The American Journal of Human Genetics, Volume *101*

Supplemental Data

Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life

Rocio Acuna-Hidalgo, Hilal Sengul, Marloes Steehouwer, Maartje van de Vorst, Sita H. Vermeulen, Lambertus A.L.M. Kiemeney, Joris A. Veltman, Christian Gilissen, and Alexander Hoischen

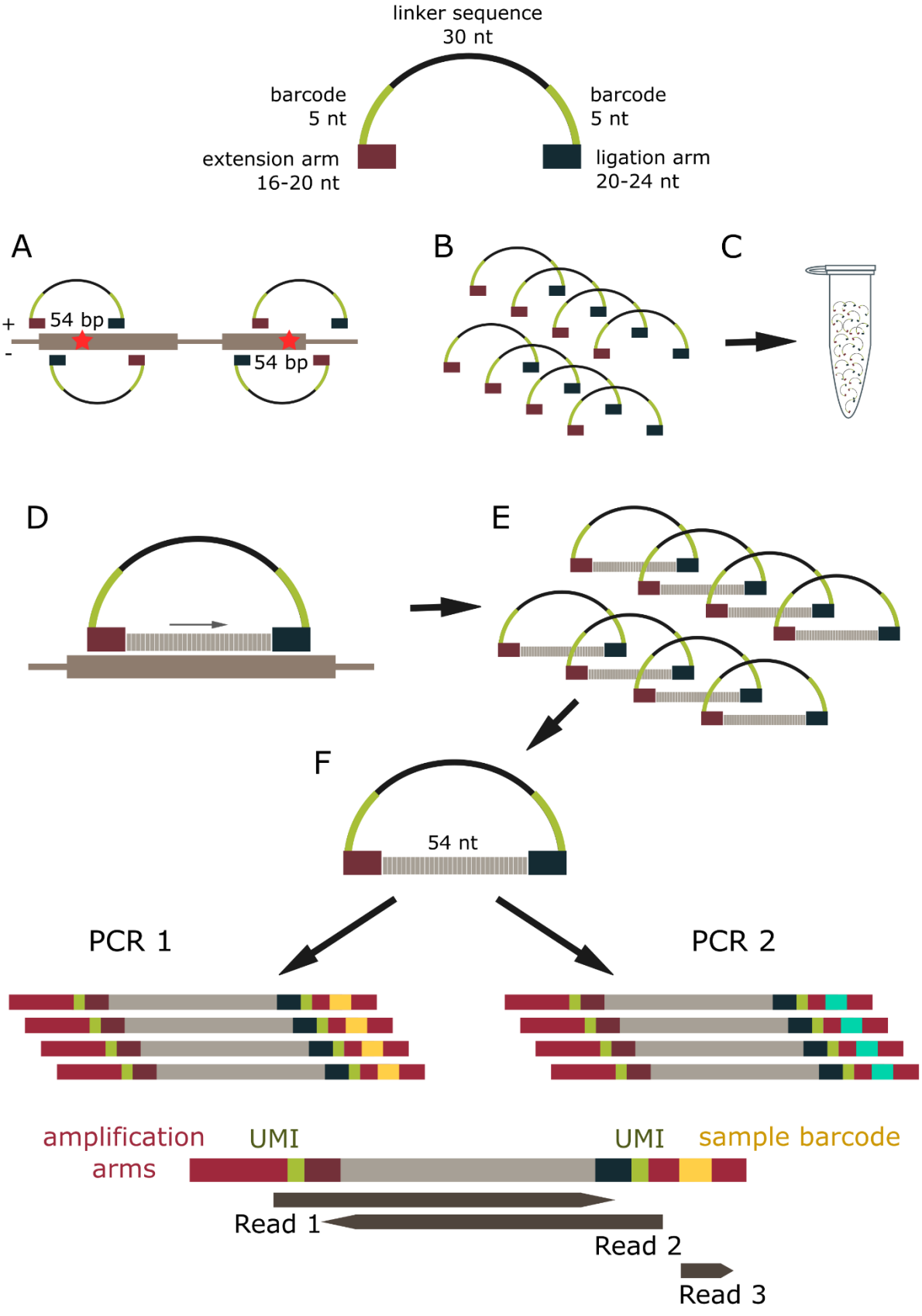
linker sequence
30 nt
barcode
5 nt
barcode
5 nt
extension arm
16-20 nt
ligation arm
20-24 nt
A
+
-
54 bp
54 bp
B
C
D
E
F
54 nt
PCR 1
PCR 2
amplification
arms
UMI
UMI
sample barcode
Read 1
Read 2
Read 3

Figure S1. Overview of smMIP protocol used in this study. Each smMIP is a single stranded 80 nucleotide-long DNA molecule consisting of the extension and the ligation arm (shown in burgundy and blue, respectively) which together are 40 nucleotides long and are designed to be complementary to the targeted region. The two arms are connected by a 30 nucleotide-long linker sequence (in black). Each smMIP contains a unique molecule identifier (UMI) composed by 2 barcodes of random 5 nucleotide sequences (shown in green). The smMIPs are designed for double tiling of regions of interest containing mutations (shown as red star) on the plus and the minus strand (A). The smMIPs are ordered as long oligonucleotides (B), after which they are pooled and phosphorylated (C). Individual smMIPs were pooled equimolarly and phosphorylated using T4 polynucleotide kinase and 10x T4 DNA ligase buffer supplemented with 10mM ATP (New England Biolabs). DNA capture is performed by mixing the phosphorylated smMIP probes with the DNA, dNTPs, polymerase and ligase to form the reverse complement of the region of interest to which the probe binds and ligate the probe into a circular single strand of DNA (D). The smMIP capture was performed on 8µl of input DNA (200 ng) supplied with 17µL of capture mixture (0.01µl Ampligase DNA ligase (100U/µl, Illumina), 2.5µl 10x ampligase buffer (Illumina), 0.28µl phosphorylated smMIP pool dilution (corresponding to a DNA:smMIP ratio of 1:8000), 0.32µl Hemo Klentaq (10U/µl, New England Biolabs), 0.03µl dNTPs (0.25mM) and 13.86µl H2O). The capture mix was incubated for 18-22 hours at 60°C. The mix is then digested with exonuclease to remove all linear DNA molecules (E). Immediately after capture, the mix was cooled and treated with exonuclease (0.5µl Exonuclease I (New England Biolabs), 0.5µl Exonuclease III (New England Biolabs), 0.2µl 10x Ampligase buffer (Illumina) and 0.8µl H2O for 45 minutes at 37°C and 2 minutes at 95°C to inactivate the exonucleases). We subsequently separated the captured and circularized molecules in two separate technical replicates and performed PCR amplification separately (F), using a sample and PCR-specific barcode (shown in yellow and cyan). Each exo-treated sample was split in two technical replicates of 10µl, which were then amplified and barcoded by PCR independently (1.25µl of barcoded reverse primer (10µM), 12.5µl 2x iProof (BioRad Laboratories), 0.125µl forward primer (100µM) and 1.8µl H2O). The PCR products were run on gel, pooled, purified using AmpureXP beads (Agencourt) and sequenced on an Illumina Nextseq platform using 2×79bp reads.

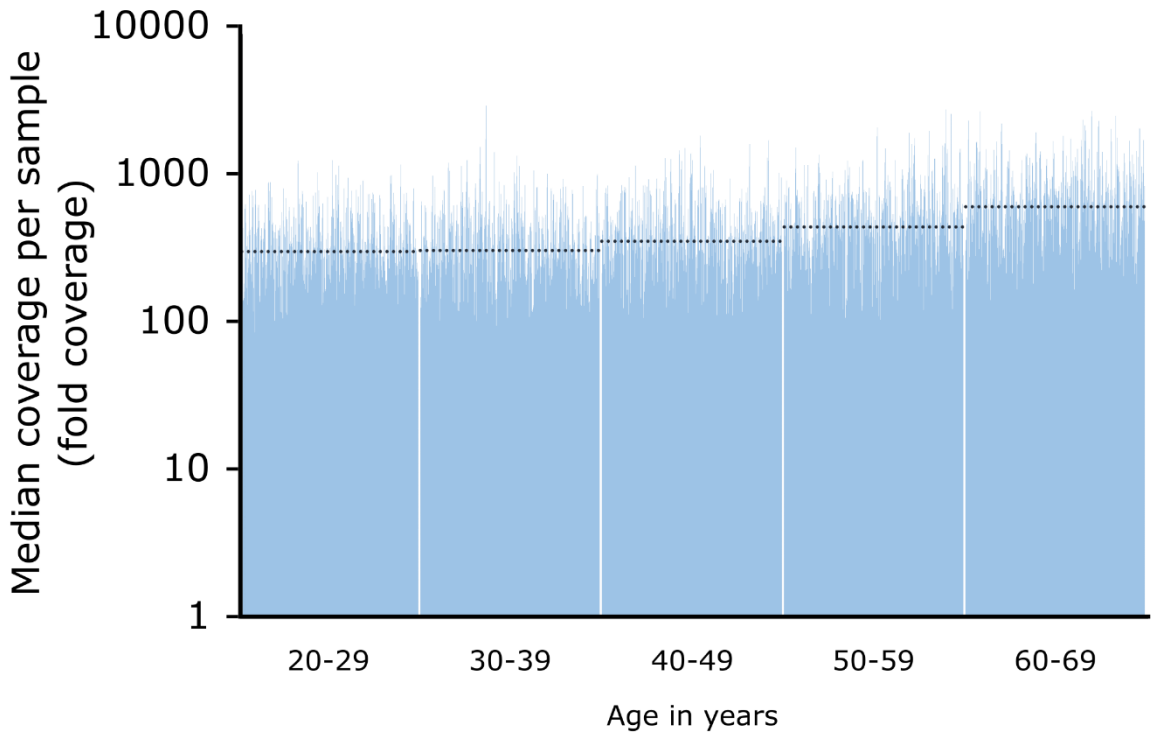

Figure S2. Median unique sequencing coverage per sample, corresponding to the number of unique DNA molecules sequenced per sample and per position after removal of PCR duplicates. Each sample is represented here by two bars, corresponding to each technical replicate. The median unique sequencing coverage per age group is shown as a horizontal dotted line. The overall median unique sequencing coverage was 418-fold per replicate and 845-fold per sample.

	Mean age	Median age	Total	Included	Male	Male (%)	Female	Female (%)
20-29	24.7	25	400	396	195	49.5	201	50.5
30-39	34.6	35	405	402	202	50.2	200	49.8
40-49	44.7	45	404	404	202	50.0	202	50.0
50-59	54.3	55	403	402	202	50.2	200	49.8
60-69	64.5	64	402	402	201	50	201	50
Total	44.6	45	2014	2006	1002	50.0	1004	50.0

Table S1. Age and sex of participants in our study.

Table S2. Selection of loci to screen for somatic mutations in blood of healthy controls. Previous reports used include Xie *et al*, Genovese *et al*, Jaiswal *et al* and McKerrell *et al*.[1-4]

Table S3. Sequencing coverage per smMIP and per age group.

Table S4. Positions used for the generation of pileup files.

Table S5. List of all somatic mutations identified in coding regions in this study. Samples with more than one mutation are marked with a superscript letter for identification.