Supplementary Information

Minimal Spatial Heterogeneity in Chronic Lymphocytic Leukemia at Diagnosis

Nadeu *et al.*

Supplementary Methods

Informed consent

Informed consent was obtained from all patients according to the International Cancer Genome Consortium guidelines.[1] The study was approved by the Hospital Clínic of Barcelona Ethics Committee.

Sample preparation and sequencing

Tumor cells from peripheral blood (PB) were purified from fresh or cryopreserved mono-nuclear cells using a cocktail of magnetically labeled antibodies as previously described (AutoMACS; Miltenyi Biotec).[2] Normal DNA was extracted from the non-tumoral fraction. DNA from lymph node (LN) tumoral cells was obtained from OCT embedded samples. All DNA extractions were performed using Qiagen kits and DNA quality was checked by SYBR-green staining on agarose gels and quantified using a Nanodrop ND-100 spectrophotometer. Paired-end whole-genome/exome sequencing (WGS/WES) was performed as previously described,[2] and sequenced in a Illumina HiSeq2000 (2x76 or 2x101 bp) or Illumina HiSeq X Ten (2x150 bp). The mean coverage obtained was 33x (range 30-41x) for WGS and 58x (range 25-150x) for WES.

Bioinformatic and statistical analyses

For WGS and WES, raw reads were mapped to the human reference genome (GRCh37) using the BWA-mem algorithm (v0.7.15),[3] and BAM files were generated, sorted, indexed and optical or PCR duplicates flagged using biobambam2 (v2.0.65) (https://gitlab.com/german.tischler/biobambam2). Quality control metrics were

extracted using Picard (v2.10.2) (https://broadinstitute.github.io/picard/). Somatic single nucleotide variants (SNVs) called by at least two algorithms (Sidrón,[2] CaVEMan (cgpCaVEManWrapper, v1.12.0),[4] Mutect2 (GATK v4.0.2.0),[5] and/or MuSE (v1.0 rc)[6]) were considered.[7] Short insertions/deletions (indels) were called by SMUFIN,[8] Pindel (cgpPindel, v2.2.3),[9] Platypus (v0.8.1),[10] and Mutect2, and retained for downstream analyses if identified by at least two algorithms. SNVs and indels only called in one tissue sample were automatically added in the second sample if at least one read with the mutation was found in the BAM file using Rsamtools (v1.30.0).[11,12] Copy number alterations (CNA) were called combining ASCAT (ascatNgs, v4.1.0)[13] and Battenberg (cgpBattenberg, v3.2.2)[14] or using FACETS (v0.5.14)[15] on WGS and WES data, respectively. To confirm the aberrations identified from WGS/WES data, CNA were also investigated using Genome-wide Human SNP Array 6.0 (Thermo Fisher Scientific) as previously described.[16] Structural variants (SV) were extracted from WGS data using SMuFin and BRASS (v6.0.5),[17] and were visually inspected on IGV.[18] Tumor purity used for downstream analysis was estimated using Battenberg (for samples with WGS) and FACETS (for WES). Tumor purity was verified (and adjusted if needed) in samples with low CNA burden based on distribution of the variant allele frequency of the clonal mutations. Tumor purities are listed in Supplementary Table 1.

The subclonal architecture of the tumors analyzed by WGS was reconstructed using a Bayesian approach. First, an Markov chain Monte Carlo sampler for a Dirichlet process mixture model was used to infer putative subclones (assignment of mutations to subclones, and estimation of the subclone frequencies in each sample), from the SNVs read count data, copy number states (from Battenberg) and tumor purities, as recently

described.[19] The phylogenetic relationships between subclones were identified following the "pigeonhole principle" excluding clusters with less than 100 mutations. The length of each tree branch in the reconstructed tree is proportional to the number of mutations assigned to the corresponding subclone.[19]

To assess the clonality of the indels as well as of the SNVs and indels identified by WES, the CCF of each mutation was calculated integrating read counts, CNA and tumor purity as described in Dentro *et al.*[20] The resulting CCF of each mutation was directly compared between each tissue to assess for spatial differences in the topographic abundance of specific mutations.

Supplementary Tables

Tables are placed in the Supplementary Tables Excel file.

Supplementary Table 1. Predicted tumor purities

Supplementary Table 2. Mutations identified by WGS

Supplementary Table 3. Copy number alterations

Supplementary Table 4. Structural variants identified by WGS

Supplementary Table 5. Mutations identified by WES

Supplementary Figures

Figure S1. WGS-based genome plots. Genome plots integrating (from the outer to the inner most layer) SNVs, indels, CNA and SV for each sample analyzed by WGS. CNA are depicted using intensity-based colors according to their CCF. A detailed representation of the CNA and SV found in chromosome 11 of patient CLL063 is shown in the bottom-left side of the figure. All CNA and SV were shared between the PB (left) and LN (right) in both cases. CN, copy number.

Figure S2. Comparison of the CNA profile of 12 PB (outer layer) and LN (inner layer) synchronous samples analyzed by WES. One case with no CNA in any of the two samples analyzed is not depicted in this figure. All CNA were shared at similar CCF between PB and LN. CNN-LOH, copy number neutral-loss of heterozygosity.

Figure S3. Density of the CCF of the tissue-specific mutations identified by WES. Density plot showing the distribution of the CCF of the tissue-specific mutations identified in the 13 cases analyzed by WES.

Figure S4. Comparison of the CCF of mutations identified by WES. Dot plots comparing the CCF of the mutations between PB (x-axis) and LN (y-axis) in the 13 cases analyzed by WES.

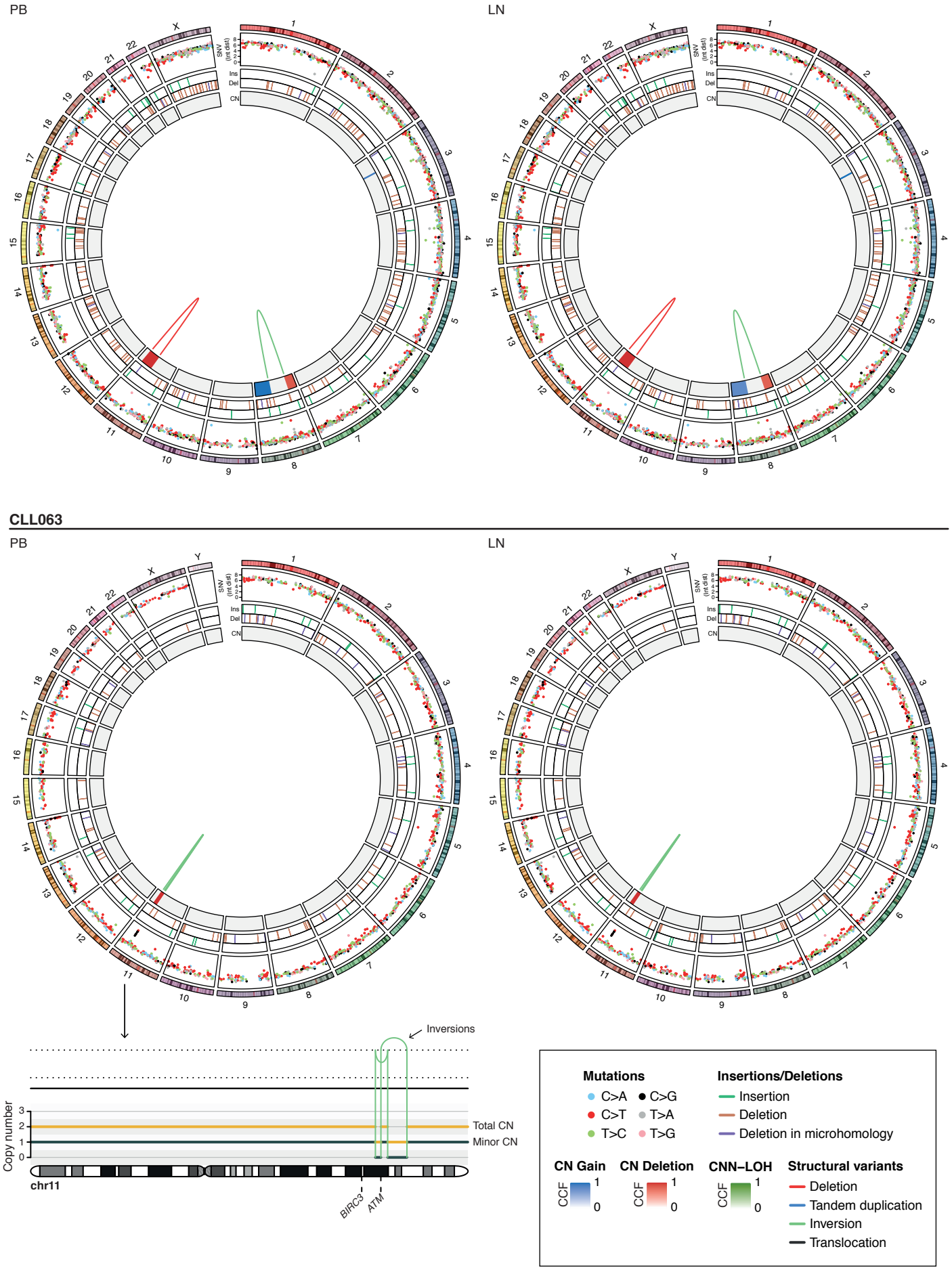
CLL290
PB
LN
CLL063
PB
LN
SNV (int dist)
Ins
Del
CN
Inversions
Copy number
Total CN
Minor CN
chr11
BIRC3
ATM
Mutations
C>A
C>G
C>T
T>A
T>C
T>G
Insertions/Deletions
Insertion
Deletion
Deletion in microhomology
CN Gain
CN Deletion
CNN–LOH
CCF
Structural variants
Deletion
Tandem duplication
Inversion
Translocation

Figure S1

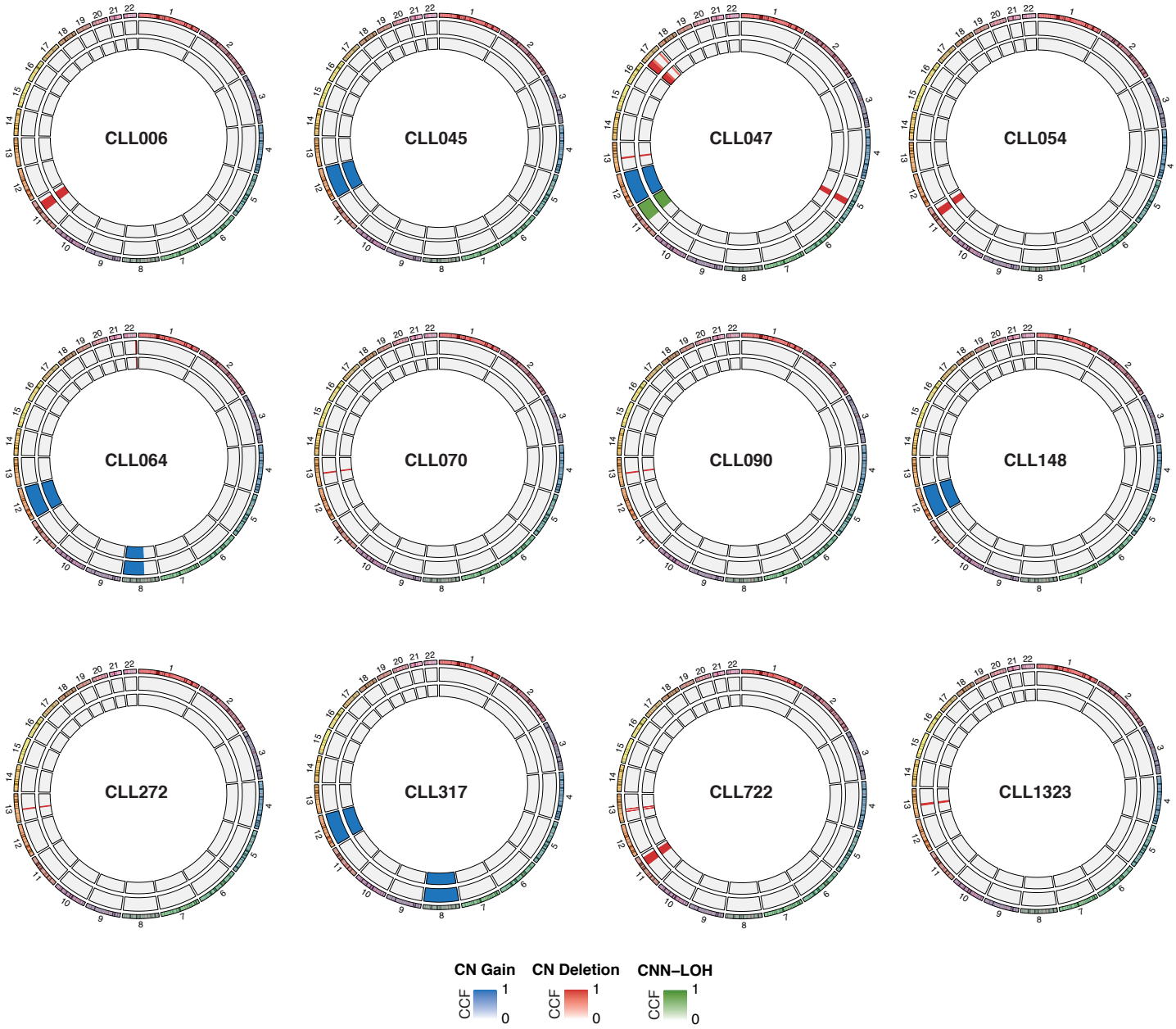
CLL006
CLL045
CLL047
CLL054
CLL064
CLL070
CLL090
CLL148
CLL272
CLL317
CLL722
CLL1323
CN Gain
CN Deletion
CNN−LOH
CCF
1
0

Figure S2

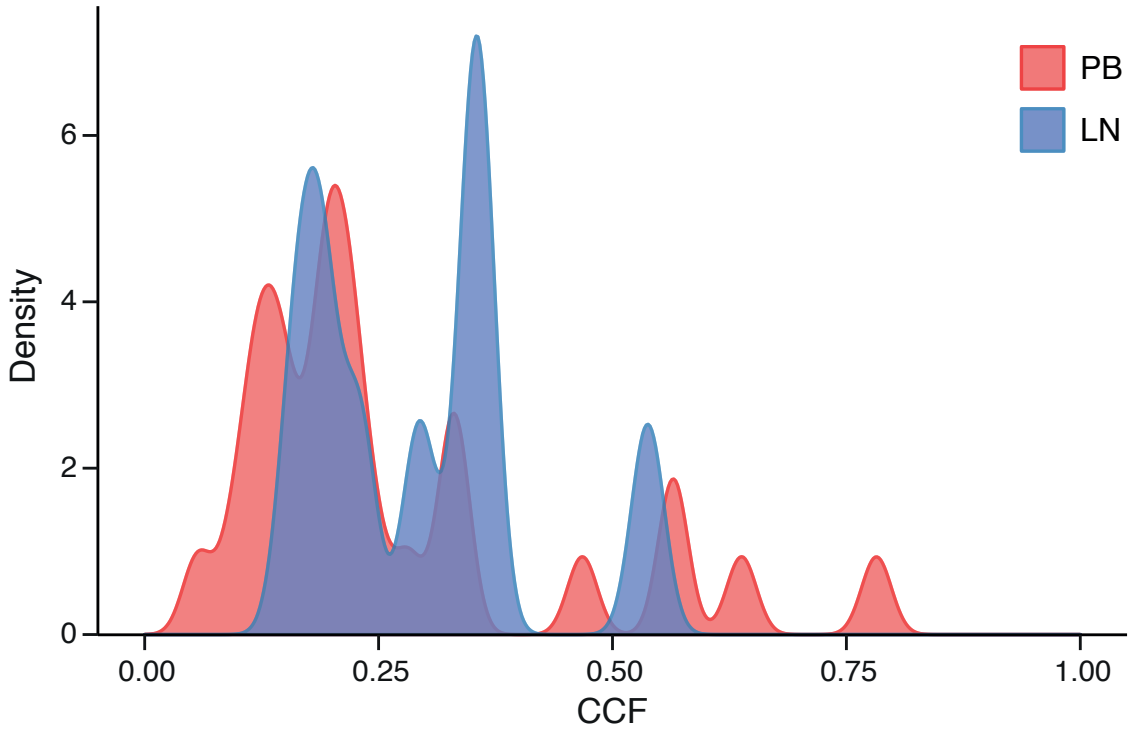

Figure S3

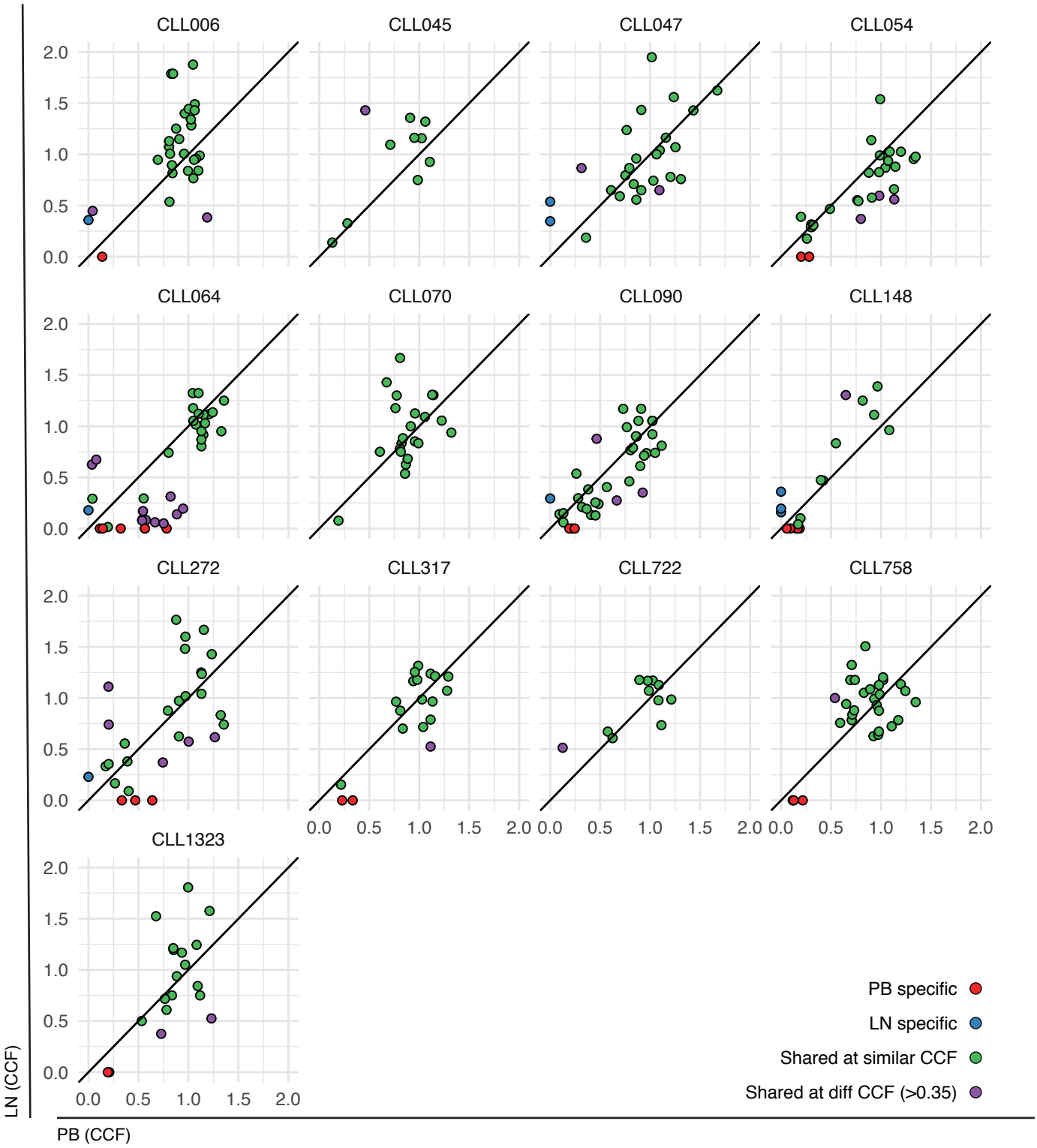
CLL006
CLL045
CLL047
CLL054
CLL064
CLL070
CLL090
CLL148
CLL272
CLL317
CLL722
CLL758
CLL1323
0.0
0.5
1.0
1.5
2.0
LN (CCF)
PB (CCF)
PB specific
LN specific
Shared at similar CCF
Shared at diff CCF (>0.35)

Figure S4

Supplementary References

1 International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C *et al.* International network of cancer genome projects. *Nature* 2010; **464**: 993–998.

2 Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JI *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2015; **526**: 519–524.

3 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60.

4 Jones D, Raine KM, Davies H, Tarpey PS, Butler AP, Teague JW *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr Protoc Bioinforma* 2016; **56**: 15.10.1-15.10.18.

5 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.

6 Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal PA *et al.* MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol* 2016; **17**: 178.

7 Maura F, Degasperi A, Nadeu F, Leongamornlert D, Davies H, Moore L *et al.* A practical guide for mutational signature analysis in hematological malignancies. *Nat Commun* 2019; **10**: 2969.

8 Moncunill V, Gonzalez S, Beà S, Andrieux LO, Salaverria I, Royo C *et al.* Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat Biotechnol* 2014; **32**: 1106–1112.

9 Raine KM, Hinton J, Butler AP, Teague JW, Davies H, Tarpey P *et al.* cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *Curr Protoc Bioinforma* 2015; **52**: 15.7.1–12.

10 Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet* 2014; **46**: 912–918.

11 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The Sequence

Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9.

12 Morgan M, Pagès H, Obenchain V HN. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. 2019.http://bioconductor.org/packages/Rsamtools.

13 Raine KM, Van Loo P, Wedge DC, Jones D, Menzies A, Butler AP *et al.* ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data. *Curr Protoc Bioinforma* 2016; **56**: 15.9.1-15.9.17.

14 Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW *et al.* The life history of 21 breast cancers. *Cell* 2012; **149**: 994–1007.

15 Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 2016; **44**: e131–e131.

16 Nadeu F, Clot G, Delgado J, Martín-García D, Baumann T, Salaverria I *et al.* Clinical impact of the subclonal architecture and mutational complexity in chronic lymphocytic leukemia. *Leukemia* 2018; **32**: 645–653.

17 Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016; **534**: 47–54.

18 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011; **29**: 24–26.

19 Maura F, Bolli N, Angelopoulos N, Dawson KJ, Leongamornlert D, Martincorena I *et al.* Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nat Commun* 2019; **10**: 3835.

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