

## **SUPPLEMENTARY FIGURES**

**Supplementary Table 1: Clinicobiological characteristics of NSCLC cohort.**

**Supplementary Table 2: Number of variants identified in TB specimens, CDXs and CDX-derived cell lines.**

**Supplementary Table 3: Number of high-confidence variants identified in CTC samples.**

**Supplementary Table 4: List of antibodies used for IHC and staining conditions.**

**Supplementary Table 5: List of antibodies used for FACS analysis.**

**Supplementary Table 6: List of antibodies used for western blot.**

**Supplementary Table 7: List of antibodies used for immunofluorescence.**

**Supplementary Figure 1: Clinical timelines of patients L1, L2, L3 and L4.** The length of each segment is proportional to the duration of treatment (1cm = 1 month). Patients L2, L3 and L4 biopsy at diagnosis was issued from the left latero-tracheal lymph node, the right inferior lobe and the left superior lobe of the lung respectively. Patient L4 biopsy at progression was sampled in the lung.

**Supplementary Figure 2: Phenotypic characterization of patient samples, CDX and CDX-derived cell lines.** (A) Number of CTCs detected by CellSearch according to tumor molecular status of the 55 NSCLC patients. (B) Representative images of cell morphology in the three CDX-derived cell lines at 10X, 20X and 40 X magnifications. Scale bar: 50 µm. (C) Immunohistochemical stainings on GR-CDXL1 CDX tumor and CDX-derived cell line (scale bar: 20 µm), patient L2 biopsy and GR-CDXL2 CDX tumor (scale bar: 30 µm) (HES, CK8/18, EpCAM, Ki67, Vimentin, TTF1,

Chromogranin A and synaptophysin) and on patient L3 biopsy, GR-CDXL3 CDX tumor and CDX-derived cell line (scale bar: 20  $\mu$ m) (HES, chromogranin A, synaptophysin, P40, CK5/6, EpCAM, Vimentin, Ki67). All images are shown at 20X magnification. **(D)** Flow cytometry analyses of epithelial (EpCAM, panCK, E-cadherin), mesenchymal (vimentin) and cancer stem cell markers (ALDH, CD133, CD90, ABCG2 and CD166) in the three CDX-derived cell lines.

**Supplementary Figure 3: In vivo drug assays.** Palpable subcutaneous tumors of GR-CDXL1 (n=10), GR-CDXL2 (n=6), GR-CDXL3 (n=6) and GR-CDXL4 (n=9) at passages P11, P5, P5 and P4 respectively were treated once every week (intraperitoneal injection) with paclitaxel (15 mg/kg) and/or cisplatin (5 mg/kg) or the vehicle as indicated by the arrows. Tumor volumes were determined by caliper according to the formula  $(\text{width}^2 \times \text{length})/2$ . Data are presented as mean tumor volumes  $\pm$  SEM.

**Supplementary Figure 4: Metastatic tumor detection in mice.** GR-CDXL1, GR-CDXL3 and GR-CDXL4 tumor FFPE sections stained with HES, and anti-human CK7 and anti-human Ki67.

**Supplementary Figure 5: Flow diagram of samples available for WES analysis for each patient.**

**Supplementary Figure 6: Comparative genomic analysis of the CDX and the CDX-derived cell lines.** **(A)** Fraction of CDX mutations shared with the corresponding cell line. **(B)** Fraction of CDX driver mutations shared with the corresponding cell line. **(C)** Mutated driver genes and their amino acid sequence variation in the CDX and the CDX-derived cell line. **(D)** Mutated driver genes and their amino acid sequence variation in the CDX only.

**Supplementary Figure 7: Dendograms resulting from hierarchical clustering analysis of all mutations detected by WES among patients L1, L3 and L4 samples.**

**Supplementary Figure 8: Comparative CNA analysis of the biopsies, the CDX and the CDX-derived cell lines.** Circos plot of CNAs detected in the biopsy, the CDX and the CDX-derived cell line (when available) for each model (GR-CDXL1, GR-CDXL2, GR-CDXL3 and GR-CDXL4).

**Supplementary Figure 9: Comparative CNA analysis of the CDX models.** Circos plot of CNAs detected in the four CDX models (gain in red, normal in black, loss in blue) (1: GR-CDXL1, 2: GR-CDXL2, 3: GR-CDXL3, 4: GR-CDXL4).

**Supplementary Figure 10: Frequency of CDX gene alterations in lung cancer histologies according to cBioPortal studies.** (A) Genes harboring truncal mutations in GR-CDXL1, GR-CDXL2 and GR-CDXL4 were found in 79% of genes altered in lung adenocarcinoma. (B) Genes harboring truncal mutations in GR-CDXL3 are found in 61% of genes altered in squamous cell carcinoma. All CDX models displayed functional inactivation of *TP53* found at 53.8%. GR-CDXL1 harbored a deleterious *KMT2C* (9.4%), *ARID1A* (6.4%), *ATRX* (5.4%), *BRCA2* mutation (3.2%) and *FANCA* deletion (0.9%). GR-CDXL2 had a *KRAS*-mutant profile which included co-occurring mutations in *KEAP1* (14.5%), *STK11* (15.7%) and *RBM10* (9.5%) genes, exclusive of adenocarcinoma. Genomic characterization revealed DDR-related *CHEK2* and *ARID1B* driver mutations found at 1.4% and 3.4% respectively (1–5). (C) Genes harboring truncal mutations in GR-CDXL1, GR-CDXL2, GR-CDXL3 and GR-CDXL4 are predominant in other types of metastatic cancers.

**Supplementary Figure 11: Statistics of allele drop-out and false-positive rate of single CTCs.** (A) and (B) Representation of allelic drop-out (ADO) in GR-CDXL1 (A. left) and GR-CDXL3 (B. left). Reliable variants (green) were defined by an equal variant allele frequency (VAF) in both germline DNA and WBC bulk samples. Variants in ADO (red) were defined by a VAF ranging from 0.2 to 0.8 in germline DNA and <0.1 or >0.9 in WBC bulk. False-positive rates in one P8 CTC (A. right) and in five P51 CTCs (B. right) are shown. To estimate false-positive rates, we divided the number of reliable somatic variants not present in bulk tumor samples (TBs and CDX) by the number of target bases covered  $\geq 8X$  in the same sample.

**Supplementary Figure 12: Chromosomal instability of GR-CDXL3 and GR-CDXL4 CDX models.** (A) Absolute copy number profiles of GR-CDXL3 and GR-CDXL4 CDX models. Yellow=normal/major copy number, red=gain, green=loss, purple=amplification. (B) Representative images of mitotic lagging chromosome analysis in GR-CDXL1 and GR-CDXL3 cells after 4'6-diamidino-2-phenylindole (DAPI, blue) and centromere staining with anti-CENPA (green). (C) Lagging DNA analysis obtained from scoring at least 50 anaphases per cell line in three independent experiments. Data are presented as mean  $\pm$  SEM, \* $p < 0.05$  (unpaired t-test with Welch's correction).

**Supplementary Figure 13: Diagram of DDR-related pathways implicated in CDX-derived cell line vulnerabilities and targeting strategies.** DNA damage in the three CDX-derived cell lines results in the activation of several repair cascades prone to defects. Olaparib-induced PARP trapping leads to synthetic lethality in *BRCA2/FANCA*-defective GR-CDXL1 cells. Dual targeting of AKT1 gain through the inhibition of PI3KA (by BYL719) and centrosome clustering via KIFC1 inhibition (by AZ82) leads to GR-CDXL3 cell death. In GR-CDXL4 cells, SLFN11 overexpression

may lead to olaparib sensitivity. SSB, single-strand DNA break; DSB, double-strand DNA break; HR, homologous recombination.

**Supplementary Figure 14: *In vivo* parameters.** (A) Western blot analysis of the levels of p-AKT and AKT in CDX-derived (GR-CDXL1, GR-CDXL3 and GR-CDXL4) and NSCLC (A427, A549 and H441) cell lines. (B) Average body weight of mice injected with GR-CDXL1 or GR-CDXL4 Luc cells measured over the course of treatment. (C) Quantitative analysis of BLI. (D) Average body weight of mice injected with GR-CDXL3 Luc cells measured over the course of treatment. (E) Quantitative analysis of BLI. Each point represents a single mouse. For **B-E**, n=5 for all groups except GR-CDXL1 NT where n=4. Data are presented as mean  $\pm$  SEM, \*\*\*p<0.001, \*\*\*\*p<0.0001 (two-way ANOVA).

## SUPPLEMENTARY METHODS

**Enrichment, detection and isolation of single CTCs.** Individual CTC isolation was performed by combining different methods. Either: (i) isolation by size of epithelial tumor cells (ISET) filtration, immunofluorescence staining and scanning of filters, followed by laser microdissection CD45<sup>+</sup> cells, or (ii) enrichment via RosetteSep, immunofluorescence staining followed by FACS isolation, or (iii) CTC detection using CellSearch followed by CTC isolation using self-seeding microwell chips. The methods are described in detail in previous reports (6, 7).

**WGA, quality control, dsDNA.** Whole genome amplification (WGA) of CTCs and CD45-positive cells was performed using the Ampli1 WGA kit (Menarini) according to manufacturer's instructions. The quality of Ampli1 WGA products was checked by multiplex PCR as described by Polzer *et al.* (8). To increase total dsDNA content in Ampli1 WGA products, ssDNA molecules were converted into dsDNA molecules using the Ampli1 ReAmp/ds kit (Menarini).

**Isolation of genomic DNA from blood, TB, CDX and CDX-derived cell line.** DNA from formalin-fixed paraffin-embedded tumor biopsies was purified with QIAamp DNA FFPE Tissue kit (Qiagen, Hilden, Germany) according to manufacturer's protocol. DNA from the CDX was extracted with AllPrep DNA/RNA kit (Qiagen) and germline DNA and cell line DNA was purified with QIAamp DNA blood kit (Qiagen).

**Sequence alignment and variant calling.** Base calling was performed using the Real-Time Analysis software sequence pipeline (2.7.7) from Illumina with default parameters. Sequence reads from amplified DNA (circulating T cells and CD45 pools) were trimmed for Ampli1 adapters with Cutadapt (1.14) (9). Human reads from xenograft samples were extracted by bamcmp (10). Reads were then aligned to the

142 human genome build hg38/GRCh38.p7 using the Burrows-Wheeler Aligner (BWA)  
143 tool (11). Duplicated reads were removed using Sambamba (12). Variant calling of  
144 single nucleotide variants (SNVs) and small insertions/deletions (indels) was  
145 performed using the Broad Institute's GATK Haplotype Caller GVCF tool (3.7) (13,  
146 14) for germline variants and MuTect2 tool (2.0, --  
147 max\_alt\_alleles\_in\_normal\_count=2; --max\_alt\_allele\_in\_normal\_fraction=0.04) (15)  
148 for somatic variants. To keep only reliable somatic variants, we then applied the  
149 following post-filtering steps:

- 150 - Variants classified as "PASS" or "t\_lod\_fstar" by MuTect2 (and not flagged  
151 as PID).
- 152 - Coverage  $\geq 8$  in the tumor and matched normal sample.
- 153 - QSS score  $\geq 20$ .
- 154 - Variant allele fraction in the tumor (VAFT)  $\geq 0.05$  with  $\geq 5$  mutated reads,  
155 variant allele fraction in the normal sample (VAFN) = 0.

156 Additional criteria were applied to generate a high-confidence set of variants from  
157 CTCs. Variants had to be present in either the primary tissue (at least 1 TB  
158 specimen) or the CDX.

159 Bam-readcount (<https://github.com/genome/bam-readcount>) was used to rescue  
160 reliable variants that were present in at least two tumor samples and were not  
161 detected by Mutect2 because of their low VAF. Ensembl's Variant Effect Predictor  
162 (VEP, release 87) (16) was used to annotate variants with respect to functional  
163 consequences (type of mutation and prediction of the functional impact on the protein  
164 by SIFT.2.2 and PolyPhen 2.2.2) and frequencies in public (dbSNP147, 1000  
165 Genomes phase 3, ExAC r3.0, COSMIC v79) and in-house databases. We used the  
166 Cancer Genome Interpreter (17) to predict driver and passenger mutations.

**Dendrogram.** SNV mutations were aggregated as a binary score (1 if a non-synonymous or frameshifting, exonic or on a splice-site was observed, else 0) at the gene level. Resulting profiles were clusterized using the binary distance and Ward's aggregation method. All computation and figure processings were performed using R v4.0.2.

**ADO and false-positive rate estimation.** CTC and CD45<sup>+</sup> pool DNA were amplified before sequencing. To estimate ADO, we selected all reliable variants in germline or CD45 DNA using HaplotypeCaller with the following post-filtering: coverage  $\geq 8$  in both samples,  $\geq 5$  variant reads representing  $\geq 5\%$  of sequenced reads at that position, genotype quality  $\geq 20$ . We then compared the proportions of normal/variant reads in the germline and CD45 DNA using Fisher's exact test. Variants with a significant difference ( $p < 0.05$ ), a variant allele fraction between 0.2 and 0.8 in germline DNA and  $< 0.1$  or  $> 0.9$  in the CD45 DNA were considered to have undergone ADO. To estimate the false-positive rates in CTC samples, we divided the number of potentially false-positive events by the number of target bases covered  $\geq 8X$  in the same sample. We adopted a conservative approach and considered as false-positive all events not found in the TB and the CDX.

**Copy number analysis.** To identify copy-number alterations (CNAs), we identified germline single-nucleotide polymorphisms (SNPs) in each sample and we calculated the coverage log-ratio (LRR) and B allele frequency (BAF) at each SNP site. Genomic profiles were divided into homogeneous segments by applying the circular binary segmentation algorithm, as implemented in the Bioconductor package *DNAcopy* (18), to both LRR and BAF values. We then used the Genome Alteration Print (GAP) method (19) to determine the ploidy of each sample, the level of

contamination with normal cells and the allele-specific copy number of each segment. Ploidy was estimated as the median copy-number across the genome. Chromosome aberrations were then defined using empirically determined thresholds as follows: gain, copy number  $>$  ploidy + 0.5; loss, copy number  $<$  ploidy – 0.5; high-level amplification, copy number  $>$  ploidy + 2; homozygous deletion, copy number  $<$  0.5. Finally, we considered a segment to have undergone LOH when the copy number of the minor allele was equal to 0. CIRCOS plots were generated using CIRCOS (20).

**Characterization of known copy number changes in CTCs.** As expected, the log-ratio (LRR) and B allele frequency (BAF) profiles of CTC samples were too noisy to obtain reliable pangenomic copy-number profiles. However, we observed that many chromosome segments displayed allelic imbalances consistent with the presence of chromosome aberrations identified in other samples, in particular PDX and cell line samples. We used these allelic imbalances to detect chromosome aberrations identified in other samples as follows:

- 1) For each CTC and each chromosome aberration, we counted the number of SNPs with consistent (e.g. BAF  $>$  0.5 in the CTC and tumor samples) and discordant allelic imbalance

- 2) We used Fisher's exact test to identify chromosome segments with a significant enrichment in consistent SNPs

- 3) We considered an aberration to be present in a CTC sample if the Fisher test was significant (p-value  $<$  0.05) with  $\geq$ 80% consistent SNPs.

**Phylogenetic inference.** Trees were built using a binary presence/absence matrix built from the VAF of each sample (present = VAF  $>$  0). The R Bioconductor package

214 phangorn v2.3.1 (21) was used to perform the parsimony ratchet method (22),  
215 generating unrooted trees. The number of mutations and driver mutations on each  
216 branch were then determined by selecting mutations present in all samples  
217 downstream the branch.

218 ***CDX-derived cell line establishment and cell culture.*** After resection, tumor  
219 fragments were conserved in RPMI 1640 medium (Life Technologies, Carlsbad, CA,  
220 USA) and immediately transferred to the laboratory. After two washes in 1X PBS (Life  
221 Technologies) and incubation for 10min in a 10 mL 1X PBS solution containing 1:10  
222 penicillin/streptomycin (penicillin 10,000 units/mL, streptomycin 10,000 µg/mL,  
223 Life Technologies), tumor fragments were first mechanically dissociated using a  
224 scalpel before enzymatic dissociation with the Tumor Dissociation Human Kit  
225 (Miltenyi Biotech, Köln, Germany) according to manufacturer's protocol. Cell  
226 suspension was then successively filtered on 100-µm and 40-µm cell strainer and  
227 washed with PBS 1X before counting. Depletion of mouse cells was performed with  
228 the Mouse Cell Depletion Kit according to manufacturer's protocol using an  
229 AutoMacs Pro Separator (Miltenyi Biotec). Tumor cells were then centrifuged and  
230 resuspended in Advanced DMEM/F12 medium (Life Technologies) supplemented  
231 with 10% FBS, 1% antibiotics (penicillin-streptomycin) and 1% ultraglutamine (Lonza,  
232 Basel, Switzerland). After counting, cells were plated in six-well plates (TPP,  
233 Trasadingen, Switzerland) coated with poly-L-lysine (Merck, Sigma-Aldrich LLC.,  
234 Saint-Louis, MO, USA). For GR-CDXL3 and GR-CDXL4 cells, culture medium is  
235 supplemented with ROCK1 inhibitor 10nM (Y-27632; Selleckchem, Houston, TX,  
236 USA). Cells were observed three times a week and passaged in tissue culture flasks  
237 for cell expansion, freezing or characterization; cells were detached using 0.005%  
238 trypsin-EDTA (Life Technologies) before centrifugation and counting. The same

239 normal-serum culture medium was used for cell expansion and permanent culture.  
240 Human adenocarcinoma NSCLC cell lines A549 (derived from primary lung tumor)  
241 and H441 (derived from pericardial fluid) were obtained through the CANCER-ID  
242 consortium and grown in RPMI 1640 medium. Lung A427 cell line (derived from  
243 primary lung tumor) was obtained from the American Type culture Collection (ATCC)  
244 and grown in MEM medium (Sigma Aldrich) supplemented with 10% FBS, 1%  
245 antibiotics, 1% ultraglutamine and 1% sodium pyruvate (Thermo Scientific). All cell  
246 lines were regularly verified for mycoplasma contamination using MycoAlert (Lonza).

247 **Immunohistochemistry.** IHC staining was performed on formalin-fixed, paraffin-  
248 embedded (FFPE) tissue from patient biopsy specimens, CDX and CDX-derived cell  
249 lines with antibodies to CK8/18, EpCAM, Ki67 antigen, Vimentin, CD44, TTF1, p40,  
250 CK5/6, Chromogranin A, Synaptophysin and SLFN11 (**Supplementary Table 4**). All  
251 primary antibody incubations were carried out at RT for 1h except incubation with  
252 CK5/6 and SLFN11, which was performed at 37°C. IHC stains were examined by an  
253 experienced pathologist (JYS).

254 **FACS analysis.** Aldehyde dehydrogenase (ALDH) activity was performed using the  
255 Aldefluor kit (StemCell Technologies) according to manufacturer's instructions.  
256 EpCAM, CD133, CD166, pan-cytokeratins and E-cadherin antibodies were used  
257 according to manufacturer's protocol. Fixation and permeabilization steps were  
258 performed for pan-cytokeratins and E-cadherin antibodies using the Fix&Perm kit  
259 (ThermoFisher) according to the manufacturer's instructions.  $2 \times 10^5$  cells were  
260 incubated with each antibody or corresponding negative control isotype antibodies  
261 (**Supplementary Table 5**) at RT for 20 min. Acquisition was performed with LSR

Fortessa cytometer (BD Biosciences) equipped with BD FACS Diva software. Data were analyzed using the Kaluza software (Beckman Coulter).

**Protein extraction and western blot analysis.** Cells were lysed in 1% NP40 lysis buffer [150 mM NaCl, 1 mM EDTA, 50 mM Tris pH7.5, 0.5% NP40, H<sub>2</sub>O] supplemented with protease and phosphatase inhibitors (Roche) for 30min on ice. Protein concentrations were measured using the Micro BCA™ protein assay kit (Thermo Scientific). Laemmli buffer (4X) containing β-mercaptoethanol was added and samples were subsequently denatured by boiling at 95°C. Proteins were resolved by 6, 8 or 10% SDS-polyacrylamide gels and transferred onto a nitrocellulose membrane (Bio-Rad), followed by a PBS-milk (5%)-blocking step (1h). Signals were detected using WesternBright ECL (Advansta) on the ChemiDoc XRS+ System (Bio-Rad). The antibodies used in this study are listed in **Supplementary Table 6**.

**RNA extraction and quantitative real-time PCR analysis.** Total RNA was isolated using the ReliaPrep RNA Cell Miniprep System (Promega) and 1 µg of RNA was used to synthesize cDNA using the High-capacity RNA-to-cDNA™ kit (Thermo Scientific) according to manufacturer's instructions. The primers used were as follows:

- *SLFN11* (forward 5'-GGCCCAGACCAAGCCTTAAT-3' and reverse 5'-CACTGAAAGCCAGGGCAAAC-3')
- *GAPDH* (forward 5'-CCTCAACGACCACTTTGTCA-3' and reverse 5'-TTCCTCTTGTGCTCTTGCTG-3')

Quantitative real-time PCR was carried out using Maxima SYBR Green/ROX qRT-PCR Master Mix (Thermo Scientific) on the Viia7 PCR system (Applied Biosystems).

Relative gene expression was calculated according to the  $\Delta\Delta C_q$  method and normalized to *GAPDH* expression.

**Drugs and chemicals.** Aphidicolin (CAS 38966-21-1, Calbiochem) was purchased from Merck. Cisplatin (Mylan Pharma) and PI3KA inhibitor BYL719 (Novartis) were kind gifts from Dr S. Postel-Vinay and Dr Fabrice Andre (INSERM U981, Gustave Roussy, Villejuif) respectively. The PARPi olaparib (AZD-2281, AstraZeneca) was purchased from Selleck Chemicals. KIFC1 inhibitor AZ82 (AOB4872-5) was obtained from CliniSciences. All stock solutions were prepared in DMSO.

**Stable cell lines expressing mCherry-Luc or GFP-Luc.** Stable tumor cell lines were established after infection with retroviral/lentiviral vectors (plasmids: MI-Luc-IRES-mCherry #75020, Addgene Teddington, UK; lentivirus: RediFect™ Red-FLuc-GFP, PerkinElmer). Production and titration of retroviral particles were performed as previously described (23). Infection was performed on retronectin-coated plates (TaKaRa Bio, CA, USA) and efficiency was assayed by testing mCherry expression using flow cytometry. When efficiency was below 98%, cell sorting was performed.

## Additional References

1. Imielinski M et al. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell* 2012;150(6):1107–1120.
2. Rizvi NA et al. Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* 2015;348(6230):124–128.
3. Caso R et al. Preoperative clinical and tumor genomic features associated with pathologic lymph node metastasis in clinical stage I and II lung adenocarcinoma. *npj Precis. Onc.* 2021;5(1):1–8.
4. Jordan NV et al. HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* 2016;537(7618):102–106.
5. Ding L et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455(7216):1069–1075.
6. Pailler E et al. Acquired Resistance Mutations to ALK Inhibitors Identified by Single Circulating Tumor Cell Sequencing in ALK-Rearranged Non–Small-Cell Lung Cancer. *Clin Cancer Res* 2019;25(22):6671–6682.
7. Faugeroux V et al. An Accessible and Unique Insight into Metastasis Mutational Content Through Whole-exome Sequencing of Circulating Tumor Cells in Metastatic Prostate Cancer. *European Urology Oncology* 2020;3(4):498–508.
8. Polzer B et al. Molecular profiling of single circulating tumor cells with diagnostic intention. *EMBO Mol Med* 2014;6(11):1371–1386.
9. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011;17(1):10–12.

- 331 10. Khandelwal G et al. Next-Generation Sequencing Analysis and Algorithms for  
332 PDX and CDX Models. *Mol Cancer Res* 2017;15(8):1012–1016.
- 333 11. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler  
334 transform. *Bioinformatics* 2009;25(14):1754–1760.
- 335 12. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing  
336 of NGS alignment formats. *Bioinformatics* 2015;31(12):2032–2034.
- 337 13. DePristo MA et al. A framework for variation discovery and genotyping using  
338 next-generation DNA sequencing data. *Nat Genet* 2011;43(5):491–498.
- 339 14. Auwera GAV der et al. From FastQ Data to High-Confidence Variant Calls: The  
340 Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics*  
341 2013;43(1):11.10.1-11.10.33.
- 342 15. Cibulskis K et al. Sensitive detection of somatic point mutations in impure and  
343 heterogeneous cancer samples. *Nat Biotechnol* 2013;31(3):213–219.
- 344 16. McLaren W et al. The Ensembl Variant Effect Predictor. *Genome Biology*  
345 2016;17(1):122.
- 346 17. Tamborero D et al. Cancer Genome Interpreter annotates the biological and  
347 clinical relevance of tumor alterations. *Genome Medicine* 2018;10(1):25.
- 348 18. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for  
349 the analysis of array CGH data. *Bioinformatics* 2007;23(6):657–663.

19. Popova T et al. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biology* 2009;10(11):R128.
20. Krzywinski M et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–1645.
21. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics* 2011;27(4):592–593.
22. Nixon KC. The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics* 1999;15(4):407–414.
23. Hamelin V, Letourneux C, Romeo P-H, Porteu F, Gaudry M. Thrombopoietin regulates IEX-1 gene expression through ERK-induced AML1 phosphorylation. *Blood* 2006;107(8):3106–3113.

373    **Abbreviations**

374

375    CTC: circulating tumor cell

376    CDX: CTC-derived eXplant

377    NSCLC: non-small cell lung cancer

378    DDR: DNA damage response

379    HR: homologous recombination

380    HRD: homologous recombination deficiency

381    NHEJ: non-homologous end joining

382    PARPi: poly-ADP polymerase inhibitor

383    TB: tumor biopsy

384    NB: nuclear body

385    CAM: chick embryo chorioallantoic membrane

386    CNA: copy number analysis

387    CIN: chromosomal instability

388    WGD: whole-genome doubling

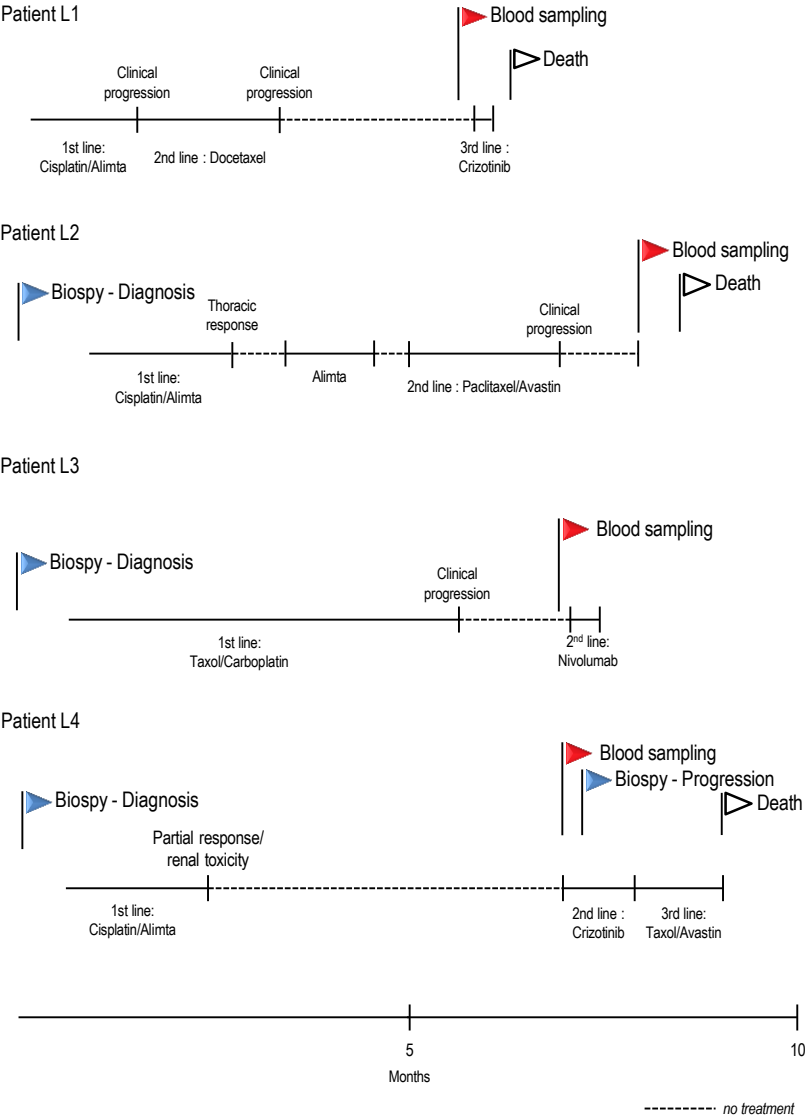
389    WES: whole-exome sequencing

390    IHC: immunohistochemistry

391    RT: room temperature

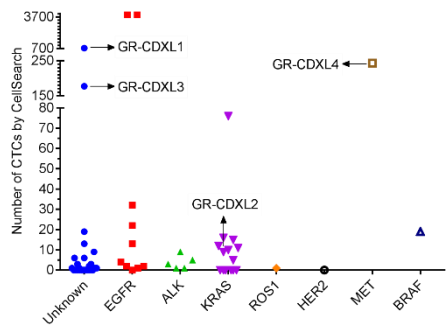
392    BLI: bioluminescence imaging

Supplementary Figure 1

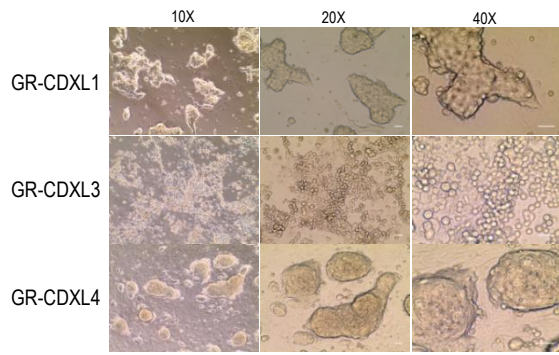


# Supplementary Figure 2

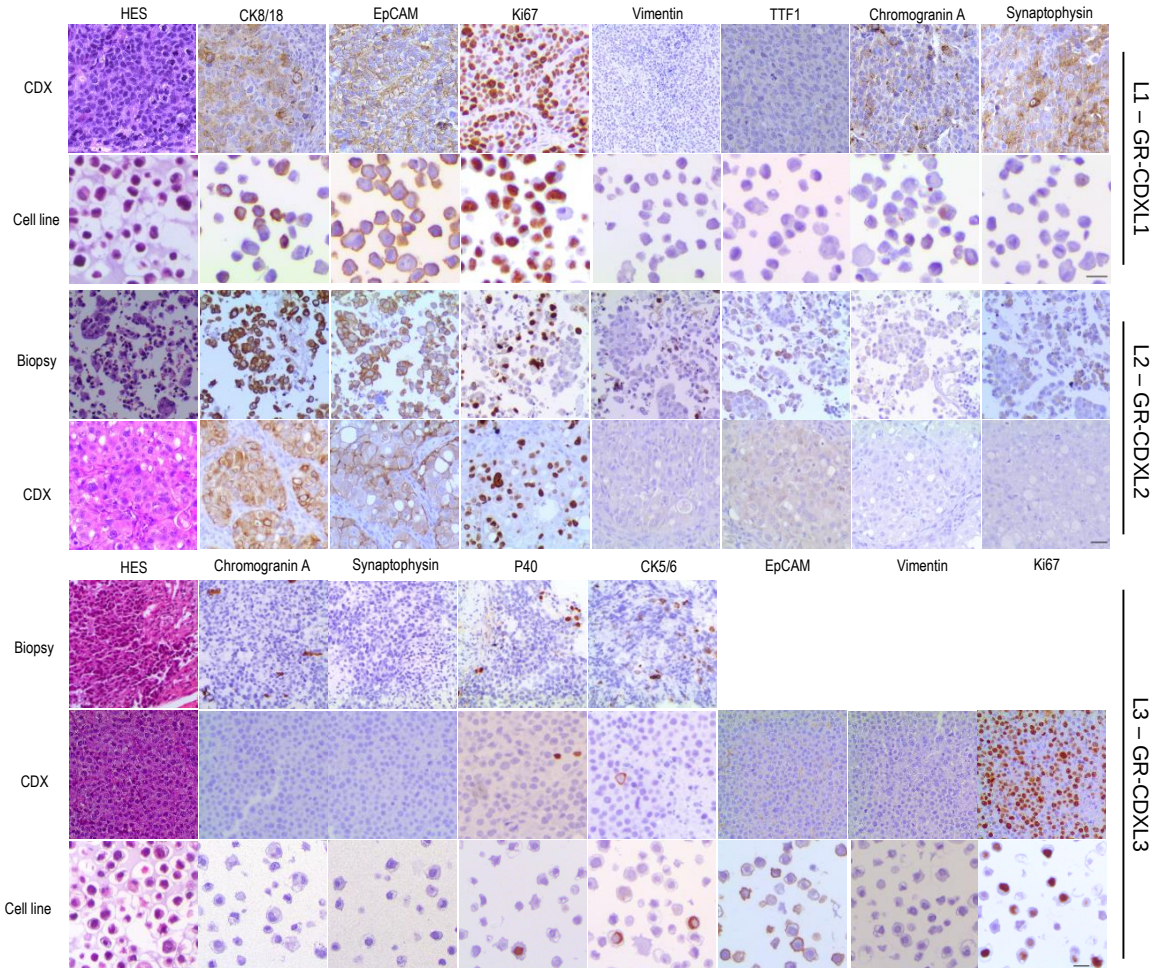
**A**



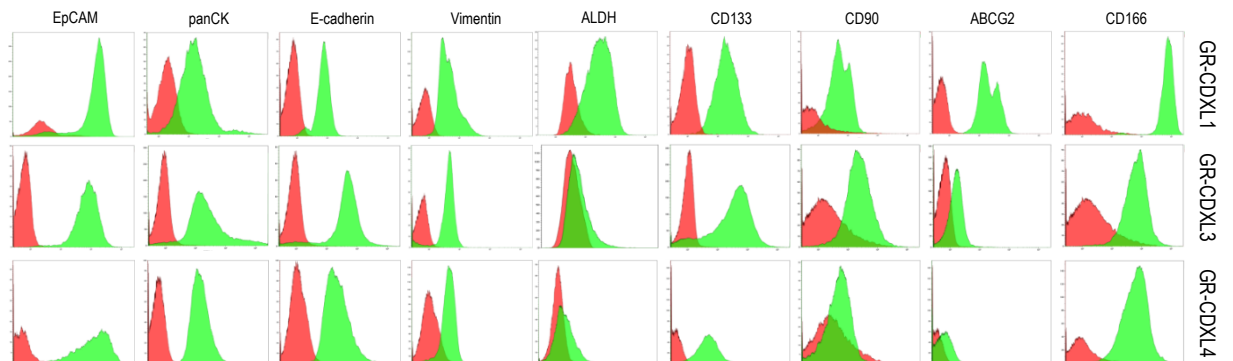
**B**



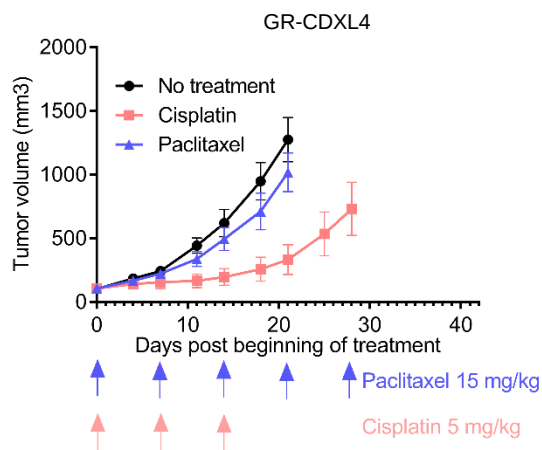
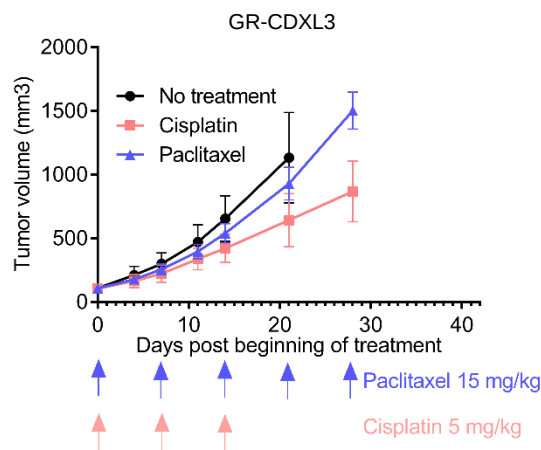
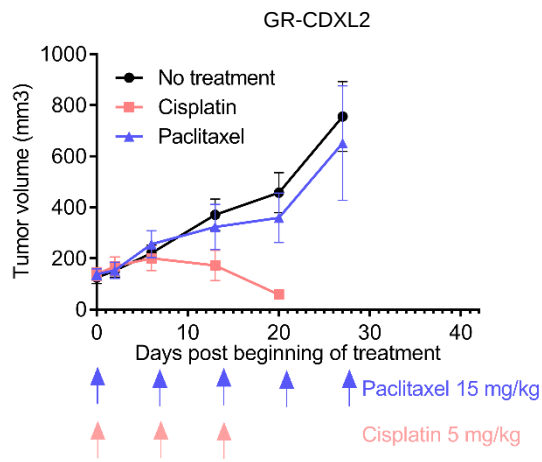
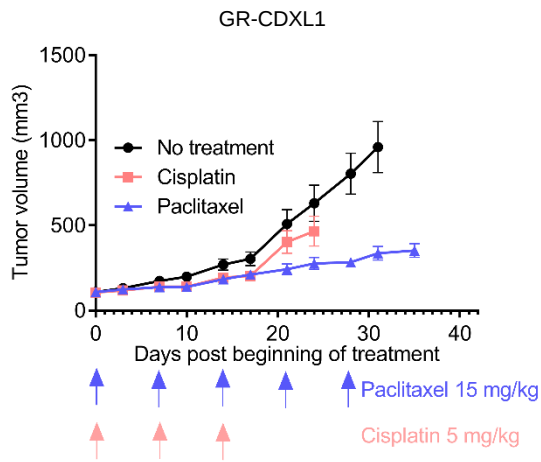
**C**



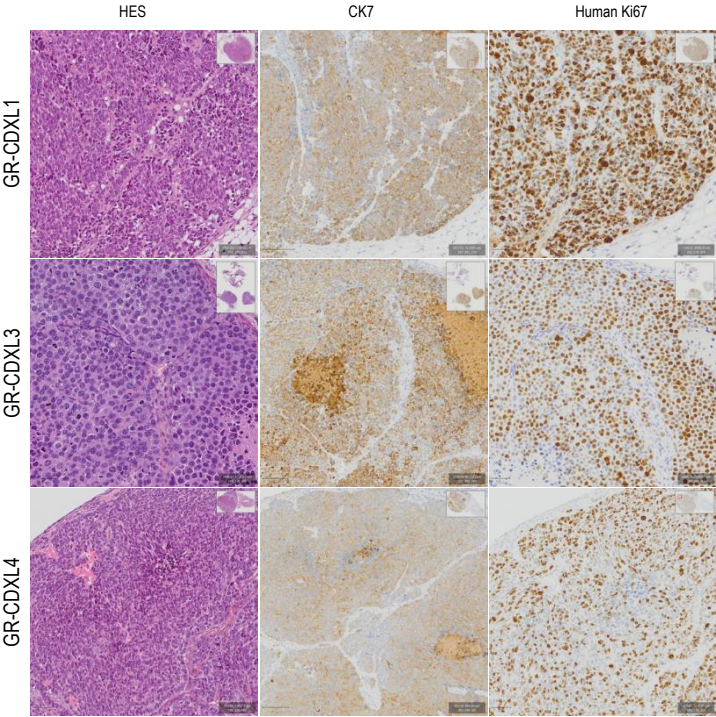
**D**



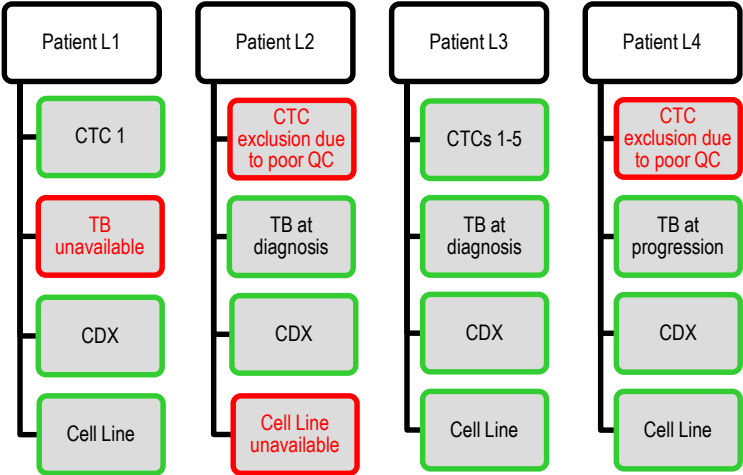
Supplementary Figure 3



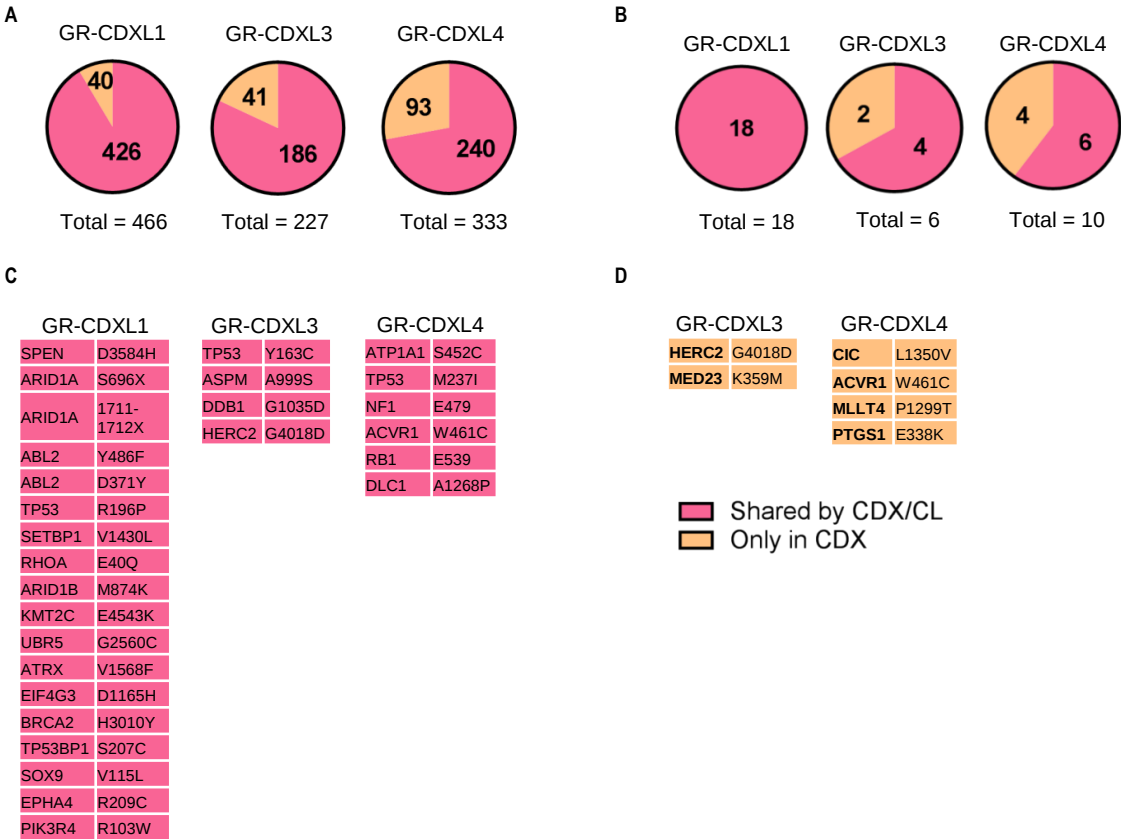
Supplementary Figure 4



Supplementary Figure 5

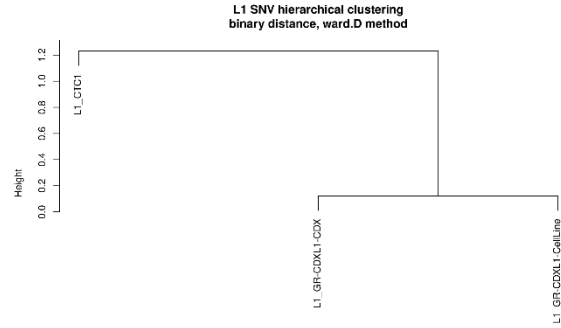


Supplementary Figure 6

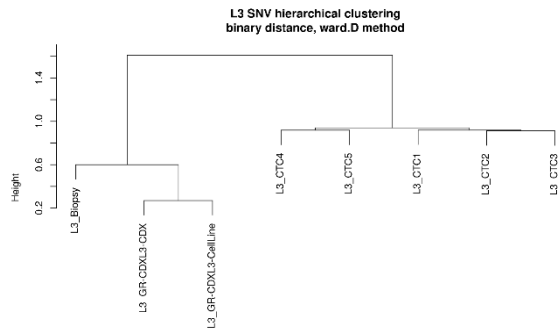


Supplementary Figure 7

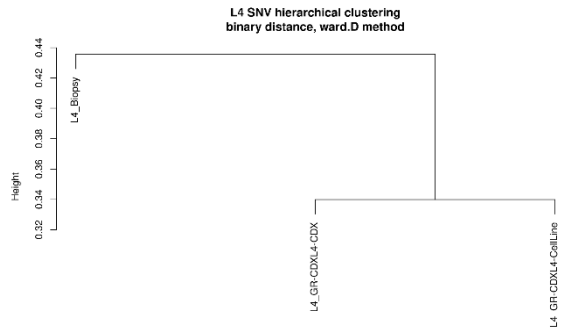
A



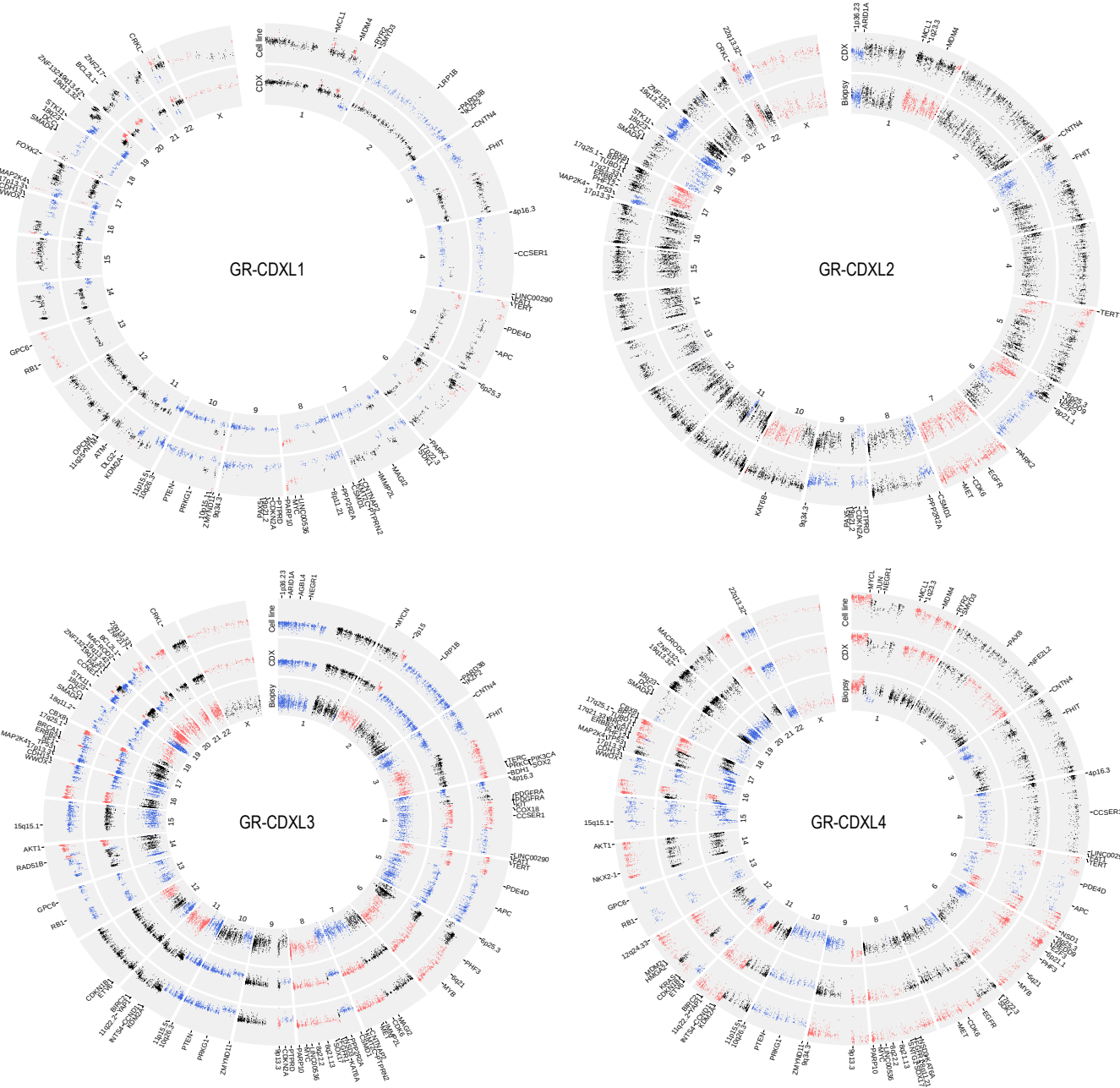
B



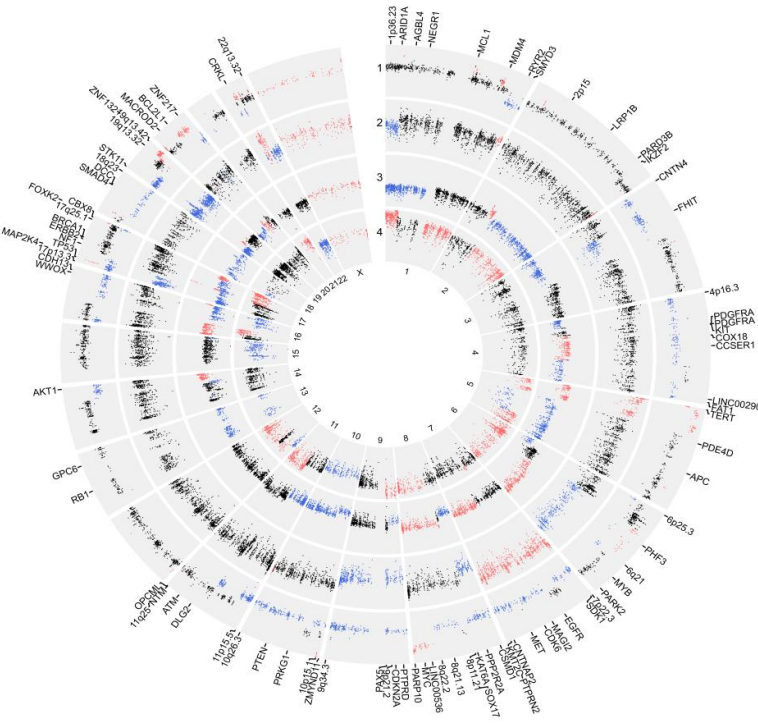
C



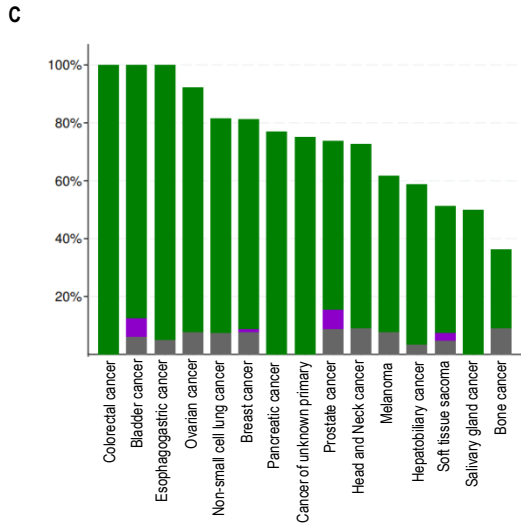
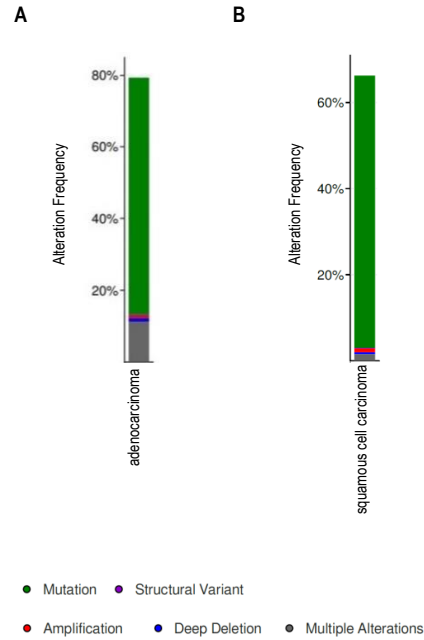
Supplementary Figure 8



### Supplementary Figure 9

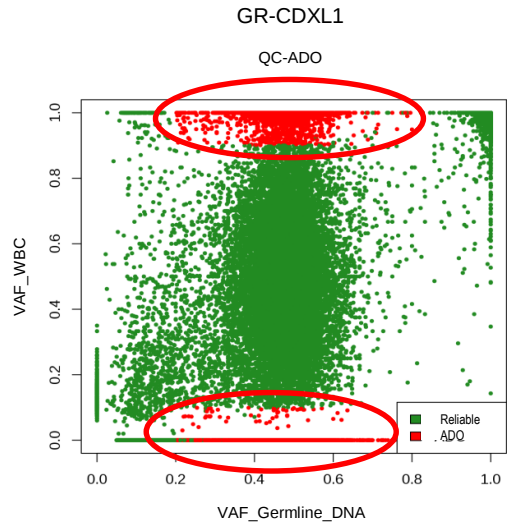


Supplementary Figure 10



Supplementary Figure 11

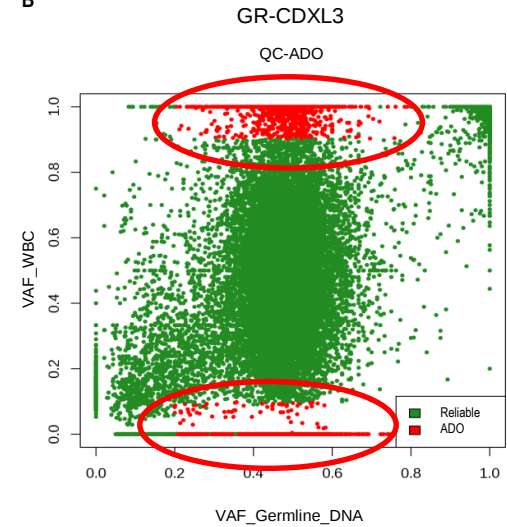
A



|                                          |  |
|------------------------------------------|--|
| Proportion of germline variants with ADO |  |
| 0,157515487                              |  |

| Sample  | Number of false-positive variants | Number of target bases covered $\geq 8X$ | False-positive rate |
|---------|-----------------------------------|------------------------------------------|---------------------|
| L1-CTC1 | 1306                              | 10465312                                 | 0,0001247           |

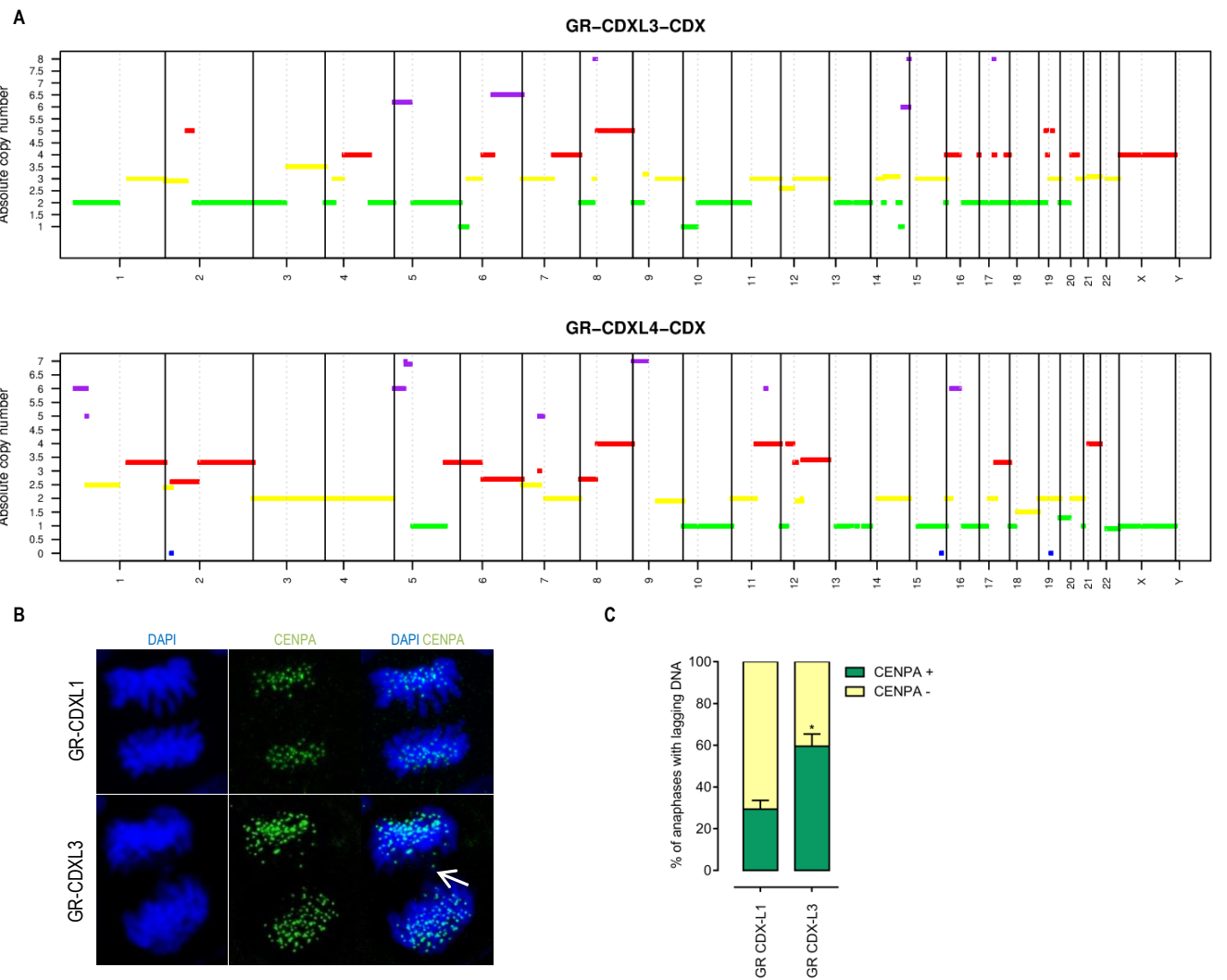
B



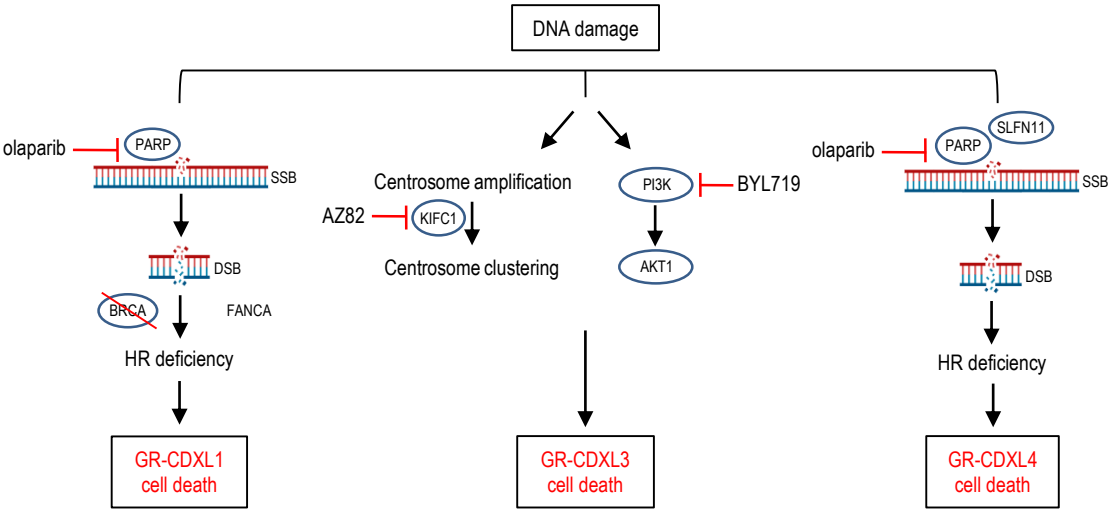
|                                          |  |
|------------------------------------------|--|
| Proportion of germline variants with ADO |  |
| 0,09321979                               |  |

| Sample  | Number of false-positive variants | Number of target bases covered $\geq 8X$ | False-positive rate |
|---------|-----------------------------------|------------------------------------------|---------------------|
| L3-CTC1 | 1647                              | 35136861                                 | 4,68739E-05         |
| L3-CTC2 | 1396                              | 29505812                                 | 4,73127E-05         |
| L3-CTC3 | 1599                              | 32123746                                 | 4,97763E-05         |
| L3-CTC4 | 1295                              | 36725416                                 | 3,52617E-05         |
| L3-CTC5 | 932                               | 32677245                                 | 2,85214E-05         |

Supplementary Figure 12



Supplementary Figure 13



Supplementary Figure 14

