Genomic Entropy confers an aggressive Phenotype in Mesothelioma

John W Oketch[1], Charlotte Poile[1], Maurice Dungey[1,2], Nathaniel Kuse[1], Jan Rogel[1], Hesam Moazzen[1], Joanna Dzialo[1], Tasneem Ibrahim[1], Kudzayi Kutywawo[1,2], Apostolos Nakas[2], Ihsaan Sabil[1], Shreya Ramalamgam[1], Nada Nusrat[1], Aleksandra Bzura[1,3], Yue Lin[1], Daniel Buckler[1],Mazin H Alhazmi[4], Alan Dawson[1,2], Francois Ng Kee Kwong[1,5,6], Panagiotis Barmpoutis[1,6], Hongji Yang[7], Huiyu Zhou[7], Richard Badge[4], Cristina Tufarelli[8], Amrita Bajaj[2], Heath Hopewell[2], Annabel J Sharkey[1,9], Sara Busacca[1,10], Jinli Luo[1,11],Catherine Jane Richards[2], Peter Wells-Jordan[2], Kevin Groot-Lipman[12], Jens C Hahne[1], Haining Yang[13], Michele Carbone[13], Aaron S Mansfield[14], Edward J Hollox[4], Min Zhang[1,15], Dean. A. Fennell[1,2]

Supplementary Information

Content:

Supplementary figure S1-S7

Supplementary Table

Link to figures' source data and extended data supporting the analyses

FIGURE S1

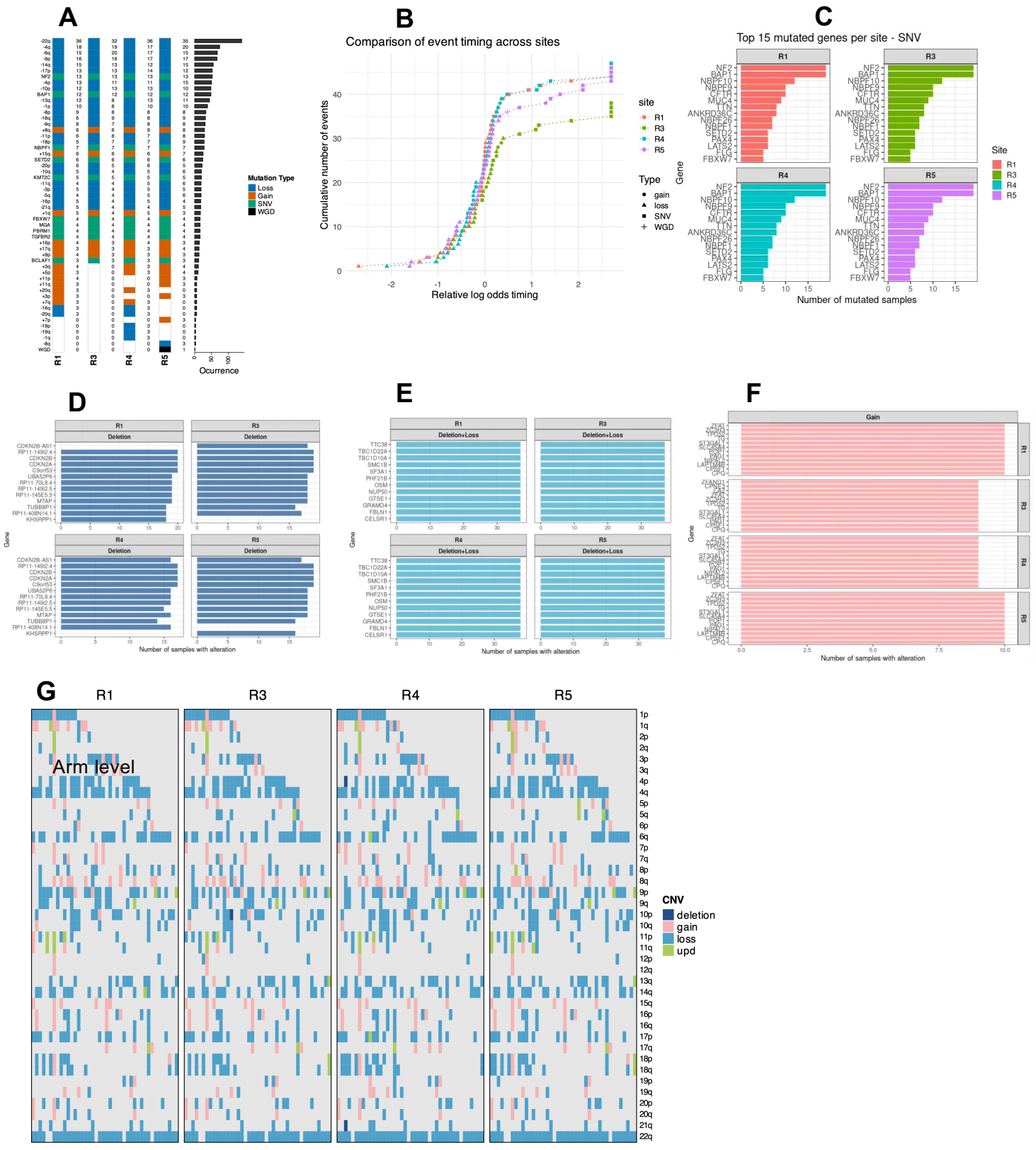


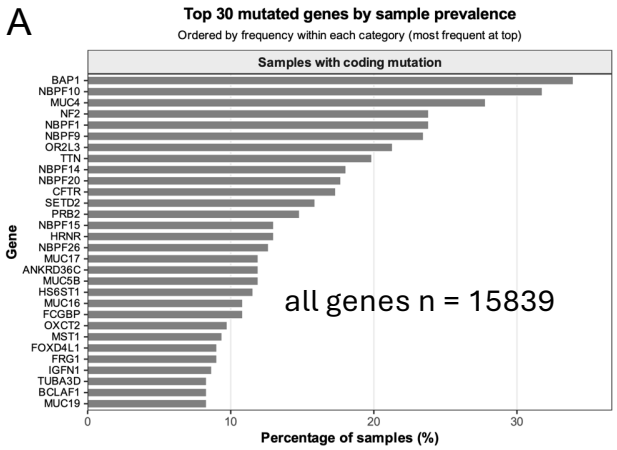
Figure S1 | Spatial genetic evolution trajectories in mesothelioma from multi-region biopsies

(A) Heatmap of recurrent mutations identified by PhylogicNDT (Timing and League modules) across 42 Medusa samples with biopsies from four sites (apex (R1), pericardium (R2), anterior (R3), posterior (R4), costophrenic angles, and oblique fissure (R5); mutations coloured by type. R2 was sequenced in only 8 samples and excluded from analysis. (B) Cumulative driver event counts stratified by median relative timing and mutation type per sampling site. (C–F) Analysis of site-specific mutation frequency based on nonsynonymous mutations and gene-level copy number alterations; no genes showed significant differential mutation across sites. (G) Heatmap of arm-level copy number calls across 42 samples and sites. Tiles coloured by status (deletion, gain, UPD); arm-level calls required >50% arm length coverage (weighted by segment length).

Figure S2

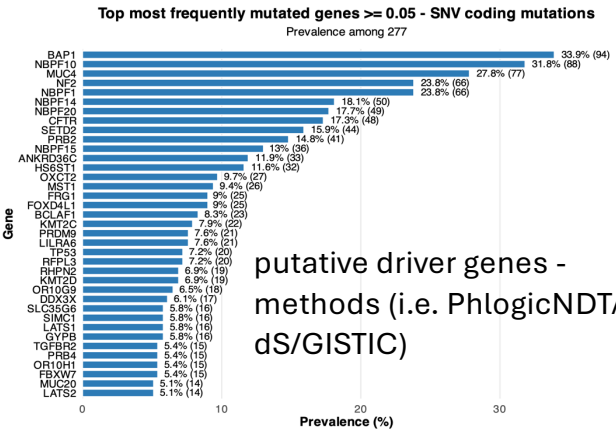
Gene-level somatic single nucleotide variant (SNV) mutations

A



all genes n = 15839

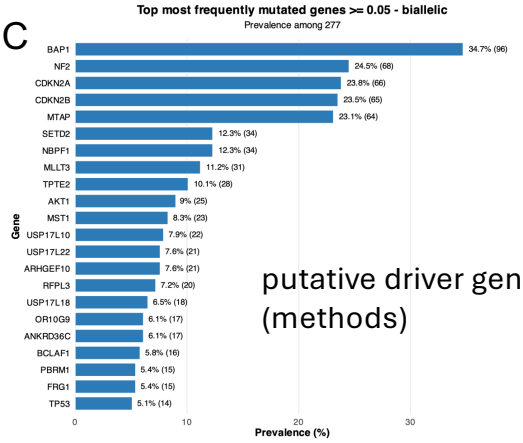
B



putative driver genes - methods (i.e. PhlogicNDT/dN-dS/GISTIC)

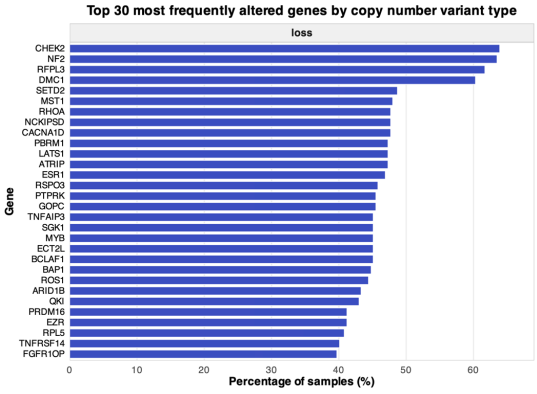
SNV/gene level copy number alterations

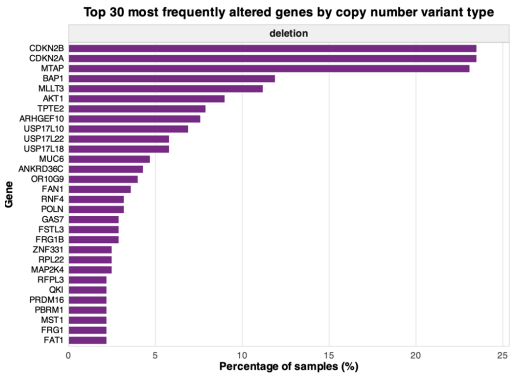
C

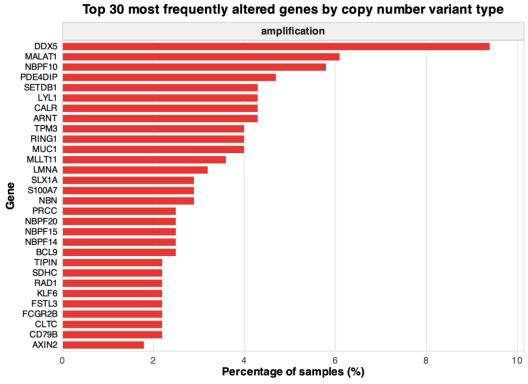


putative driver genes (methods)

gene level copy number alterations among putative driver genes







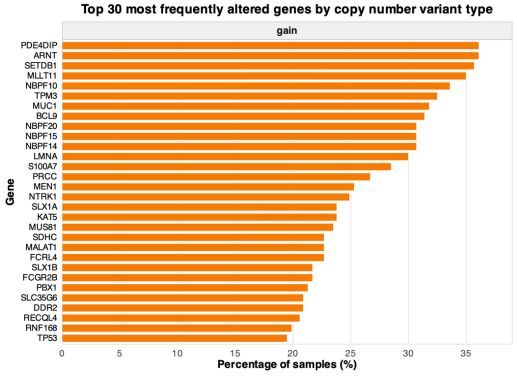


Figure S2 | Recurrent somatic alterations

(A–B) Frequently mutated genes based on non-synonymous SNVs (nonsense, missense, nonstop, frameshift, splice-site, start-site, and in-frame indels) across (A) all 15,839 genes and (B) 297 putative driver genes identified by PhylogicNDT, dN/dS, and GISTIC analyses. (C) Genes with biallelic inactivation, defined as homozygous deletion or heterozygous loss with a non-synonymous SNV on the remaining allele.(E–G) Most frequently altered putative driver genes based on gene-level copy number status.

Figure S3

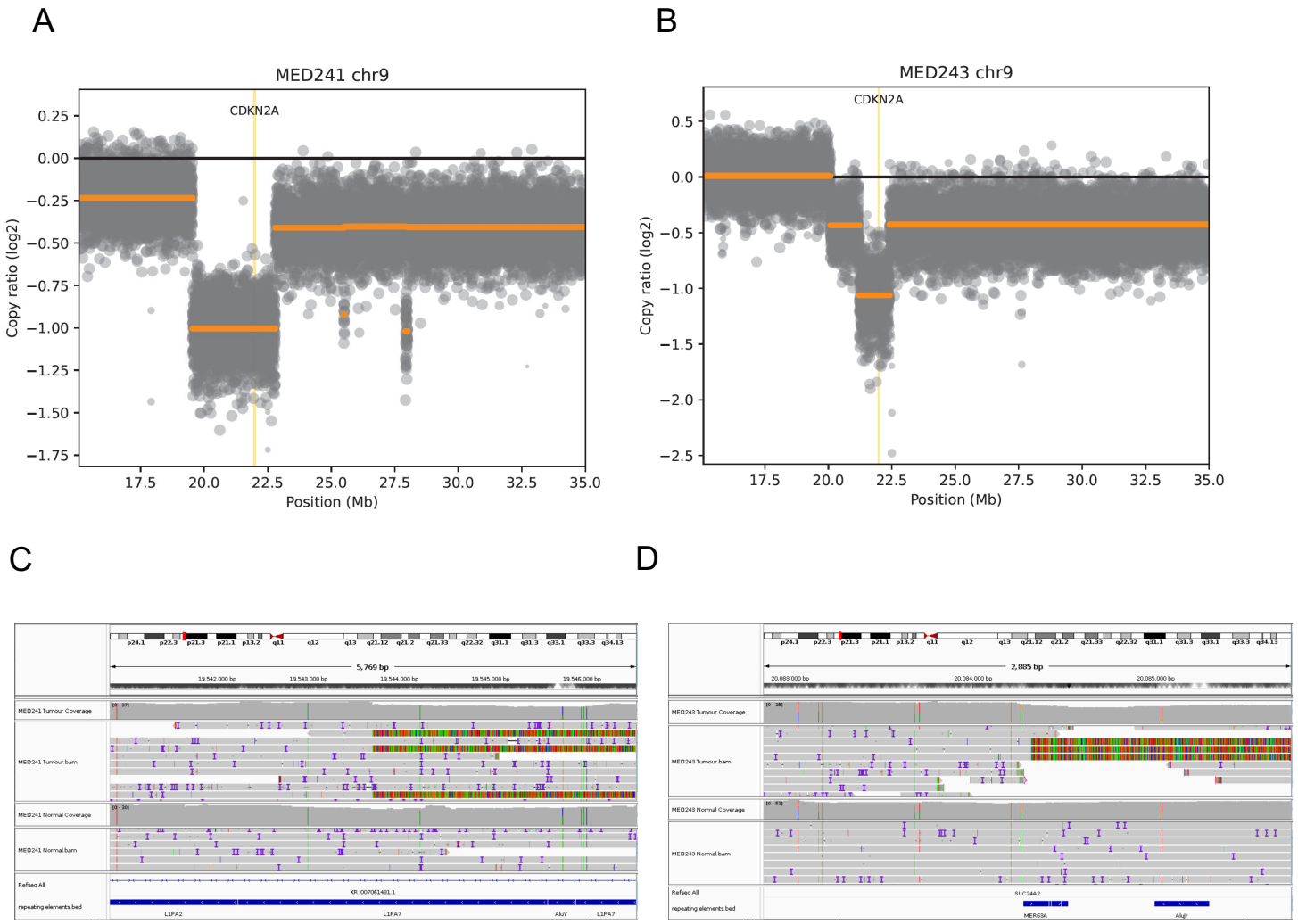


Figure S3 | Identification of L1 retro-transposition at CDKN2A deletion breakpoints

Read depth analysis of tumours from patients MED241 (A) and MED243 (B) showing homozygous deletion of CDKN2A. Identification of soft-clipped long sequencing ONT reads from MED241 (C) and MED243 (D) using the Integrative Genomics Viewer.

Figure S4

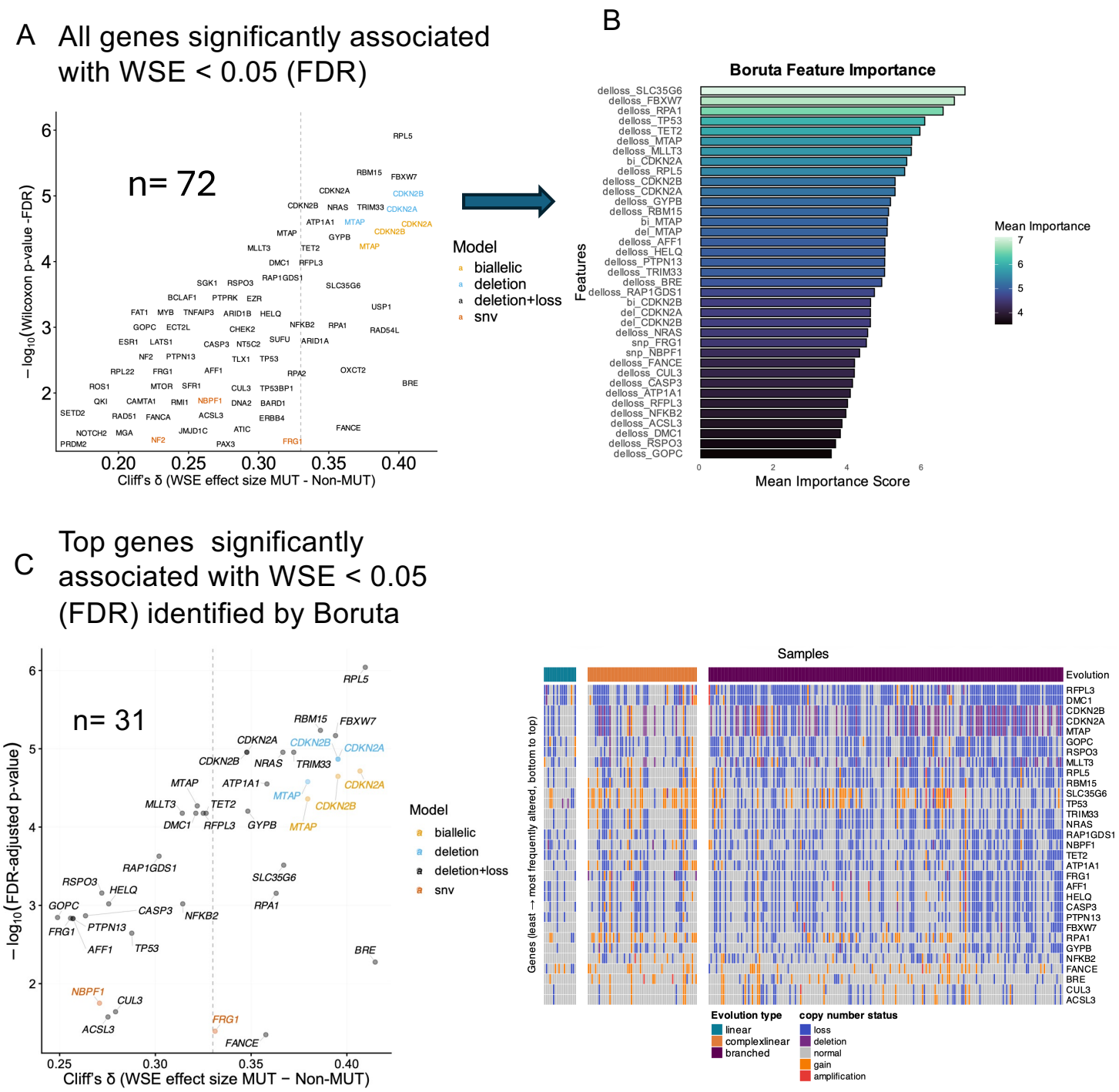


Figure S4 | Associations between driver gene alterations and intratumor heterogeneity (WSE)

(A,B) Gene-level associations between genomic alterations and whole-sample entropy (WSE) were assessed using Wilcoxon rank-sum tests across four alteration classes (SNVs, homozygous deletions, loss/deletions, gain/amplifications, and biallelic inactivation) for a targeted set of 297 genes, including genes under positive selection by dN/dS analysis and focal somatic copy number alterations (SCNAs) from GISTIC2.0 intersected with the COSMIC Cancer Gene Census and DNA damage response genes. (A) One-sided, FDR-adjusted P values for WSE in mutant versus non-mutant tumours across 277 samples, grouped by alteration type (biallelic inactivation, deletion, deletion plus loss, gain, and coding SNVs); only genes with P < 0.05 are shown, and the vertical dotted line indicates the median Cliff's delta effect-size threshold (0.3). Genes are coloured by alteration model. (B) Boruta feature selection highlighting top gene alterations associated with WSE. (C) As in (A) but restricted to top-ranked genes. (D) Heatmap of copy number status for top genes across 277 tumours stratified by phylogenetic topology (L, CL, B)

FIGURE S5

A

SCNA-1

sample	chr	startpos	endpos	BAF	pval	LogR	ntot	nMaj1_A	nMin1_A	frac1_A	nMaj2_A	nMin2_A	frac2_A	cnv -WES	predicted clonality -WES
MED220	3	1214847	197581147	0.500	1.000	0.013	2.067	1	1	1.000	NA	NA	NA	normal	clonal
MED216	3	42218095	53843843	0.556	0.000	-0.200	0.995	1	0	0.805	1	1	0.195	loss	subclonal
MIST2_014	3	49834571	128622742	0.500	1.000	0.180	2.108	1	1	1.000	NA	NA	NA	normal	clonal
MED099	3	39111140	78719577	0.752	1.000	-0.420	1.090	1	0	1.000	NA	NA	NA	loss	clonal
MIST4_019	3	3080509	13679688	0.605	1.000	-0.170	1.480	1	0	1.000	NA	NA	NA	loss	clonal

SCNA-2

sample	cnv -WES	IHC/pathology	Phylogeny
MED220	normal	loss	linear
MED216	loss	loss	linear
MIST2_014	normal	NA	branched
MED099	loss	loss	branched
MIST4_019	loss	loss	complexlinear

B

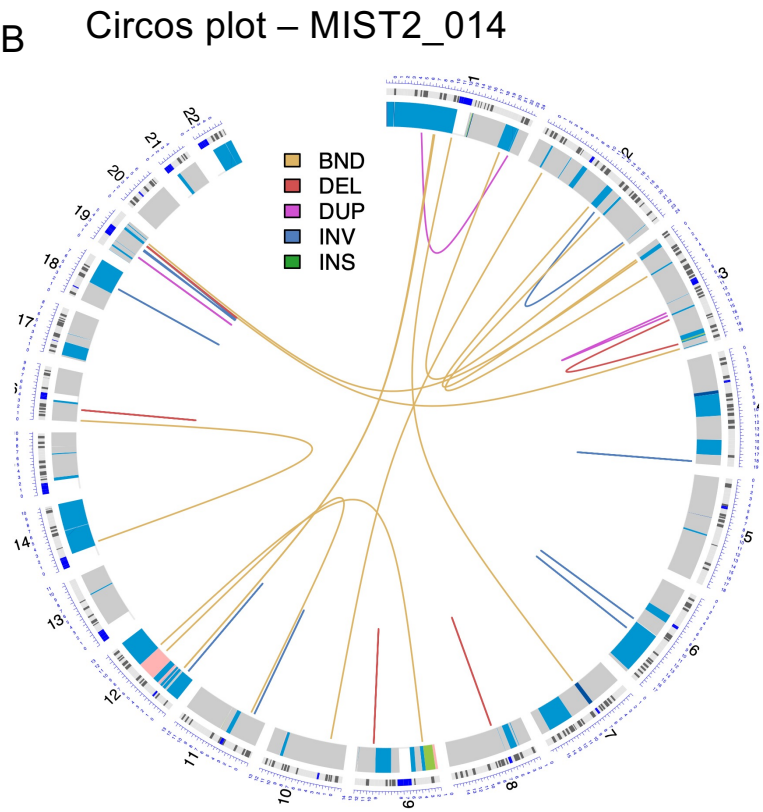


Figure S5 | Somatic copy number alterations and structural variation in BAP1 germline mutant mesothelioma tumours

(A) Summary of somatic copy-number alteration (SCNA) calls on chromosome 3 encompassing the *BAP1* locus across five germline carriers. Clonality was inferred using Battenberg, and phylogenetic reconstruction was performed using the PhylogicNDT *Cluster* and *BuildTree* modules. (B), Circos plot for sample MIST2_014 illustrating extensive structural variation, including inter- and intra-chromosomal rearrangements (breakpoints, deletions, duplications, inversions and insertions), consistent with highly complex genome architecture. The density and connectivity of rearrangements indicate catastrophic chromosomal restructuring, compatible with chromothripsis-like events

Figure S6

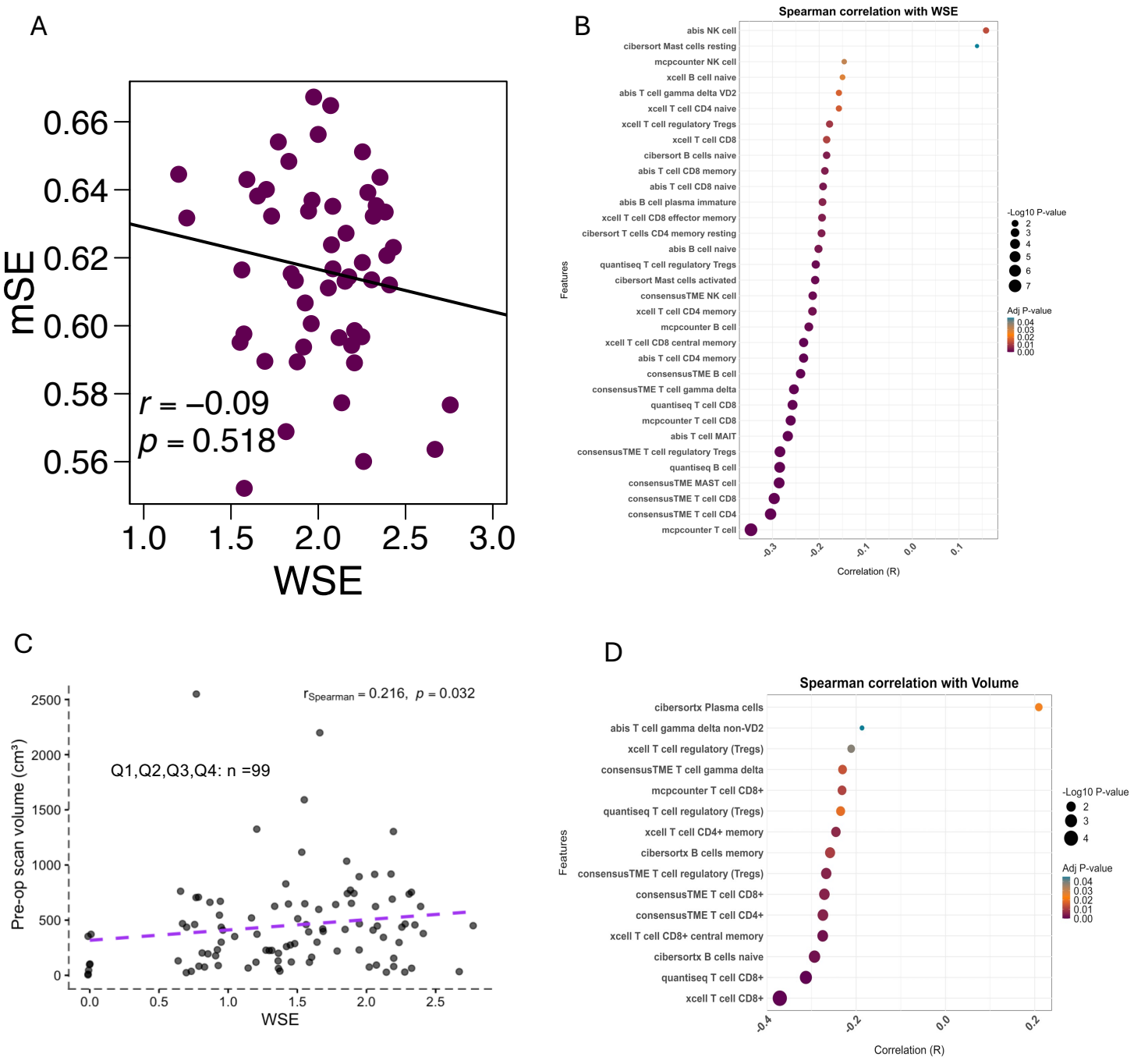


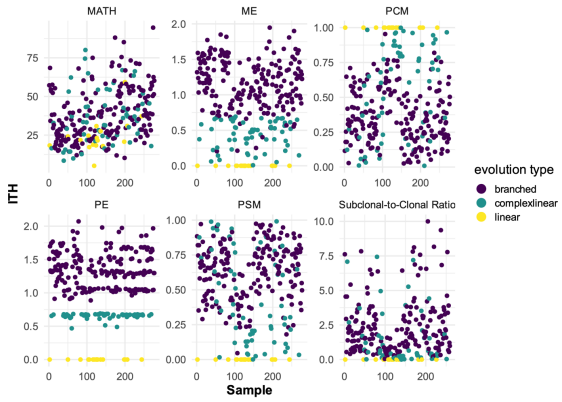
Figure S6 | WSE associations between methylation entropy, tumour volume, and immune composition

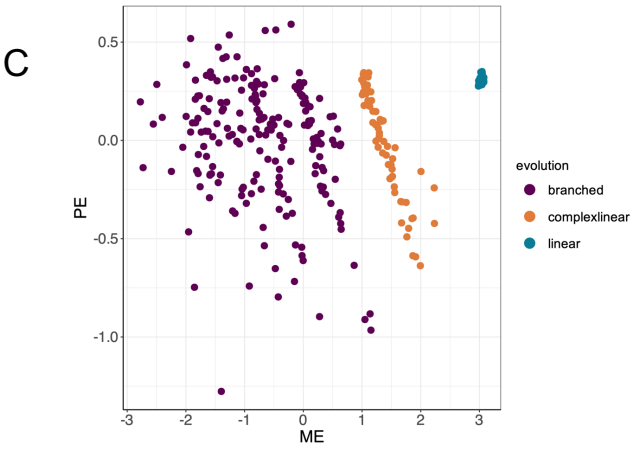
(A) Scatter plot showing methylation Shannon entropy (mSE) versus weighted Shannon entropy (WSE) in mesothelioma tumors (n = 52) . Each point represents a tumor sample. Spearman's rank correlation coefficient (*r*) and permutation-based *p* value (10,000 permutations) are shown on the plot. .(B) Bubble plot showing Spearman correlation between WSE and estimated fractions of $CD8^+$ T cells, NK cells, and B cells. Both $CD8^+$ T cells and B cells exhibited consistent negative correlations with WSE across independent deconvolution methods. (C) Spearman correlation between WSE and AI tumour volume showing a positive association between the two measures across Q1 to Q4 tumors. (D) Bubble plot showing Spearman correlation between tumour AI volume and immune cell fractions. Larger volumes were negatively correlated with $CD8^+$ T cells and B cells across deconvolution methods.

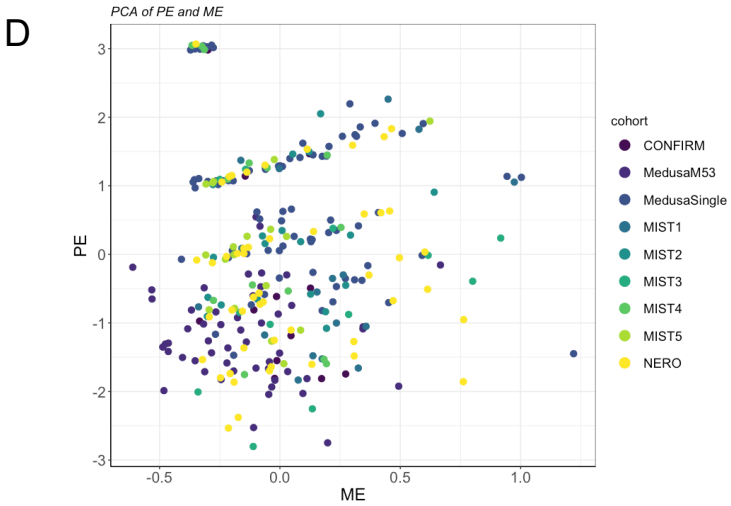
FIGURE S7

A

Cohort	Counts
CONFIRM	12
MedusaM53	52
MedusaSingle	83
MIST1	17
MIST2	19
MIST3	12
MIST4	14
MIST5	18
NERO	50







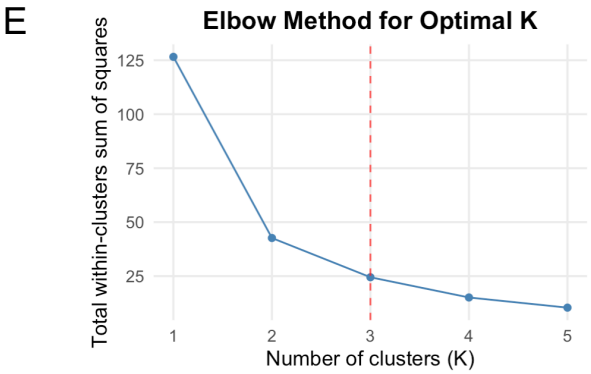


Figure S7 | Cohort composition and validation of intratumor heterogeneity metrics.

(A) Number of participants analysed per cohort. (B) Scatter plots of 277 tumours across intratumour heterogeneity (ITH) matrices, with each point representing a single tumour coloured by phylogenetic topology. (C–D) Principal component analysis (PCA) of the 277 tumours based on proportion entropy (PE) and mutation entropy (ME), stratified by phylogenetic topology (C) and cohort (D). (E) Elbow plot of k-means clustering on phylogenetic metrics showing total within-cluster sum of squares (WSS) versus number of clusters (K = 1–5). A clear inflection at K = 3 indicates three main groups. Combined PE and ME better resolved phylogenetic topologies compared with MATH score or PSM and demonstrated a homogeneous distribution across the cohorts. Phylogenetic trees are available at https://doi.org/10.25392/leicester.data.31073389.

Table S1 : Genes within recurrent somatic copy number alteration peaks identified by GISTIC2.0 that map to early evolutionary epochs inferred by PhylogicNDT

Gene	Event	ROLE_IN_CANCER	TIER	Arm
BRCA2	deletion	TSG	1	13q
FOXO1	deletion	oncogene, TSG, fusion	1	13q
LCP1	deletion	fusion	2	13q
RB1	deletion	TSG	1	13q
NBEA	deletion		2	13q
CYSLTR2	deletion	oncogene	2	13q
ARNT	deletion	oncogene, TSG, fusion	1	1p
ATP1A1	deletion	oncogene, TSG	1	1p
BCL9	deletion	oncogene, fusion	1	1p
FCGR2B	deletion	oncogene, fusion	1	1p
LMNA	deletion	fusion	1	1p
MUC1	deletion	fusion	1	1p
NOTCH2	deletion	oncogene, TSG	1	1p
NRAS	deletion	oncogene	1	1p
NTRK1	deletion	oncogene, TSG, fusion	1	1p
DDR2	deletion	oncogene	1	1p
PBX1	deletion	oncogene, fusion	1	1p
PRCC	deletion	fusion	1	1p
RPL5	deletion	TSG	1	1p
S100A7	deletion	fusion	2	1p
SDHC	deletion	TSG	1	1p
TPM3	deletion	TSG, fusion	1	1p
PDE4DIP	deletion	fusion	1	1p
SETDB1	deletion	oncogene	2	1p
MLLT11	deletion	fusion	1	1p
TRIM33	deletion	TSG, fusion	1	1p
RBM15	deletion	fusion	1	1p
FCRL4	deletion	oncogene, fusion	1	1p
MTOR	deletion	oncogene	1	1p
RPL22	deletion	TSG, fusion	1	1p
SKI	deletion	oncogene	2	1p
PRDM2	deletion	TSG	2	1p
TNFRSF14	deletion	TSG	1	1p
CAMTA1	deletion	TSG, fusion	1	1p
PRDM16	deletion	oncogene, fusion	1	1p
RHOA	deletion	oncogene, TSG	1	3p
CACNA1D	deletion	oncogene	1	3p
BAP1	deletion	TSG	1	3p
SETD2	deletion	TSG	1	3p
NCKIPSD	deletion	fusion	2	3p
PBRM1	deletion	TSG	1	3p
CTNNB1	deletion	oncogene, fusion	1	3p
ESR1	deletion	oncogene, TSG, fusion	1	6q
MYB	deletion	oncogene, fusion	1	6q
PTPRK	deletion	TSG, fusion	1	6q
ROS1	deletion	oncogene, fusion	1	6q
SGK1	deletion	oncogene	2	6q
TNFAIP3	deletion	TSG	1	6q
EZR	deletion	fusion	1	6q
LATS1	deletion	TSG	1	6q
QKI	deletion	oncogene, TSG	1	6q
BCLAF1	deletion		2	6q
FGFR1OP	deletion	fusion	1	6q
GOPC	deletion	fusion	1	6q
ARID1B	deletion	TSG	1	6q
RSPO3	deletion	oncogene, fusion	1	6q
ECT2L	deletion		2	6q
FGFR1	deletion	oncogene, fusion	1	8p
NRG1	deletion	TSG, fusion	1	8p
PCM1	deletion	fusion	1	8p
WRN	deletion	TSG	1	8p
ARHGEF10	deletion	TSG	2	8p
LEPROTL1	deletion	TSG	2	8p

Extended data

- Figures' source data can also be obtained at – https://doi.org/10.25392/leicester.data.31073389
- Extended data - https://doi.org/10.25392/leicester.data.31073389