Supplementary Appendix

Clonal hematopoiesis and therapy-related myeloid neoplasms following neuroblastoma treatment

Tim H.H. Coorens*[1], Grace Collord*[1,2,3], Wanhua Lu[4], Emily Mitchell[1,4,5], Jannat Ijaz[1], Thomas Roberts[4], Thomas R.W. Oliver[1], G.A. Amos Burke[4], Michael Gattens[4], Emmy Dickens[4], Jyoti Nangalia[1,4,5,6], Marc Tischkowitz[7], John Anderson[8,9], Adam Shlien[10], Anna L Godfrey[4], Matthew J Murray[4,11#], Sam Behjati[1,4,12#‡]

[1]Wellcome Sanger Institute, Hinxton, CB10 1SA, UK.

[2]Department of Haematology, University College London Hospital, London, UK.

[3]Department of Haematology, University College London Cancer Institute, London, UK.

[4]Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK.

[5]Department of Haematology, University of Cambridge, Cambridge, CB2 0QQ, UK.

[6]Cambridge Stem Cell Institute, Cambridge, Cambridgeshire, UK CB2 2XY

[7]Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, CB2 0QQ, UK.

[8]UCL Great Ormond Street Institute of Child Health, London, WC1N 1EH, UK.

[9]Great Ormond Street Hospital for Children NHS Foundation Trust, London, WC1N 3JH, UK.

[10]The Hospital for Sick Children, Toronto, Ontario, Canada.

[11]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, UK.

[12]Department of Paediatrics, University of Cambridge, Cambridge, CB2 0QQ, UK.

*These authors contributed equally

[#]Co-directing

[‡]Corresponding author

Contents
Page

Supplemental methods

1. Samples and ethical approval

All human material was obtained from participants in the study "Investigating how childhood tumors and congenital disease develop" (NHS National Research Ethics Service reference 16/EE/0394). Written informed consent for research, including comprehensive genetic analysis, was obtained from patients' guardians in accordance with the Declaration of Helsinki and the study protocols approved by the relevant ethics committees. Data on demographic and diagnostic features (including age, previous cancer treatment, cytogenetics and hematological indices) were clinically available at the time of patient recruitment. Genomic DNA was extracted from peripheral blood, bone marrow mononuclear cells or mobilised peripheral blood mononuclear cells, and DNA integrity was verified by visualization on a 0.8% agarose gel.

2. Library preparation and sequencing

2.1 Whole genome sequencing

Short insert (500bp) genomic libraries were constructed and 150 base pair paired-end sequencing clusters were generated on the Illumina HiSeq X Ten or NovaSeq 6000 platform according to Illumina no-PCR library protocols. An overview of samples is shown in **Table S1**.

2.2 Targeted sequencing

We designed a custom cRNA bait set (SureSelect, Agilent, UK, ELID ID: 3129591) complementary to all coding exons of 95 genes recurrently mutated in clonal hematopoiesis and hematological cancers, including the genes most frequently implicated in paediatric myeloid neoplasms (**Table S2**). Genomic DNA was sheared using the Covaris M220. Equimolar pools of 10 libraries were prepared and sequenced on the Illumina HiSeq 2000 using 75 base paired-end sequencing as per Illumina and Agilent SureSelect protocols.

2.3 Sequencing data deposition

Targeted and whole genome sequencing data are deposited at the European Genome Phenome Archive (EGA) under accessions EGAD00001006423 and EGAD00001006424 (https://www.ebi/ac.uk/ega).

3. Variant calling

3.1 Variant calling from whole genome sequences

Whole genome sequences were mapped to the GRCh37d5 reference genome using the Burroughs-Wheeler Aligner (BWA-mem).[1]

Single-nucleotide variants (SNVs) were called using an extensively validated in-house version of Cancer Variants through Expectation Maximization (CaVEMan, 1.11.2)[2,3] against an *in silico* unmatched normal based on the reference genome. CaVEMan compares sequencing reads between study and nominated normal samples and uses a naïve Bayesian model and expectation-maximization approach to calculate the probability of a somatic variant at each

base (https://github.com/cancerit/CaVEMan). Beyond the standard post-processing filters of CaVEMan, we removed variants resulting from artifactual mapping by setting the median alignment score of reads supporting a mutation as greater than or equal to 140 (ASMD>=140) and requiring that fewer than half of the reads were clipped (CLPM=0). Across all samples from one patient and, where applicable, their parents, we re-genotyped the SNVs and indels using a cut-off for read mapping quality (30) and base quality (25).

Small insertions and deletions (indels) were identified using an in-house version of Pindel (v2.2.5; https://github.com/cancerit/cgpPindel).[4]

Germline and somatic point mutations were differentiated based on a one-sided exact binomial test. For this test, the null hypothesis is that the number of reads supporting the variants across copy number normal samples is drawn from a binomial with p=0.5 (p=0.95 for copy number equal to one), and the alternative hypothesis drawn from a distribution with p<0.5 (or p<0.95). Resulting p-values were corrected for multiple testing with the Benjamini-Hochberg method[3] and a cut-off was set at $q < 10^{-5}$ to minimize false positives.

Putative variants were also removed if they were present in a region of the genome with a consistently low or high coverage across all copy number-neutral samples from the same patient. This amounted to a mean coverage of between 50 and 110 for Patients 1 (PD31013) and 2 (PD42747), and between 50 and 150 for Patient 3 (PD34954) (**Figure S2**).

When an SNV in the t-MDS or t-AML clone was shared with blood or bone marrow samples, we distinguished true sharing of somatic variants from noise using a beta-binomial model of a site-specific error rate as previously described.[5] For this, we constructed an unmatched normal panel that was specific to the sequencing platform used (i.e. Illumina HiSeq X Ten or NovaSeq 6000). The resulting probabilities were again corrected for multiple testing using the Benjamini-Hochberg method. All shared SNVs were further visually inspected using the genome browser, Jbrowse.[5]

Copy number variants (CNVs) were called using the ASCAT (v4.1.2; https://github.com/cancerit/ascatNgs)[6] and Battenberg (v3.3.0; https://github.com/cancerit/cgpBattenberg)[7] algorithms. In addition, for Patient 1 (PD31013), the availability of parental whole genome sequences allowed for full phasing of the copy number variants and identification of the specific parental allele that was affected by each chromosomal abnormality.

For Patient 1, we formally quantified the sensitivity of Battenberg to determine the smallest clone size for which the algorithm is able to detect del(7q) or monosomy 7 given our sequencing depth and information from parental DNA (and hence complete phasing). We simulated allele counts for informative SNPs within the 1000Genomes single nucleotide polymorphism (SNP) locations in Patient 1 (PD31013), which amounted to 1,654,457 SNPs across her 22 autosomes. Allele counts were derived from a binomial distribution (with p=0.5) given a Poisson distribution of coverage (N):

$$allele\ counts \sim Bin(n \sim Pois(N), p = 0.5)$$

We introduced a del(7q) or monosomy 7 to assess the sensitivity for a range of clone sizes (s). The paternal counts (the allele that retained a copy) were simulated as follows:

$$allele\ counts \sim Bin(n \sim Pois(N/(1-\frac{s}{2}), p = 0.5 * (1-\frac{s}{2}))$$

The maternal counts followed as the difference between the coverage and the paternal counts. These allele counts were then fed into the standard Battenberg pipeline (without the statistical phasing implemented in the base algorithm). We used N=(25, 50, 100) and s=(0, 0.001, 0.002, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.06, 0.065, 0.07, 0.075, 0.08, 0.085, 0.09, 0.095, 0.1) as our coverage and clone size range, respectively. At the relevant coverage, the minimum clone size at which del(7q) or monosomy 7 was reliably detectable amounted to 4%.

Structural variants (SVs) were called using BRASS (v6.1.2; https://github.com/cancerit/BRASS)[8] and SVs were retained if having a paired end read support of at least 5 or where reconstruction of the breakpoint via assembly was possible (see **Table S3**).
All filtered SNVs (**Table S4** and **Table S5**) will be made in excel format available on github along with relevant code.

3.2 Variant calling from targeted sequencing data

Targeted sequencing reads were aligned to the reference genome (GRCh37d5) using the short read Burrows-Wheeler aligner (BWA-ALN).[8] Unmapped reads, PCR duplicates and reads mapping to regions outside the target regions (merged exonic regions + 10bp either side of each exon) were excluded from analysis. Sequencing depth at each base was assessed using Bedtools coverage v2.24.0.[9]

Somatic single nucleotide variants (SNVs) were called using Shearwater, an algorithm developed for detecting subclonal mutations in deep sequencing experiments (https://github.com/gerstung-lab/deepSNV v1.21.5)[10–12] considering only reads with minimum nucleotide and mapping quality of 25 and 40, respectively. This algorithm models the error rate at individual loci using information from multiple unrelated samples.

To seek subclonal copy number variants in targeted sequencing data, we applied FACETS (Fraction and Allele-Specific Copy Number Estimates from Tumor Sequencing), an allele-specific copy number analysis (ASCN) method.[13]

3.3 Annotation of candidate driver mutations

Substitutions and indels were annotated according to ENSEMBL version 58 using VAGrENT (http://cancerit.github.io/VAGrENT/)[14] for transcript and protein effects.

To identify potential driver events, we considered variants presenting in established cancer genes (Tate et al., 2019).[15] Tumour suppressor coding variants were considered if they were

annotated as functionally deleterious by VAGrENT, or alternatively if they were disruptive rearrangement breakpoints or homozygous deletions. Additionally, homozygous deletions were required to be focal (<1 Mb in size) or constitute a known contiguous gene syndrome implicated in t-MN.[16] Mutations in oncogenes were considered driver events if they were located at previously reported canonical hot spots (point mutations) or amplified the intact gene. Amplifications also had to be focal (<1 Mb) and increase the copy number of oncogenes to a minimum of 5 copies.

To identify potential cancer predisposition germline variants, all non-somatic variants were screened against the ClinVar database[17] and Human Gene Mutation Database (HGMD).[18]

4. Mutational signature analysis

We fitted reference mutational signatures[19] to 96 trinucleotide mutation count data using the SigFit package[20].

We assigned the probability of a certain mutation arising from certain signature as the weighted fraction of the product of signature contribution and mutational probability for that signature:

$$p_{C[C>T]C,A} = \frac{C_A P_{C[C>T]C,A}}{\sum_{i}^{N} C_i P_{C[C>T]C,i}}$$

In other words, the odds of signature A causing a C[C>T]C mutation is the product of its contribution (C_A) to the sample multiplied by the probability that signature A generates a C[C>T]C change. This is normalised by the sum of these products for all signatures in the sample of interest. For example, the *PTPN11* G503E mutation in the t-MN of PD31013 is a C[C>T]C change, so given the contribution of signatures, the probability breakdown is the following:

Signature	Contribution	P of C[C>T]C	Normalised probability	Etiology
SBS1	0.023	0.0003	$<10^{-4}$	Spontaneous deamination of 5-methylcytosine (clock-like signature)
SBS5	0.058	0.019	0.01	Unknown (clock-like signature)
SBS18	0.039	0.007	0.003	Oxidative stress
SBS31	0.370	0.164	0.58	Platinum chemotherapy
SBS35	0.510	0.083	0.41	Platinum chemotherapy

Hence, there is a 99% probability of the *PTPN11* G503E mutation arising due to platinum-induced mutagenesis.

The *SETBP1* D868G arose long after cisplatin and carboplatin treatment had been completed and only has a 66% probability of arising due to platinum-associated signatures:

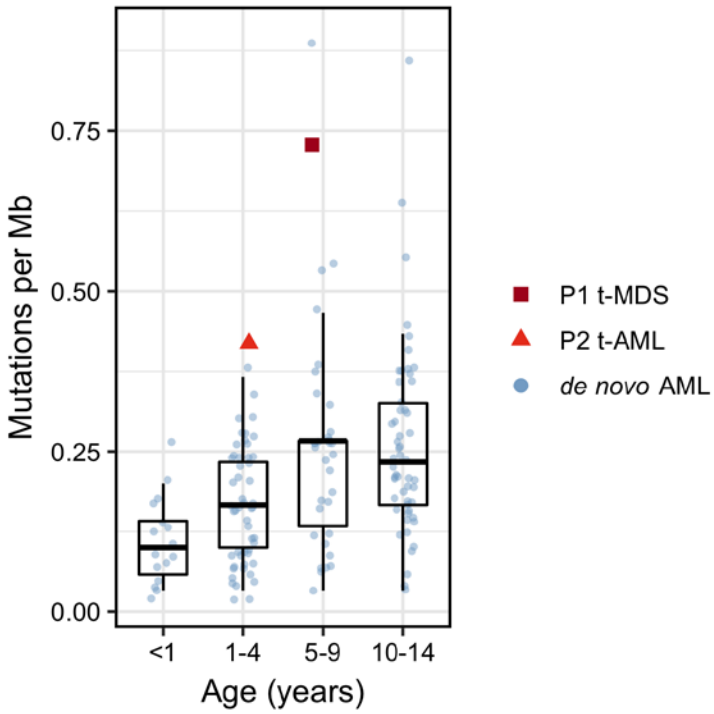
Signature	Contribution	P of C[T>C]G	Normalised probability	Etiology
SBS1	0.023	0.00036	0.0022	Spontaneous deamination of 5-methylcytosine (clock-like signature)
SBS5	0.058	0.021	0.32	Unknown (clock-like signature)
SBS18	0.039	0.0016	0.017	Oxidative stress
SBS31	0.370	0.0045	0.44	Platinum chemotherapy
SBS35	0.510	0.0016	0.22	Platinum chemotherapy

Full signature extraction results for *de novo* germline mutations and blood, bone marrow and tumor mutations will be deposited on github in excel format (**Table S6** and **Table S7**).

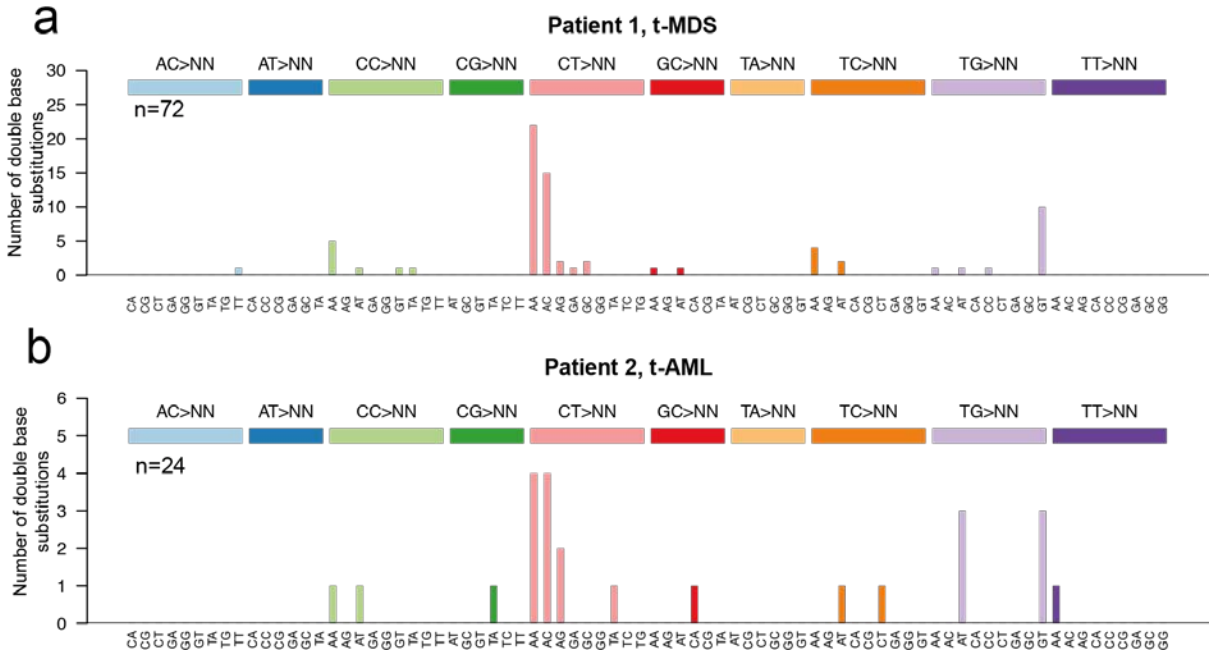
5. Code availability

The scripts containing bespoke code for filtering, analysis and visualization described above is available from GitHub at https://github.com/TimCoorens/Paed_CH_chemo

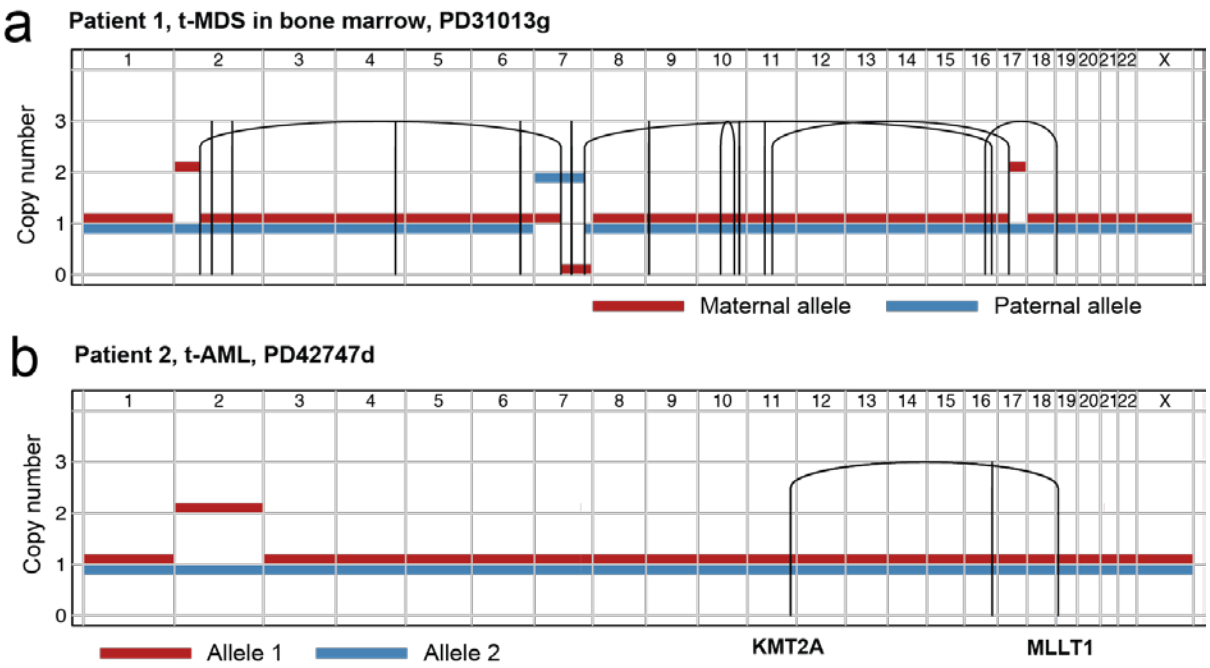
Supplemental Figures



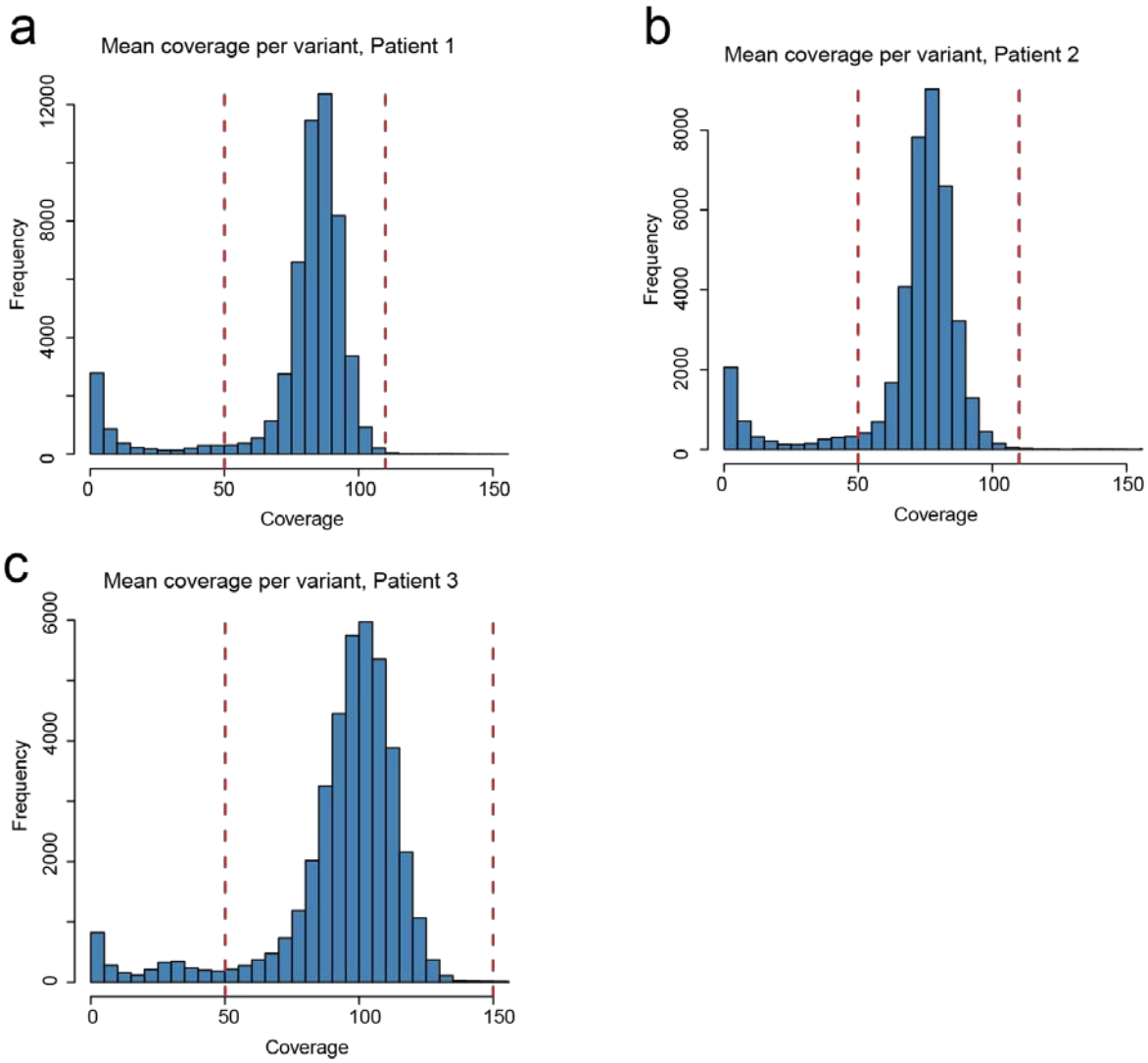
Supplemental Figure 1. Mutation burden in *de novo* pediatric AML and TMN cases
Point mutation burden per megabase of genome is indicated on the y-axis and patient age in years on the x-axis. The red square and orange triangle indicate the mutation burdens of Patient 1's t-MDS and Patient 2's t-AML, respectively. Mutation counts for individual *de novo* pediatric AML samples from published data[21] are represented by blue dots. Box plot centres, hinges and whiskers represent the median, first and third quartiles and 1.5×interquartile range, respectively. AML, acute myeloid leukemia; Mb, megabasee; P1, Patient 1; P2, Patient 2; t-MDS, therapy-related myelodysplastic syndrome; t-AML, therapy-related acute myeloid leukemia.



Supplemental Figure 2. Doublet-base substitution mutational profiles in TMN genomes. The x-axis shows the 78 classes of doublet-base substitutions (DBS). The y-axis shows the number of DBS in each class. The total number of DBS per genome is indicated at the upper left of each plot. Panel A shows the DBS profile in Patient 1's t-MDS. Panel B shows the DBS profile of Patient 2's t-AML. In both genomes, the doublet-base substitutions are primarily accounted for by doublet-base substitution signature 5 (DBS5), which has been linked to platinum chemotherapy exposure. t-MN, therapy-related myeloid neoplasm; t-MDS, therapy-related myelodysplastic syndrome; t-AML, therapy-related acute myeloid leukemia.



Supplemental Figure 3. Structural variant profiles of TMN genomes.
Allele-specific copy number profiles and chromosomal rearrangements are shown with genomic position on the x-axis and absolute copy number on the y-axis. Panel A shows the structural variant profile for Patient 1's t-MDS, with maternal and paternal alleles indicated in red and blue, respectively. Panel B shows the structural variant profile of Patient 2's t-AML, with the *KMT2A-MLL1* balanced translocation annotated. TMN, therapy-related myeloid neoplasm; t-MDS, therapy-related myelodysplastic syndrome; t-AML, therapy-related acute myeloid leukemia.



Supplemental Figure 4. Exclusion of sites with consistently low or high coverage.
The x-axis shows the whole genome sequencing coverage and the y-axis shows the number of point mutations. Panel A shows the distribution of point mutations according to whole genome sequencing coverage in copy number normal samples from Patient 1 (PD31013b, PD31013d, PD31013h). Panel B shows the distribution of point mutations according to whole genome sequencing coverage in copy number normal samples from Patient 2 (PD42747a, PD42747c). Panel C shows the distribution of point mutations according to whole genome sequencing coverage in the copy number normal sample from Patient 3 (PD31012b). The variants between the vertical red lines intersecting the x-axis at 50 and 110 (Patients 1 and 2) or 50 and 150 (Patient 3) demarcate the mutations included in downstream analysis. Variants at positions with coverage outside of these bounds were excluded due to enrichment for mapping errors.

Supplementary Tables

Table S1: Overview of study cohort

PatientID	Cancer Diagnosis	Germline Cancer Predisposition	Blood Samples	Tumor Samples	Bone Marrow Samples	Normal Renal Tissue Samples	Parental Samples	Sibling ID
PD34276	High-risk neuroblastoma*		PD34276b				PD34957b; PD34958b	PD34277b
PD34954	High-risk neuroblastoma*	PHOX2B mutation	PD34954b	PD34954a; PD34954c			PD34952b; PD34953b	-
PD36156	Osteosarcoma*	RECQL4 mutation	PD36156b	PD36156a			PD36157b; PD36158b	-
PD36159	Wilms tumor		PD36159b	PD36159c		PD36159g; PD36159h	PD36160b; PD36161b	-
PD36165	Wilms tumor	Overgrowth syndrome	PD36165b	PD36165a		PD36165f	PD36166b; PD36167b	-
PD36812	High-risk neuroblastoma*	CHEK2, SMARCA4 mutations	PD36812b; PD36812b2	PD36812a			PD34955b; PD34956b	-
PD37272	Wilms tumor		PD37272g	PD37272a		PD37272b; PD37272c; PD37272d	PD37286b; PD37287b	-
PD37276	Wilms tumor		PD37276g	PD37276a		PD37276b; PD37276c	PD37288b; PD37289b	-
PD40306	Infantile fibrosarcoma		PD40306b	PD40306a			PD40307b; PD40308b	-
PD40309	Unclassifiable renal tumor		PD40309b	PD40309a		PD40309d	PD40310b; PD40311b	-
PD40312	Osteosarcoma*		PD40312b; PD40312n	PD40312e			PD40313b; PD40314b	-
PD40641	Wilms tumor	Germline H19 hypermethylation	PD40641b			PD40641d; PD40641e	PD40642b; PD40643b	-
PD40735	Wilms tumor		PD40735k; PD40735m	PD40735a		PD40735b; PD40735h	PD40736b; PD40737b	-
PD40738	Wilms tumor		PD40738b; PD40738j	PD40738a		PD40738g	PD41065b; PD41066b	-
PD42181	Ewing sarcoma		PD42181b	PD42181a			PD41848b; PD41847b	-
PD46693	High-risk neuroblastoma*		PD46693b	PD46693a			PD46694a; PD46695a	-
PD46692	Rhabdomyosarcoma		PD46692b	PD46692a			PD46697b; PD46698b	-
PD31013	High-risk neuroblastoma*	BARD1 mutation	PD31013b; PD31013c; PD31013h	PD31013a	PD31013d; PD31013e; PD31013f; PD31013g		PD34278b; PD34279b	PD31012
PD31012	Low-risk neuroblastoma*	BARD1 mutation	PD31012b	PD31012c			PD34278b; PD34279b	PD31013
PD42747	High-risk neuroblastoma*		PD42747b; PD42747c		PD42747d			

* in the Cancer Diagnosis column indicate that samples were taken after platinum agent treatment

Table S2: Genes used for the design of RNA baits

TargetID	Interval	Regions	Size
ARID1A	chr1:27022885-27107257	21	7364
ASXL1	chr20:30946569-31025151	17	5060
ATM	chr11:108098342-108236245	62	10411
ATP6AP1	chrX:153657029-153664247	10	1730
ATP6V1B2	chr8:20054908-20082849	16	1980
B2M	chr15:45003735-45008550	3	443
BCL10	chr1:85733300-85742045	3	762
BCL2	chr18:60795848-60985909	2	793
BCL6	chr3:187440236-187451491	8	2281
BCOR	chrX:39909159-39937192	15	5648
BCORL1	chrX:129139198-129190121	14	5690
BRAF	chr7:140426284-140624513	21	2799
CALR	chr19:13049484-13054805	10	1494
CARD11	chr7:2946262-2998150	24	3945
CBL	chr11:119077118-119170501	16	3041
CCND3	chr6:41903579-41909397	6	1151
CD58	chr1:117057425-117113604	7	922
CD79B	chr17:62006576-62009631	6	813
CDKN2A	chr9:21968198-21994463	6	1248
CEBPA	chr19:33792234-33793435	1	1202
CREBBP	chr16:3777709-3929927	31	7988
CSF1R	chr5:149433622-149466000	22	3443

CSF3R	chr1:36931687-36945107	16	3018
CUX1	chr7:101459301-101926392	34	6113
DNMT3A	chr2:25457138-25536863	25	3388
EBF1	chr5:158126109-158526496	16	2099
EP300	chr22:41488999-41574970	31	7865
ETNK1	chr12:22778088-22837898	11	1750
ETV6	chr12:11803052-12044545	10	1643
EZH2	chr7:148504728-148544400	21	2876
FBXW7	chr4:153244023-153332965	14	2898
FLT3	chr13:28578179-28674657	25	3504
FOXO1	chr13:41133650-41240359	2	2008
GATA2	chr3:128199852-128205884	5	1543
GNA13	chr17:63010365-63052721	4	1214
GNAS	chr20:57415152-57485894	17	4436
GNB1	chr1:1718760-1756902	9	1217
H3F3A	chr1:226252043-226259190	3	561
HIST1H1B	chr6:27834617-27835317	1	701
HIST1H1C	chr6:26056005-26056666	1	662
HIST1H1D	chr6:26234486-26235171	1	686
HIST1H1E	chr6:26156609-26157288	1	680
IDH1	chr2:209101793-209116285	8	1408
IDH2	chr15:90627488-90645632	11	1579
IKZF3	chr17:37922033-38020389	9	1793
IL7R	chr5:35857070-35876598	8	1716
IRF8	chr16:85936612-85954898	8	1466

JAK2	chr9:5021978-5126801	23	3859
KDM6A	chrX:44732788-44970666	31	5090
KIT	chr4:55524172-55604733	21	3351
KMT2C	chr7:151833907-152132881	60	16122
KMT2D	chr12:49415553-49449117	55	17762
KRAS	chr12:25362719-25398328	6	828
MBD1	chr18:47793948-47807596	20	2712
MEF2B	chr19:19256493-19267955	9	1565
MPL	chr1:43803510-43818453	12	2233
MYC	chr8:128748830-128753214	3	1425
MYD88	chr3:38180143-38182787	5	1054
NF1	chr17:29422216-29705959	63	10271
NOTCH1	chr9:139390513-139440248	34	8348
NOTCH2	chr1:120457919-120612030	35	8243
NPM1	chr5:170814943-170837579	12	1134
NRAS	chr1:115251146-115258791	4	650
PAX5	chr9:36840516-37034038	11	1484
PDGFRA	chr4:55106210-55161449	24	3930
PHF6	chrX:133511638-133559370	9	1387
PIM1	chr6:37138069-37141877	6	1335
POT1	chr7:124462572-124537237	17	2444
POU2F2	chr19:42595399-42636573	16	2803
PPM1D	chr17:58677766-58740923	8	2054
PRDM1	chr6:106534419-106555371	9	2715
PTEN	chr10:89623697-89725239	9	1912

PTPN11	chr12:112856906-112942578	16	2142
RAD21	chr8:117859729-117878978	13	2156
RRAGC	chr1:39305215-39325328	7	1340
RUNX1	chr21:36164422-36421206	11	1804
SETBP1	chr18:42281302-42643673	6	5100
SETD2	chr3:47058573-47205424	23	8200
SF3B1	chr2:198257017-198299733	27	4585
SMC3	chr10:112327565-112364070	29	4234
SOCS1	chr16:11348690-11349345	1	656
SRSF2	chr17:74732233-74733252	2	706
STAG2	chrX:123156468-123234457	34	4541
STAT3	chr17:40467753-40526091	24	2799
STAT6	chr12:57490345-57502071	22	3038
TCF3	chr19:1611696-1650257	18	2813
TET2	chr4:106111617-106197686	10	6365
TNFAIP3	chr6:138192355-138202466	8	2533
TNFRSF14	chr1:2488094-2494722	9	1063
TP53	chr17:7565247-7579922	14	1658
U2AF1	chr21:44513202-44527614	9	970
WT1	chr11:32410594-32456901	11	1788
XPO1	chr2:61705945-61761042	24	3717
ZEB1	chr10:31608154-31816202	14	4047
ZRSR2	chrX:15808609-15841375	12	1930

Table S3: Rearrangements in TMN genomes

Sample	Chr1	Start1	End1	Chr2	Start2	End2	Str1	Str 2	Class
PD31013g	10	64274893	64274894	10	101807999	101808000	+	+	deletion
PD31013g	10	115352535	115352539	10	115358137	115358141	+	+	deletion
PD31013g	11	49470897	49470898	11	49501833	49501834	-	-	tandem-duplication
PD31013g	11	69668285	69668286	17	32592699	32592700	+	+	translocation
PD31013g	16	58389087	58389088	19	4477792	4477793	-	+	translocation
PD31013g	2	71397155	71397156	7	74049196	74049197	+	-	translocation
PD31013g	2	103606305	103606307	2	103690218	103690220	-	-	tandem-duplication
PD31013g	2	159499285	159499291	2	159521109	159521115	-	-	tandem-duplication
PD31013g	4	165182548	165182550	4	165229311	165229313	+	+	deletion
PD31013g	6	134271315	134271317	6	134448413	134448415	-	-	tandem-duplication
PD31013g	7	102606970	102606973	7	102663898	102663901	+	+	deletion
PD31013g	7	102720030	102720031	7	102728596	102728597	+	+	deletion
PD31013g	7	139004681	139004682	16	76154792	76154793	-	-	translocation
PD31013g	9	8912316	8912320	9	10088906	10088910	+	+	deletion
PD31013c	2	71397176	71397177	7	74049164	74049165	+	-	translocation
PD42747d	11	50676665	50676666	11	50676464	50676465	+	+	deletion
PD42747d	11	118353359	118353360	19	6247186	6247187	+	-	translocation
PD42747d	11	118353365	118353366	19	6247215	6247216	-	+	translocation
PD42747d	16	75844202	75844217	16	75844207	75844222	+	+	deletion
PD42747d	17	44223916	44223917	17	44224321	44224322	+	+	deletion
PD42747d	9	6751209	6751210	9	6752054	6752055	+	+	deletion

Supplemental References

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–595.
2. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012;486(7403):400–404.
3. Jones D, Raine KM, Davies H, et al. cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr. Protoc. Bioinformatics*. 2016;56:15.10.1–15.10.18.
4. Raine KM, Hinton J, Butler AP, et al. cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *Curr. Protoc. Bioinformatics*. 2015;52:15.7.1–15.7.12.
5. Coorens THH, Treger TD, Al-Saadi R, et al. Embryonal precursors of Wilms tumor. *Science*. 2019;366(6470):1247–1251.
6. Van Loo P, Nordgard SH, Lingjærde OC, et al. Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* 2010;107(39):16910–16915.
7. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994–1007.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
9. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–842.
10. Gerstung M, Beisel C, Rechsteiner M, et al. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nat. Commun.* 2012;3:811.
11. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics*. 2014;30(9):1198–1204.
12. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;348(6237):880–886.
13. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* 2016;44(16):e131.
14. Menzies A, Teague JW, Butler AP, et al. VAGrENT: Variation Annotation Generator. *Current Protocols in Bioinformatics*. 2015;52(1.):
15. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 2019;47(D1):D941–D947.
16. McNerney ME, Godley LA, Le Beau MM. Therapy-related myeloid neoplasms: when genetics and environment collide. *Nat. Rev. Cancer*. 2017;17(9):513–527.
17. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016;44(D1):D862–8.
18. Stenson PD, Ball EV, Mort M, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* 2003;21(6):577–581.
19. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94–101.
20. Gori K, Baez-Ortega A. sigfit: flexible Bayesian inference of mutational signatures. *bioRxiv*. 2018;
21. Bolouri H, Farrar JE, Triche T Jr, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* 2018;24(1):103–112.