Supplemental Materials

Genomic and epigenomic insights into the origin, pathogenesis and clinical behavior of mantle cell lymphoma subtypes

Nadeu F, Martín-Garcia D, *et al.*

SUPPLEMENTAL METHODS

Cohort and sample preparation

We studied 82 patients with mantle cell lymphoma (MCL) by whole-genome sequencing (WGS) (n=61) and whole-exome sequencing (WES) (n=21) including 74% conventional MCL (cMCL) and 26% leukemic non-nodal MCL (nnMCL) subtypes. We also performed gene expression arrays, Nanostring, and DNA methylation arrays, as described below (supplemental Table 1). We classified all cases as cMCL and nnMCL based on gene expression signatures and/or SOX11 immunohistochemistry depending on the available material. Clinical, morphological, IGHV mutations or genetic alterations were not used for the classification. In 64 cases with leukemic involvement the classification was determined using the gene expression profiling signatures detected by the L-MCL-16 Nanostring assay or Affymetrix expression arrays previously published.[1] In 6 additional leukemic cases in which the quantity of the RNA was not sufficient for the array study we used the 3-gene reduced signature (*SOX11, HDGFRP3, DBN1*) detected by reverse transcription quantitative PCR, as previously described.[2] In the 12 cases in which only a tissue sample was available we based the diagnosis on SOX11 protein expression detected by immunohistochemistry using two different anti-SOX11 antibodies (anti-SOX11 MRQ-58, Cell Marque, and anti-SOX11 143, clone CL0143, Atlas Antibodies), as previously described.[3] Most cases were studied with several of these techniques and, as previously reported, the results among them were highly concordant.[1,2,4] Informed consent was obtained according to the Institutional Review Board of the Hospital Clínic of Barcelona following the International Cancer Genome Consortium guidelines.[5] The study was approved by the Hospital Clínic of Barcelona Ethics Committee. Tumor cells were purified from peripheral blood (PB) as previously described.[6] Tumor cell purity was ≥80% as assessed by flow cytometry. Normal cells were purified from PB in all patients and had <4% tumor cells also assessed by flow cytometry. Tumor DNA was extracted using appropriate Qiagen kits (Qiagen), as previously described.[7] Samples analyzed by WES were described in a previous publication of our group.[7] Samples analyzed in this study by WGS were collected at diagnosis (n=57) or at relapse post-treatment (n=4) from PB (n=46), lymph nodes (n=12), bone marrow (n=1), tonsil (n=1), and skin (n=1) (Table 1).

Whole-genome/exome sequencing

WES data of 21 tumor/normal pairs and WGS data of 4 tumor/normal pairs was described in a previous study (European Genome-Phenome Archive, EGA accession number EGAS00001000510).[7] Additional WGS were performed on 57 pairs using a PCR-free library preparation protocol (TruSeq DNA PCR Free), sequenced in an Illumina HiSeq X Ten (2x150bp), and deposited at EGA (EGAS00001004165). Tumor and normal WGS raw reads were mapped to the human reference genome (GRCh37) using the BWA-mem algorithm (v0.7.15).[8] BAM files were sorted, indexed, and optical and PCR duplicates were flagged using biobambam2 (v2.0.65, https://gitlab.com/german.tischler/biobambam2). Quality metrics were extracted using Picard (https://broadinstitute.github.io/picard, v2.10.2). All algorithms were executed with default options. Mean coverage from tumor and normal samples were summarized in supplemental Table 2.

Somatic and germline variant calling

Somatic single nucleotide variants (SNVs) and indels were called using Sidrón, as previously described.[6] Kataegis were defined by the presence of 5 o more SNVs with a mutational interdistance <1000 bp. Short insertions/deletions (indels) were detected integrating Sidrón and Pindel (cgpPindel v2.2.3, with default arguments) algorithms. All coding indels were manually reviewed on Integrative Genomic Viewer (IGV).[9] Mutational burden was calculated based on the number of non-N bases in the reference genome used. Germline variant calling was performed following Genome-Analysis Toolkit (GATK) best practices. Briefly, germline SNPs and indels from WGS and WES data were called using HaplotypeCaller (GATK, v4.0.2.0), and filtering of low-quality variants was performed using VariantRecalibrator and ApplyRecalibration.[10–12] We searched for germline mutations in MCL driver genes and in cancer predisposing genes previously described.[13,14] SNVs were functionally predicted using CADD, Mutation Assessor, PolyPhen and SIFT (dbNSFP v3.5a).

Detection of copy number alterations

Copy number alterations (CNA) were analyzed using Battenberg algorithm (cgpBattenberg, v3.2.2)[15] in the 61 samples analyzed by WGS. CNA were also analyzed using the Affymetrix Genome-wide Human SNP Array 6.0 (n=78) and Cytoscan (n=4)

(Thermo Fisher Scientific) in the 82 MCL cases as previously described.[16] CNA from SNP-arrays were used to verify and complement the results obtained from Battenberg. CNA smaller than 100 Kb and physiological CNAs at immunoglobulin (IG) and T-cell receptor (TCR) loci were filtered out. To allow the integration of the 82 cases, CNA present in <15% of the tumor cells called by Battenberg (cases with WGS) were removed as a result of their low sensitivity of the SNP-array (cases with WES and SNP-array). CNA located at terminal regions were obtained from SNP-arrays due to the inherent difficulty to study these regions from WGS data. Tumor purity and ploidy were extracted from Battenberg in the 61 cases with WGS. Predicted ploidies were manually compared with SNP-array and FISH data, and adjusted when necessary. The tumor cell content measured by flow cytometry analyses was used in the 21 cases analyzed by WES.

Detection of structural variants

Structural variants (SV) were extracted from WGS data using SMuFin[17] and LUMPY (v0.2.13),[18] and were visually inspected on IGV. The integration of the two algorithms was performed using in-house scripts merging SV with breakpoints closer than 150 base pairs. SV were annotated using ANNOVAR[19] at breakpoint positions and classified as "disrupted gene" if the breakpoint was located inside the gene. SV detected by a single pipeline that were covered by less than 7 paired-end reads and/or less than 3 split-reads, and did not overlap with CNA were removed. Small SV (<500 bp) not overlapping with CNA were also removed. Physiological SV at the IG and TCR loci were removed. We considered chromothriptic events those defined by the presence of seven or more oscillating changes between two different copy number states and/or the presence of more than six breakpoints occurring in a single chromosome. Chromoplexic events, defined by the ligation of fragments from three or more chromosomes by translocations, were inferred from WGS data. Similar to chromoplexia, templated insertions were considered when, in addition to ligation of three or more chromosome fragments, concomitant small duplications were found in all breakpoints. Breakage-fusion bridge cycles were inferred from WGS and SNP-array data based on the identification of patterns of focal copy number increases of chromosomal segments as well as fold-back inversions, together with the presence of telomeric deletions.

Telomere length

Telomere length was determined from WGS data using qMotif (v1.2, https://sourceforge.net/p/adamajava/wiki/qMotif%201.2) in the 61 tumor and normal samples. Briefly, reads mapping to the telomeric regions and containing 3 consecutive repeats of the most common telomeric motif (TTAGGG) or 2 adjacent repeats of this motif with variations in the first 3 positions were considered. We did not find any relationship between the age of the patients and the telomere length of their normal sample.

WGS-based characterization of the IG/*CCND1* translocation

We used a bespoke algorithm to characterize the IG/*CCND1* translocation at a base-pair resolution from WGS data. Briefly, the algorithm searched for paired reads in which one read mapped to any of the IG loci (heavy, kappa or lambda) and its pair mapped to the region surrounding *CCND1*. We also considered split reads in which a fragment of the read mapped to one of the IG loci and another fragment of the same read mapped near *CCND1*. By integrating both types of reads we extracted the most likely rearrangement, and the exact breakpoint was obtained from split reads. Of note, the intermediate fragment of the split reads that did not map to any of the two locations (neither IG nor *CCND1*) was considered as N nucleotides. We next used the gene annotation from wgEncodeGencodeBasicV19 to determine the gene found at/near the breakpoints obtained within the IG loci. Class switch regions have been described elsewhere (Huebschmann et al., in preparation). To determine the potential activity of AID in the breakpoints identified we searched for its three described motifs (WRCY, RGYW, WGCW). For recombination-activating genes (RAG) activity we searched for the RAG motifs observed near the J and D genes of the IG heavy chain locus (supplemental Table 8).

Immunoglobulin gene rearrangements and IGHV mutational status

Immunoglobulin gene rearrangements (heavy and light chain rearrangements and class switch recombination [CSR]) and IGHV mutational status were analyzed from WGS data using IgCaller (https://github.com/ferrannadeu/IgCaller) in the 61 cases with WGS available, as previously described (Nadeu *et al.* under review). On the other hand,

for the 21 cases without WGS, the IG heavy chain rearrangement and IGHV mutational status was previously analyzed by Sanger sequencing.[7]

Mutational signatures

Mutational signatures were analyzed for SNV, double base substitution (DBS), indels and SV using the 61 cases with WGS available. SNV were first classified in the 96 classes according to their 5' and 3' base. To enable the detection of mutational processes active in specific target regions but contributing to a small number of mutations we further divided SNV as clustered (5 or more mutations in less than 1000 bp intermutational distance) or unclustered, and according to the location of the mutations (within or outside of IG loci). Mutational signatures were extracted de novo using the MutationalPatterns R package (v1.4.3) using a non-negative matrix factorization (NMF).[20] The extracted signatures were compared to the signatures described in COSMIC (https://cancer.sanger.ac.uk/cosmic/signatures/SBS), and those with a cosine similarity >0.7 were considered. To measure the contribution of each signature in each tumor we used a fitting approach (MutationalPatterns) and iteratively removed the less contributing signature if removal of the signature decreases the cosine similarity between the original and reconstructed 96-profile <0.01, as previously described.[21] To study the profile of DBS and indels we classified them in 78 and 84 classes, as recently described (https://github.com/ferrannadeu/indelsClassification).[21] The number of DBS and indels in MCL was relatively small to properly perform a mutational signature analysis, and therefore we only compared the profile of DBS and indels between cMCL and nnMCL. The Palimpsest R package (1.0.0)[22] was used to study SV signatures. SVs were classified into 38 categories according to the SV type (deletions, duplications, inversions or translocations) and their size, and signatures were extracted using NMF.

Identification of driver alterations

Using the integration of the data, an in-house workflow was performed to identify driver alterations: i) mutated driver genes were defined considering type of mutation, gene size, coverage, and local density of mutations as previously described.[6] ii) GISTIC algorithm (2.0.23)[23] was used to identify driver CNA from the 61 cases from WGS and the 21 cases with WES. iii) Driver altered genes and CNA were considered when they

were recurrently altered in 3 or more cases, and statistically significant (false discovery rate 18%). Genes mutated at lower frequency but carrying known driver or truncating mutations were also considered as drivers. CNA not significant by GISTIC but with high frequency (>20%) or with focal homozygous deletion or amplification and previously published in more than one MCL series were also considered as drivers. Recurrently mutated driver genes in the present series but that were not expressed in any normal B cell[24] nor in MCL were flagged as "not expressed" and not considered as driver alterations in subsequent analyses. These genes corresponded to large genes located in late-replicating regions.

Clonal representation and temporal order of the driver alterations

The cancer cell fraction (CCF) of the SNV and indels was calculated integrating read counts, copy-number status of each loci and tumor purity, as previously described.[25] The CCF of the CNA was obtained from Battenberg (samples with WGS) or from SNP-array data (samples with only WES), as previously described.[26] To assess the temporal order of the driver alterations we tested for each alteration the enrichment of out-going edges compared to in-going edges, where out-going edges corresponded to the number of instances where the alteration was present at higher CCF than other alteration in the same tumor sample, and in-going when it was present at lower CCF, as previously described.[27] Driver alterations were classified as early, late (false discovery rate <0.2) or intermediate/not powered.

Gene expression profiling

Global gene-expression profiling of RNA extracted from 44 samples was performed using Affymetrix U219 microarrays (Thermo Fisher Scientific, Waltham, MA). The raw data were normalized using the RMA method. The expression of each gene was summarized with the mean of all the probesets mapping the corresponding gene, except for the *CCND1* gene which all probesets were independently included in the analyses. We also determined the gene expression signatures of cMCL and nnMCL using the previously described Nanostring L-MCL16 Assay.[1]

Real-time quantitative PCR

The gene expression levels of *TERT* (Hs99999022_m1, Thermo Fisher Scientific) were investigated by quantitative PCR (qPCR) using GUSB gene as endogenous control and Universal Human Reference (UHR) RNA as a calibrator, as previously described.[2] Experiments were performed in triplicate.

DNA methylation analyses

We generated DNA methylation profiles for 70 MCL with genetic data using Infinium MethylationEPIC BeadChip (Illumina) following manufacturer's instructions. We used minfi R package to mine and normalize raw idat files.[28] Several probes were excluded including those with bad detection p-value (≥0.01), non-CpG probes, CpGs falling in SNPs with MAF=0, chrY and chrX probes, and individual-specific probes. We then removed samples with bad intensity signal and/or bad probe conversions (data not shown). Finally, SWAN algorithm was used to normalize the data.[29]

In order to construct the epiCMIT score,[30] we first selected CpGs residing in repressive chromatin context in normal and neoplastic B cells using ChIP-seq from Blueprint.[31] Specifically, we mapped CpGs falling at poised promoters (H3K4me1, H3K4me3 and H3K27me3), H3K27me3, H3K9me3 and low signal heterochromatin chromatin states, in 15 normal B cell samples, 5 MCL, 7 chronic lymphocytic leukemia, and 5 multiple myeloma cases. Since normal B-cell development involves active proliferation in normal physiological conditions, we thus excluded CpGs not modulated during normal B-cell maturation. To have a clear ground state, we took unmethylated (≤0.1) or methylated (≥0.9) CpGs in hematopoietic precursor cells (HPC), which have a low proliferative history. We selected those CpGs showing at least 0.5 methylation difference between HPC and terminally-differentiated bone marrow plasma cells to retain CpGs with a clear modulation upon cell division, which have accumulated the proliferative history of the entire B-cell differentiation. We separated our analyses in CpGs gaining methylation in poised promoters and H3K27me3 regions and CpGs losing methylation in H3K9me3 and low signal heterochromatin regions, obtaining 184 CpGs and 1164, respectively. Next, we calculated epiCMIT-hyper and epiCMIT-hypo as the average methylation values of those 184 and 1164 CpGs. To have epiCMIT-hyper and epiCMIT-hypo in the same scale, the former was computed as 1 minus epiCMIT-hypo.

Finally, the greater of both scores was considered to obtain the final epiCMIT score. Further information regarding the epiCMIT score might be found in Reference [30].

Reference epigenomes

We used the chromatin states of normal B cell subsets (n=15) and MCL cases (n=5) that we generated within the Blueprint Consortium.[31] We measured the contribution of each SNV mutational signature in each chromatin state region. As we observed that the signature enrichments were similar within both groups (i.e. normal and neoplastic B cells), we represented as bar plots the merged results of mutational signature association to different chromatin states. Similarly, we annotated all SV breakpoints identified (with the exception of IG/*CCDN1* translocations) to the available chromatin states and calculated the enrichments of distinct chromatin states in the breakpoints as compared to the distribution of chromatin states within whole genome. Furthermore, in a similar way, we calculated chromatin states enrichments of specific SV types (translocations, deletions, duplications and inversions). Log2 mean fold change of both MCL and normal B cells were represented as a heatmap. We also used the chromatin state data of normal B cells and MCL cases to characterize the region surrounding the IG/*CCDN1* translocation.

In situ Hi-C

In situ Hi-C was obtained for two sorted human B-cell subpopulations, naive (NBC) and memory (MBC) B cells, and the five MCL cases, two cMCL and three nnMCL, in which we had the reference epigenome.[32] Contact matrices figures were generated using the 3D Genome Browser online platform.[33]

Topologically associating domain (TAD) borders at the *CCND1* region were assessed using TADpole[34] on the 20Kb raw Hi-C interaction matrices of MCL cases as well as naive and memory B cells.

PCR and Sanger sequencing verification

A subset of SNV affecting driver genes and a subset of SV from cases with available material were selected for PCR and Sanger sequencing verification. Specific primers were designed using Primer 3 software (supplemental Tables 3-4). PCR products were

treated using ExoSap IT (USB Corporation) and sequenced with ABI Prism BigDye terminator v3.1 (Applied Biosystems) and 5 pmol of each primer. Sequencing reactions were run on an ABI-3730 automated sequencer (Applied Biosystems). All sequences were examined with the Mutation Surveyor DNA Variant Analysis Software (Softgenetics). We verified 26/26 (100%) SNV and 66/75 (88%) SV that were selected according to DNA availability (supplemental Tables 5-6).

Cytogenetics and fluorescence in situ hybridization (FISH)

Conventional cytogenetics was performed on Giemsa-banded chromosomes (G-banding) obtained after a 72-hour culture using tetradecanoyl-phorbol-acetate stimulation. When possible, at least 20 metaphases per sample were analyzed. Karyotypes were described according to the International System for Human Cytogenetic Nomenclature.[35] A complex karyotype was defined by the presence of at least three clonal chromosomal aberrations, including at least one structural aberration was detected.[36] FISH analyses were performed using probes that interrogated for *CCND1* rearrangement (locus specific (LSI) t(11;14) CCND1/IGH dual fusion (DF) [Vysis, Abbolt Molecular], and XL CCND1 dual color (DC) break-apart (BA) [MetaSystems]), LSI BCL6 (3q27) DC BA, LSI MYC (8q24) DC BA, LSI ATM (11q22) DC, CEP 12, LSI DLEU (13q14.3) DC, LSI TP53 (17p13) DC, and LSI BCL2 (18q21) DC BA (Vysis, Abbot Molecular). A custom, non-commercial *TERT* three-color BA probe was applied to verify *TERT* alterations identified by WGS or copy number arrays.[37] At least 200 nuclei were examined for each probe, whenever possible. Whole chromosomal paintings (WCP) were performed in 5 cases with complex rearrangements to verify/elucidate the alterations identified by WGS and G-banding. The analyses were performed using standard fluorescence microscopes (Nikon Eclipse 50i or Zeiss) equipped with appropriate filter sets. ISIS FISH Imaging System (MetaSystems) was used for acquisition and processing of digital images.

Statistical analyses

Pearson and Spearman correlations were used to measure the association between continuous variables, as appropriate. Student t, Mann–Whitney U and Kruskal–Wallis tests were used to compare continuous variables between two or more groups, as

appropriate. Fisher exact test was used to compare categorical variables between groups. Multivariate linear regression was used to control for the cell-of-origin (COO) in the epiCMIT analyses. The endpoint of the study was overall survival (OS). OS curves were estimated using the Kaplan-Meier (KM) method. The log-rank test and Cox regression were used to assess the relationship of the OS with categorical and continuous variables, respectively. Cox regression was used to estimate hazard ratios (HR) and to evaluate the independent prognostic value of multiple variables. *P* values were adjusted using the Benjamini-Hochberg method (*Q* values).

SUPPLEMENTAL TABLES

Supplemental Tables are placed in the Supplemental Tables Excel file.

Table S1. Techniques applied to each MCL sample

Table S2. Statistics and details of samples subjected to WGS

Table S3. Primers used for the verification of gene mutations

Table S4. Primers used for the verification of structural variants

Table S5. Mutations verified by Sanger sequencing

Table S6. Structural variants verified by Sanger sequencing

Table S7. IG gene characterization using IgCaller

Table S8. RAG motifs used within the analysis of the IG/*CCND1* rearrangement

Table S9. Somatic mutations identified in 61 MCL WGS *(.txt file)*

Table S10. Structural variants identified by WGS in 61 MCL

Table S11. Copy number alterations identified by WGS in 61 MCL

Table S12. WGS-based characterization of the t(11;14)

Table S13. Clinicopathologic and molecular characteristics of the 5 MCL patients with a non-RAG mediated t(11;14)

Table S14. Coding somatic mutations in 82 MCL analyzed by WGS/WES

Table S15. Copy number alterations identified by SNP-array in 21 MCL with WES data

Table S16. Significant driver gene alterations

Table S17. Significant and recurrent driver copy number alterations

Table S18. Number and type of alteration of MCL drivers

Table S19. Temporal order of acquisition of driver alterations

Table S20. Germline mutations in *ATM* and *CHEK2*

SUPPLEMENTAL FIGURES

Figure S1. Correlation between mutation types in MCL subtypes

Box plots and scatter plots of the total number of mutations separated by type (single nucleotide variant [SNV], double base substitution [DBS], and short insertions/deletions [indel]) in cMCL (blue) and nnMCL (yellow). Pairwise correlations are shown independently for each molecular subtype. One nnMCL (M364) had a remarkable high number of indels (n=712), although no microsatellite instability was detected in this case.

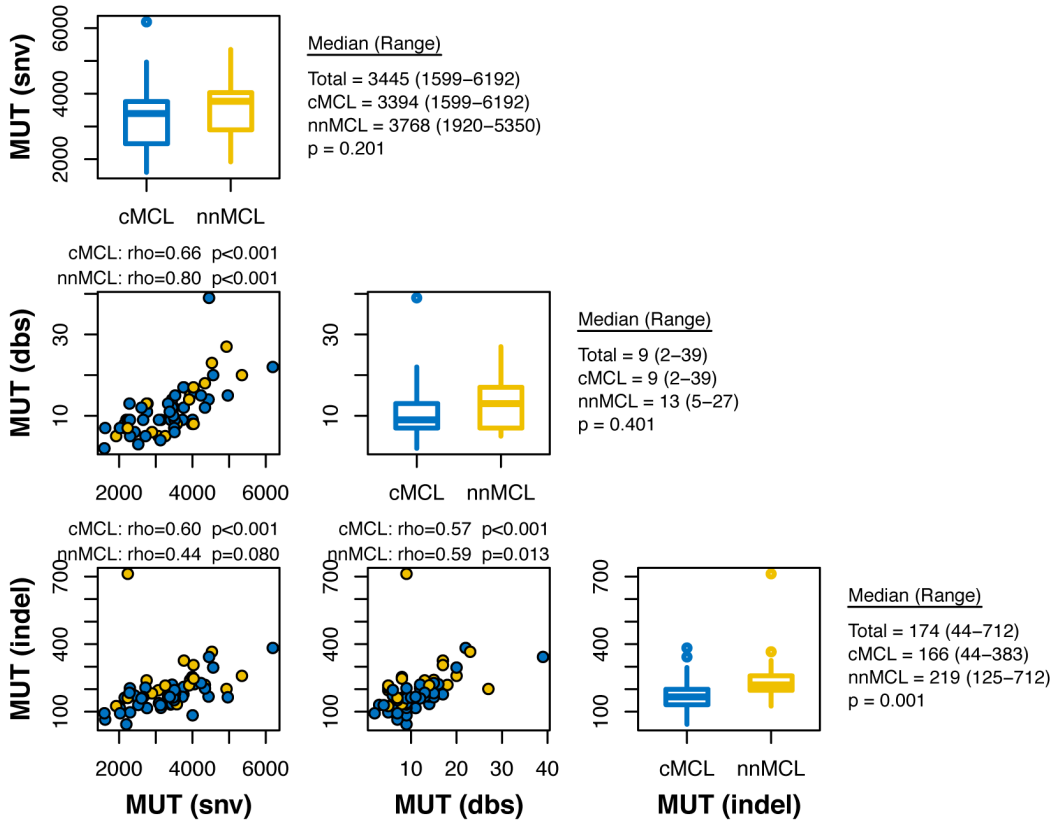


Figure S2. Correlation between copy number alterations in MCL subtypes

Box plots and scatter plots of the total number of structural variants (SV) (intrachromosomal and interchromosomal) and copy number alterations (CNA) (gains and losses) in cMCL (blue) and nnMCL (yellow). Pairwise correlations are shown independently for each molecular subtype.

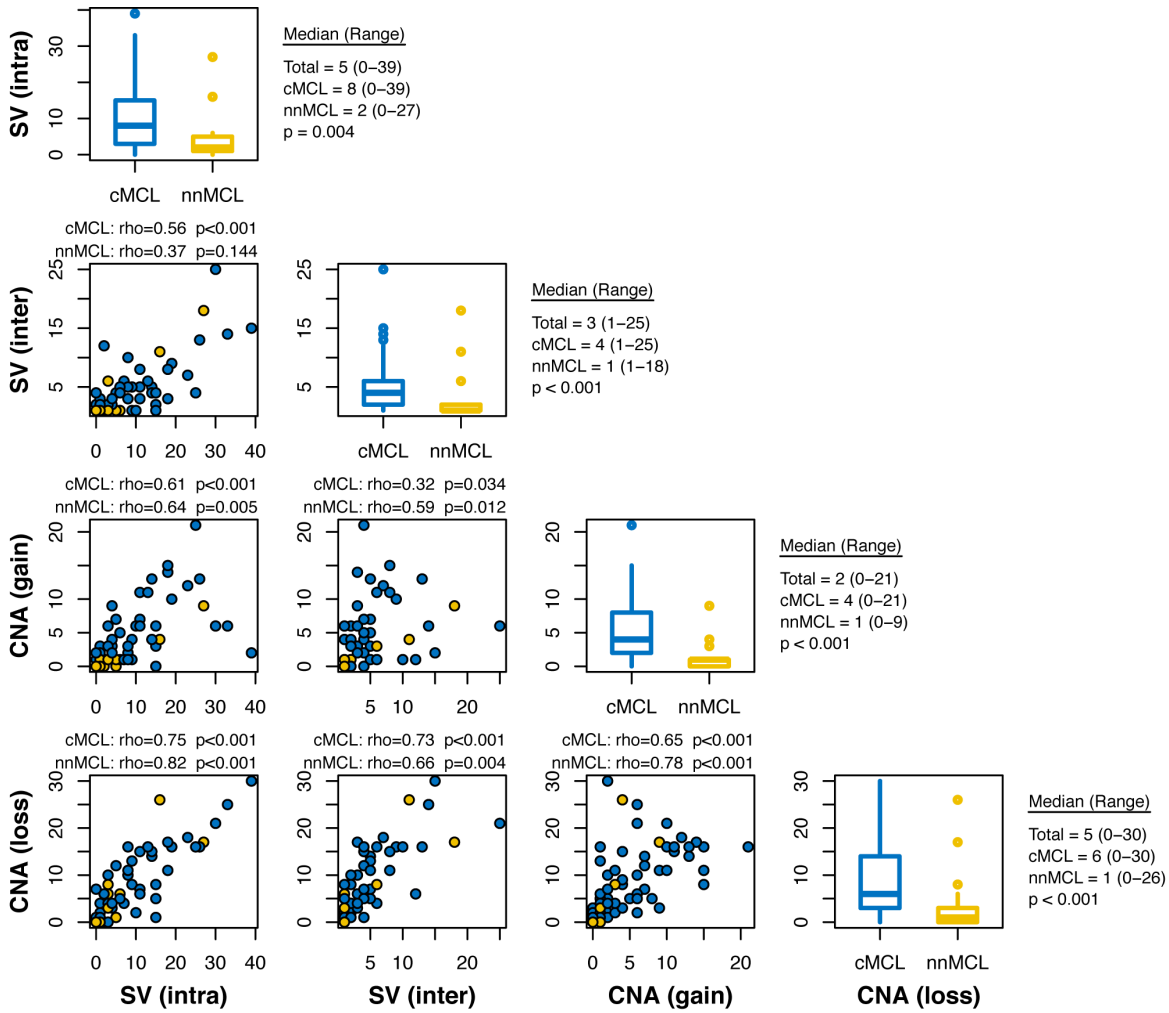


Figure S3. Correlation between alterations in MCL subtypes

Box plots and scatter plots of the total number of mutations, structural variants (SV), copy number alterations (CNA), and driver alterations in cMCL (blue) and nnMCL (yellow). Pairwise correlations are shown independently for each molecular subtype. Only one nnMCL lacked a driver alteration (note that the primary IG/*CCND1* translocation was not considered as driver).

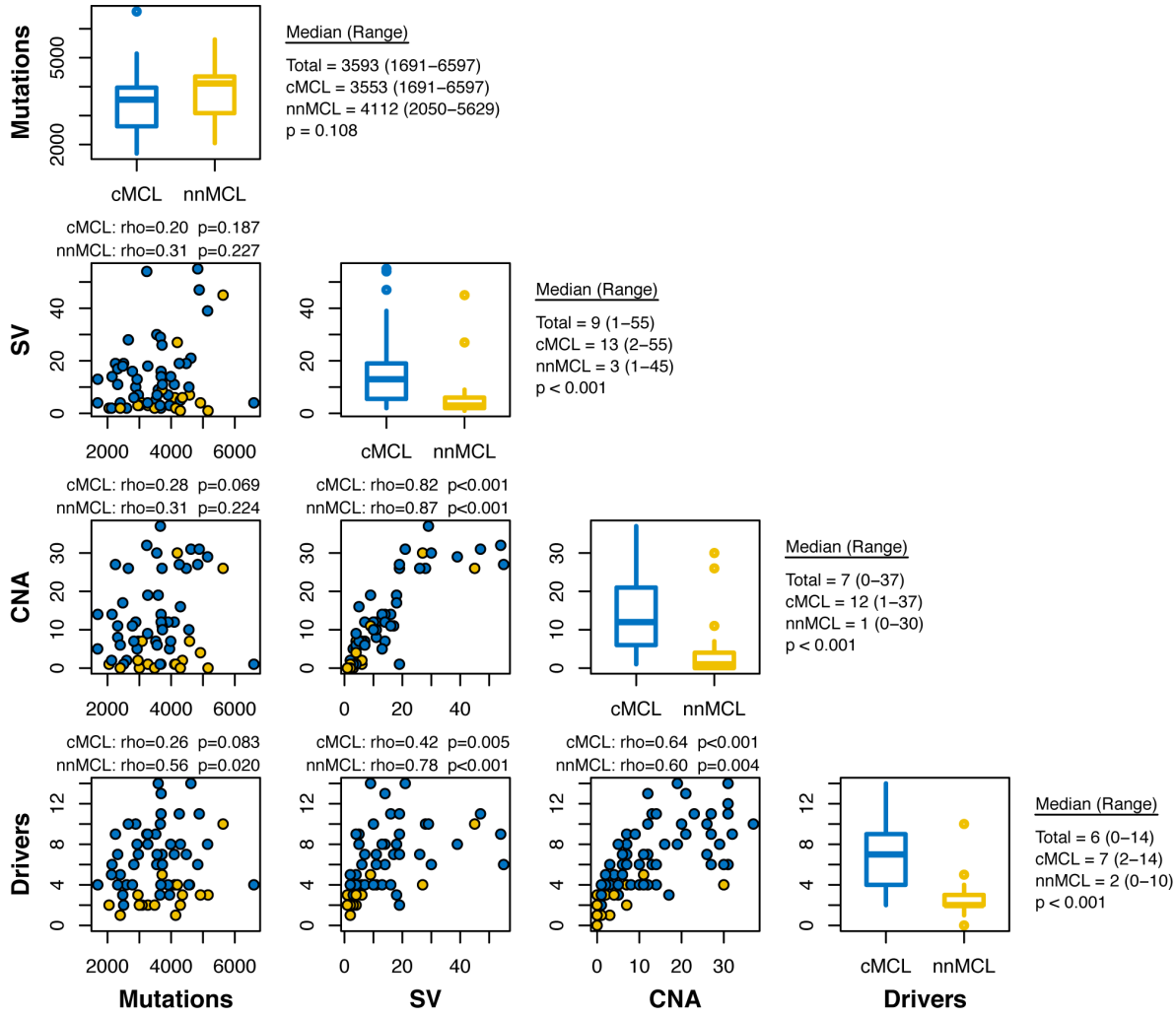


Figure S4. SNV mutational signatures, and DBS and indel profiles

A. Mutational signatures identified de novo in MCL samples in unclustered (*left*) and clustered (*right*) SNVs. The signature from COSMIC with the highest cosine similarity is indicated on the right. Note that we identified the potential presence of SBS37 in clustered mutations which is not reported in the main text. Considering its unknown etiology and the small number of mutations attributed to this signature we do not feel confident to conclude that this signature might be present/relevant in MCL. **B.** Profile of DBS in cMCL and nnMCL. **C.** Indels profile in cMCL and nnMCL.

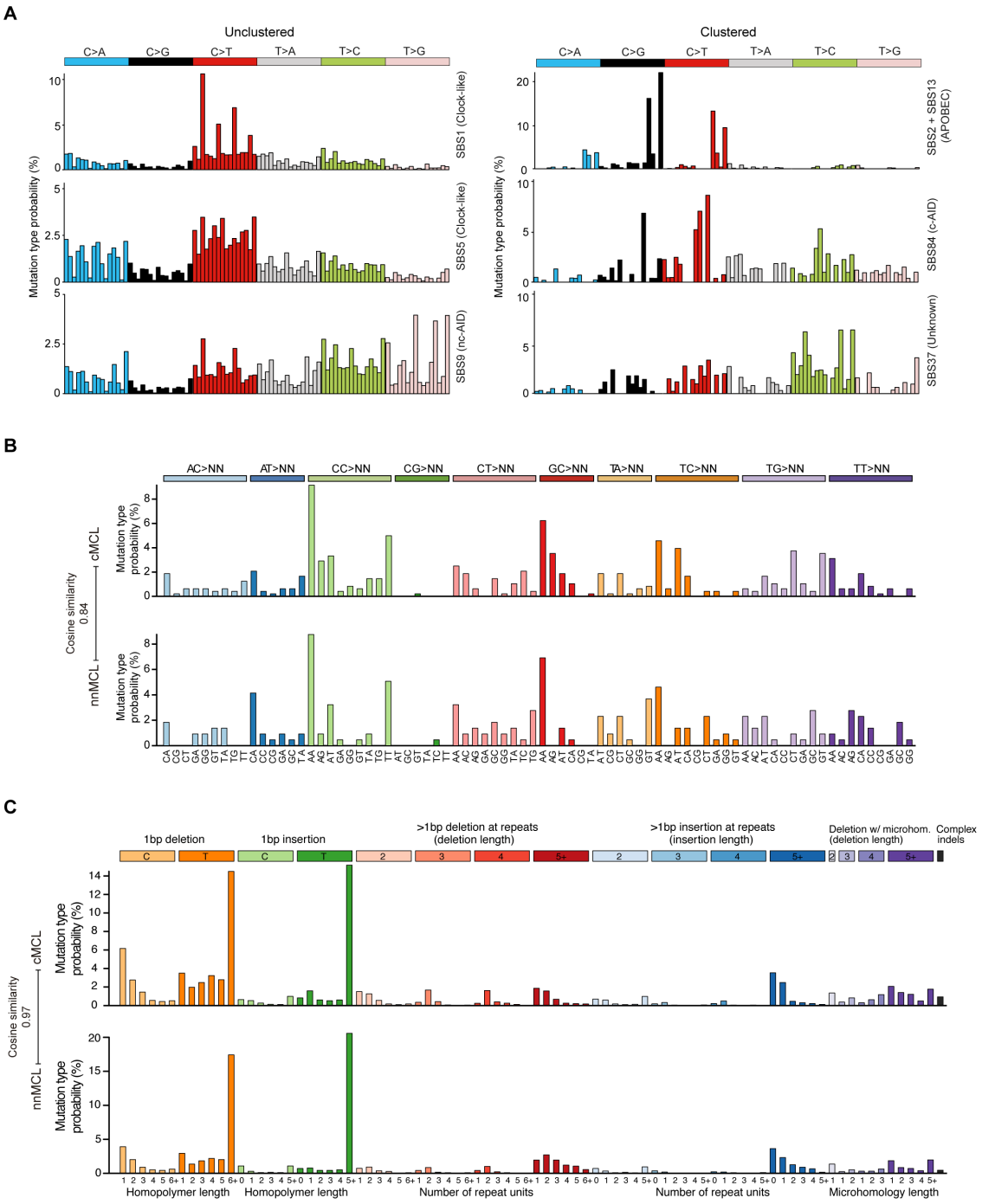


Figure S5. Contribution of SNV mutational signatures in specific chromatin states in MCL subtypes

Mutations detected in cMCL (*top*) and in nnMCL (*bottom*) and the contribution of each SNV mutational signature in each of the defined chromatin states.

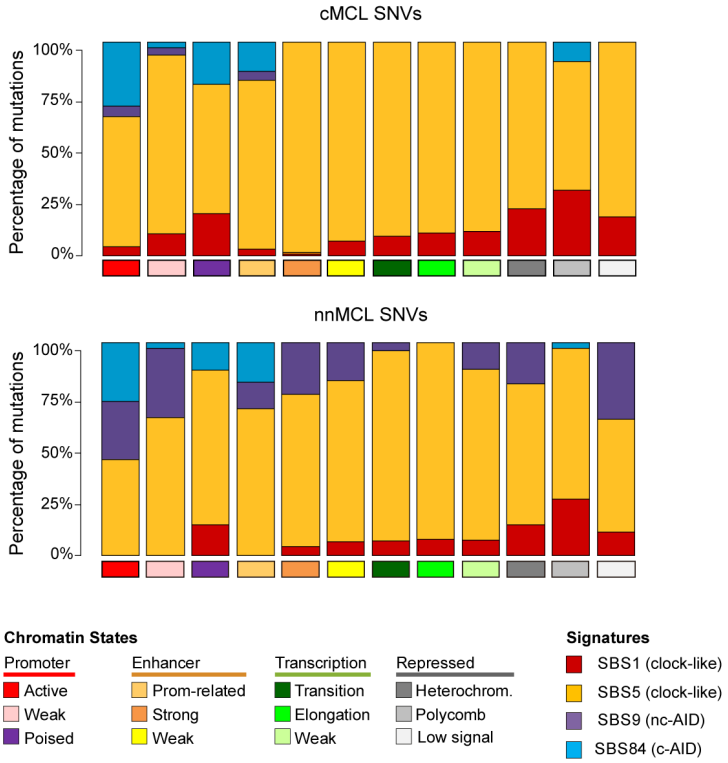


Figure S6. Characterization of mutational signatures related to AID and APOBEC activity in MCL subtypes

A. Contribution of AID-related mutational signatures per case. **B.** Enrichment of AID-related mutations in nnMCL. **C.** Contribution of APOBEC-related mutational signatures in cMCL and nnMCL. **D.** Contribution of the mutational signatures identified in MCL according to their cluster or unclustered nature, and the IG loci. **E.** Examples of APOBEC-related mutations in kataegic sites overlapping with structural variants (SV) in chromosomes 8 and 12 for case M244 (nnMCL).

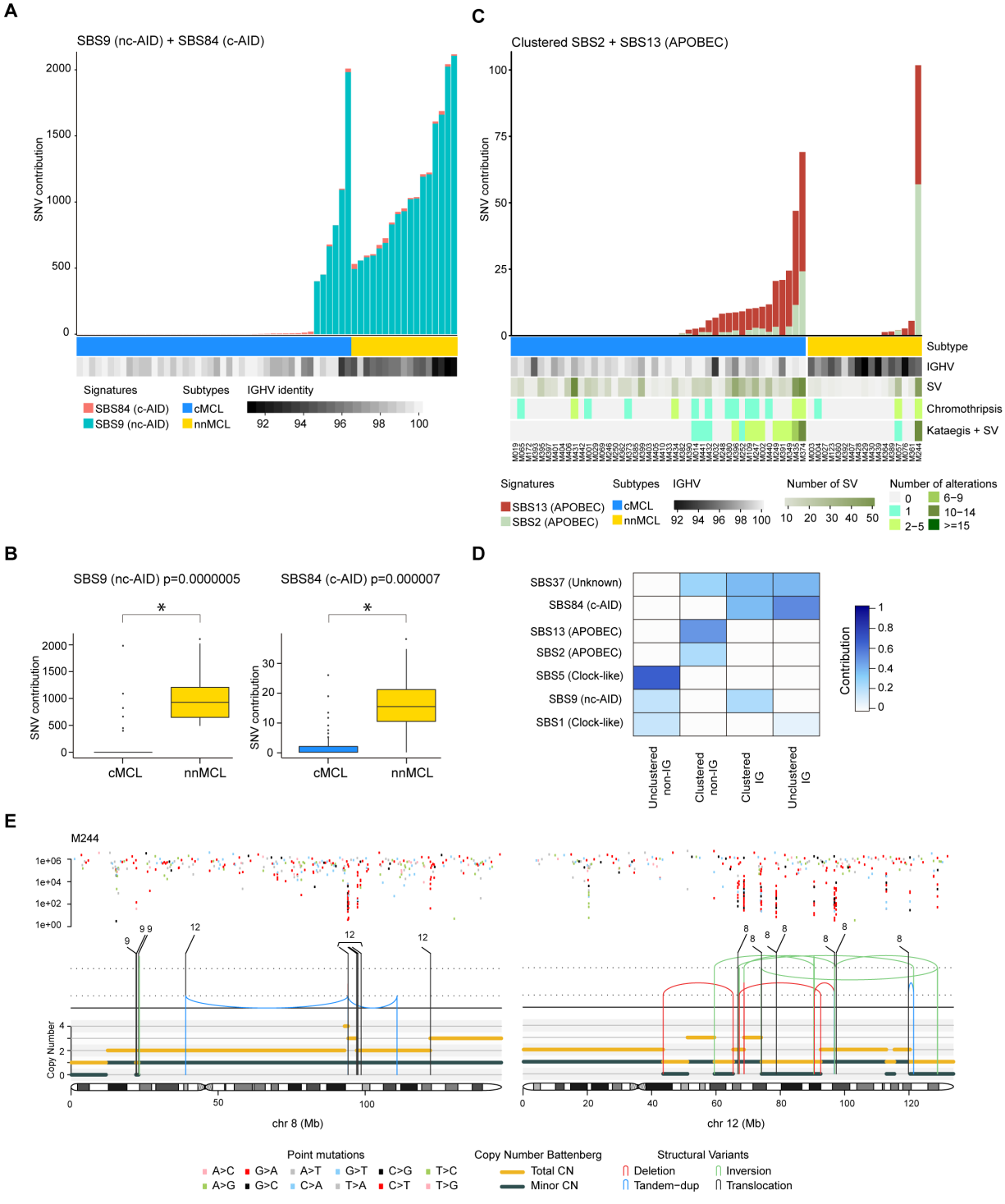


Figure S7. Distribution of kataegis

A. Distribution of kataegic sites according to chromosome and with/without overlapping structural variants (SV). **B.** Frequency of APOBEC-related mutations in kataegic sites according to the presence/absence of overlapping SV. **C.** Distribution of SV types with overlapping kataegis. Each SV type is adjusted by its number of alterations.

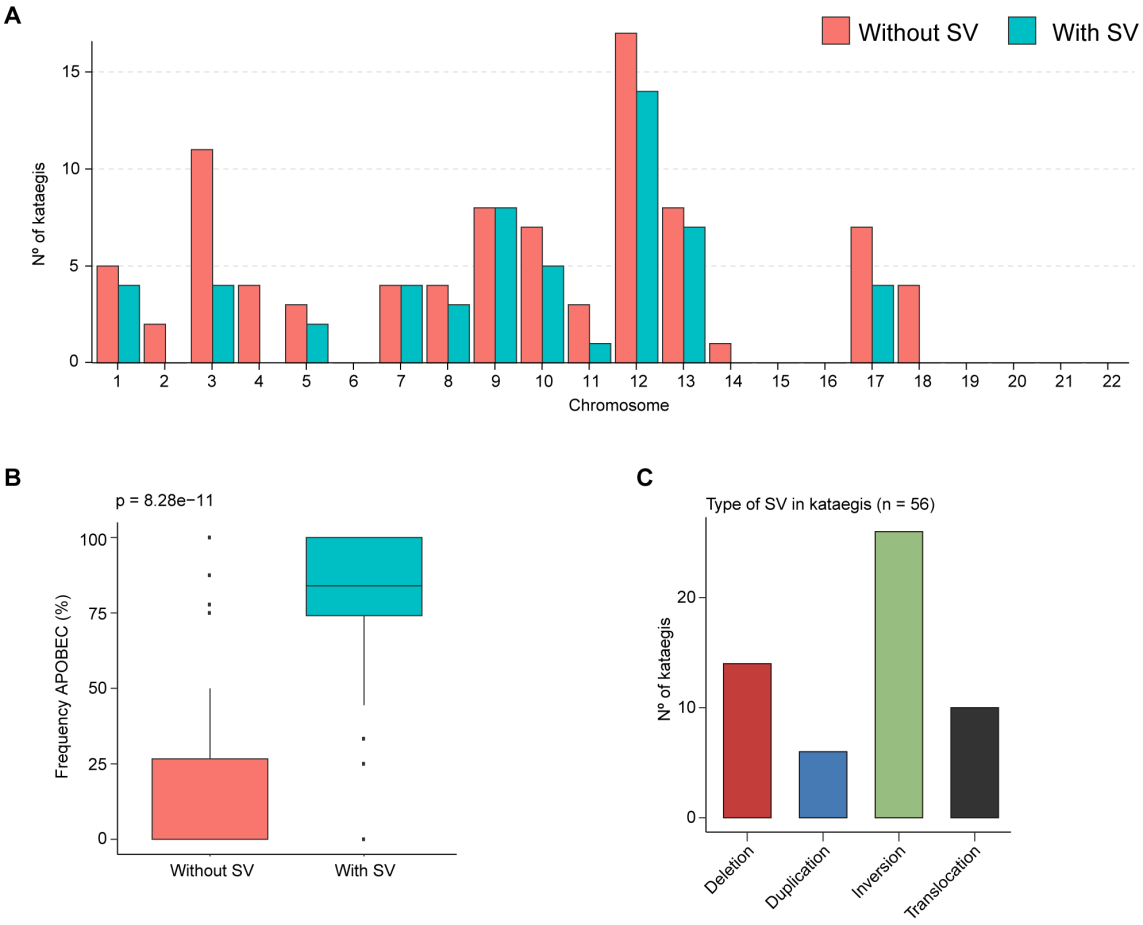


Figure S8. IGHJ and IGHD genes involved in t(11;14) and V(D)J rearrangements

Distribution of IGHJ (*top*) and IGHD (*bottom*) genes involved in the t(11;14) in cMCL and nnMCL as well as in the productive V(D)J rearrangement of both MCL and normal B cells. J and D usage of normal B cells were obtained from a previous study.[38]

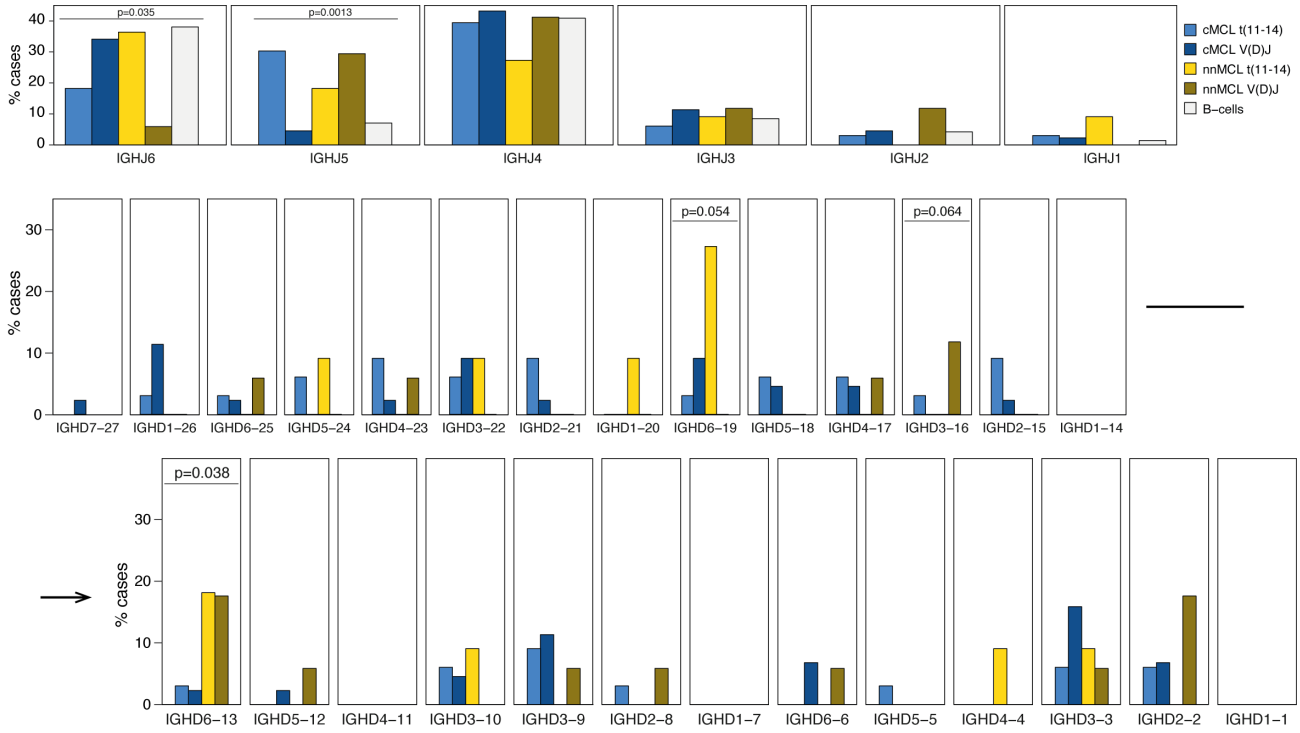


Figure S9. Representation of non-D-J translocations

A. Representation of three cases with IG breaks at canonical RSS sites of IGHD and IGHV genes, suggesting that these translocations were acquired during the second step of the V(D)J recombination process. Note that the IGHV3-30-0 gene was not annotated in the reference used, but obtained from IMGT database.[39] Patients' labels are depicted in blue for cMCL and yellow for nnMCL. **B.** Schema of the three cases with IG breaks at cryptic RSS sites or at canonical RSS sites but followed by exonucleolytic removal of several bases. **C.** Representation of three cases with complex rearrangements involving additional chromosomes. AIDm, AID motif close to the breakpoint. CpG, CpG close to the breakpoint. RSS, RAG signal sequence. N_x, numer of N nucleotides observed in the junction. **D.** Case with unbalanced translocation due to the insertion of the IGK locus upstream of *CCND1*.

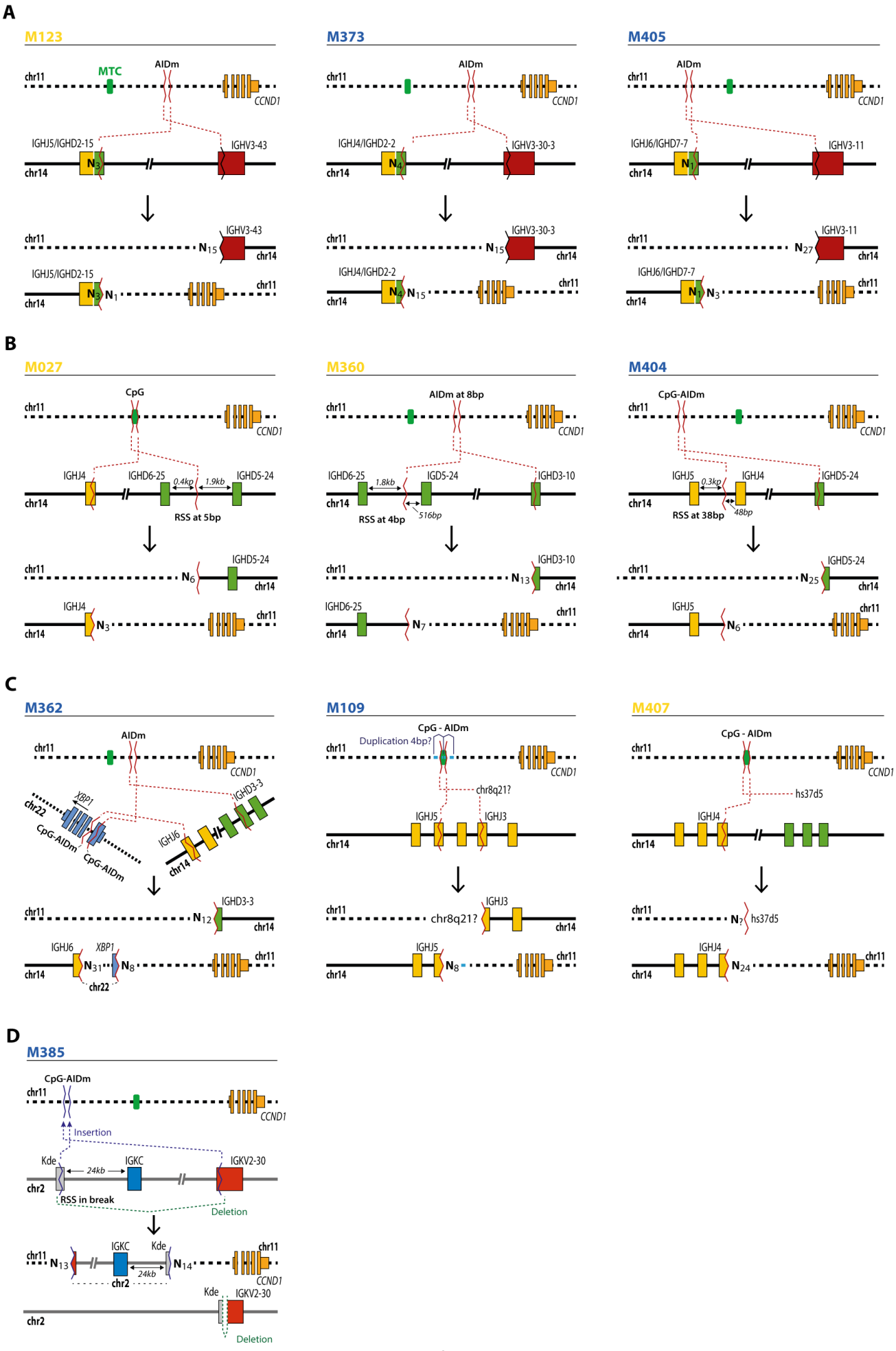

A
M123
MTC
AIDm
chr11
CCND1
IGHJ5/IGHD2-15
IGHV3-43
chr14
N3
N15
N1
M373
IGHJ4/IGHD2-2
IGHV3-30-3
N4
M405
IGHJ6/IGHD7-7
IGHV3-11
N27
B
M027
CpG
IGHJ4
IGHD6-25
0.4kp
1.9kb
IGHD5-24
RSS at 5bp
N6
N3
M360
AIDm at 8bp
1.8kb
IGD5-24
IGHD3-10
RSS at 4bp
516bp
N13
N7
M404
CpG-AIDm
IGHJ5
0.3kp
48bp
N25
N6
C
M362
XBP1
chr22
CpG-AIDm
IGHD3-3
IGHJ6
N12
N31
N8
M109
Duplication 4bp?
chr8q21?
IGHJ3
IGHJ5
M407
hs37d5
N7
N24
D
M385
Insertion
Kde
IGKC
24kb
IGKV2-30
chr2
RSS in break
Deletion
N13
N14

Figure S10. Reads spanning the t(11;14) and V(D)J rearrangements (M172)

Whole-genome sequencing reads from case M172 (cMCL) spanning the t(11;14) are shown in green, while reads supporting the IGHJ6-IGHV3-13 (productive) rearrangement and the IGHJ6-IGHV4-34 (productive and expressed) rearrangement are depicted in pink and blue, respectively. By the lack of mutations in the t(11;14) and IGHJ6-IGHV3-13 reads we might accept that the translocation occurred in the IGHJ6-IGHV3-13 allele. Histograms on top represent the coverage of this region both for the normal and tumor sample. If a mutation is present it is highlighted using non-gray colors.

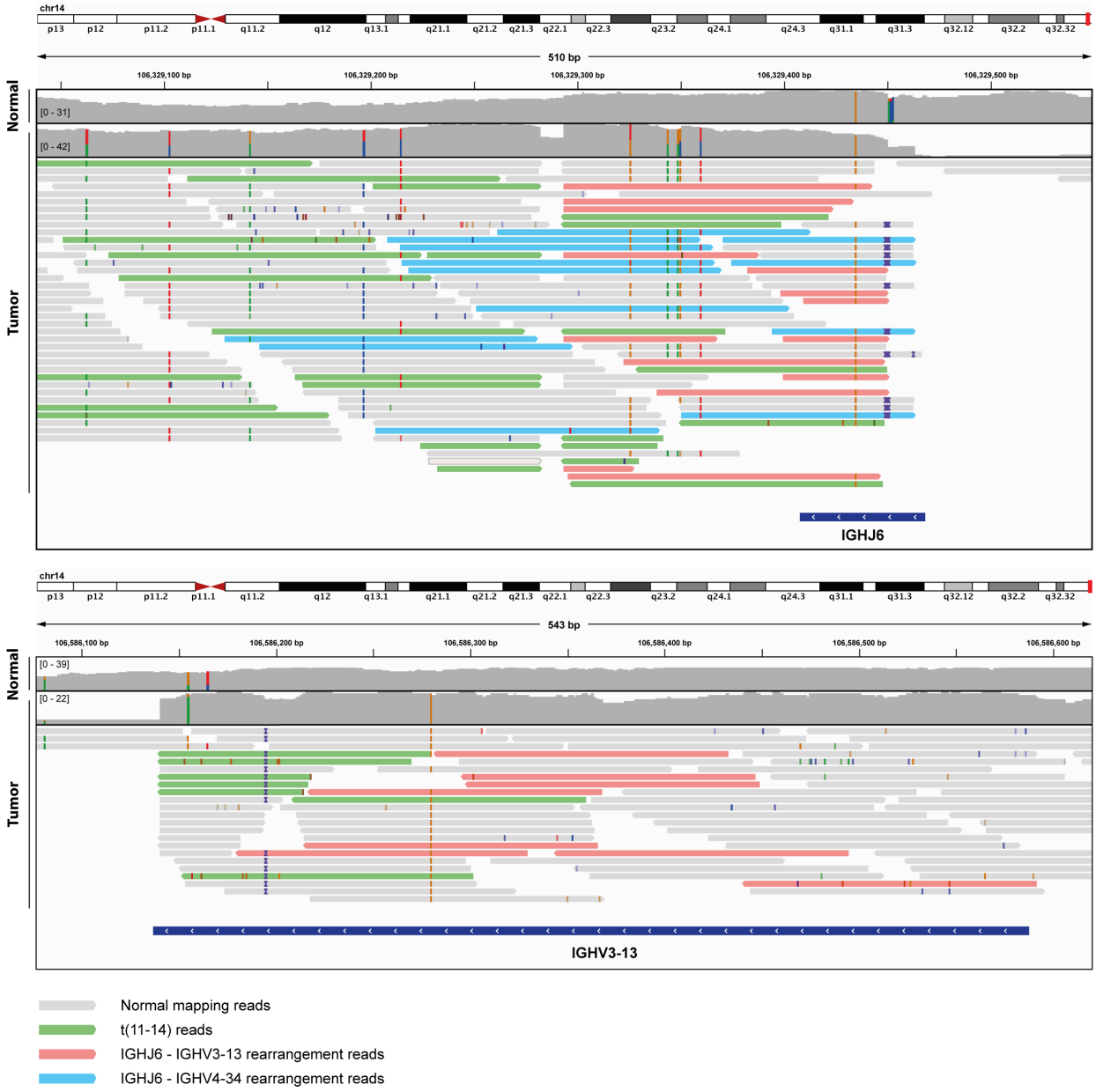


Figure S11. Reads spanning the t(11;14) and V(D)J rearrangements (M361)

Whole-genome sequencing reads from case M361 (nnMCL) spanning the t(11;14) are shown in green, while reads supporting the IGHJ3-IGHV4-31 (unproductive) and IGHJ2-IGHV2-26 (productive rearrangement) are depicted in pink and blue, respectively. The presence of t(11;14) supporting reads both between IGHJ4-IGHJ3 and within the IGHV4-31 gene argue for this translocation being occurred in the rearranged IGHJ3-IGHV4-31 allele.

Figure S12. *CCND1* expression

A. Long and short *CCND1* isoform expression by microarray depicted based on the underlying t(11;14) mechanism observed, the presence of *CCND1* secondary events, and the total number of CNA and SV. **B.** *CCND1* expression based on the location of the *CCND1* break**.** MTC, major translocation cluster. **C.** Box plot distribution of *CCND1* expression according to the t(11;14) type. **D.** *CCND1* expression based on the presence of *CCND1* secondary activating alterations (i.e. *CCND1* secondary event).

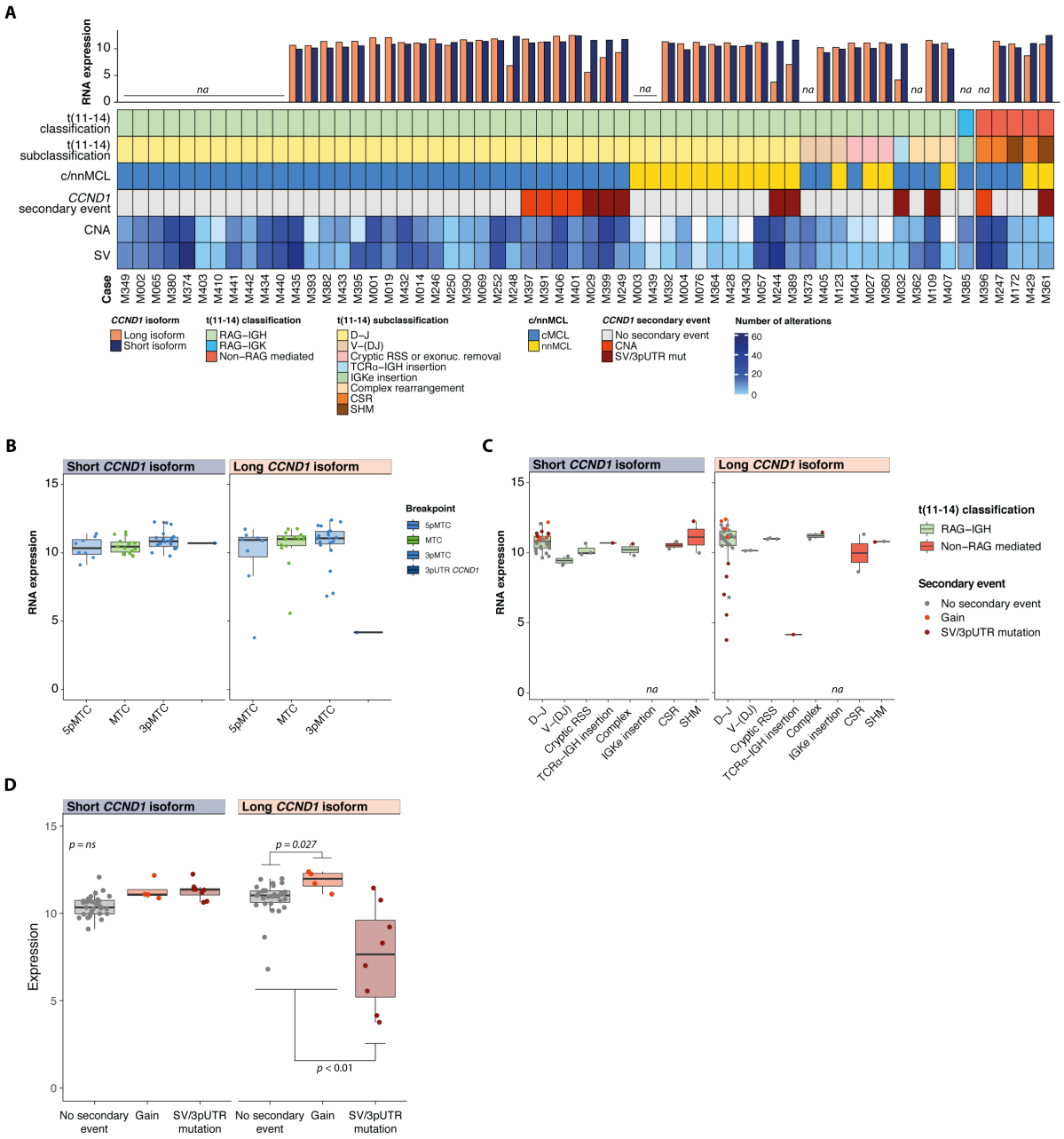


Figure S13. Representation of 6 MCL cases showing duplications of the region nearby chr11 breakpoints

AIDm, AID motif close to the breakpoint. CpG, CpG close to the breakpoint. RSS, RAG signal sequence. N_x, numer of N nucleotides observed in the junction. Dup, duplication. MTC, major translocation cluster. Patients' labels are depicted in blue for cMCL and yellow for nnMCL.

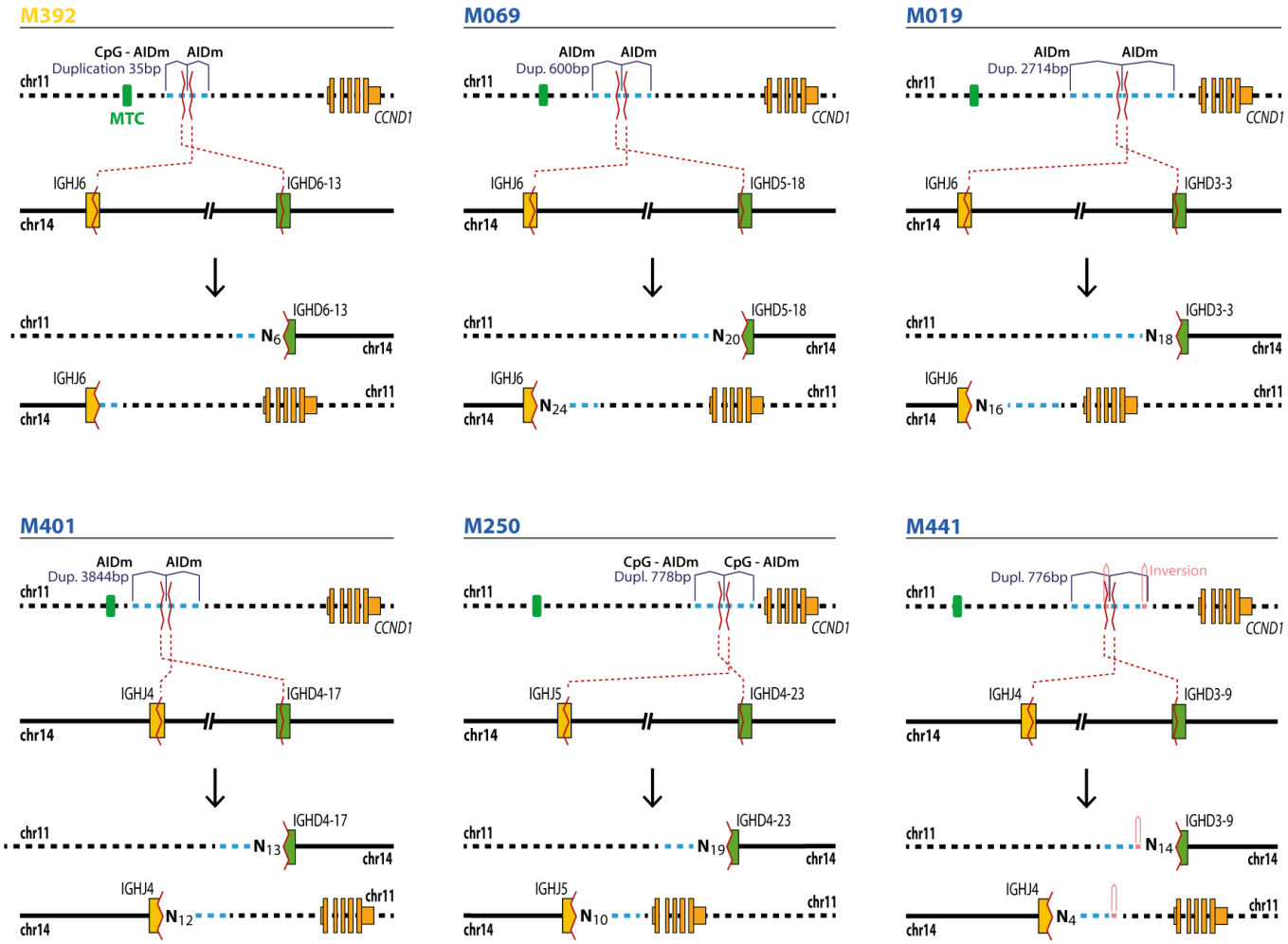


Figure S14. *CCND1* breakpoints

A. Distribution of the *CCND1* breakpoints in cMCL (*top*) and nnMCL (*bottom*). The major translocation cluster (MTC) region is highlighted in green. Of note, nnMCL were more often kappa compared to cMCL (11/17, 65% vs 21/43, 49%). Nonetheless, this difference was not statistically significant (*P*=0.16). This observation is consistent with other studies in which "indolent MCL" (probably corresponding to our nnMCL) express more frequently kappa than lambda.[40] **B.** Breakpoints are depicted based on the light chain expression of each case. The two cases expressing both kappa and lambda had a productive IGK as well as a productive IGL rearrangement identified by IgCaller, but only one IGH productive rearrangement, and were considered as single clones expressing both light chains. **C.** Breakpoints are depicted based on the presence of CpG or AID motifs (WRCY, RGYW, WGCW) within 10bp. Bar plots on the right of each section summarizes breaks according to their location.

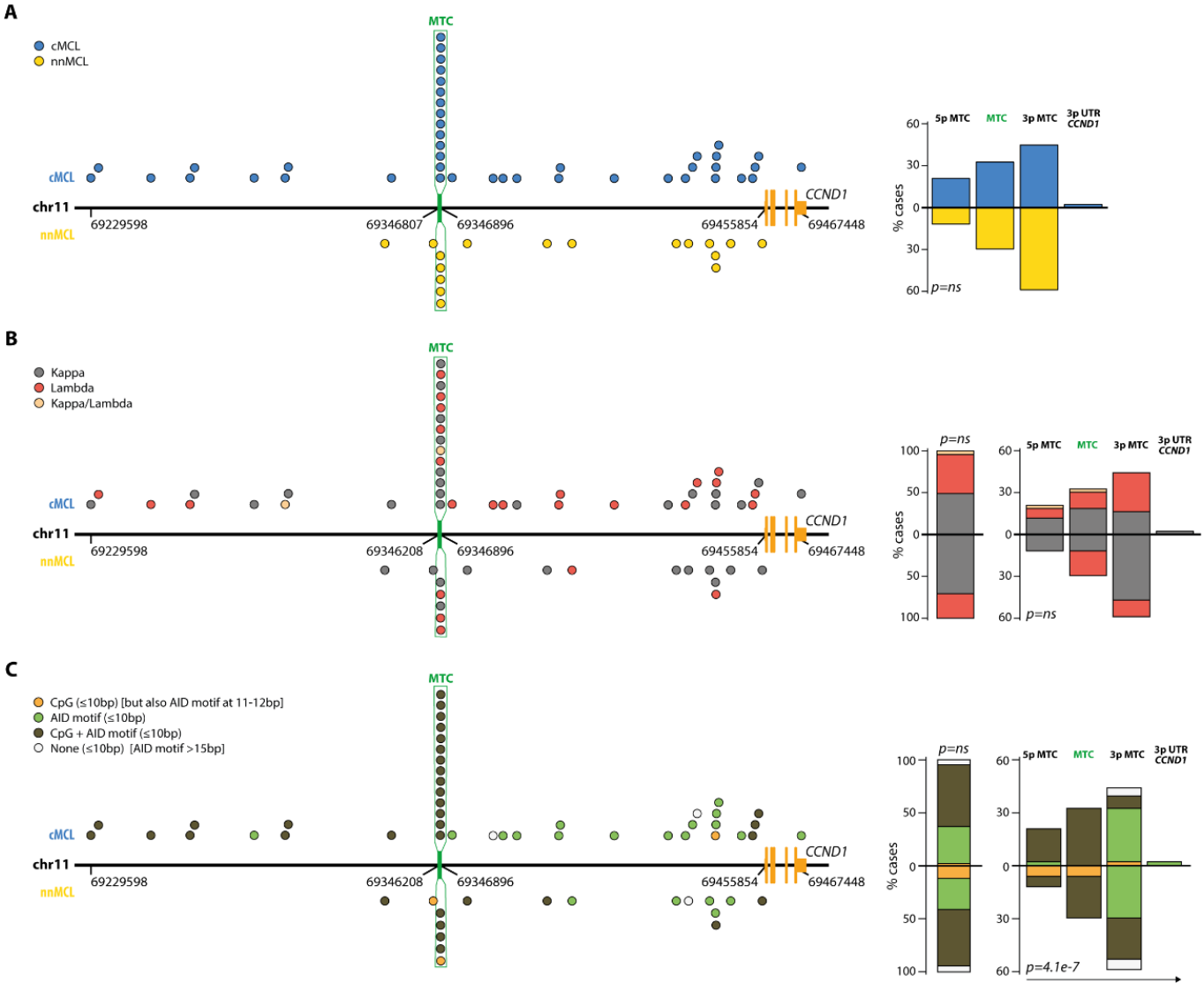


Figure S15. Chromatin conformation of the *CCND1* loci

Hi-C contact matrices for naive B cells (NBC) and the two cMCL cases (M019 and M001) (*left*), and memory B cells (MBC) and one nnMCL case (M076) (*right*). Color code indicates Hi-C interactions. The location of the *CCND1* gene, TAD borders, t(11;14) breakpoint and major translocation cluster (MTC) are marked. The represented region corresponds to chr11:68,987,467-70,066,106 (hg19).

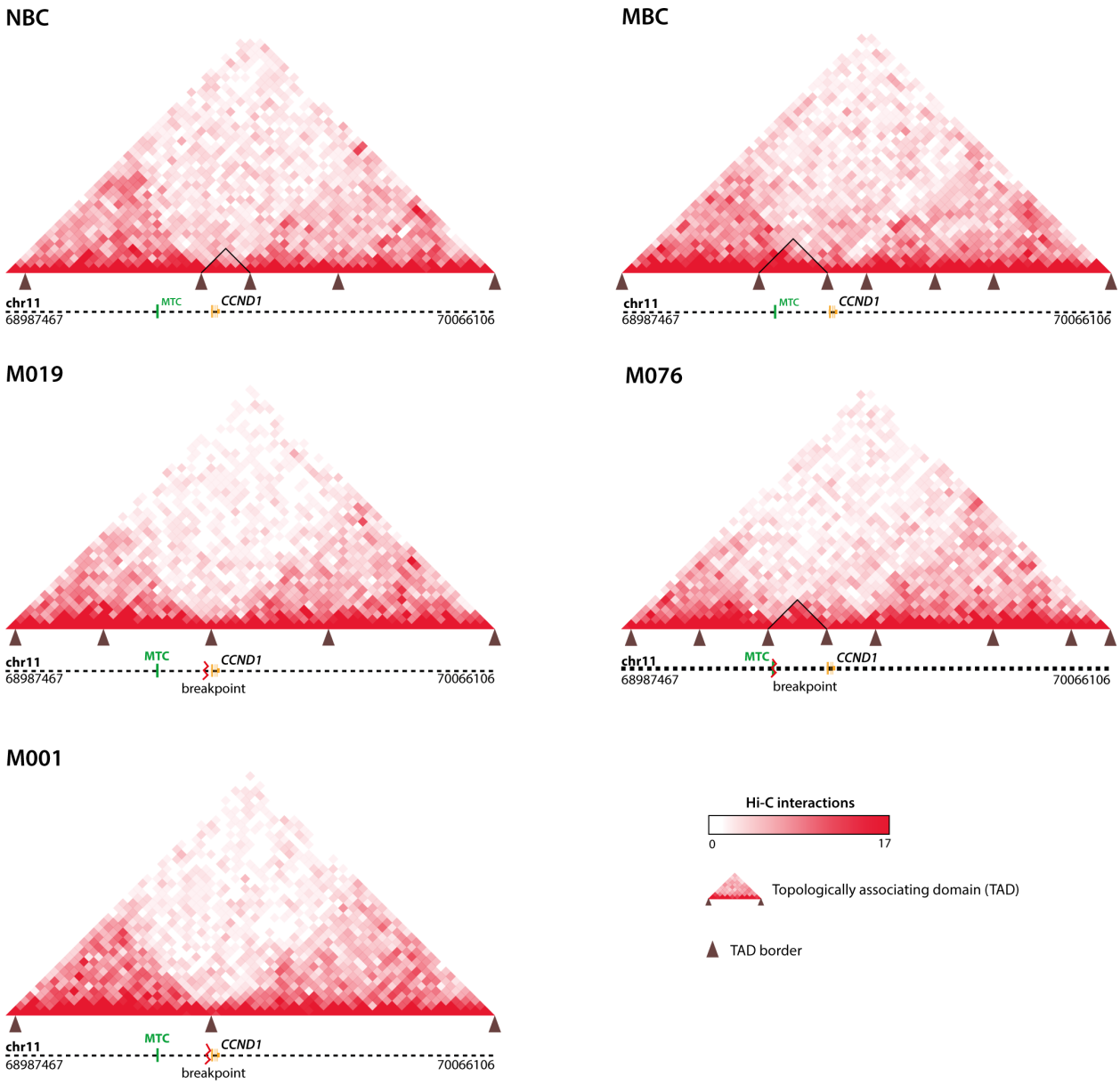


Figure S16. Structural variants affecting driver genes: *BIRC3*, *UBR5* and *PAX5*

(*Top*) Linear chromosomal representation depicting copy number states of total copy number (orange) and minor copy number (grey). The structural variants are represented by color-coded arcs. (*Bottom*) Zoomed image of the driver gene affected. **A.** Three overlapping deletions truncating *BIRC2/3* genes in case M361 (nnMCL). **B.** Focal deletion of *UBR5* found in M380 (cMCL). **C.** *PAX5* was disrupted in two cases by translocation in case M246 (cMCL) (upper plot) or by an inversion in M252 (cMCL) (bottom plot).

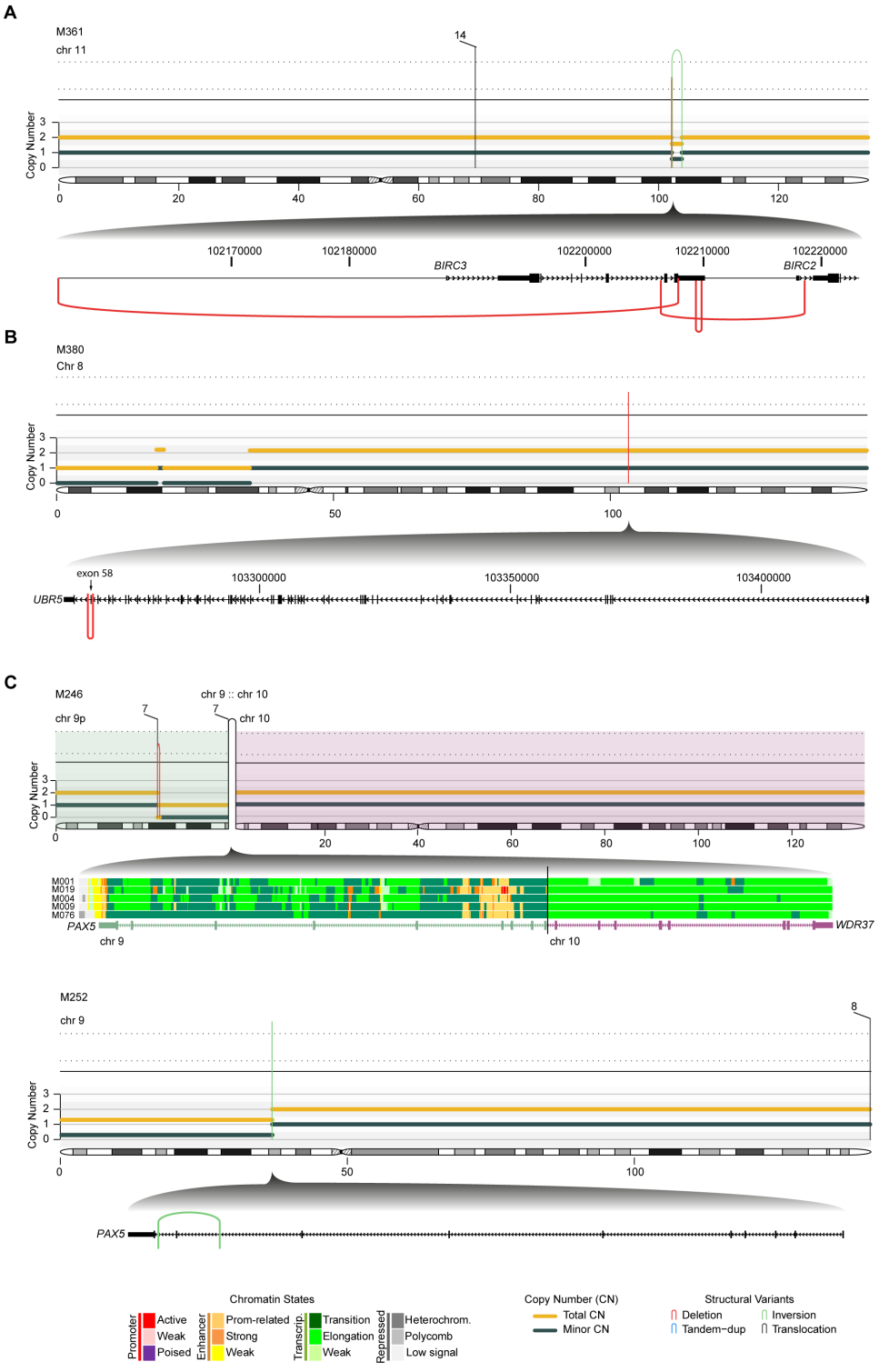


Figure S17. 2p inversion activating *MYCN* gene

A. (*Top*) Linear chromosomal representation of chromosome 2 depicting an inversion in the short arm (green arc) for case M406 (cMCL). (*Bottom*) Representation of the inverted 2p region resulting in the juxtaposition of *MYCN* near the IGK strong enhancer region (orange). **B.** Sanger sequencing verification of the inversion juxtaposing *MYCN* and IGK. **C.** *MYCN* gene expression levels by microarray. The case with the inversion (M406) shows a high *MYCN* expression level.

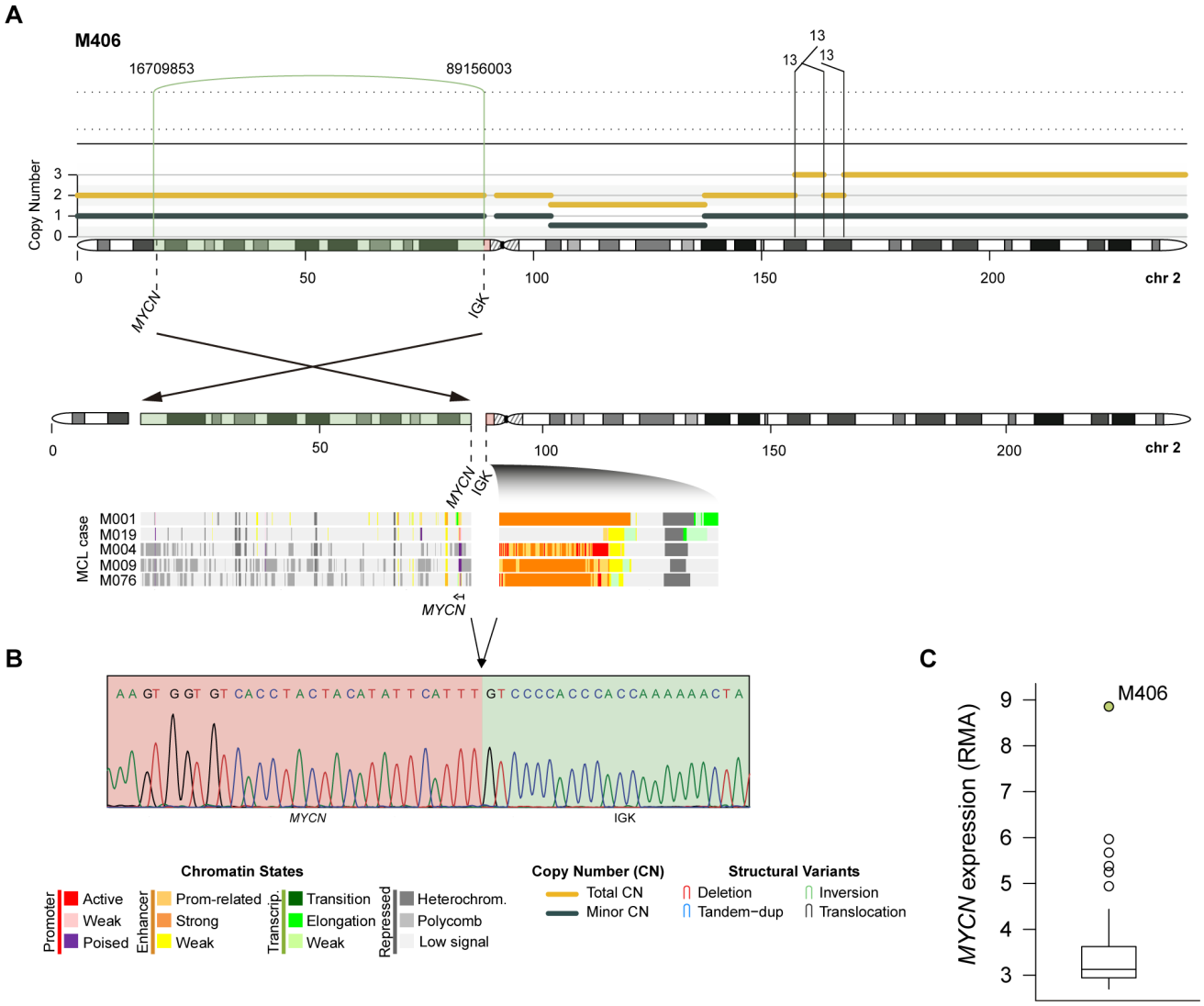
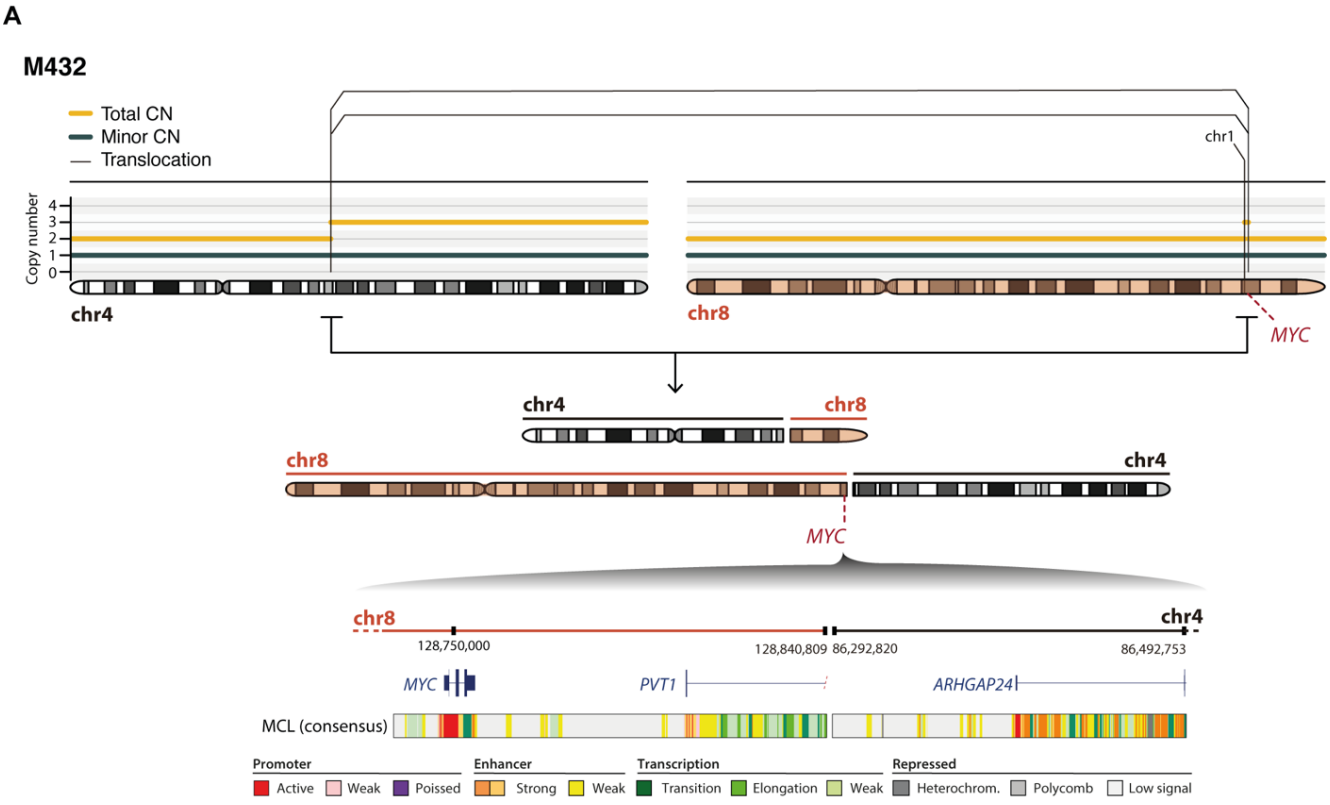


Figure S18. *MYC* non-IG rearrangement

A. WGS representation of chromosomes 4 and 8 of M432 (cMCL) with a focal copy number gain of *MYC* and a reciprocal translocation between both chromosomes that truncated *PVT1* and places *MYC* near enhancer regions of chr4. **B.** Metaphase FISH verification of *MYC* gain (red) and rearrangement with chr4 using a *MYC* break-apart (BA) probe.



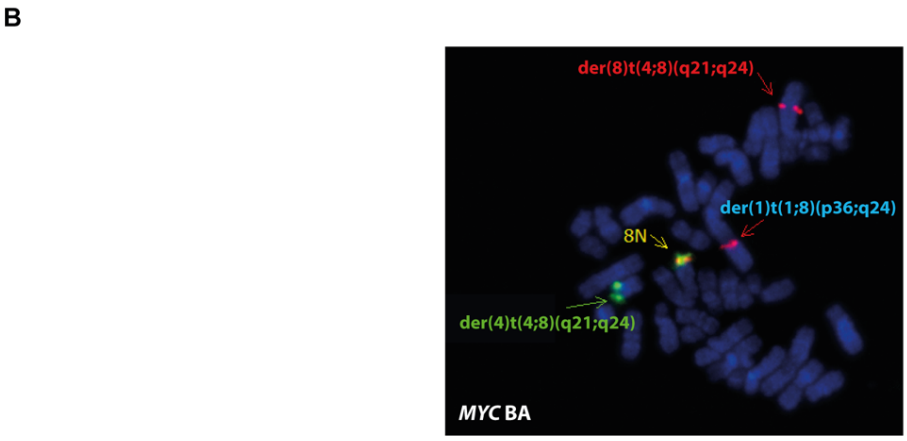
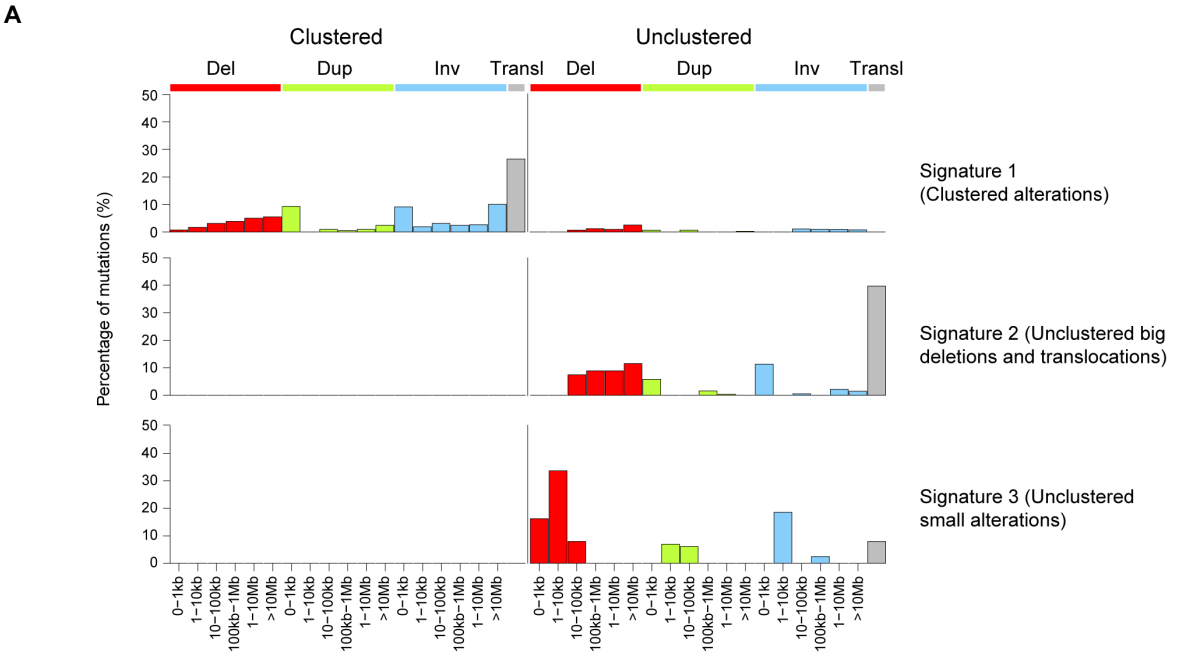
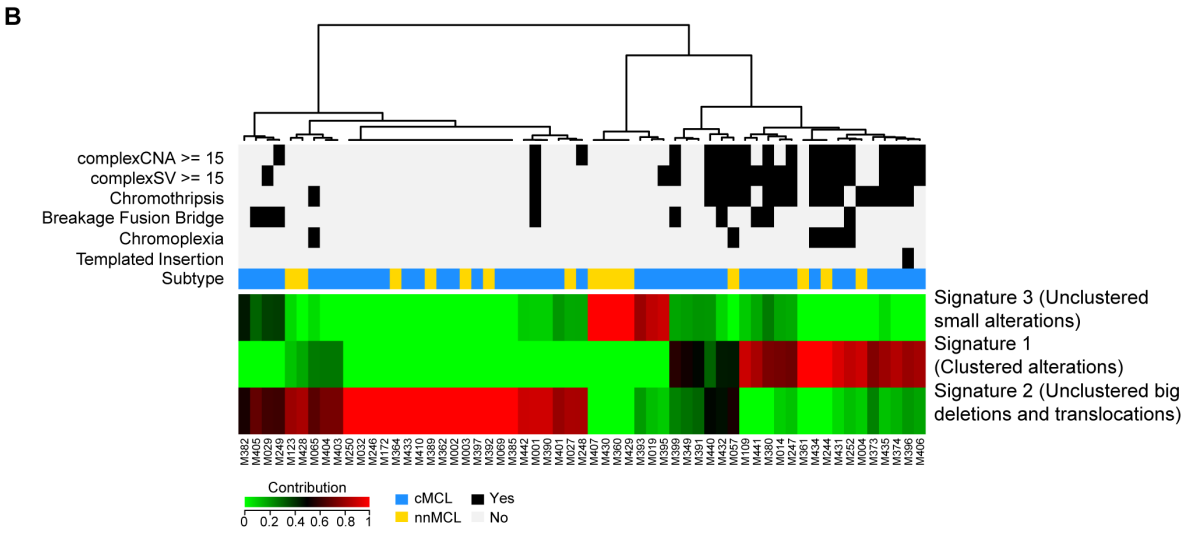


Figure S19. Signatures of structural variants

A. Three de novo signatures of structural variants (SV) present in MCL. **B.** Contribution of each signature in each case. **C.** Integration of SV breakpoints and chromatin states according to SV type.





C

		Translocations		Deletions		Duplications		Inversions	
		MCL	B cells	MCL	B cells	MCL	B cells	MCL	B cells
Promoter	Active	2.0	2.2	1.4	1.7	0.2	0.5	1.5	1.1
	Weak	1.0	0.2	0.3	0.7	-0.5	-0.2	-0.2	0.5
	Poised	0.0	-1.0	-1.0	0.4	0.6	0.7	0.2	0.4
Enhancer	Prom-related	0.7	0.7	1.6	1.4	-0.1	-0.4	0.1	0.0
	Strong	0.9	-0.2	1.0	1.4	-1.3	-2.4	-0.7	-0.2
	Weak	0.2	0.7	0.4	0.5	-1.1	-0.9	-0.5	-0.2
Transcrip.	Transition	1.5	1.7	1.4	2.1	0.9	1.2	1.0	0.8
	Elongation	0.5	0.6	1.0	1.2	1.1	1.2	0.0	0.2
	Weak	-0.1	-0.1	0.1	0.2	-0.3	-0.1	-0.1	-0.2
Repressed	Heterochrom.	0.0	-0.2	-0.1	-0.4	0.0	-0.4	-0.1	0.0
	Polycomb	0.1	-0.1	0.0	0.0	-0.2	0.1	-0.3	-0.5
	Low signal	-0.2	-0.1	-0.3	-0.3	-0.1	-0.1	0.0	0.0

Log2FC: <-2 … >2

Figure S20. *CDKN2A/B* deletions and structural variants

(*Top*) Accumulated copy number plot of chromosome 9 depicting frequent deletions (red) and focal homozygous deletion (bordeaux) in the minimal region containing *CDKN2A/B* genes. Patients' labels are depicted in blue for cMCL and yellow for nnMCL. (*Bottom*) Zoomed image of two different cases with deletion concomitant with structural variants. (*Left*) Case M246 with a *CDKN2A/B* homozygous deletion resulting from a narrow deletion of one allele and a larger deletion of the other allele flanked by rearrangement at both breakpoints with chromosomes 7 and 10. (*Right*) Case M374 with chromothripsis at 9p resulting in two homozygous deletions and four different rearrangements with 11p arm with the both flanking *CDKN2A/B* breaks.

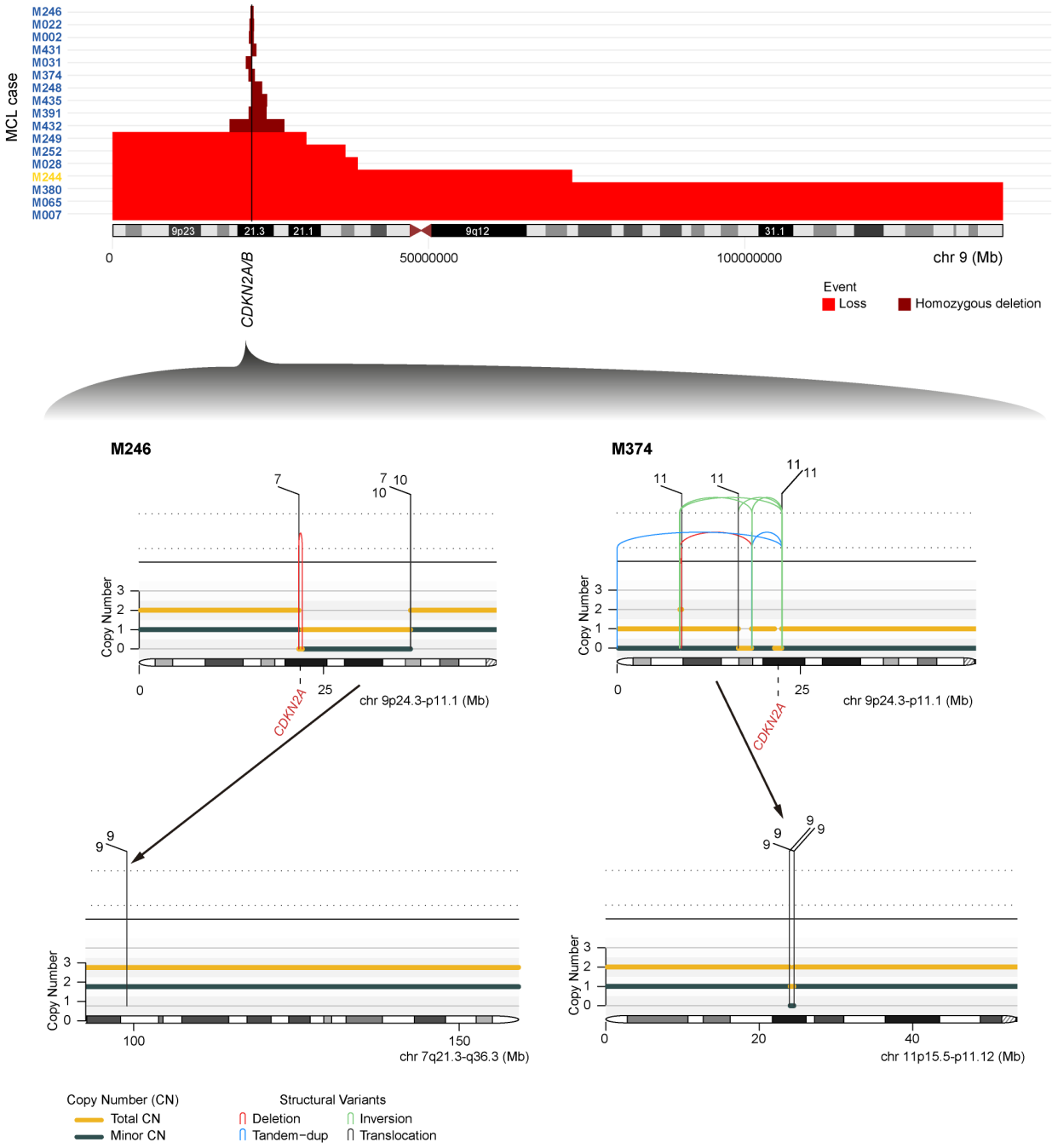


Figure S21. Copy number alterations affecting *MYC*, *BMI1*, *MIR17HG* and *BCL2*

Accumulated copy number plot of chromosomes 10, 13, 18 and 8 frequent gains and amplifications, with the minimal regions containing *BMI1*, *MIR17HG*, *BCL2 and MYC* genes, respectively. Patients' labels are depicted in blue for cMCL and yellow for nnMCL.

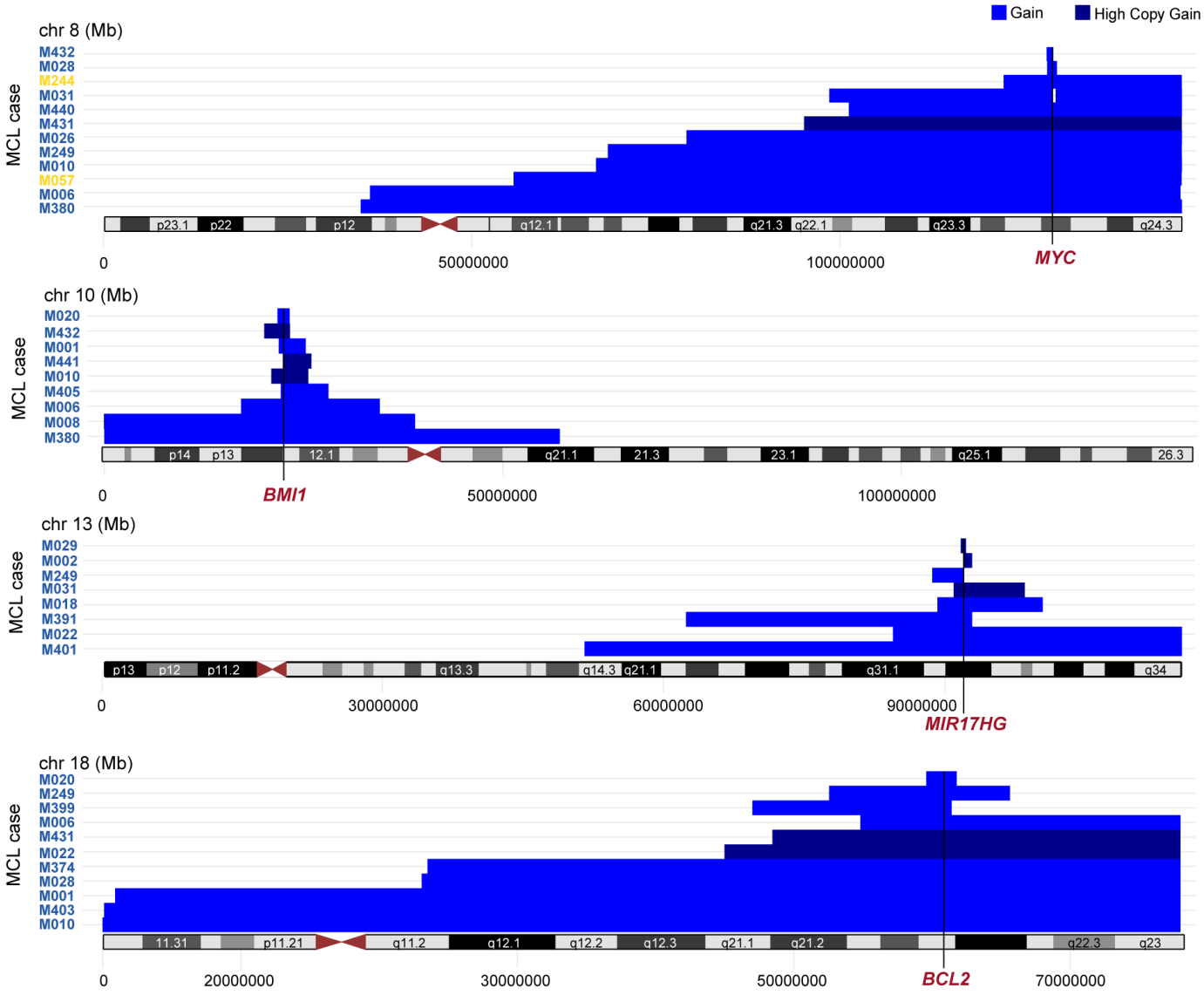
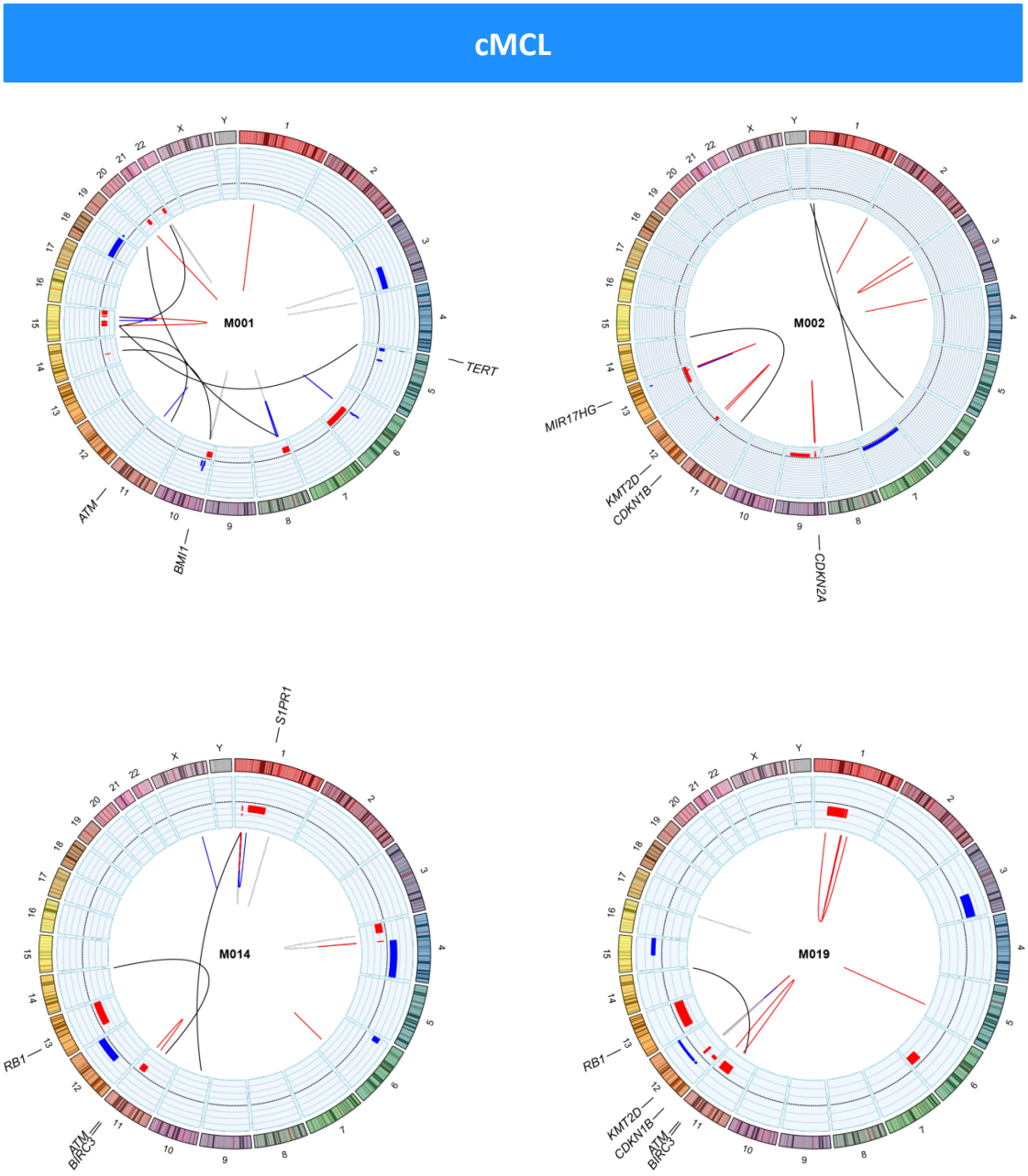
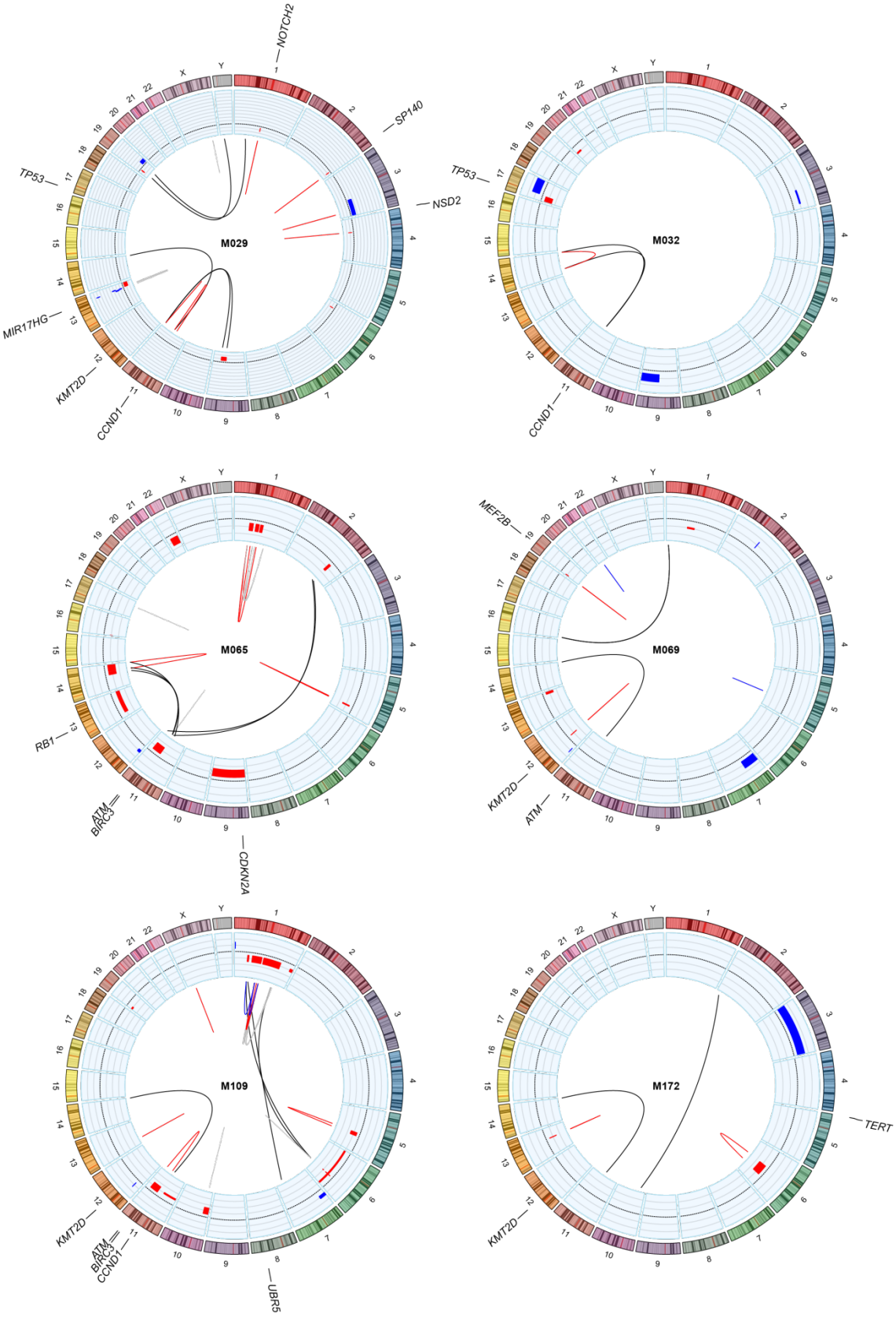
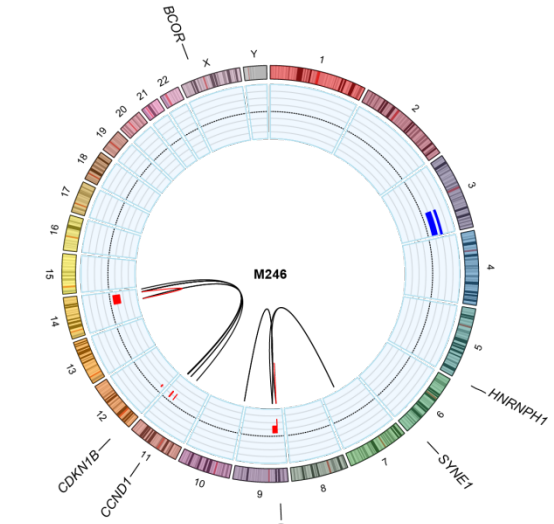


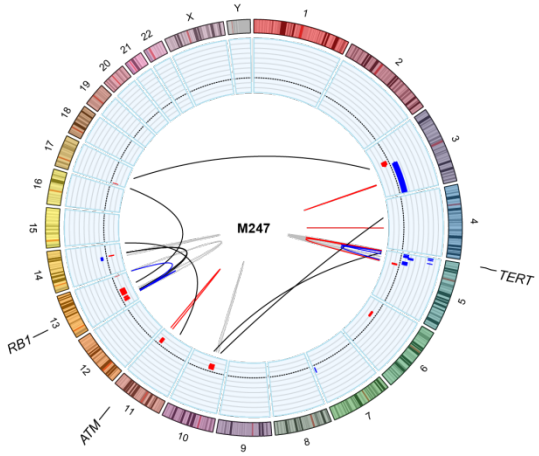
Figure S22. Circular representation of SV and CNA of the 61 MCL

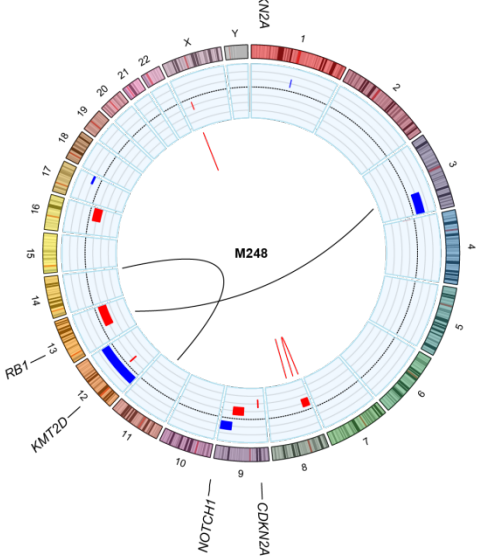
Circular representation of each case (44 cMCL and 17 nnMCL) depicting from the outer to the inner side, driver genes affected by any alteration (mutations, CNA or SV), chromosomes, copy number plot (gains in blue and deletions in red) and chromosomal rearrangements in arcs colored in black, red, blue and dark green, for translocations, deletions, duplications and inversions, respectively.

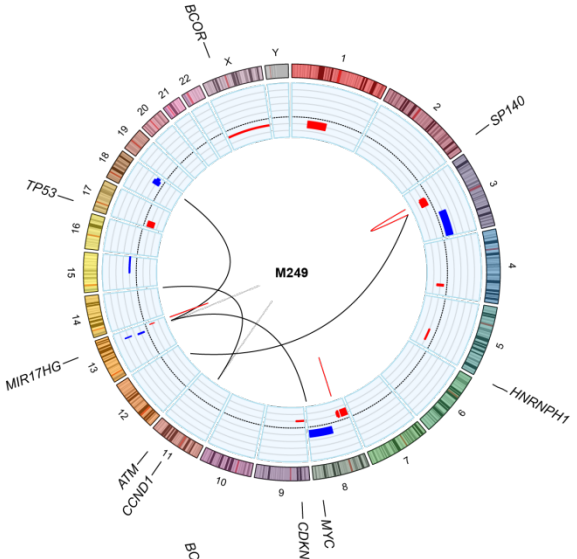


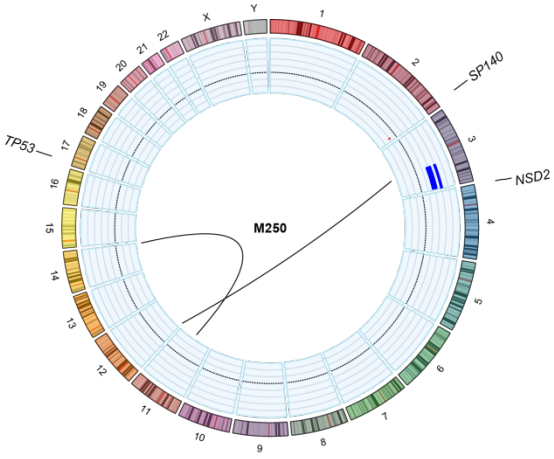

M029
NOTCH2
SP140
NSD2
TP53
MIR17HG
KMT2D
CCND1
M032
TP53
CCND1
M065
RB1
ATM
BIRC3
CDKN2A
M069
MEF2B
KMT2D
ATM
M109
KMT2D
ATM
BIRC3
CCND1
UBR5
M172
TERT
KMT2D

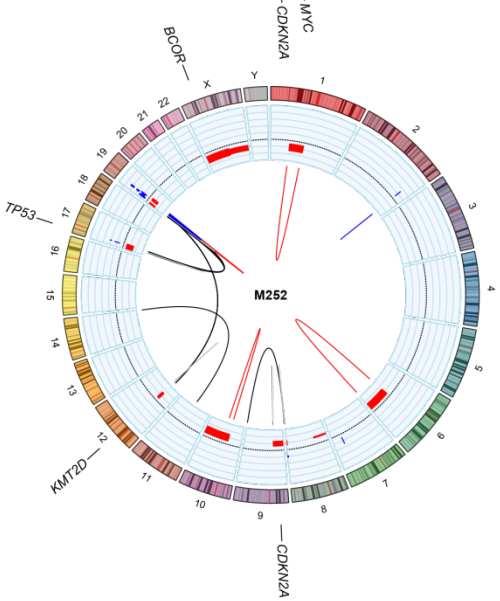

BCOR
M246
HNRNPH1
SYNE1
CDKN2A
CCND1
CDKN1B

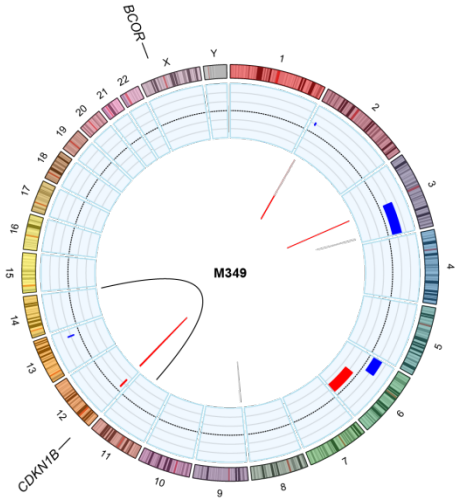

M247
TERT
RB1
ATM

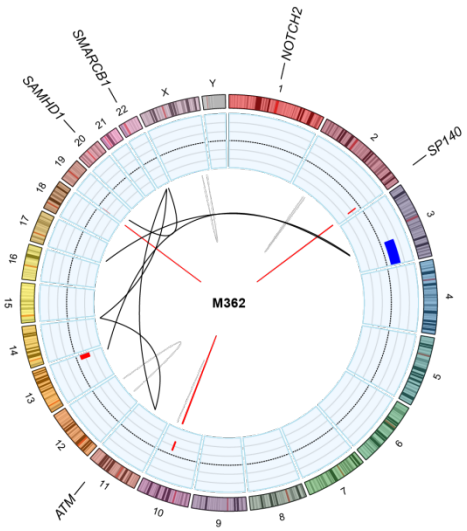

M248
RB1
KMT2D
NOTCH1
CDKN2A

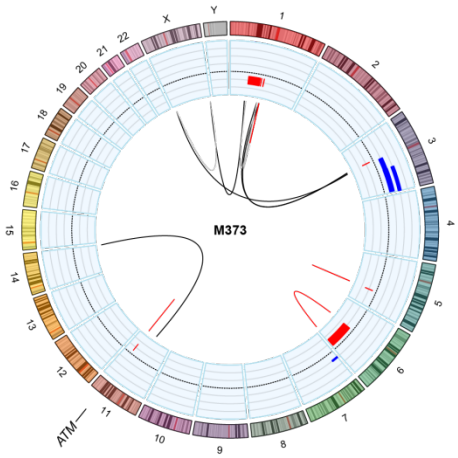

BCOR
SP140
TP53
M249
HNRNPH1
MIR17HG
ATM
CCND1
CDKN2A
MYC

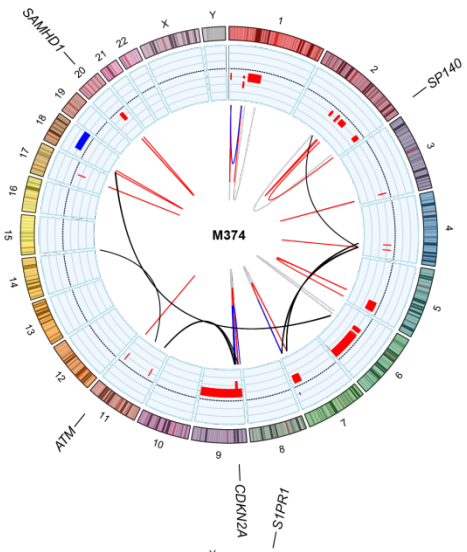

SP140
TP53
NSD2
M250

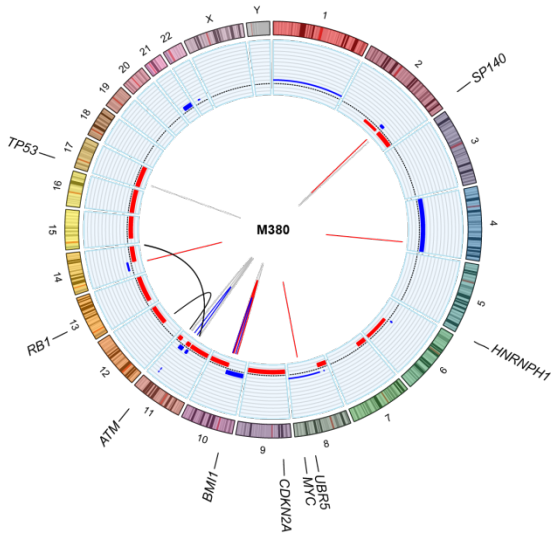

BCOR
TP53
M252
KMT2D
CDKN2A

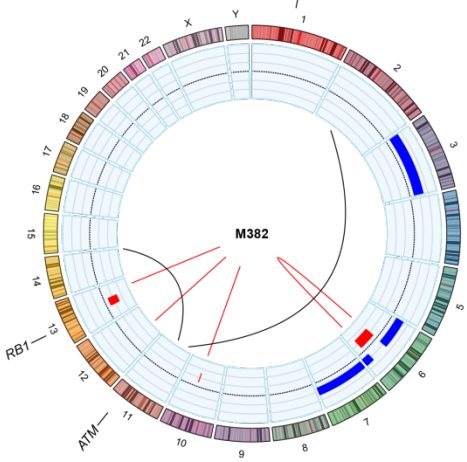

BCOR
CDKN1B
M349

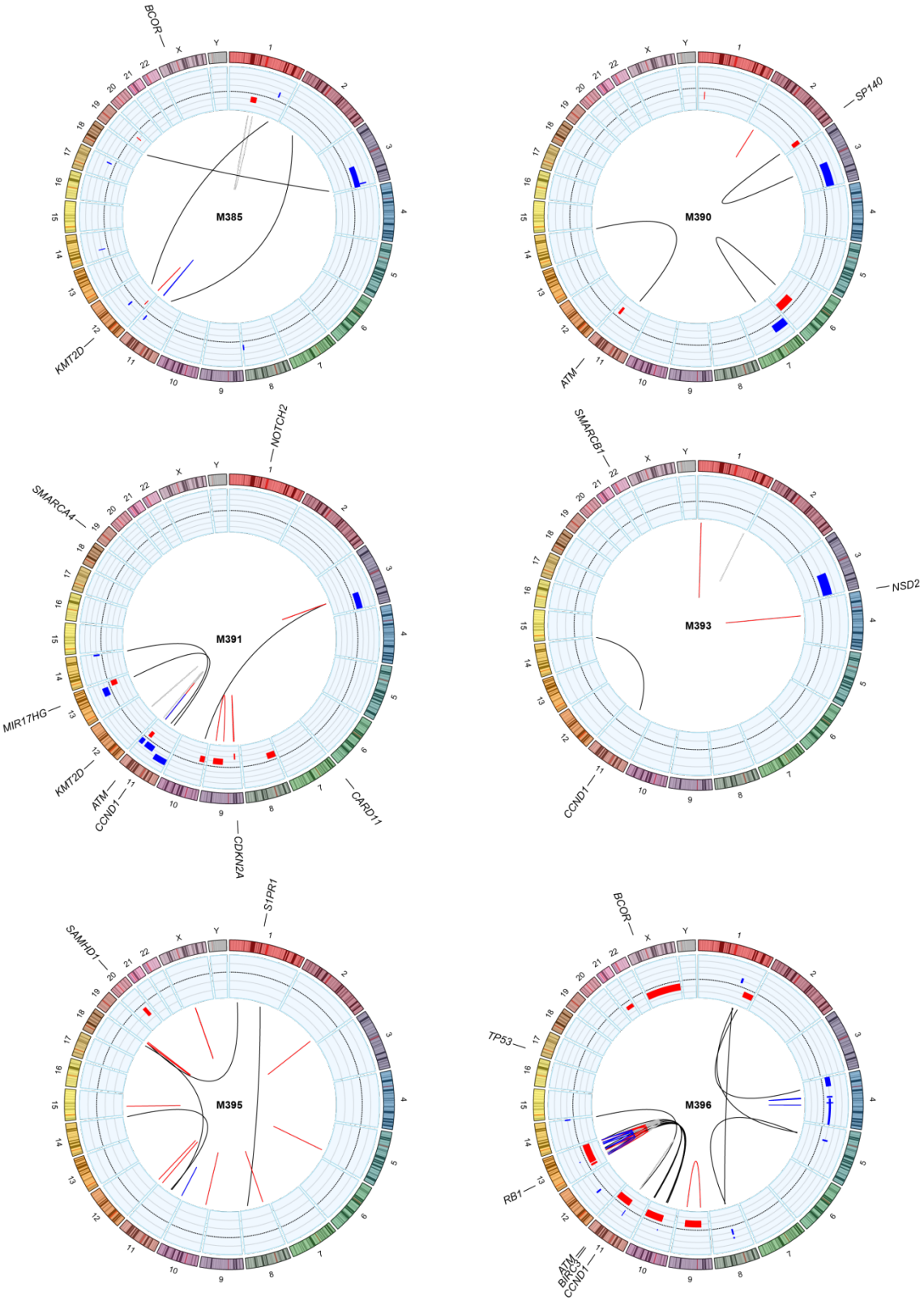

NOTCH2
SMARCB1
SAMHD1
SP140
ATM
M362

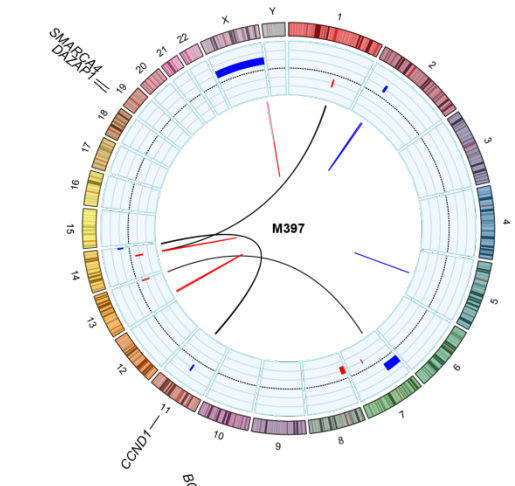

ATM
M373

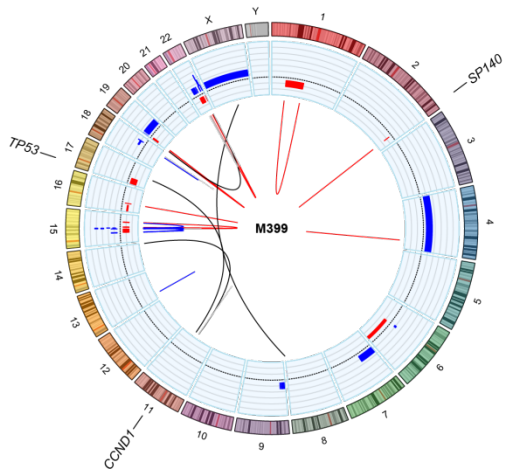

SAMHD1
SP140
ATM
CDKN2A
S1PR1
M374

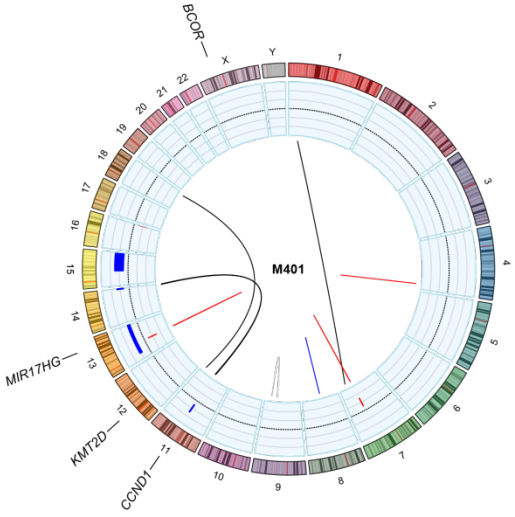

SP140
TP53
RB1
HNRNPH1
ATM
BMI1
CDKN2A
UBR5
MYC
M380

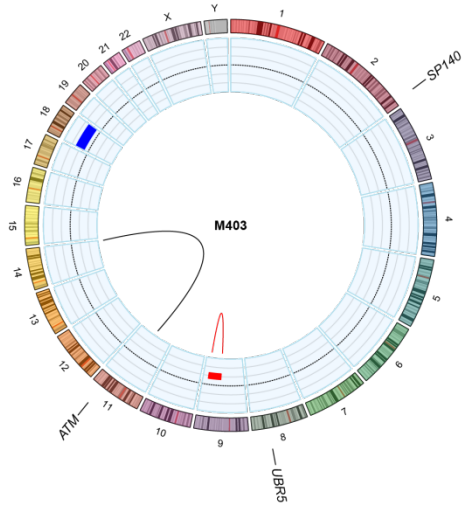

RB1
ATM
M382

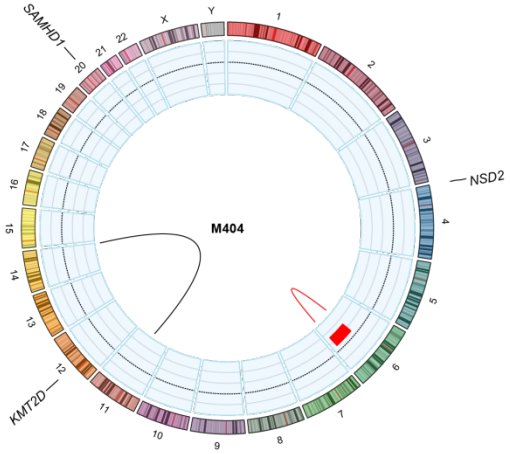

BCOR
KMT2D
M385
SP140
ATM
M390
NOTCH2
SMARCA4
MIR17HG
KMT2D
ATM
CCND1
CDKN2A
CARD11
M391
SMARCB1
NSD2
CCND1
M393
S1PR1
SAMHD1
M395
BCOR
TP53
RB1
ATM
BIRC3
CCND1
M396

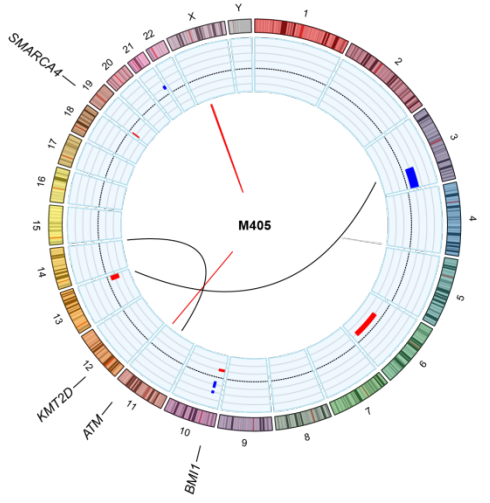

SMARCA4
DAXX
M397
CCND1

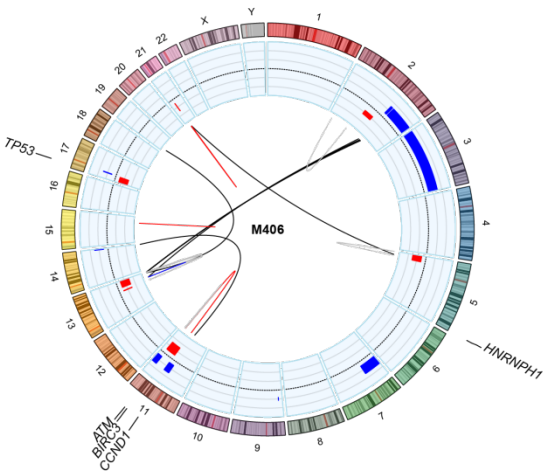

TP53
SP140
M399
CCND1

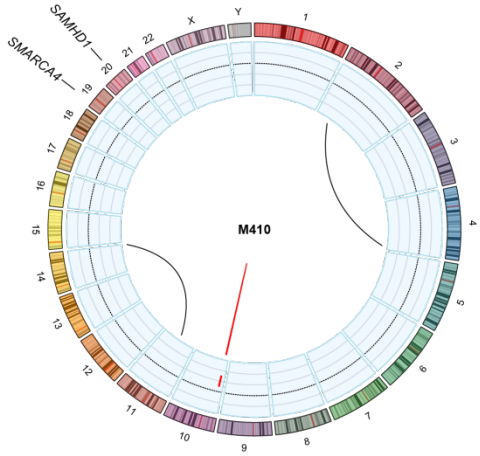

BCOR
M401
MIR17HG
KMT2D
CCND1

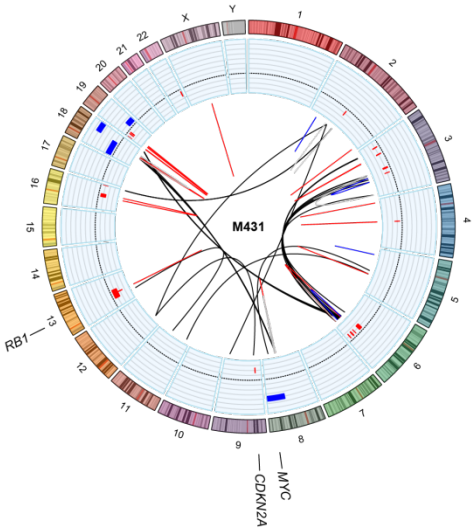

SP140
M403
ATM
UBR5

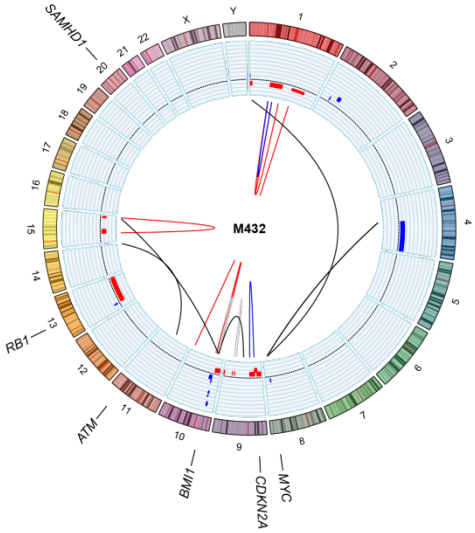

SAMHD1
NSD2
M404
KMT2D

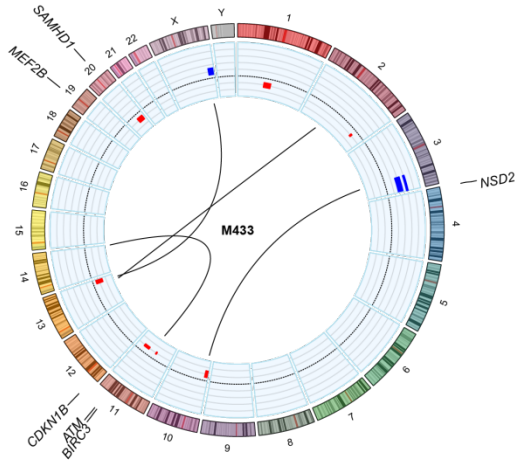

SMARCA4
M405
KMT2D
ATM
BMI1

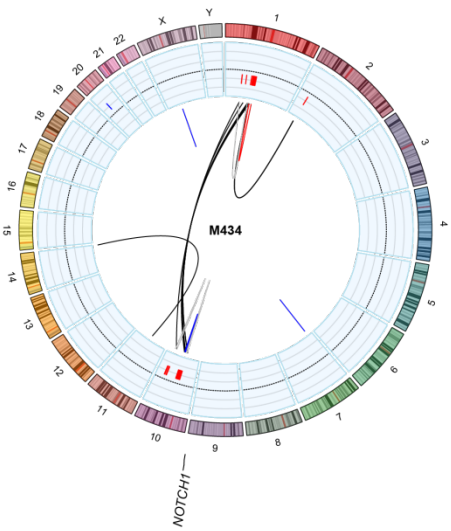

M406
TP53
HNRNPH1
ATM
BIRC3
CCND1

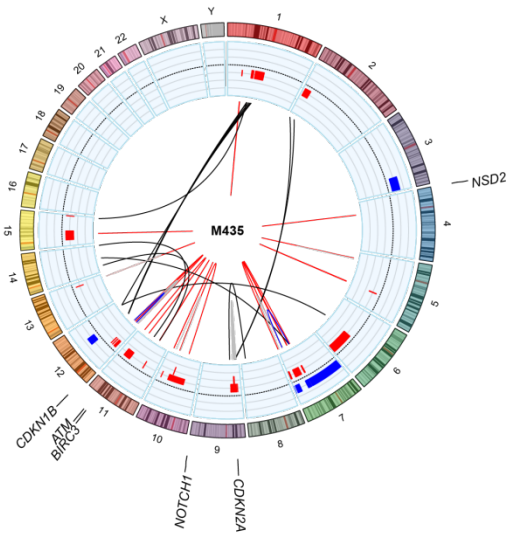

M410
SAMHD1
SMARCA4

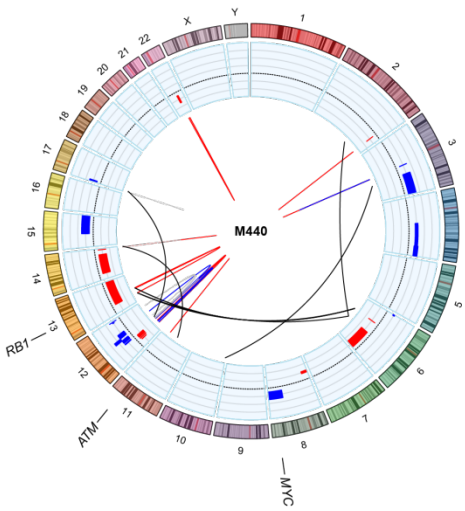

M431
RB1
CDKN2A
MYC

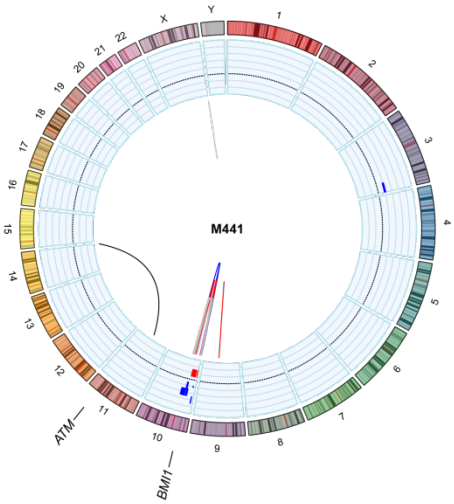

M432
SAMHD1
RB1
ATM
BMI1
CDKN2A
MYC

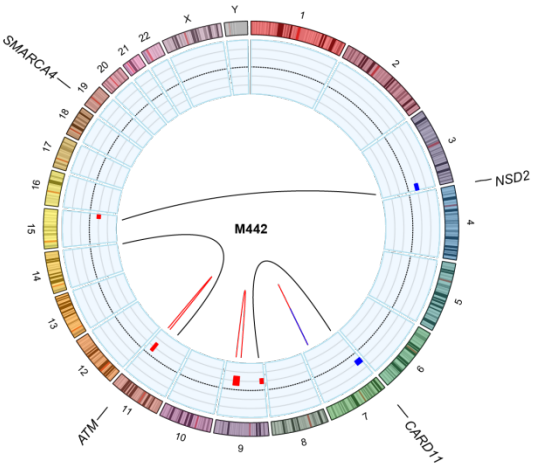

M433
SAMHD1
MEF2B
NSD2
CDKN1B
ATM
BIRC3


M434
NOTCH1

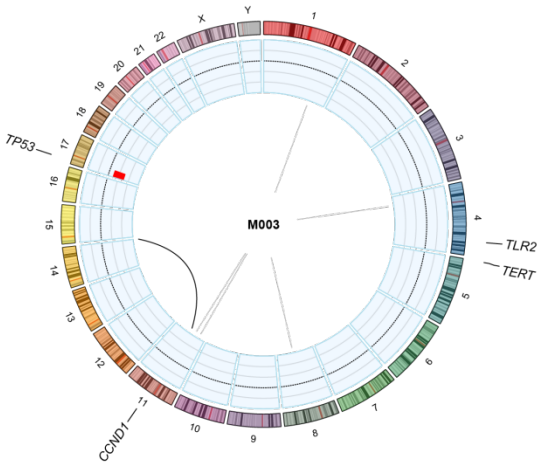

M435
NSD2
CDKN1B
ATM
BIRC3
NOTCH1
CDKN2A

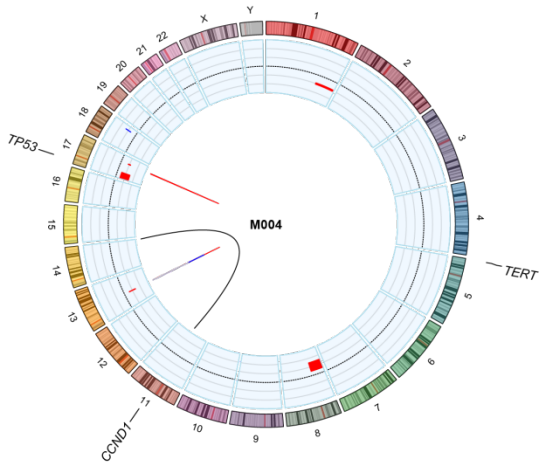

M440
RB1
ATM
MYC

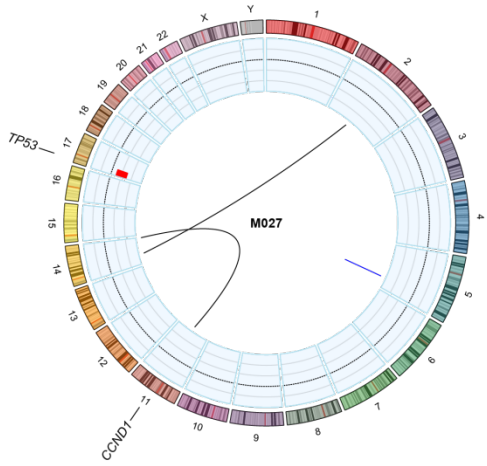

M441
ATM
BMI1

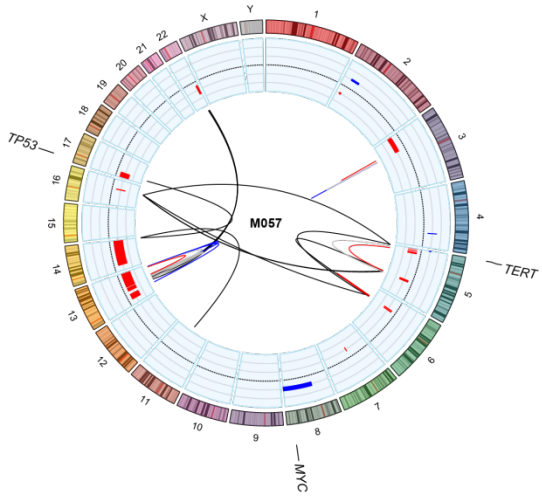

M442
SMARCA4
NSD2
ATM
CARD11

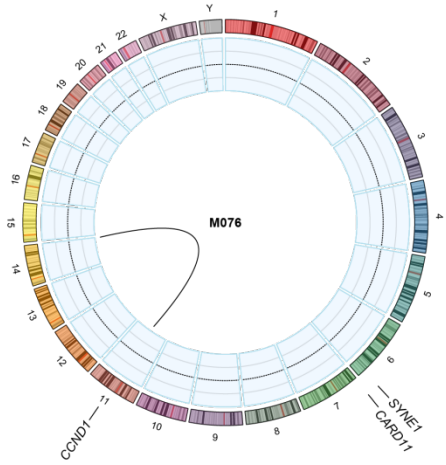
nnMCL

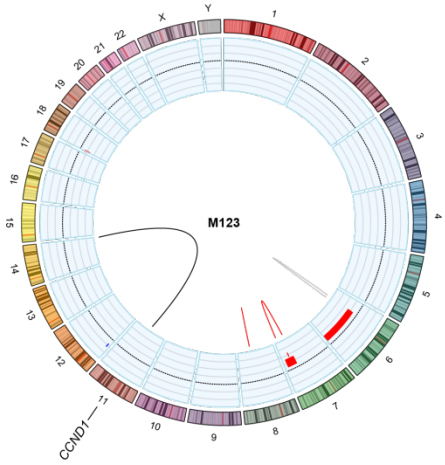

M003
TP53
TLR2
TERT
CCND1

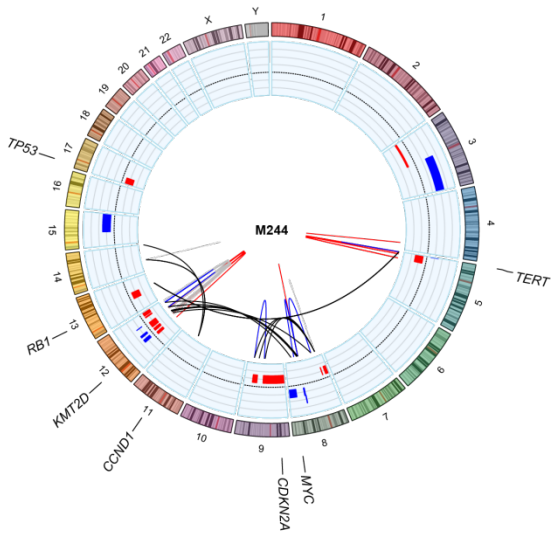

M004
TP53
TERT
CCND1

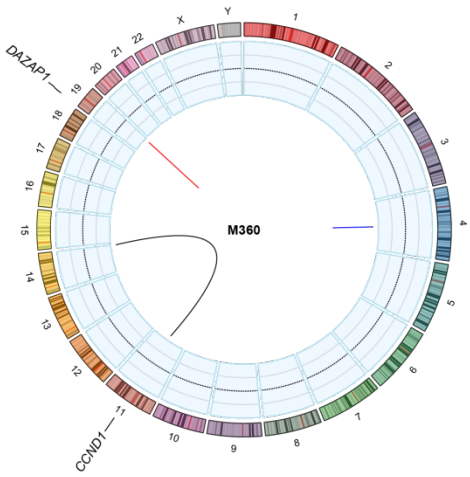

M027
TP53
CCND1

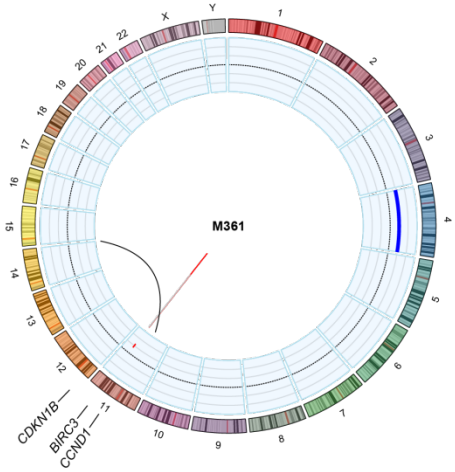

M057
TP53
TERT
MYC

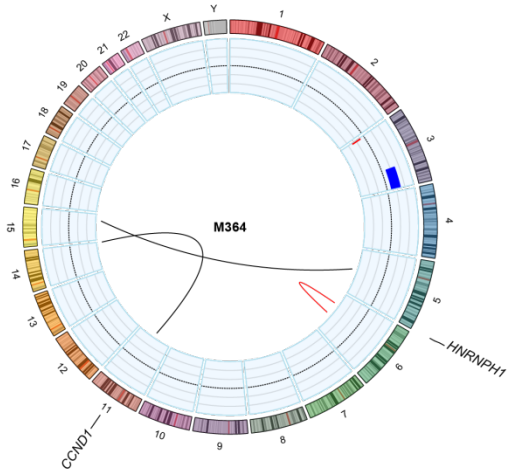

M076
CCND1
SYNE1
CARD11

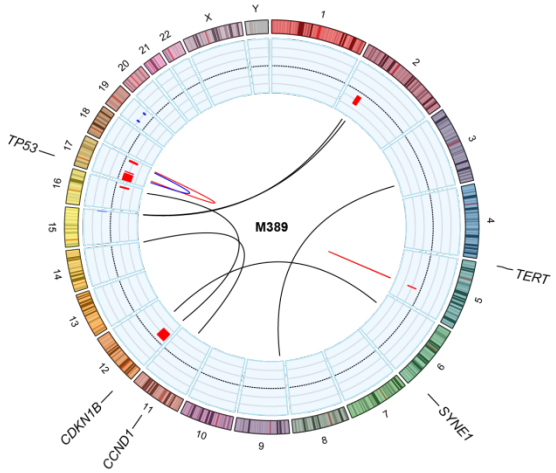

M123
CCND1

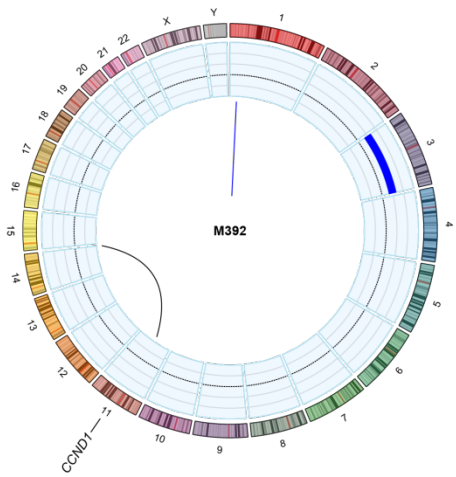

M244
TP53
TERT
RB1
KMT2D
CCND1
CDKN2A
MYC

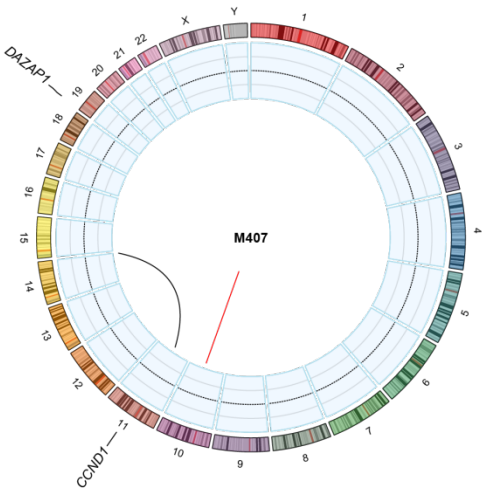

M360
DAZAP1
CCND1

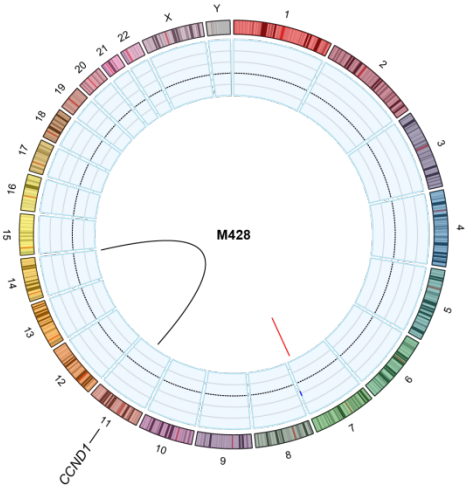

M361
CDKN1B
BIRC3
CCND1

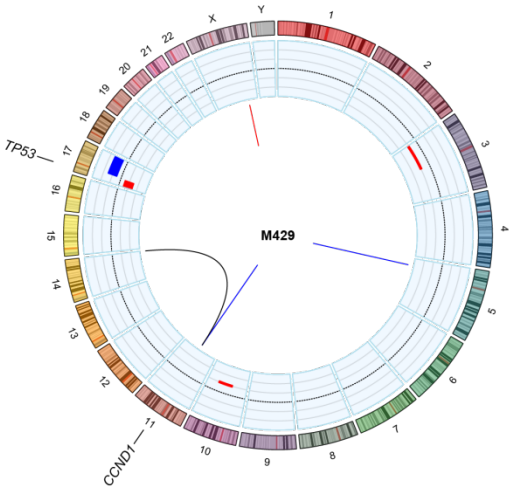

M364
HNRNPH1
CCND1

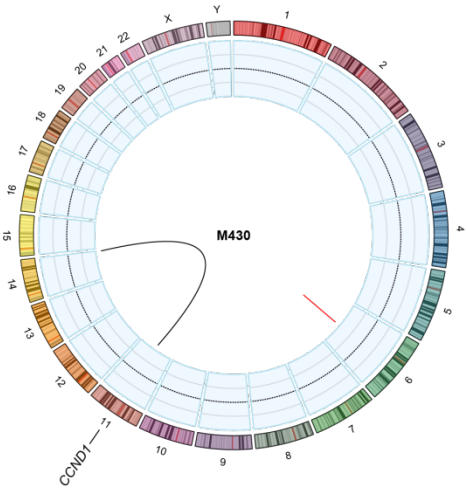

M389
TP53
TERT
CDKN1B
CCND1
SYNE1


M392
CCND1


M407
DAZAP1
CCND1


M428
CCND1


M429
TP53
CCND1


M430
CCND1

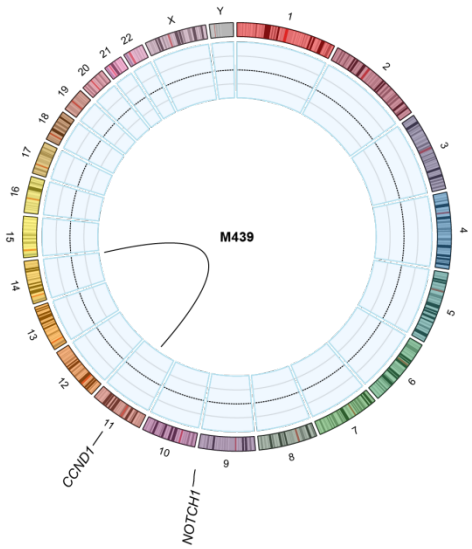

M439
CCND1
NOTCH1

Figure S23. *TERT* alterations

A. Schematic representation of non-coding *TERT* mutations and Sanger sequencing verification of 3 cases. **B.** Accumulated copy number plot of the short arm of chr5 depicting cases with gains and focal amplifications affecting *TERT*. Patients' labels are depicted in blue for cMCL and yellow for nnMCL. **C.** (*Top*) Linear representation of partial chr5 and chr16 involved in a rearrangement affecting *TERT* in case M057 (nnMCL). (*Bottom*) Zoomed image of both regions including the chromatin states. **D.** Representative metaphase of case M172 (cMCL) hybridized with WCP probes for chr5 and chr11 showing a derivative chr5 with addition of material at the *TERT* locus, the additional material corresponded to chr11. **E.** *TERT* expression levels by qPCR in cases with different types of *TERT* alterations, cMCL and nnMCL cases were colored in blue and yellow, respectively. **F.** (*Left*) Design of *TERT* probe used for FISH. (*Right*) Case M006 (cMCL, only WES data) with a gain by SNP-array shows multiple copies of the central and telomeric part (green and aqua), confirming the gain detected by SNP-array, and showing an additional rearrangement with loss of the centromeric part (red). **G.** Oncoprint plot of *TERT* alterations according to type of *TERT* alteration, MCL molecular subtype and presence of *TP53* alterations. *SV could not be analyzed in cases with only WES data (white).

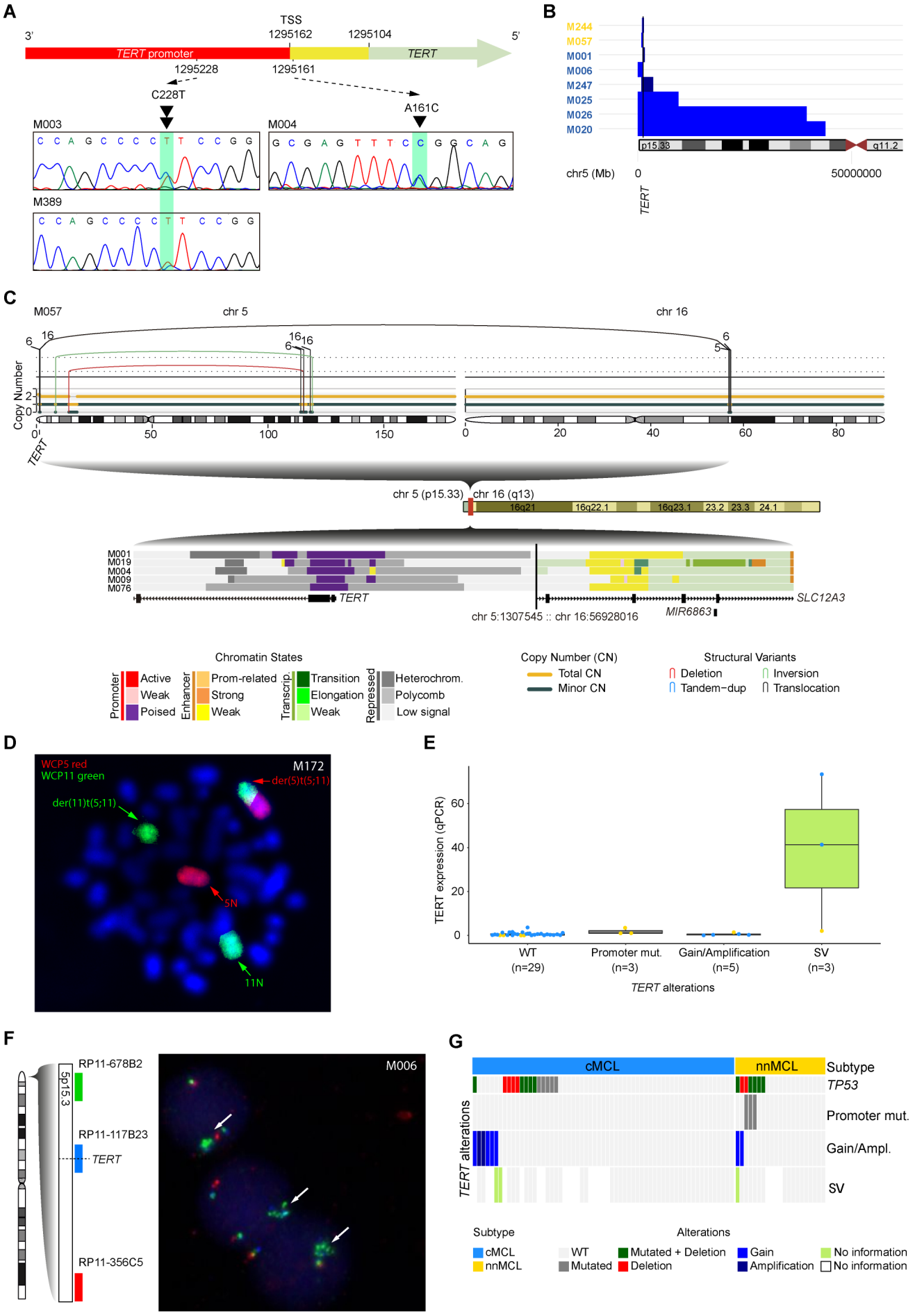

A
TSS
1295162
1295104
3'
5'
TERT promoter
TERT
1295228
1295161
C228T
A161C
M003
M004
M389
B
M244
M057
M001
M006
M247
M025
M026
M020
p15.33
q11.2
chr5 (Mb)
0
50000000
TERT
C
M057
chr 5
chr 16
Copy Number
TERT
chr 5 (p15.33)
chr 16 (q13)
16q21
16q22.1
16q23.1
23.2
23.3
24.1
M001
M019
M004
M009
M076
TERT
SLC12A3
MIR6863
chr 5:1307545 :: chr 16:56928016
Chromatin States
Promoter
Active
Weak
Poised
Enhancer
Prom-related
Strong
Weak
Transcrip.
Transition
Elongation
Weak
Repressed
Heterochrom.
Polycomb
Low signal
Copy Number (CN)
Total CN
Minor CN
Structural Variants
Deletion
Tandem-dup
Inversion
Translocation
D
WCP5 red
WCP11 green
M172
der(5)t(5;11)
der(11)t(5;11)
5N
11N
E
TERT expression (qPCR)
WT
(n=29)
Promoter mut.
(n=3)
Gain/Amplification
(n=5)
SV
(n=3)
TERT alterations
F
RP11-678B2
5p15.3
RP11-117B23
TERT
RP11-356C5
M006
G
cMCL
nnMCL
Subtype
TP53
Promoter mut.
Gain/Ampl.
SV
TERT alterations
Subtype
cMCL
nnMCL
Alterations
WT
Mutated
Mutated + Deletion
Deletion
Gain
Amplification
No information
No information

Figure S24. Telomere length of MCL tumors according to *ATM* status

Distribution of the telomere length of MCL tumors according to *ATM* status. Tbases, number of telomeric bases in the tumor sample; Cbases, number of telomeric bases in the normal sample.

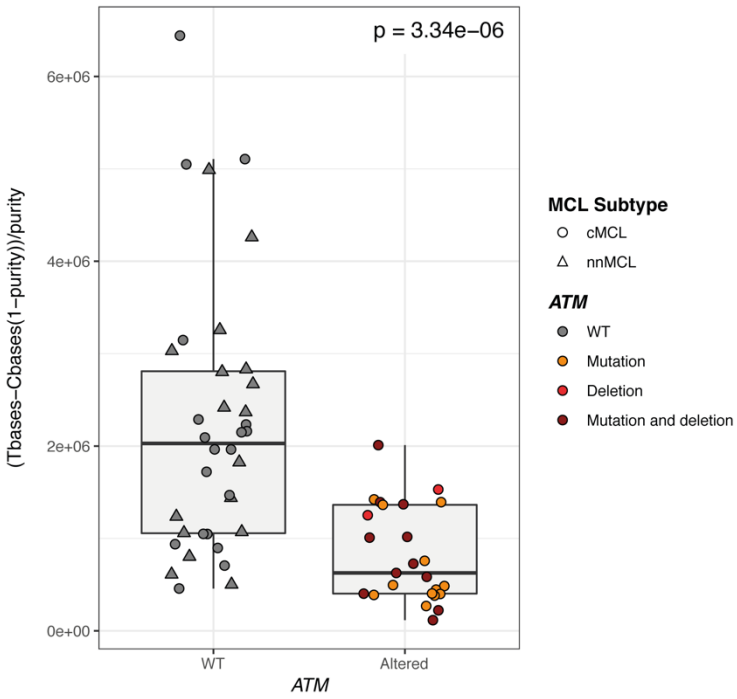


Figure S25. Relationship between recurrent somatic alterations in MCL

Pairwise associations between recurrent somatic alterations of driver MCL genes/regions found in at least 5 MCL tumors. Breakage-fusion bridge (BFB) cycles were also included in the comparisons. Mutually exclusive relationships are shown in red and co-ocurrence in blue, with a color shade reflecting the odds ratio. Q-value range is indicated (***Q<0.001; **Q<0.01; *Q<0.1).

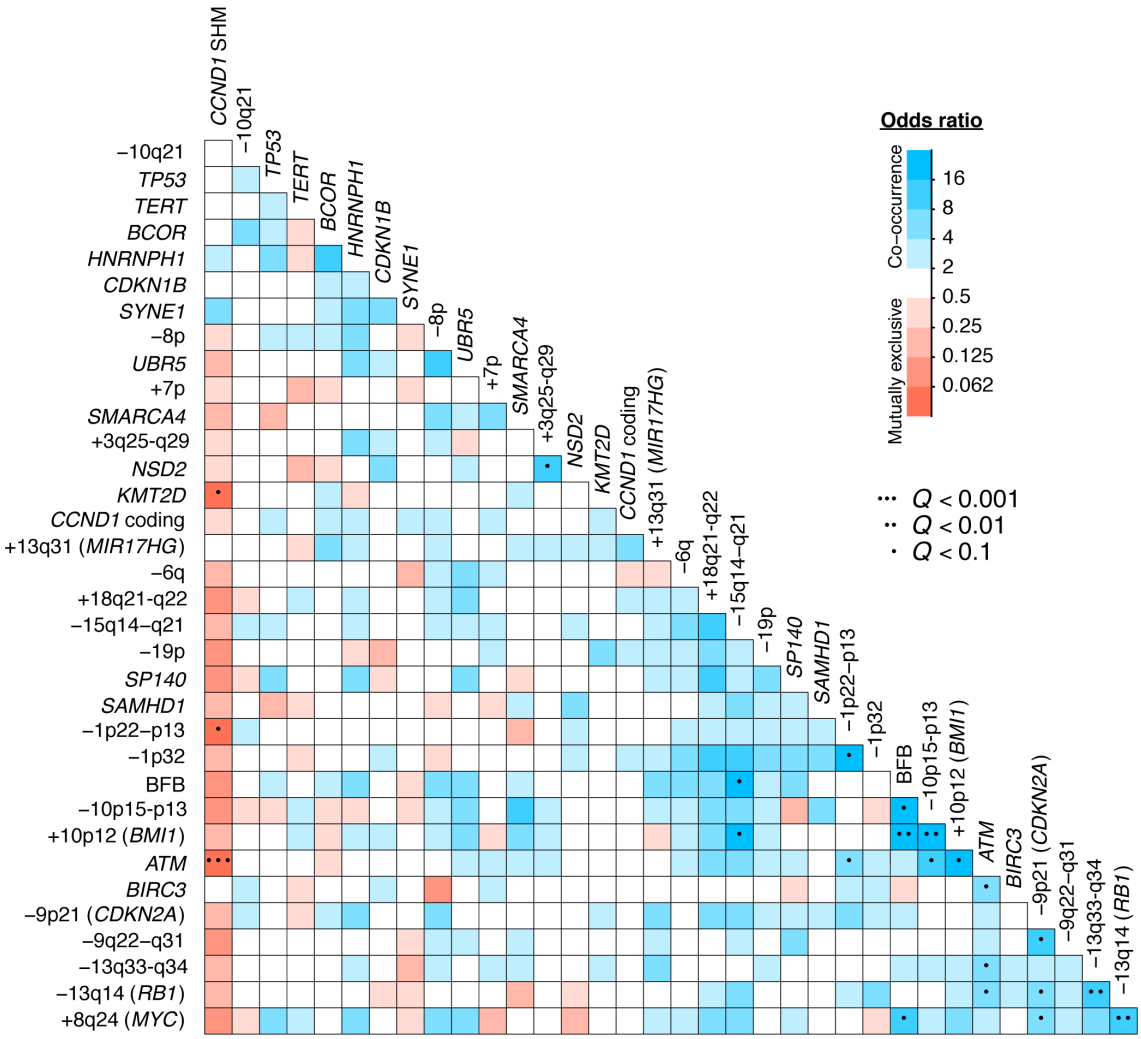


Figure S26. Sources of DNA methylation variability in MCL

A. Principal component analysis (PCA) of DNA methylation data for 70 MCL (first and second components are represented). Shape represents MCL subtypes, and colors epigenetic groups. **B.** Association of 10th first components of the PCA with clinic-biological variables as well as driver genetic alterations ordered by recurrence. **C.** epiCMIT is associated with high Ki67 in cMCL and with SBS1. **D.** epiCMIT correlates with number of CNA, particularly within cMCL.

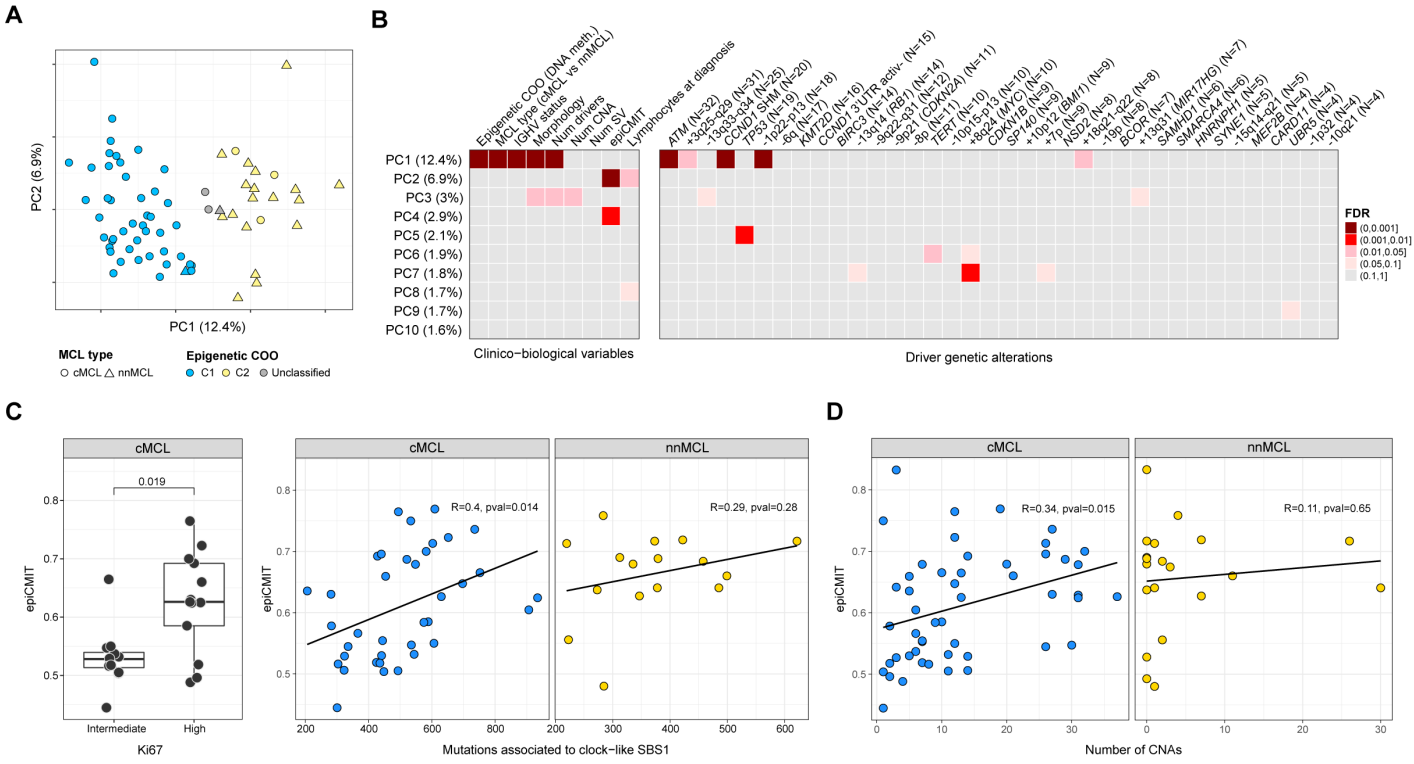
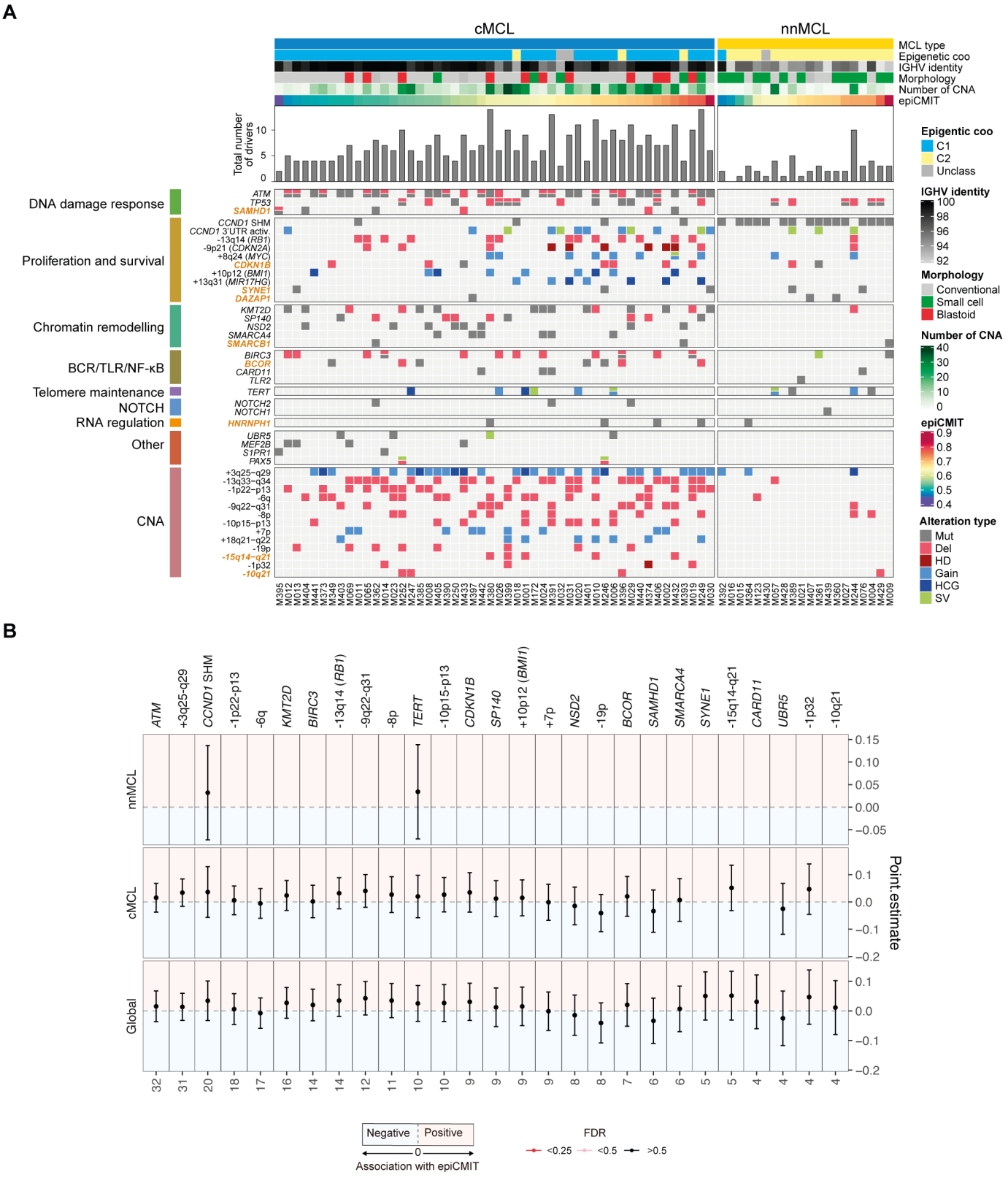


Figure S27. Relationship between driver genetic alterations and epiCMIT

A. Driver genetic alterations in cMCL and nnMCL ordered by epiCMIT. **B.** 95% confidence intervals for the mean epiCMIT difference between the presence and absence of the alterations without a clear association with epiCMIT.



SUPPLEMENTAL REFERENCES

1. Clot G, Jares P, Giné E, et al. A gene signature that distinguishes conventional and leukemic nonnodal mantle cell lymphoma helps predict outcome. *Blood*. 2018;132(4):413–422.
2. Royo C, Navarro A, Clot G, et al. Non-nodal type of mantle cell lymphoma is a specific biological and clinical subgroup of the disease. *Leukemia*. 2012;26(8):1895–1898.
3. Soldini D, Valera A, Solé C, et al. Assessment of SOX11 expression in routine lymphoma tissue sections: characterization of new monoclonal antibodies for diagnosis of mantle cell lymphoma. *Am. J. Surg. Pathol.* 2014;38(1):86–93.
4. Croci GA, Hoster E, Beà S, et al. Reproducibility of histologic prognostic parameters for mantle cell lymphoma: cytology, Ki67, p53 and SOX11. *Virchows Arch.* 2020;
5. International Cancer Genome Consortium, Hudson TJ, Anderson W, et al. International network of cancer genome projects. *Nature*. 2010;464(7291):993–998.
6. Puente XS, Beà S, Valdés-Mas R, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526(7574):519–524.
7. Bea S, Valdes-Mas R, Navarro A, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc. Natl. Acad. Sci.* 2013;110(45):18250–18255.
8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
9. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat. Biotechnol.* 2011;29(1):24–26.
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. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 2011;43(5):491–498.
12. Van der Auwera GA, Carneiro MO, Hartl C, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* 2013;11(SUPL.43):11.10.1-11.10.33.
13. Boultwood J. Ataxia telangiectasia gene mutations in leukaemia and lymphoma. *J. Clin. Pathol.* 2001;54(7):512–516.
14. Tort F, Hernàndez S, Beà S, et al. CHK2-decreased protein expression and infrequent genetic alterations mainly occur in aggressive types of non-Hodgkin lymphomas. *Blood*. 2002;100(13):4602–4608.
15. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994–1007.
16. Martín-Garcia D, Navarro A, Valdés-Mas R, et al. CCND2 and CCND3 hijack immunoglobulin light-chain enhancers in cyclin D1- mantle cell lymphoma. *Blood*. 2019;133(9):940–951.
17. Moncunill V, Gonzalez S, Beà S, et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat. Biotechnol.* 2014;32(11):1106–1112.
18. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 2014;15(6):R84.
19. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
20. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* 2018;10(1):33.
21. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94–101.
22. Shinde J, Bayard Q, Imbeaud S, et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics*. 2018;34(19):3380–3381.

23. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.

24. Beekman R, Chapaprieta V, Russiñol N, et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* 2018;24(6):868–880.

25. Dentro SC, Wedge DC, Van Loo P. Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb. Perspect. Med.* 2017;7(8):a026625.

26. Nadeu F, Clot G, Delgado J, et al. Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia*. 2018;32(3):645–653.

27. Landau DA, Tausch E, Taylor-Weiner AN, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526(7574):525–530.

28. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–1369.

29. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biol.* 2012;13(6):R44.

30. Duran-Ferrer M, Clot G, Nadeu F, et al. The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *bioRxiv*. 2020;2020.02.06.937383.

31. Stunnenberg HG, Hirst M, Abrignani S, et al. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell*. 2016;167(5):1145–1149.

32. Vilarrasa-Blasi R, Soler-Vila P, Verdaguer-Dot N, et al. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *bioRxiv*. 2019;764910.

33. Wang Y, Song F, Zhang B, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* 2018;19(1):151.

34. Soler-Vila P, Pons PC, Farabella I, Stefano M Di, Marti-Renom MA. Hierarchical chromatin organization detected by TADpole. *bioRxiv*. 2019;698720.

35. McGowan-Jordan J, Simons A, Schmid M (Eds). An international system for human cytogenomic nomenclature. Karger: Basel 2016.

36. Göhring G, Michalova K, Beverloo HB, et al. Complex karyotype newly defined: the strongest prognostic factor in advanced childhood myelodysplastic syndrome. *Blood*. 2010;116(19):3766–3769.

37. Nagel I, Szczepanowski M, Martín-Subero JI, et al. Deregulation of the telomerase reverse transcriptase (TERT) gene by chromosomal translocations in B-cell malignancies. *Blood*. 2010;116(8):1317–1320.

38. Brezinschek HP, Brezinschek RI, Lipsky PE. Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. *J. Immunol.* 1995;155(1):190–202.

39. Lefranc M-P, Giudicelli V, Ginestoux C, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* 2009;37(Database issue):D1006–D1012.

40. Ye H, Desai A, Zeng D, et al. Smoldering mantle cell lymphoma. *J. Exp. Clin. Cancer Res.* 2017;36(1):185.