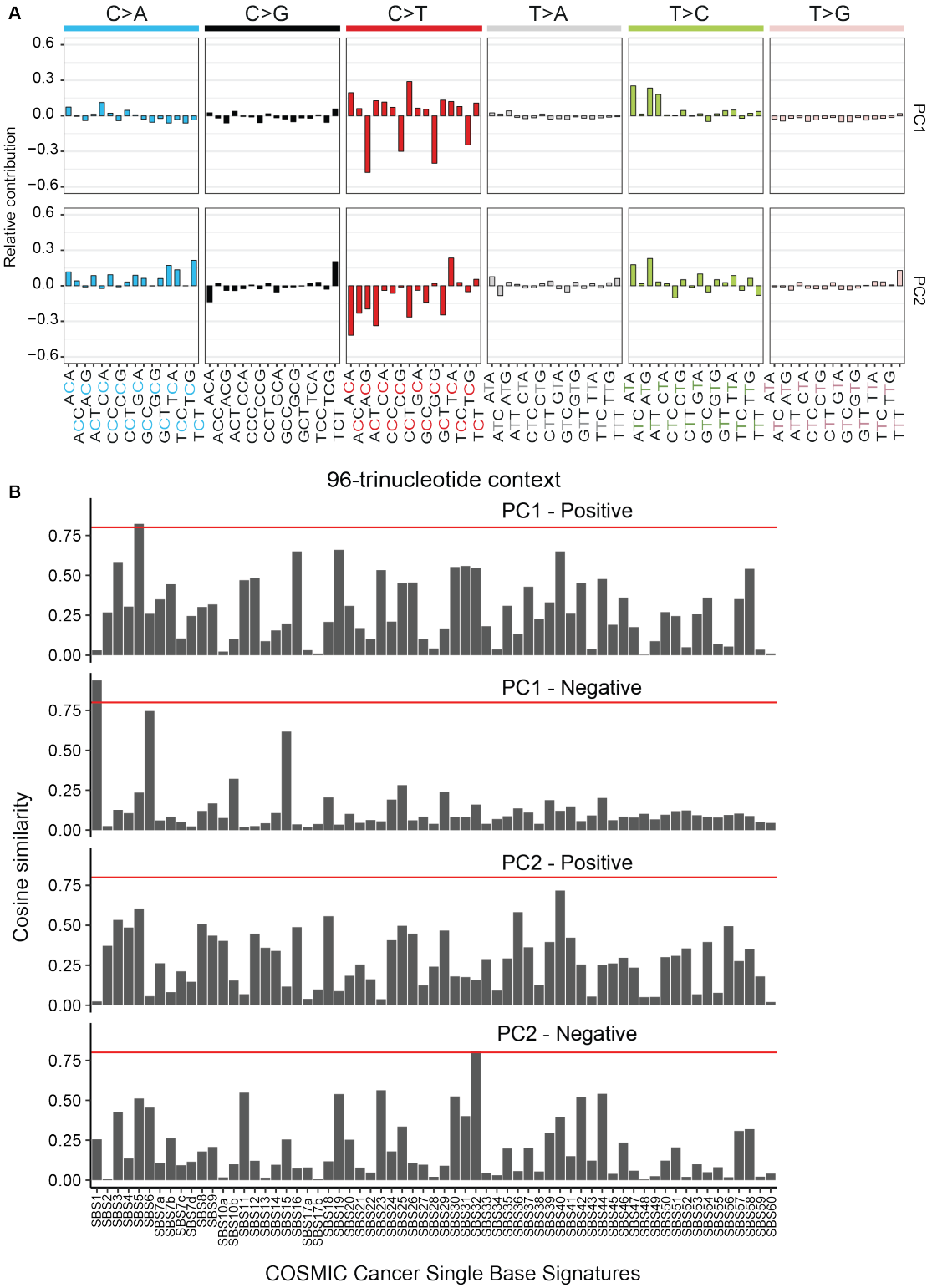

Figure S3. Principal components in PCA correspond to signature 1, 5 and 32. Related to Figure 3A.

(**A**) Contribution of each base substitution to the first two principal components. Values acquired from the *prcom* rotation values of the PCA performed in Figure 3A.

(**B**) Cosine similarity for all of the COSMIC single base substitution signatures. PC1 Positive, PC1-Negative and PC2-Negative have a cosine similarity of < 0.8 with Signature 5,1 and 32 respectively and were indicated on the axes on Figure 3A.

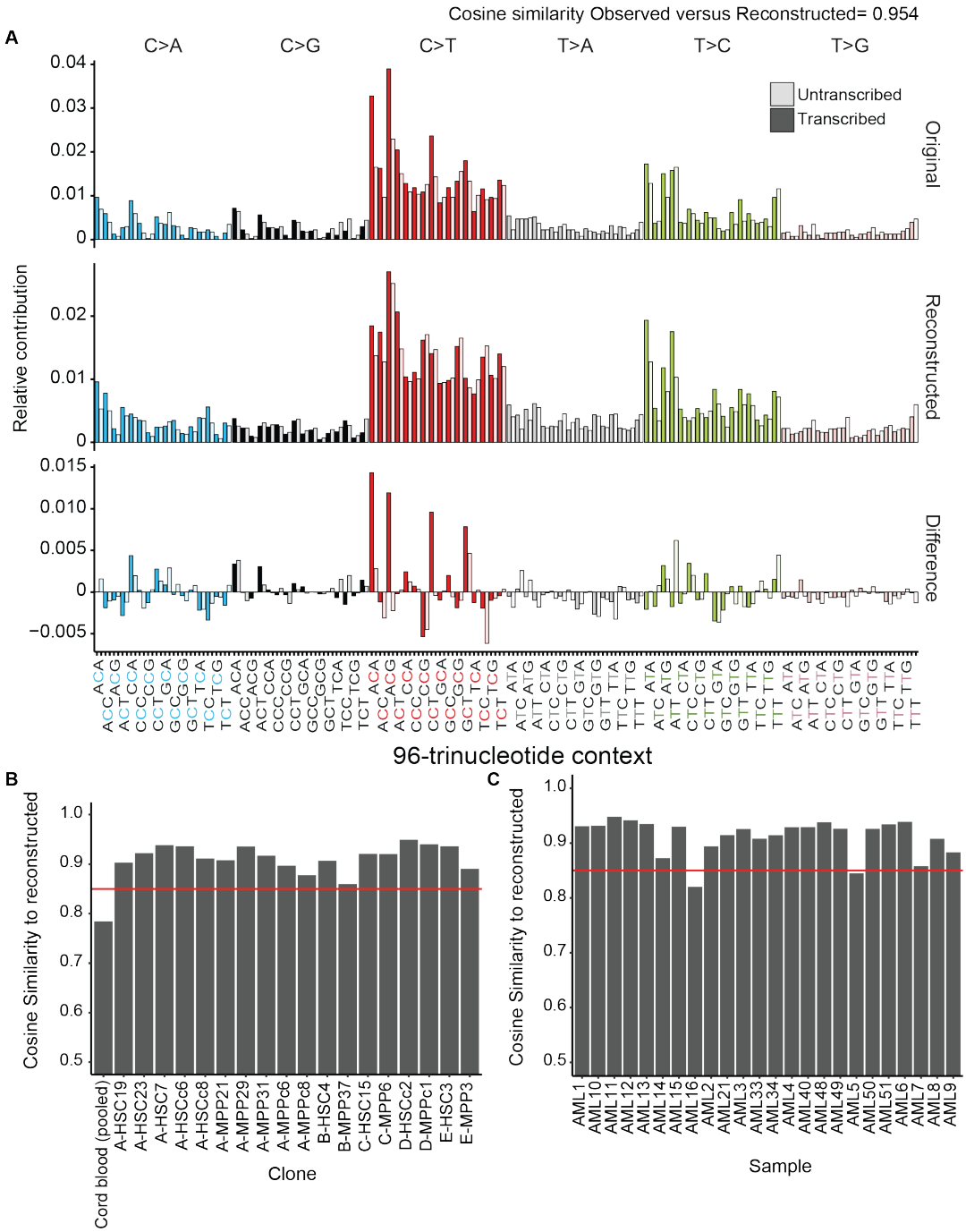

Figure S4. Absolute and transcriptional strand mutational profiles and cosine similarity of HSPC and AML mutation profiles to reconstructed profiles. Related to Figure 3D-E.

(**A**) Upper graph depicts transcriptional strand bias profiles from HSPCs (pooled) compared with the reconstructed 192-nucleotide matrix using transcriptional-strand signatures. (middle graph). Lower graph depicts relative difference between observed and expected profiles. Cosine similarity between observed and reconstructed mutational profiles indicated above figure.

(**B**) Cosine similarity of HSPC and cord blood mutation profiles to their respective reconstructed mutation profiles with Signature 1, 5 and 32. All four cord blood samples were pooled, as mutation load in these samples was low.

(**C**) Cosine similarity of AML mutation profiles to their respective reconstructed mutation profiles with Signature 1, 5 and 32.

Table S1. Overview of sample and donor information. Related to Figure 2.

HPSC	Donor	Age (years)	Gender	Cell type	Surveyed genome (%)*	No. base substitutions†	No. of unique base substitutions ‡
A-HSCc6	A	33	Male	HSC	96.6	499	498
A-HSCc8	A	33	Male	HSC	96.7	543	542
A-HSCc7	A	33	Male	HSC	90.8	464	462
A-HSCc19	A	33	Male	HSC	96.1	554	553
A-HSCc23	A	33	Male	HSC	95.4	527	527
A-MPPc6	A	33	Male	MPP	96.1	536	534
A-MPPc8	A	33	Male	MPP	96.8	505	505
A-MPPc21	A	33	Male	MPP	96.1	592	592
A-MPPc29	A	33	Male	MPP	96.1	548	545
A-MPPc31	A	33	Male	MPP	95.3	433	432
B-HSCc4	B	26	Male	HSC	96.4	504	502
B-MPPCc37	B	26	Male	MPP	96.8	423	421
C-HSCc15	C	55	Female	HSC	96.5	910	906
C-MPPc6	C	55	Female	MPP	96.4	1018	1014
D-HSCc2	D	63	Male	HSC	92.4	859	859
D-MPPc1	D	63	Male	MPP	92.4	818	818
E-HSC3	E	41	Female	HSC	92.8	611	611
E-MPP3	E	41	Female	MPP	92.7	580	580
CB1-MPPA4	CB1	0	Female	MPP	96.4	37	-
CB2-MPPC9	CB2	0	Male	MPP	89.0	44	44
CB2-MPPC17	CB2	0	Male	MPP	94.1	32	32
CB2-MPPO6	CB2	0	Male	MPP	91.4	45	45

* Percentage of the non-N autosomal genome with 20x coverage in both blood progenitor and reference sample.

† Number of somatic base substitutions detected within surveyed genome.

‡ Number of unique somatic base substitutions compared to other clones sequenced from donor.

Table S2. Validation Rates of SNVs in HSPCs. Related to STAR methods section 'smMIP analysis of SNVs'.

HSPC	Number of Amplicon Loci*	Validated Amplicon Loci	Validation rate
AC63HSC	54	48	0.888888889
AC63MPP	41	37	0.902439024
BCHHSC	65	61	0.938461538

* Amplicon Loci sequenced at least 20x depth.

Table S3. Non-synonymous base substitutions in sequenced HSPCs. Related to Figure 2.
HSPC clone, clone identifier (see Table S1), Chromosome of the variant, Position on the chromosome, Reference allele and mutated alternative allele, Mutation type of the variant, Gene in which variant is located, Amino acid change, allelic fraction of the variant.

Table S4. Oligonucleotide sequences for amplicon sequencing of shared mutations and smMIP validation. Related to Figure 4 and Table S2.
(**A**) Variants and primer pairs used to sequence additional HSPC clones and mature bood populations for shared mutations in HSPCs of donor A.
(**B**) Primer sequences used for smMIP validation of variants.