UNIVERSITÀ DEL PIEMONTE ORIENTALE

SCUOLA DI MEDICINA

Dipartimento di Medicina Traslazionale

PhD Program in Sciences and Medical Biotechnology

XXXI CYCLE

DIFFUSE LARGE B-CELL LYMPHOMA

GENOTYPING ON THE LIQUID BIOPSY

Tutor:

Chiar.mo Prof. Gianluca Gaidano

Coordinator:

Prof.ssa Marisa Gariglio

Candidate: Dr Fary Diop

Matricula: 20005154

Accademic year 2017/2018

INDEX

SUMMARY

Accessible and real-time genotyping for diagnostic, prognostic, or treatment purposes is increasingly impelling in diffuse large B-cell lymphoma (DLBCL). Cell-free DNA (cfDNA) is shed into the blood by tumor cells undergoing apoptosis and can be used as source of tumor DNA for the identification of DLBCL mutations, clonal evolution, and genetic mechanisms of resistance. In this study, we aimed at tracking the basal DLBCL genetic profile and its modification upon treatment using plasma cfDNA. Ultra-deep targeted next generation sequencing of pretreatment plasma cfDNA from DLBCL patients discovered DLBCL-associated mutations that were represented in >20% of the alleles of the tumor biopsy with >90% sensitivity and∼100% specificity. Plasma cfDNA genotyping also allowed for the recovery of mutations that were undetectable in the tissue biopsy, conceivably because, they were restricted to clones that were anatomically distant from the biopsy site.

Longitudinal analysis of plasma samples collected under (R-CHOP) chemotherapy showed a rapid clearance of DLBCL mutations from cfDNA among responding patients. Conversely, among patients who were resistant to R-CHOP, basal DLBCL mutations did not disappear from cfDNA and, moreover, among treatment-resistant patients, new mutations were acquired in cfDNA that marked resistant clones selected during the clonal evolution. These results demonstrate that cfDNA genotyping of DLBCL is as accurate as genotyping of the diagnostic biopsy to detect clonally

represented somatic tumor mutations and is a real-time and noninvasive approach to tracking clonal evolution and the emergence of treatment-resistant clones.

SOMMARIO

Il DNA libero circolante (cell-free DNA, cfDNA) viene rilasciato nel sangue dalle cellule tumorali che vanno incontro a meccanismi di apoptosi o necrosi, e può essere utilizzato come fonte di DNA tumorale per l'identificazione delle mutazioni nel linfoma diffuso a grandi cellule B (DLBCL), dell'evoluzione clonale e dei meccanismi genetici di resistenza. In questo studio, abbiamo mirato a rintracciare il profilo genetico alla diagnosi del DLBCL e la sua modifica dopo trattamento, utilizzando il cfDNA estratto da plasma. Il cfDNA e il DNA genomico ottenuto dalla biopsia tessutale di 50 pazienti affetti da DLBCL sono stati sottoposti a ultra-deep next generation sequencing, rilevando mutazioni tipicamente associate al DLBCL e rappresentate in> 20% degli alleli della biopsia tumorale, con una sensibilità> 90% e una specificità del 100%. La genotipizzazione del cfDNA al plasma ha permesso di identificare mutazioni che non erano rilevabili nella biopsia tissutale, presumibilmente perché, a causa dell'eterogeneità spaziale del tumore, erano limitate ai cloni che erano anatomicamente distanti dal sito della biopsia. L'analisi longitudinale dei campioni di plasma raccolti durante la chemioterapia con Rituximab-ciclofosfamide-doxorubicina-vincristina-prednisone (R-CHOP) ha mostrato una rapida scomparsa delle mutazioni dal cfDNA tra i pazienti rispondenti al trattamento. Al contrario, nel cfDNA dei pazienti resistenti a R-CHOP, non solo le mutazioni identificate alla diagnosi non scomparivano dal cfDNA, ma emergevano nuove mutazioni acquisite, indicando la presenza di cloni resistenti selezionati durante l'evoluzione clonale.

1. INTRODUCTION

1.1. Liquid biopsy and cell-free DNA

Rarity of neoplastic cells in the biopsy imposes major technical hurdles that prevented large genomic studies in aggressive lymphomas, such as diffuse large B-cell lymphoma (DLBCL). Limitations in accessing fresh tumor material from DLBCL tissue biopsies has prevented the rapid translation of DLBCL gene mutations into prognostic or predictive tools for the clinical practice. Also, serial sampling of tumors to track the acquisition of drug-resistance mutations requires a re-biopsy, which may not be routinely feasible in the clinical practice. Therefore, alternative accessible sources of tumor DNA may help to complement the molecular diagnostic analyses that are routinely carried out on formalin-fixed paraffin-embedded (FFPE) tissue biopsies.[1] Liquid biopsy-based genotyping from serum/plasma or other body fluids coupled with advanced molecular technologies can provide non-invasive method for genetic analyses. These samples contain cell-free nucleic acids (cfNAs) which are valuable markers in different diagnostic protocols.[2] Cell-free fragments of DNA (cfDNA) are shed into the bloodstream by tumor cells undergoing apoptosis.[3] Accessing tumor cfDNA through the bloodstream has clear sampling advantages and allows serial monitoring of disease genetics in real time. cfDNA is also representative of the entire tumor heterogeneity, thus enabling to bypass the anatomical biases

imposed by tissue biopsies in the reconstruction of the entire cancer clonal architecture, and to identify resistant clones that are dormant in non-accessible tumor sites.[3]

In DLBCL, cfDNA has been quantified or used to track the tumor clonotypic immunoglobulin gene rearrangement for minimal residual disease monitoring.[4-7] Conversely, the evidence that cfDNA mirrors the underlying tumor genetics in DLBCL and the proof that cfDNA can be used to track in real time clonal evolution-driven resistance upon treatment are currently limited to retrospective series.[8,9]

1.2. Diffuse large B-cell lymphoma (DLBCL)

DLBCL is the most common B cell non-Hodgkin lymphoma (B-NHL) in the adult, is an aggressive disease that remains incurable in approximately 30% of patients.[10] DLBCL arises from the clonal expansion of B cells in the germinal center (GC), a specialized microenvironment that forms in secondary lymphoid organs upon encounter of a naïve B cell with its cognate antigen, in the context of T-cell dependent co-stimulation.[10]

Using gene expression profiling (GEP), two molecularly distinct forms of DLBCL were identified, resembling either germinal center B-cells (GCB) or activated B-cells (ABC), while additional 15-30% of cases remain unclassified.[11] These two forms were morphologically overlapping but had gene expression patterns indicative of different stages of B-cell differentiation.[11]

Consistent with their putative "cell-of-origin" (COO), GCB DLBCLs display high-level expression of the master regulator *BCL6* and harbor hypermutated immunoglobulin genes with ongoing somatic hypermutation, whereas ABC DLBCLs show activation of NF-κB and BCR signaling pathways, and upregulation of genes required for plasmacytic differentiation.[12] The COO classification has been shown to identify distinct DLBCL prognostic subgroups, with GCB DLBCLs being associated with a significantly better outcome as compared with the ABC subgroup.[13,14]

1.3. Genetic lesions common to GCB and ABC DLBCL

A large number of genetic lesion are involved in DLBCL pathogenesis, and some are commonly shared among GCB and ABC DLBCL. In particular, one commonly disrupted program in DLBCL is represented by epigenetic remodeling. The genetic lesions include *i)* alterations of histone modification genes; *ii)* alterations deregulating *BCL6*; *iii)* loss of immune surveillance mechanisms, and *iv)* other lesions.

i) Alterations of histone modification genes: up to 30% of cases, with some preference for GCB-DLBCL, harbor mutations and/or deletions inactivating CREB- Binding Protein (*CREBBP*) and, more rarely, E1A Binding Protein P300 (*EP300*), two ubiquitously expressed acetyltransferases that modify lysine residues on both histone and non-histone nuclear proteins, modulating the activity of a large number of DNA-binding transcription factors.[15] *CREBBP* mutations include truncating events that

remove the C-terminal HAT domain and amino acid changes that impair its affinity for Acetyl-CoA, severely reducing its enzymatic activity.

At least one third of DLBCLs feature mutations in the Lysine Methyltransferase 2D (*KMT2D*) gene. *KMT2D* encodes for a methyltrasferase that controls epigenetic transcriptional regulation by mono-, di- and trimethylating the lysine 4 position of histone 3 (H3K4). While the consequences of *KMT2D* mutations in DLBCL have not been elucidated yet, most events are predicted to generate severely truncated proteins lacking the catalytic SET domain, which is required for its methyltransferase activity.[10]

ii) Alterations deregulating BCL6: chromosomal rearrangements of the *BCL6* (B-cell Lymphoma 6 protein) locus characterize as many as 35% of DLBCL patients, although with two- to three-fold higher frequencies in ABC-DLBCLs. These balanced, reciprocal recombination events juxtapose the coding domain of *BCL6* downstream to heterologous promoters derived from alternative chromosomal partners, leading to deregulated expression of an intact protein, in part by preventing its downregulation during post-GC differentiation.[10,16]

In addition to genetic lesions directly affecting the *BCL6* gene, DLBCL have devised a number of ways to deregulate the *BCL6* function indirectly. Indeed, about 10-15% of DLBCL patients harbor gain-of-function somatic mutations in the Myocyte Enhancer Binding Factor 2B (*MEF2B*) transcription factor, a protein highly expressed in the GC and involved in *BCL6* transcriptional activation.[17,18]

Moreover, in 5% of DLBCL cases, loss-of-function mutations/deletions of F-Box Protein 11 (*FBXO11*) impair proteasomal-mediated degradation of the BCL6 protein, which is controlled by this E3 ubiquitin ligase.[19]

iii) Loss of immune surveillance mechanisms: in 29% of cases, the beta-2-microglobulin (*B2M*) gene is lost because of structurally disruptive mutations and/or deletions, and another 30% of cases lack *B2M* expression in the absence of genetic lesions, suggesting the existence of additional genetic or epigenetic mechanisms of inactivation.[20] *B2M* encodes for an invariant subunit of the HLA class I (HLA-I) complex, which is expressed on the surface of all nucleated cells and is required for recognition by cytotoxic T lymphocytes. As a result, over 60% of DLBCL lack surface HLA-I expression, which in turn may favor lymphomagenesis by allowing evasion from immune surveillance.[21]

iv) Other lesions: Mutations and deletions of *TP53* (Tumor Protein 53) remain an important pathogenic lesion in ~20% of all DLBCL.[22,23] In particular, *TP53* mutations affecting its DNA binding domain are most important from a prognostic standpoint.[24] Also shared across both DLBCL subtypes are mutations of the Forkhead box protein O1 (*FOXO1*) transcription factor. These events cluster around a phosphorylation site required for AKT-mediated nuclear-cytoplasmic translocation and inactivation of *FOXO1*, and were suggested to enhance its activity by preventing its nuclear export following PI3K signaling.[25]

1.4. Genetic lesions in GCB DLBCL

Few lesions had been found preferentially associated with GCB-DLBCL, including *i)* chromosomal translocations of *BCL2* and *MYC*; *ii) PTEN* alterations; *iii)* mutations of *EZH2* and, *iv)* mutations in the Gα13 pathway.

i) Chromosomal translocations of BCL2 and MYC: BCL2 (B-cell lymphoma 2) is deregulated in diffuse large B-cell lymphoma. The t(14;18) translocation causes constitutive overexpression of *BCL2* by juxtaposing it to immunoglobulin heavy chain gene enhancer elements. This translocation is found in about 35-45% of GCB cases.[26] *MYC* rearrangements have been detected in approximately 5% to 14% of DLBCL and these alterations are frequently associated with *BCL2* or *BCL6* rearrangements.[10,27] Both *BCL2* and *MYC* translocations lead to ectopic expression of the involved protein, in part by allowing escape from *BCL6*-mediated transcriptional repression.[10]

ii) PTEN alterations: Phosphatase and tensin homolog (*PTEN*) is a tumor suppressor gene that can resist the function of PI3K and negatively regulate AKT activity.[28] *PTEN* deletions (mostly heterozygous) are detected in 11.3% of DLBCL, and showed opposite prognostic effects in patients with AKT hyperactivation and in *MYC* rearranged DLBCL patients.[29] *PTEN* mutations, detected in 10.6% of patients, are associated with upregulation of genes involved in central nervous system function, metabolism, and AKT/mTOR signaling regulation. Multi-levels of *PTEN* abnormalities and dysregulation

play important roles in *PTEN* expression and loss, and that loss of *PTEN* tumor-suppressor function contributes to the poor survival of DLBCL patients with AKT hyperactivation.[29]

iii) Mutations of EZH2: Histone methyltransferase enhancer of Zeste homolog 2 (*EZH2*) encodes a subunit of polycomb-repressive complex 2 (*PRC2*) and is responsible for the trimethylation of lysine 27 of histone H3 (H3K27me3).[30] The *EZH2* gene is mutated in about 23% of GCB-DLBCLs. The most frequent mutation is a missense affecting the amino acid tyrosine at position 641 (p.Y641), which encodes the catalytic site of the SET domain, resulting in a mutant protein that acts synergistically with the wild-type enzyme to increase histone H3 trimethylation.[31,32] Increased *EZH2* expression of tumor cells is associated with better prognosis in GCB DLBCL.[33]

iv) Mutations in the Gα13 pathway: Deep sequencing studies of GCB-DLBCL have revealed mutations in G Protein Subunit Alpha 13 (*GNA13*),[34] that encodes for Gα13. Approximately 20% of GCB-DLBCLs are characterized by structurally damaging mutations in various components of a G-protein coupled inhibitory circuit that regulates the growth and local confinement of GCB cells.[10] Mutations in the Gα13 signaling pathway in DLBCL result in loss of function, and that restoration and/or activation of this signaling pathway help reduce tumor growth and progression.[35]

1.5. Genetic lesions in ABC DLBCL

The genomic landscape of ABC-DLBCL is associated with genetic lesions leading to constitutive activation of the NF-κB transcription factor,[36] and it involves *i)* Mutations activating the BCR signaling pathway; *ii)* Mutations activating the Toll-Like Receptor (TLR) pathway and, *iii)* Mutations inactivating negative regulators of NF-κB.

i) Mutations activating the BCR signaling pathway: ABC-DLBCL cells were found to display a chronic, active form of BCR (B-cell receptor) signaling.[37] Over 20% of patients harbor somatic mutations in the Ig superfamily members *CD79B* (Cluster of Differentiation 79 B) and, at lower frequencies, *CD79A*.[10,37,38] In most cases, the mutations replace the first tyrosine residue (p.Y196) in the cytoplasmic immunoreceptor tyrosine-based activation motifs (ITAMs).

These events are thought to circumvent negative feedback circuits that attenuate BCR signaling, thus maintaining it chronically active. In ~9% of ABC-DLBCL, activation of BCR and NF-κB is sustained by oncogenic mutations of the Caspase Recruitment Domain Family Member 11 (*CARD11*) gene. *CARD11* is a major component of the "signalosome" complex, the coordinated recruitment of which is required for proper transduction of BCR signaling.[39] These events cluster in the exons encoding for the protein coiled-coil domain and enhance the ability of *CARD11* to transactivate NF-κB target genes.[37,39]

ii) Mutations activating the Toll-Like Receptor (TLR) pathway: MYD88 (Myeloid differentiation primary response 88) mutations are found in one third of ABC-DLBCLs, where they target an invariant

residue within the TIR (Toll/IL1 receptor) domain, leading to a L265P substitution.[38,40] This mutation induces *IRAK4* (Interleukin-1 receptor-associated kinase 4) kinase activity and phosphorylation through the spontaneous assembly of a protein complex containing IRAK1 and IRAK4, which in turn can activate NF-κB and JAK/STAT3 transcriptional responses.[41]

iii) Mutations inactivating negative regulators of NF-κB: Almost one third of ABC-DLBCL harbor biallelic *TNFAIP3* (TNF Alpha Induced Protein 3) truncating mutations and/or deletions.[42] *TNFAIP3* encodes for a dual function ubiquitin-modification enzyme involved in the termination of NF-κB responses triggered by TLR and BCR stimulation. *TNFAIP3* mutations are thought to induce inappropriately prolonged NF-κB responses.[42,43]

1.6. DLBCL treatment

At present, the combination of rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone (R-CHOP) is the gold standard treatment for DLBCL.[44] Rituximab is an antibody directed against the CD20 protein, which is primarily found on the surface of B cells and is present on many lymphoma cells,[45] while cyclophosphamide, doxorubicin, vincristine and prednisone are chemotherapy agents.[46] About 50% to 70% of patients may be cured by R-CHOP chemotherapy. Nevertheless, R-CHOP is found to be inadequate in 30% to 40% of patients.[47] For these patients, different processes may account for their lack of response to R-CHOP. Death related to R-CHOP toxicities, although it is a rare

event in young patients, may be observed in 5% of patients older than age 70 years. R-CHOP failures are principally due to either primary refractoriness or relapse after reaching a complete response (CR).[48]

Mutations of clinical importance in DLBCL affect the *TP53* gene, whose variants are consistently associated with poor prognosis among patients treated with R-CHOP,[49] and the *CARD11, CD79A, CD79B*, and *MYD88* genes, whose variants predict the benefit or no benefit from ibrutinib, the inhibitor of Bruton's tyrosine kinase (BTK), monotherapy.[50]

2. AIM OF THE STUDY

In this study we aim to demonstrate that cfDNA genotyping of DLBCL:

i) Is as accurate as genotyping of the diagnostic biopsy to detect somatic mutations of allelic abundance >20%;

ii) Is a non-invasive tool to track the emergence of treatment-resistant clones;

3. METHODS

3.1. Patients

The study had a prospective, observational, non-intervention, uni-centered design and consisted in the longitudinal collection of peripheral blood (PB) samples and clinical data from DLBCL patients treated with R-CHOP at the University of Eastern Piedmont.

Inclusion criteria were: *i)* male or female adults ≥18 years; *ii)* diagnosis of untreated DLBCL after pathological revision; *iii)* treatment with R-CHOP; *iv)* evidence of a signed informed consent. A total of 50 previously untreated DLBCL patients fulfilled the inclusion criteria were recruited in the study from November 2013 to August 2015 as training series (n=30) and from September 2016 to October 2016 as validation cohort (n=20) (Table 1). All patients received R-CHOP treatment.

The following biological material was collected: *i)* cfDNA isolated from plasma at diagnosis before treatment start, during R-CHOP courses (on day 1 of each course before treatment infusion), at the end of treatment and at progression, and *ii)* normal germline genomic DNA (gDNA) extracted from PB granulocytes after Ficoll separation.

For comparative purposes, tumor gDNA from the paired DLBCL diagnostic tissue biopsy was available for 36 patients (extracted from fresh specimens in 25 cases and from FFPE specimens in 11 cases). In the remaining 14 cases, the leftover of the FFPE diagnostic biopsy was not available or gave insufficient amount of DNA for CAPP-seq. In all instances, basal plasma samples were collected in close

temporal proximity of the tumor tissue biopsy (7-14 days after diagnostic tissue biopsy) and before starting treatment.

Clinical information were prospectively maintained in the University of Eastern Piedmont lymphoma database. Disease response was assessed by PET/CT.[51] Patients provided informed consent in accordance with local IRB requirements and Declaration of Helsinki.

Paired plasma cfDNA and normal gDNA from granulocytes collected from 6 healthy donors were used to set the experimental and biological background of the ultra-deep next generation sequencing (NGS) approach. The study was approved by the Ethical Committee of the Ospedale Maggiore della Carità di Novara affiliated with the University of Eastern Piedmont (Protocol Code CE 112/15).

3.2. Genomic DNA (gDNA) extraction

PB granulocytes were separated by Ficoll gradient density centrifugation as source of normal germline gDNA. Tumor gDNA was isolated from the fresh or FFPE diagnostic tissue biopsies containing >70% of tumor cells as estimated by morphology and immunohistochemistry.

Tumor and normal gDNA were extracted by using the “salting out” protocol.[52] PB was diluted 1:2 with physiological solution (NaCl 0.9%) and then centrifuged in a gradient differentiation Sigma Diagnostic™ Histopaque®-1077 Cell Separation Medium (Sigma-Aldrich, St. Louis, MO, USA) solution

to obtain mononuclear white blood cells and granulocytes. Cells were lysed with Lysis Buffer (Tris-HCl 1M, pH 8.2, NaCl 5M, EDTA 0.5M), SDS 20% and digested with pronase E (20 mg/mL).

Samples were incubated at 37°C overnight in a shaking incubator. Proteins were precipitated with 6M NaCl, and subsequently discarded after centrifugation at 3200 rpm for 20 minutes. DNA was isolated by precipitation with pure ethanol and washed three times with 75% ethanol. The excess ethanol was evaporated, and the DNA was dissolved with TE buffer (Tris-HCl 1M, pH 8.2 and EDTA 0.5M).

3.3. Plasma cfDNA extraction

30 ml of PB samples were collected in EDTA tubes and centrifuged at 820 g for 10 min to separate plasma from cells within one hour from collection. Plasma was then further centrifuged at 13000 g for 10 min to pellet and remove any remaining cells and stored at -80°C until DNA extraction. cfDNA was extracted from 1 ml aliquots of plasma immediately after thawing by using the QIAamp circulating nucleic acid kit (Qiagen, Hilden, Germany) and quantified using a SYBR green-based real-time qPCR assay for the β-globin gene (180-bp amplicon) carried out on StepOnePlus Real-Time PCR System (Step One software 2.0; Applied Biosystems, Foster City, CA, USA).

3.4. Library design for hybrid selection

A targeted resequencing gene panel including coding exons and splice sites of 59 genes (target region: 207299bp) that are recurrently mutated in DLBCL and other mature B-cell tumors has been specifically designed for this project. The gene panel allowed a priori the recovery of at least one clonal mutation in 92.6% (95% CI: 83.6-97.5%) of DLBCL patients, as documented by *in silico* validation against public genomic datasets of DLBCL.[53-56]

3.5. CAPP-seq library preparation and ultra-deep next generation sequencing

In the training cohort (n=30), the gene panel was analyzed in plasma cfDNA collected at diagnosis, during R-CHOP courses, at the end of treatment, and at progression (total cfDNA samples=127), and, for comparative purposes to filter out polymorphisms, in germline gDNA from the paired granulocytes. The tumor gDNA from the paired tissue biopsy was also investigated to test the accuracy of cfDNA genotyping.

In the validation cohort (n=20), the gene panel was investigated in plasma cfDNA collected at diagnosis, and in normal and tumor gDNA from paired granulocytes and tissue biopsy, respectively. Tumor and germline gDNA from tissues (median=318 ng) were sheared through sonication (Covaris M220 focused-ultrasonicator, Woburn, MA, USA) before library construction to obtain 200-bp fragments. For plasma cfDNA, which is naturally fragmented, 2-717 ng (median=17 ng) of DNA were used for library construction without additional fragmentation.

The next generation sequencing (NGS) libraries were constructed using the KAPA Library Preparation Kit (Kapa Biosystems, Wilmington, MA, USA), and hybrid selection was performed with the custom SeqCap EZ Choice Library (Roche NimbleGen, Madison, WI, USA). Multiplexed libraries (n=6 per run) were sequenced using 300-bp paired-end runs on a MiSeq sequencer (Illumina, San Diego, CA, USA).

3.6. Bioinformatic pipeline for variant calling

Non-synonymous somatic mutation calling in plasma cfDNA was performed separately and in blind from mutation calling in tumor gDNA. After CAPP-seq, FASTQ sequencing reads were initially deduped through FastUniq v1.1. Then, the deduped FASTQ sequencing reads were locally aligned to the hg19 version of the human genome assembly using BWA v.0.6.2 and assembled into a mpileup file using SAMtools v.1. Single nucleotide variations and indels were called in plasma cfDNA vs germline gDNA, and tumor gDNA vs germline gDNA, respectively, with the somatic function of VarScan2.

Variants annotated as SNPs according to dbSNP 138 (with the exception of *TP53* variants that were manually curated and scored as SNPs according to the IARC TP53 database), intronic variants mapping >2 bp before the start or after the end of coding exons, and synonymous variants were filtered out. Two independent statistical approaches (Fisher's exact test and Z-test) were then used to filter out variants below the base-pair resolution background frequencies in cfDNA across the target

region. Only variants that had a significant call in both test were retained (Bonferroni adjusted test $p<6x10^{-8}$).

3.7. Sanger sequencing

Sanger sequencing was also used to validate the most abundant plasma cfDNA mutations detected by CAPP-seq. PCR primers were designed in the Primer 3 program (http://frodo.wi.mit.edu/primer3/). Purified amplicons were subjected to Sanger sequencing and compared to the corresponding germline sequences using the Mutation Surveyor Version 4.0.8 software package (SoftGenetics, State College, PA, USA; http://www.softgenetics.com) after automated and/or manual curation. cfDNA from plasma were subjected to Sanger sequencing and compared to the corresponding germline. Mutations were then confirmed from both strands on independent PCR products.

3.8. Statistical analysis

Sensitivity and specificity of plasma cfDNA genotyping were calculated in comparison with tumor gDNA genotyping as the gold standard. The analysis was performed with SPSS v.22.0

4. RESULTS

4.1. Characteristics of the study cohorts

The study was based on a prospectively collected consecutive series of 30 newly diagnosed DLBCL patients (training cohort) whose characteristics were consistent with an unselected cohort of DLBCL (Table 1). Upon R-CHOP treatment, 83.3% (95% CI: 65.9-93.1%; n=25/30) of patients achieved a PET/CT negative complete remission, while 16.7% (95% CI: 6.8-34.0%; n=5/30) failed to achieve a complete remission. Among patients that achieved complete remission (median follow-up 6 months), two relapsed. The median number of cfDNA molecules per ml of plasma at disease presentation was 771.7 (range: 137.2-18742.5). An independent validation series of 20 consecutive DLBCL patients was also assessed to confirm the accuracy of plasma cfDNA genotyping (Table 1).

4.2. Plasma cfDNA genotyping discloses somatic mutations in DLBCL-associated genes

To provide the proof of principle that plasma could function as a liquid biopsy for tracking recurrently mutated genes in DLBCL, plasma cfDNA collected at presentation from the training cohort was genotyped by using CAPP-seq, a targeted ultra-deep NGS approach for plasma cfDNA genotyping already validated in solid tumors. In our cohort, 80% or more of the target region covered >1000x in all DLBCL and 80% or more of the target region covered >2000x in 17/30 DLBCL.[57] Paired normal gDNA was also analyzed to confirm the somatic origin of mutations. Overall, within the interrogated genes,

66.6% (95% CI: 47.1-82.7%; n=20/30) of patients harbored somatic mutations (total number 129; range: 2-13 mutations per patient) that were detectable in plasma cfDNA (Table 2).

In order to validate the NGS results, cfDNA was extracted from a second aliquot of plasma, and then subjected to a second CAPP-seq and ultra-deep-NGS sequencing. Robustness of the plasma cfDNA CAPP-seq approach and of the bioinformatics analysis was documented by the high concordance (R^2 0.916) of variant calling from the independent duplicate experiments (Figure S1), which consistently confirmed all the variants initially discovered in plasma cfDNA, including those of low allele frequency, thus excluding their origin from a batch-specific experimental noise.

Sanger sequencing consistently detected all plasma cfDNA mutations showing a representation within the sensitivity range of this approach (allele frequency >10%) (Figure S2), thus validating the CAPP-seq results on a different experimental platform.

Consistent with the typical spectrum of mutated genes in DLBCL, plasma cfDNA genotyping discovered somatic variants of *KMT2D* in 30.0% (95% CI: 14.7-49.4%; n=9/30) of cases, *TP53* in 23.3% (95% CI: 9.9-42.2%; n=7/30), *CREBBP* in 20.0% (95% CI: 7.7-38.5%; n=6/30), *PIM1* and *TNFAIP3* in 16.6% (95% CI: 5.6-34.7%; n=5/30), *EZH2*, *STAT6*, and *TBL1XR1* in 13.3% (95% CI: 3.7-30.7%; n=4/30), *B2M*, *BCL2*, *CARD11*, *CCND3* and *FBXW7* in 10.0% (95% CI: 2.1-26.5%; n=3/30), *CD58*, *CD79B* and *MYC* in 6.6% (95% CI: 0.8-22.0%; n=2/30), *EP300*, *GNA13*, *MEF2B*, *MYD88* and *TNFRSF14* in 3.3% (95% CI: 0-17.2%; n=1/30) (Figure 1A). Though the sample size was unpowered to show significant enrichments

of mutated genes within cell of origin categories of DLBCL, *EZH2* and *BCL2* mutations were as expected more frequent in germinal-center (GC) DLBCL, while *TNFAIP3* and *PIM1* mutations were more frequent in non-GC DLBCL (Figure 1A).[10]

Among recurrently affected genes, the molecular spectrum of mutations identified in plasma cfDNA was highly consistent with that of variants that have been detected in tumor gDNA of published DLBCL series and reported in the COSMIC database (Figure 1B and 1C). *PIM1* and *MYC* were affected by multiple hotspot mutations that were suggestive of AID (Activation-induced cytidine deaminase)-related events (Figure 1C; Table 2).[58] Notably, patient ID23, who had the t(8;14) *MYC* translocation, also harbored multiple *MYC* mutations in cfDNA, consistent with the known accumulation of mutations in the translocated *MYC* gene due to the IGH enhancer-driven misfire of somatic hypermutation (Table 2).[59]

Analysis of 6 healthy donors by CAPP-seq did not disclose any somatic mutations in plasma cfDNA (Figure S3), suggesting that the ultra-deep NGS and variant calling approaches used in this study did not pick up biological or analytical background noises in cfDNA.

4.3. Validation of biopsy-free DLBCL genotyping

The fresh or FFPE tissue biopsy of 20 DLBCL patients from the training cohort was genotyped by CAPP-seq in blind of the mutational profile recovered in the paired plasma cfDNA. To validate our

sequencing approach of FFPE samples, gDNA from paired fresh/FFPE samples processed from the same DLBCL biopsy (n=4) was subjected to CAPP-seq. Pairwise analysis of data showed high (96%) concordance in variant recovery from FFPE vs fresh samples (Figure S4).

In order to systematically derive the accuracy of cfDNA genotyping, the results of plasma cfDNA genotyping and tumor gDNA genotyping (gold standard; mutation spectrum shown in Figure 2A) were then compared (Figures 3A-C) in 18 DLBCL training cases provided with the paired plasma/tissue samples and informative because mutated within the target region. Sequencing coverage of the tumor gDNA was comparable to that of plasma cfDNA (Figure S5).

Genotyping of plasma cfDNA collected at diagnosis identified a total of 108 somatic mutations, while genotyping of the gDNA from the diagnostic tissue biopsy identified 105 somatic mutations. Biopsy-confirmed tumor mutations were detectable with 82.8% (n=87/105, 95% CI: 74.4-88.9%) sensitivity in pretreatment plasma cfDNA samples (Figure 3B). Biopsy-confirmed tumor mutations not discovered in the cfDNA (n=18/105) generally had a low representation in the tissue diagnostic biopsy (median allele frequency in the tumor biopsy=5.7%) (Figure 4A and 4B). Consistently, by ROC analysis, cfDNA genotyping showed the highest sensitivity (97.1%; 95% CI: 89.5-99.8%; n=68/70) in discovering mutations that were represented in $\geq$20% of the alleles of the tumor biopsy (Figure 4B), thus demonstrating that plasma cfDNA can accurately mirror the profiles of the most abundant clones found in tumor tissues. Plasma cfDNA was uninformative on tumor biopsy mutational

status in 10% (95% CI: 1.5-31.3%; n=2/20) of DLBCL cases (Figure 3C). The representation of the variants in plasma correlated with the LDH levels at DLBCL diagnosis, but not with disease stage (Figure S6).

Plasma cfDNA genotyping disclosed additional 21 somatic mutations that were not detectable in the tissue biopsy, including 7 variants affecting *PIM1*, a known target of somatic hypermutation (Figure 2A and 3; Table S4).[58] Repeated ultra-deep NGS consistently confirmed these variants, thus excluding their origin from a batch-specific experimental noise (Figure S2). Because the target region that has been sequenced in the compiled 20 DLBCL patients might support a total of 18,011,440 potential non-synonymous variants if they were randomly distributed, and considering that only 21 non-synonymous mutations were detected in plasma cfDNA but not in the tumor biopsy, the false positive rate of plasma cfDNA genotyping was 1.16×10^{-6} corresponding to a specificity >99.99% compared to tumor gDNA genotyping.

As previously reported in other cancer types,[60] non-synonymous mutations occurring only in plasma cfDNA conceivably represented tumor mutations restricted to clones that were anatomically distant from the biopsy site, rather than false positive calls. Though the study lacked a systematic multiregional sequencing of tumor samples from multiple anatomical sites, in support of the above interpretation are the observations that: *i)* the plasma cfDNA *FBXW7* p.S668G mutation of patient ID9, though undetectable in the diagnostic tissue biopsy, was identified in tumor cells from the cytospin of

the cerebrospinal fluid collected at the time of isolated meningeal relapse (Figures 5 and 7); and *ii)* all mutations lacking in the diagnostic tissue biopsy disappeared from cfDNA upon achieving complete remission of DLBCL.

Plasma cfDNA was genotyped by CAPP-seq in an independent cohort of 20 consecutive DLBCL. Overall, within the interrogated genes, 85.0% (95% CI: 63.1-95.6%; n=17/20) of validation patients harbored somatic mutations (total number 83; range: 1-12 mutations per patient) that were detectable in plasma cfDNA (Figure 1D; Table S4). The molecular spectrum of mutations identified in plasma cfDNA of the validation series was highly consistent with that observed in the training series and more in general with that of published DLBCL cohorts (Figure 1E and F). CAPP-seq of paired tumor and plasma samples from 16 DLBCL of the validation series confirmed the sensitivity of plasma cfDNA genotyping in recovering biopsy-confirmed mutations (Figure 3 D-E). Biopsy-confirmed tumor mutations were detectable with 82.8% (n=58/70, 95% CI: 72.2-90.0%) sensitivity in pretreatment plasma cfDNA samples. Most of the biopsy-confirmed tumor mutations not discovered in the cfDNA had a low representation in the tissue diagnostic biopsy <20% (Figure 4C). Consistently, by ROC analysis, cfDNA genotyping showed the highest sensitivity (91.3%; 95% CI: 79.1-97.1%; n=42/46) in discovering mutations that were represented in $\geq$20% of the alleles of the tumor biopsy (Figure 4D).

4.4. Longitudinal monitoring of DLBCL genotype by using plasma cfDNA

A total of 127 plasma cfDNA samples were sequentially evaluated to assess the dynamics of mutations in plasma upon treatment with R-CHOP. Longitudinal analysis showed a rapid clearance of DLBCL mutations in the cfDNA among responding patients (Figure 6A). Among patients who were primary resistant to R-CHOP (ID12, ID13), DLBCL mutations did not disappear from cfDNA (Figure 6B). In one of the two patients who responded to R-CHOP but ultimately had an early relapse in the central nervous system (ID9), mutations were still detectable in the plasma cfDNA sample collected while the patient was in complete remission (1-month prior relapse). In addition, among patients that were primary refractory to R-CHOP or relapsed after treatment, new mutations appeared in the cfDNA that conceivably marked resistant clones that were selected during the clonal evolution process taking place under the selective pressure of treatment (Figure 7).

5. DISCUSSION

By applying a training-validation approach, this study shows that plasma cfDNA genotyping: *i)* is as accurate as genotyping of the diagnostic biopsy to detect somatic mutations of allelic abundance >20% in DLBCL; *ii)* allows the identification of mutations that are undetectable in the tissue biopsy conceivably because restricted to clones that are anatomically distant from the biopsy site; and *iii)* is a real-time and non-invasive tool to track clonal evolution and emergence of treatment resistant clones.

The identification of genomic alterations with clinical relevance in hematologic malignancies is increasing the need for assays that can routinely identify tumor mutational profile. Among hematologic malignancies with a leukemic component, accessibility of tumor cells in the PB has allowed the fast incorporation of gene mutations into genetic prognostic and predictive algorithms.[61-63] At variance with other hematologic malignancies, DLBCL typically lacks a leukemic involvement, and bone marrow dissemination is infrequent.[64] Limited access to the tumor material has therefore hampered the development and validation of molecular prognostic models in DLBCL, whose molecular stratification represents an unmet medical need.[65] On these bases, among hematologic malignancies, DLBCL is an ideal model in which the liquid biopsy may allow a step forward in the translation of disease genetics into clinically useful markers, especially in the era of novel agents that are active in molecular subgroup of the disease.[50]

Because of its >90% sensitivity and ~100% specificity, plasma cfDNA is an easily accessible source of tumor DNA that allows to accurately profile DLBCL patients for cancer gene mutations represented in >20% of the tumor alleles, which is the sensitivity threshold of conventional Sanger sequencing methods that are broadly used to characterize tumor tissue specimens. On these bases, plasma cfDNA is an effective surrogate of direct tumor genotyping by conventional sequencing for the detection of clonally abundant mutations.

Plasma cfDNA genotyping also informs on variants that are subclonally represented in the tumor biopsy, though a proportion (~50%) of low abundance mutations (i.e. allele frequency <20% in the tumor biopsy) has been missed by our ultra-deep NGS approach conceivably because of its chemistry-dependent sensitivity limit of (~10^{-3}). Indeed, the allelic fractions of mutations in tumor biopsies and plasma samples are generally correlated, indicating that low abundant mutations in the tumor are scantly represented in plasma. The increasing evidence that small subclones have a clinical impact on treatment resistance and outcome in B-cell tumors prompts the development of sensitive approaches for cfDNA that are capable of exactly mirroring both the clonal and subclonal composition of the tumor.[66] For example, incorporation of molecular barcoding in the library preparation chemistry along with the application of in silico bioinformatics algorithms to suppress background NGS artifacts allow to increase the sensitivity of CAPP-seq of up to 15 times.[67]

Plasma cfDNA represents a complementary source of tumor DNA for DLBCL genotyping compared to the tissue biopsy. On the one hand, the complete molecular heterogeneity of a tumor cannot be adequately assessed by single or even multiple tissue biopsies, whereas cfDNA genotyping captures genetic information shed from all sites of the disease. Also, plasma is an accessible source of tumor DNA when DNA cannot be retrieved from the diagnostic tissue biopsy (i.e. because of the limited size or poor conservation of the specimen). On the other hand, plasma cfDNA genotyping misses a proportion of small subclonal mutations. Also, plasma cfDNA genotyping cannot accurately differentiate of de novo DLBCL vs transformed tumors, precisely define the DLBCL cell of origin, and fully detect tumor chromosomal translocations [9] which can instead be all routinely scored by analyzing the tissue biopsy. These notions suggest that the liquid biopsy is not a substitute of the tumor biopsy, but instead provides complementary information in DLBCL.

Treatment for DLBCL is currently undergoing a shift from chemotherapy towards regimens incorporating targeted agents.[44,50] Along with the clinical development of novel targeted agents in DLBCL, mutation-driven mechanisms of resistance to these drugs are increasingly emerging.[50] On these bases, cfDNA genotyping may be used as a strategy of molecular monitoring to inform on the acquisition of targeted drug-resistance in DLBCL. cfDNA analysis can be easily repeated at multiple timepoints on plasma samples obtained before, during and after treatment to provide a picture of the

changes in tumor genetics, including dynamic changes in the mutation profile that occur during therapy.[59]

Beside clonal evolution, longitudinal cfDNA genotyping can also inform on residual disease in cancer.[49] Our CAPP-seq approach does not reach the sensitivity of typical minimal residual disease assays, and the number of informative cases (i.e. cases that achieved a radiological remission but ultimately relapsed) is small in the present study cohort. Despite such limitations, the observation that, among early relapsing patients, low level somatic mutations persisted in plasma while the disease was apparently in remission support the development of plasma cfDNA mutational profiling as a tool for residual disease surveillance in DLBCL. A CAPP-seq incorporating molecular barcoding coupled with bioinformatics that allows the simultaneously tracking of multiple somatic mutations in cfDNA have been developed to outperform immunoglobulin sequencing and radiographic imaging for the detection of minimal residual disease in DLBCL.[9]

There are relatively few studies investigating the importance of the liquid biopsy in DLBCL that have shown very promising results with different approaches. In two studies the IGHV-D-J gene segment of the rearranged immunoglobulin was used as plasma cfDNA biomarker to identify patients at high risk of treatment failure, while in two additional studies cfDNA was used to detect somatic mutations.[5,6,8,9] The retrospective nature of the previous studies that relied on archival material conceivably obtained and handled with different protocols might have precluded a tight control of the

pre-analytic factors that have a substantial effect on plasma cfDNA analysis, including the amount of "contaminating" DNA coming from circulating cells. The prospective design and the standardized protocols for sample processing used in our study allowed to stringently control pre-analytic factors that might influence the results of plasma cfDNA genotyping. In less controlled situations such as multicentered studies, the use of Cell-Free DNA BCT tubes, that avoid storage-related pre-analytic biases, can provide a broadly validated way to obtain stable cfDNA samples.[68]

An immediate clinical application of cfDNA genotyping in DLBCL patients is the incorporation of this assay within clinical trials to support post-hoc patient stratification according to baseline disease genetics, to develop treatment-specific prognostic, predictive or actionable genomic biomarkers, and to track treatment resistance in a clinically relevant time-frame and quality standard.

REFERENCES

1. Rossi D, Diop F, Spaccarotella E, Monti S, Zanni M, Rasi S, Deambrogi C, Spina V, Bruscaggin A, Favini C, Serra R, Ramponi A, Boldorini R, Foà R, Gaidano G. Diffuse large B-cell lymphoma genotyping on the liquid biopsy. *Blood*. 2017;129(14):1947-1957.

2. Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. *Eur J Hum Genet*. 2018;26(7):937-945.

3. Diaz LA Jr, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol*. 2014;32(6):579-586.

4. Hohaus S, Giachelia M, Massini G, Mansueto G, Vannata B, Bozzoli V, Criscuolo M, D'Alò F, Martini M, Larocca LM, Voso MT, Leone G. Cell-free circulating DNA in Hodgkin's and non-Hodgkin's lymphomas. *Ann Oncol.* 2009;20(8):1408-1413.

5. Kurtz DM, Green MR, Bratman SV, Scherer F, Liu CL, Kunder CA, Takahashi K, Glover C, Keane C, Kihira S, Visser B, Callahan J, Kong KA, Faham M, Corbelli KS, Miklos D, Advani RH, Levy R, Hicks RJ, Hertzberg M, Ohgami RS, Gandhi MK, Diehn M, Alizadeh AA. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. *Blood*. 2015;125(24):3679-3687.

6. Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, Shovlin M, Jaffe ES, Staudt LM, Lai C, Steinberg SM, Chen CC, Zheng J, Willis TD, Faham M, Wilson WH. Circulating

tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. *Lancet Oncol*. 2015;16(5):541-549.

7. Roschewski M, Staudt LM, Wilson WH. Dynamic monitoring of circulating tumor DNA in non-Hodgkin lymphoma. *Blood*. 2016;127(25):3127-3132.

8. Bohers E, Viailly PJ, Dubois S, Bertrand P, Maingonnat C, Mareschal S, Ruminy P, Picquenot JM, Bastard C, Desmots F, Fest T, Leroy K, Tilly H, Jardin F. Somatic mutations of cell-free circulating DNA detected by next-generation sequencing reflect the genetic changes in both germinal center B-cell-like and activated B-cell-like diffuse large B-cell lymphomas at the time of diagnosis. *Haematologica*. 2015;100(7):e280-e284.

9. Scherer F, Kurtz DM, Newman AM, Stehr H, Craig AF, Esfahani MS, Lovejoy AF, Chabon JJ, Klass DM, Liu CL, Zhou L, Glover C, Visser BC, Poultsides GA, Advani RH, Maeda LS, Gupta NK, Levy R, Ohgami RS, Kunder CA, Diehn M, Alizadeh AA. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci Transl Med*. 2016;8(364):64ra155.

10. Pasqualucci L, Dalla-Favera R. The Genetic Landscape of Diffuse Large B Cell Lymphoma. *Semin Hematol*.2015;52(2): 67–76.

11. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever

MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503-511.

12. Shaffer AL 3rd, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol*. 2012; 30:565-610.

13. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, López-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM. Lymphoma/Leukemia Molecular Profiling Project. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(25):1937-1947.

14. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A*. 2003;100(17):9991-9996.

15. Goodman RH, Smolik S. CBP/p300 in cell growth, transformation, and development. *Genes & development*. 2000;14(13):1553–1577.

16. Iqbal J, Greiner TC, Patel K, Dave BJ, Smith L, Ji J, Wright G, Sanger WG, Pickering DL, Jain S, Horsman DE, Shen Y, Fu K, Weisenburger DD, Hans CP, Campo E, Gascoyne RD, Rosenwald A, Jaffe ES, Delabie J, Rimsza L, Ott G, Müller-Hermelink HK, Connors JM, Vose JM, McKeithan

T, Staudt LM, Chan WC; Leukemia/Lymphoma Molecular Profiling Project. Distinctive patterns of BCL6 molecular alterations and their functional consequences in different subgroups of diffuse large B-cell lymphoma. *Leukemia*.2007;21(11):2332–2343.

17. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, Califano A, Pasqualucci L, Basso K, Dalla-Favera R. MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nature immunology*.2013;14(10):1084–1092.

18. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, Marra MA. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298–303.

19. Duan S, Cermak L, Pagan JK, Rossi M, Martinengo C, di Celle PF, Chapuy B, Shipp M, Chiarle R, Pagano M. FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. *Nature*.2012;481(7379):90–93.

20. Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, Dominguez-Sola D, Pasqualucci L, Dalla-Favera R. Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer cell*.2011;20(6):728–740.

21. Miyashita K, Tomita N, Taguri M, Suzuki T, Ishiyama Y, Ishii Y, Nakajima Y, Numata A, Hattori Y, Yamamoto W, Miyazaki T, Tachibana T, Takasaki H, Matsumoto K, Hashimoto C, Takemura S, Yamazaki E, Fujimaki K, Sakai R, Motomura S, Ishigatsubo Y. Beta-2 microglobulin is a strong prognostic factor in patients with DLBCL receiving R-CHOP therapy. *Leuk Res*.2015;S0145-2126(15)30368-30374.

22. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, Chan J, Bhagat G, Chadburn A, Gaidano G, Mullighan CG, Rabadan R, Dalla-Favera R. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nature genetics*.2011;43(9):830–837.

23. Monti S, Chapuy B, Takeyama K, Rodig SJ, Hao Y, Yeda KT, Inguilizian H, Mermel C, Currie T, Dogan A, Kutok JL, Beroukhim R, Neuberg D, Habermann TM, Getz G, Kung AL, Golub TR, Shipp MA. Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large B cell lymphoma. *Cancer cell*.2012;22(3):359–372.

24. Lu TX, Young KH, Xu W, Li JY. TP53 dysfunction in diffuse large B-cell lymphoma. *Crit Rev Oncol Hematol*.2016;97:47-55.

25. Trinh DL, Scott DW, Morin RD, Mendez-Lago M, An J, Jones SJ, Mungall AJ, Zhao Y, Schein J, Steidl C, Connors JM, Gascoyne RD, Marra MA. Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. *Blood*.2013;121(18):3666–3674.

26. Schuetz JM, Johnson NA, Morin RD, Scott DW, Tan K, Ben-Nierah S, Boyle M, Slack GW, Marra MA, Connors JM, Brooks-Wilson AR, Gascoyne RD. BCL2 mutations in diffuse large B-cell lymphoma. *Leukemia*.2012;26(6):1383-1390.

27. Karube K, Campo E. MYC alterations in diffuse large B-cell lymphomas. *Semin Hematol*. 2015;52(2):97-106.

28. Ma Y, Zhang P, Gao Y, Fan H, Zhang M, Wu J. Evaluation of AKT phosphorylation and PTEN loss and their correlation with the resistance of rituximab in DLBCL. *Int J Clin Exp Pathol*. 2015;8(11):14875-14884.

29. Wang X, Cao X, Sun R, Tang C, Tzankov A, Zhang J, Manyam GC, Xiao M, Miao Y, Jabbar K, Tan X, Pang Y, Visco C, Xie Y, Dybkaer K, Chiu A, Orazi A, Zu Y, Bhagat G, Richards KL, Hsi ED, Choi WWL, van Krieken JH, Huh J, Ponzoni M, Ferreri AJM, Møller MB, Parsons BM, Winter JN, Piris MA, Li S, Miranda RN, Medeiros LJ, Li Y, Xu-Monette ZY, Young KH. Clinical Significance of PTEN Deletion, Mutation, and Loss of PTEN Expression in De Novo Diffuse Large B-Cell Lymphoma. *Neoplasia*. 2018;20(6):574-593.

30. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, Shah S, Zhu H, Kimbara M, Shashkin P, Charlot JF, Tcherpakov M, Corbett R, Tam A, Varhol R, Smailus D, Moksa M, Zhao Y, Delaney A, Qian H, Birol I, Schein J, Moore R, Holt R, Horsman DE, Connors JM, Jones S, Aparicio S, Hirst M, Gascoyne RD, Marra MA. Somatic mutations altering EZH2 (Tyr641) in follicular and diff use large B-cell lymphomas of germinal-center origin. *Nat Genet*. 2010; 42: 181 – 185.

31. Elodie Bohers, Sylvain Mareschal, Philippe Bertrand, Pierre Julien Viailly, Sydney Dubois, Catherine Maingonnat, Philippe Ruminy, Hervé Tilly & Fabrice Jardin. Activating somatic mutations in diffuse large B-cell lymphomas: lessons from next generation sequencing and key elements in the precision medicine era. *Leukemia & Lymphoma*.2015;56(5):1213-1222.

32. Wigle TJ, Knutson SK, Jin L, Kuntz KW, Pollock RM, Richon VM, Copeland RA, Scott MP. The Y641C mutation of EZH2 alters substrate specificity for histone H3 lysine 27 methylation states. *FEBS Lett*. 2011;585(19):3011-3014.

33. Lee HJ, Shin DH, Kim KB, Shin N, Park WY, Lee JH, Choi KU, Kim JY, Lee CH, Sol MY. Polycomb protein EZH2 expression in diff use large B-cell lymphoma is associated with better prognosis in patients treated with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone. *Leukemia & Lymphoma*. 2014;55(9): 2056–2063.

34. Muppidi JR, Schmitz R, Green JA, Xiao W, Larsen AB, Braun SE, An J, Xu Y, Rosenwald A, Ott G, Gascoyne RD, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Vaidehi N, Staudt LM, Cyster JG. Loss of signalling via Ga13 in germinal center B-cell-derived lymphoma. *Nature*. 2014; 516(7530):254-258.

35. O'Hayre M, Inoue A, Kufareva I, Wang Z, Mikelis CM, Drummond RA, Avino S, Finkel K, Kalim KW, DiPasquale G, Guo F, Aoki J, Zheng Y, Lionakis MS, Molinolo AA, Gutkind JS. Inactivating Mutations in GNA13 and RHOA in Burkitt's Lymphoma and Diffuse Large B-cell Lymphoma: A Tumor Suppressor Function for the Gα13/RhoA Axis in B Cell. *Oncogene*. 2016; 35(29): 3771–3780.

36. Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W, Powell J, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Campo E, Jaffe ES, Delabie J, Smeland EB, Rimsza LM, Fisher RI, Weisenburger DD, Chan WC, Staudt LM. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A*. 2008;105(36):13520–13525.

37. Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, Kohlhammer H, Lamy L, Zhao H, Yang Y, Xu W, Shaffer AL, Wright G, Xiao W, Powell J, Jiang JK, Thomas CJ, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Johnson NA, Rimsza LM, Campo E, Jaffe ES, Wilson WH, Delabie J, Smeland EB, Fisher RI, Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Pierce SK, Staudt LM. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature*. 2010;463(7277):88–92.

38. Kim Y, Ju H, Kim DH, Yoo HY, Kim SJ, Kim WS, Ko YH. CD79B and MYD88 mutations in diffuse large B-cell lymphoma. *Hum Pathol*. 2014;45(3):556-64.

39. Knies N, Alankus B, Weilemann A, Tzankov A, Brunner K, Ruff T, Kremer M, Keller UB, Lenz G, Ruland J. Lymphomagenic CARD11/BCL10/MALT1 signaling drives malignant B-cell proliferation via cooperative NF-κB and JNK activation. *Proc Natl Acad Sci U S A*. 2015;112(52):E7230-7238.

40. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, Kohlhammer H, Xu W, Yang Y, Zhao H, Shaffer AL, Romesser P, Wright G, Powell J, Rosenwald A, Muller-Hermelink HK, Ott G, Gascoyne RD, Connors JM, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Fisher RI,

Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Staudt LM. Oncogenically active MYD88 mutations in human lymphoma. *Nature*.2011;470(7332):115–119.

41. Rhyasen GW, Starczynowski DT. IRAK signalling in cancer. *British Journal of Cancer*. 2015;112(2):232-237.

42. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, Bernal-Mizrachi L, Mann KP, Flowers C, Naresh K, Evens A, Gordon LI, Czader M, Gill JI, Hsi ED, Liu Q, Fan A, Walsh K, Jima D, Smith LL, Johnson AJ, Byrd JC, Luftig MA, Ni T, Zhu J, Chadburn A, Levy S, Dunson D, Dave SS. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A.* 2013;110(4):1398-1403.

43. Compagno M, Lim WK, Grunn A, Nandula SV, Brahmachary M, Shen Q, Bertoni F, Ponzoni M, Scandurra M, Califano A, Bhagat G, Chadburn A, Dalla-Favera R, Pasqualucci L. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*. 2009;459(7247):717-721.

44. Younes A, Thieblemont C, Morschhauser F, Flinn I, Friedberg JW, Amorim S, Hivert B, Westin J, Vermeulen J, Bandyopadhyay N, de Vries R, Balasubramanian S, Hellemans P, Smit JW, Fourneau N, Oki Y. Combination of ibrutinib with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) for treatment-naive patients with CD20-positive B-cell non-Hodgkin lymphoma: a non-randomised, phase 1b study. *Lancet Oncol*. 2014;15(9):1019-1026.

45. Kwak J-Y. Treatment of Diffuse Large B Cell Lymphoma. *The Korean Journal of Internal Medicine*. 2012;27(4):369-377.

46. Chiappella A, Tucci A, Castellino A, Pavone V, Baldi I, Carella AM, Orsucci L, Zanni M, Salvi F, Liberati AM, Gaidano G, Bottelli C, Rossini B, Perticone S, De Masi P, Ladetto M, Ciccone G, Palumbo A, Rossi G, Vitolo U. Fondazione Italiana Linfomi. Lenalidomide plus cyclophosphamide, doxorubicin, vincristine, prednisone and rituximab is safe and effective in untreated, elderly patients with diffuse large B-cell lymphoma: a phase I study by the Fondazione Italiana Linfomi. *Haematologica*. 2013;98(11):1732-1738.

47. Tomita N, Takasaki H, Fujisawa S, Miyashita K, Ogusa E, Kishimoto K, Matsuura S, Sakai R, Koharazawa H, Yamamoto W, Fujimaki K, Fujita H, Ishii Y, Taguchi J, Kuwabara H, Motomura S, Ishigatsubo Y. Standard R-CHOP therapy in follicular lymphoma and diffuse large B-cell lymphoma. *J Clin Exp Hematop*. 2013;53(2):121-125.

48. Coiffier B, Sarkozy C. Diffuse large B-cell lymphoma: R-CHOP failure-what to do?. *Hematology Am Soc Hematol Educ Program*. 2016;2016(1):366-378.

49. Xu-Monette ZY, Wu L, Visco C, Tai YC, Tzankov A, Liu WM, Montes-Moreno S, Dybkaer K, Chiu A, Orazi A, Zu Y, Bhagat G, Richards KL, Hsi ED, Zhao XF, Choi WW, Zhao X, van Krieken JH, Huang Q, Huh J, Ai W, Ponzoni M, Ferreri AJ, Zhou F, Kahl BS, Winter JN, Xu W, Li J, Go RS, Li Y, Piris MA, Møller MB, Miranda RN, Abruzzo LV, Medeiros LJ, Young KH. Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study. *Blood*. 2012;120(19):3986-3996.

50. Wilson WH, Young RM, Schmitz R, Yang Y, Pittaluga S, Wright G, Lih CJ, Williams PM, Shaffer AL, Gerecitano J, de Vos S, Goy A, Kenkre VP, Barr PM, Blum KA, Shustov A, Advani R, Fowler NH, Vose JM, Elstrom RL, Habermann TM, Barrientos JC, McGreivy J, Fardis M, Chang BY, Clow F, Munneke B, Moussa D, Beaupre DM, Staudt LM. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat Med*. 2015;21(8):922-926.

51. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, Lister TA;Alliance, Australasian Leukaemia and Lymphoma Group; Eastern Cooperative Oncology Group; European Mantle Cell Lymphoma Consortium; Italian Lymphoma Foundation; European Organisation for Research; Treatment of Cancer/ Dutch Hemato-Oncology Group; Grupo Espanol de Me´dula O´ sea; German High-Grade Lymphoma Study Group; German Hodgkin's Study Group; Japanese Lymphoma Study Group; Lymphoma Study Association; NCIC Clinical Trials Group; Nordic Lymphoma Study Group; Southwest Oncology Group; United Kingdom National Cancer Research Institute. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014;32(27):3059-3067.

52. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids Res*. 1988; 16(3):1215.

53. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, Chan J, Bhagat G, Chadburn A, Gaidano G, Mullighan CG, Rabadan R, Dalla-Favera R. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011;43(9):830-837.

54. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL, Novak AJ, Dogan A, Ansell SM, Link BK, Zou L, Gould J, Saksena G, Stransky N, Rangel-Escareño C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Hernández-Lemus E, Schwarz-Cruz y Celis A, Imaz-Rosshandler I, Ojesina AI, Jung J, Pedamallu CS, Lander ES, Habermann TM, Cerhan JR, Shipp MA, Getz G, Golub TR. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci* USA. 2012;109(10):3879-3884.

55. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, Dunphy C, Choi W, Au WY, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, Bernal-Mizrachi L, Mann KP, Flowers C, Naresh K, Evens A, Gordon LI, Czader M, Gill JI, Hsi ED, Liu Q, Fan A, Walsh K, Jima D, Smith LL, Johnson AJ, Byrd JC, Luftig MA, Ni T, Zhu J, Chadburn A, Levy S, Dunson D, Dave SS. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci USA*. 2013;110(4):1398-1403.

56. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, Scott DW, Ding J, Roth A, Chiu R, Corbett RD, Chan FC, Mendez-Lago M, Trinh DL, Bolger-Munro M, Taylor G, Hadj Khodabakhshi A, Ben-Neriah S, Pon J, Meissner B, Woolcock B, Farnoud N, Rogic S, Lim EL, Johnson NA, Shah S, Jones S, Steidl C, Holt R, Birol I, Moore R, Connors JM, Gascoyne RD, Marra MA. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood*. 2013;122(7):1256-1265.

57. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, Shrager JB, Loo BW Jr, Alizadeh AA, Diehn M. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20(5):548-554.

58. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Küppers R, Dalla-Favera R. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 2001; 412(6844):341-346.

59. Rabbitts TH, Hamlyn PH, Baer R. Altered nucleotide sequences of a translocated c-myc gene in Burkitt lymphoma. *Nature*. 1983;306(5945):760-765.

60. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin SF, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD, Caldas C, Rosenfeld N. Noninvasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*.2013;497(7447):108-112.

61. Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. *N Engl J Med*. 2015;373(12):1136-1152.

62. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. *N Engl J Med*. 2015;373(16):1541-1552.

63. Rossi D, Gaidano G. The clinical implications of gene mutations in chronic lymphocytic leukaemia. *Br J Cancer*. 2016;114(8):849-854.

64. Sehn LH, Scott DW, Chhanabhai M, Berry B, Ruskova A, Berkahn L, Connors JM, Gascoyne RD. Impact of concordant and discordant bone marrow involvement on outcome in diffuse large B-cell lymphoma treated with R-CHOP*. J Clin Oncol*. 2011;29(11):1452-1457.

65. Weinstock DM, Dalla-Favera R, Gascoyne RD, Leonard JP, Levy R, Lossos IS, Melnick AM, Nowakowski GS, Press OW, Savage KJ, Shipp MA, Staudt LM. A roadmap for discovery and translation in lymphoma. *Blood*. 2015;125(13):2175-2177.

66. Rossi D, Khiabanian H, Spina V, Ciardullo C, Bruscaggin A, Famà R, Rasi S, Monti S, Deambrogi C, De Paoli L, Wang J, Gattei V, Guarini A, Foà R, Rabadan R, Gaidano G. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood*. 2014;123(14):2139-2147.

67. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, Stehr H, Liu CL, Bratman SV, Say C, Zhou L, Carter JN, West RB, Sledge GW, Shrager JB, Loo BW Jr, Neal JW, Wakelee HA, Diehn M, Alizadeh AA. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol*. 2016;34(5):547-555.

68. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. *Clin Chim Acta*. 2013; 424:222-230.

Table 1: Clinical data of the 50 previously untreated DLBCL patients included in the study. CR, complete remission; DLBCL, diffuse large B-cell lymphoma; ECOG, Eastern Cooperative Oncology Group; F, female; M, male; FFPE, formalin-fixed paraffin embedded; NA, not assessable; GC, germinal center; IPI, international prognostic index; LDH, lactate dehydrogenase.

ID	Cohort	Tissue biopsy	Sex	Age at diagnosis	Ann Arbor stage	LDH	ECOG performance status	Extranodal sites	IPI	Cell of origin	Treatment response	Follow-up (months)
ID1	Training	Frozen	F	64	IVX	high	<2	<2	3	GC	CR	11 (progression)
ID2	Training	Frozen	F	35	II	normal	<2	<2	0	non- GC	CR	21
ID3	Training	Frozen	M	72	IVX	high	<2	<2	3	GC	CR	15
ID4	Training	Frozen	F	64	I	normal	<2	<2	1	non- GC	CR	23
ID5	Training	Frozen	F	83	IIX	high	<2	<2	2	GC	CR	24
ID6	Training	Frozen	F	64	IV	high	<2	<2	3	non- GC	CR	2 (death)
ID7	Training	Frozen	M	84	IEX	normal	<2	<2	1	GC	CR	17
ID8	Training	Frozen	F	80	III	normal	<2	<2	2	non- GC	CR	21
ID9	Training	Frozen	F	61	IVX	high	<2	>1	4	non- GC	CR	4 (progression)
ID10	Training	Frozen	M	26	IIX	high	<2	<2	1	non- GC	CR	21
ID11	Training	Frozen	M	38	IVX	normal	<2	<2	1	GC	CR	25
ID12	Training	Frozen	M	75	IIX	normal	<2	<2	1	non- GC	Less than CR	1 (death)
ID13	Training	Frozen	F	74	IVX	high	<2	>1	4	non- GC	Less than CR	10 (death)
ID14	Training	Frozen	F	77	IV	high	<2	>1	4	GC	Less than CR	5 (savage treatment)
ID15	Training	Frozen	F	41	I	normal	<2	<2	0	non- GC	CR	25
ID16	Training	Frozen	M	62	IV	high	<2	>1	4	non- GC	CR	6
ID17	Training	Frozen	F	76	III	normal	<2	<2	2	GC	CR	6
ID18	Training	NA	M	83	IE	normal	<2	<2	1	non- GC	CR	16
ID19	Training	NA	F	52	II	normal	<2	<2	0	GC	CR	14
ID20	Training	NA	M	55	III	high	<2	<2	2	GC	CR	10
ID21	Training	FFPE	F	64	III	high	<2	<2	3	GC	CR	10
ID22	Training	NA	M	82	IV	normal	<2	>1	3	non- GC	CR	10
ID23	Training	FFPE	F	76	IV	high	<2	<2	3	non- GC	CR	4
ID24	Training	NA	F	66	I	normal	<2	<2	1	non- GC	CR	16
ID25	Training	NA	M	76	IV	normal	<2	>1	3	non- GC	CR	12
ID26	Training	NA	M	79	IV	normal	<2	<2	2	non- GC	CR	13
ID27	Training	NA	M	55	III	normal	<2	>1	2	non- GC	Less than CR	8 (savage treatment)
ID28	Training	FFPE	F	81	IV	high	<2	>1	4	non- GC	CR	9
ID29	Training	NA	M	54	IE	high	<2	<2	1	non- GC	CR	14
ID30	Training	NA	M	63	IIX	high	<2	<2	3	non- GC	Less than CR	5 (savage treatment)
ID31	Validation	Frozen	F	59	III	high	<2	<2	2	non- GC	CR	12
ID32	Validation	Frozen	F	68	IE	high	≥2	<2	3	GC	CR	12
ID34	Validation	Frozen	F	49	III	high	<2	<2	2	non- GC	CR	1

ID35	Validation	Frozen	M	39	III	normal	<2	<2	1	non- GC	CR	6
ID36	Validation	Frozen	F	78	II	normal	<2	<2	1	non- GC	CR	11
ID37	Validation	Frozen	M	69	III	high	<2	<2	3	GC	CR	11
ID38	Validation	Frozen	M	36	IX	high	<2	<2	1	GC	Less than CR	7 (death)
ID39	Validation	FFPE	F	36	IVX	high	<2	>2	3	non- GC	PD	9 (death)
ID41	Validation	FFPE	F	85	III	high	>2	<2	4	GC	PD	8 (death)
ID42	Validation	FFPE	F	77	IV	high	<2	>1	4	non- GC	CR	31
ID43	Validation	FFPE	F	80	IV	normal	<2	>2	3	non- GC	CR	5
ID44	Validation	NA	M	73	IV	normal	<2	<2	2	GC	CR	17
ID46	Validation	NA	F	48	IVX	high	<2	<2	2	non- GC	CR	23
ID47	Validation	Frozen	F	78	IVA	high	<2	>1	4	non- GC	NA	1
ID48	Validation	FFPE	F	72	IIX	normal	<2	<2	1	non- GC	CR	22
ID49	Validation	FFPE	M	72	IIX	normal	<2	<2	1	GC	NA	4
ID50	Validation	NA	F	68	IV	high	<2	>2	4	GC	NA	2
ID51	Validation	FFPE	M	78	IV	normal	<2	<2	2	non- GC	CR	26
ID52	Validation	FFPE	F	46	IIX	high	<2	<2	1	non- GC	CR	26
ID53	Validation	NA	F	56	IV	normal	<2	<2	1	GC	Less than CR	5

Table 2: Somatic non-synonymous mutations discovered by cfDNA genotyping and their validation in tumor gDNA. CHROM, chromosome; POS, position; REF, reference allele; VAR, variant allele; AA, aminoacidic change; VAF, variant allele frequency; gDNA, genomic DNA; cfDNA, cell-free DNA.

ID Sample	Cohort	Gene	RefSeq	CHROM	POS	REF	VAR	cDNA	AA	VAF in tumor gDNA (mutated reads/total reads)	VAF in cfDNA (mutated reads/total reads)
ID1	Training	*CREBBP*	NM_004380.2	chr16	3781353	G	A	c.5012C>T	p.A1671V	74.79 (795/1063)	53.93 (618/1146)
ID1	Training	*STAT6*	NM_003153.4	chr12	57498345	C	T	c.1114G>A	p.E372K	49.01 (373/761)	29.87 (253/847)
ID1	Training	*STAT6*	NM_003153.4	chr12	57492817	T	G	c.1936A>C	p.I646L	-	3.12 (54/1731)
ID1	Training	*TP53*	NM_000546.5	chr17	7578454	G	A	c.476C>T	p.A159V	60.58 (355/586)	34.69 (273/787)
ID3	Training	*BCL2*	NM_000633.2	chr18	60985877	C	T	c.23G>A	p.G8E	1.76 (48/2732)	-
ID3	Training	*CARD11*	NM_032415.4	chr7	2976810	T	A	c.1202A>T	p.D401V	43.77 (1935/4421)	8.25 (357/4329)
ID3	Training	*EZH2*	NM_004456.4	chr7	148506437	G	A	c.2075C>T	p.A692V	13.03 (873/5004)	1.9 (60/3151)
ID3	Training	*GNA13*	NM_006572.4	chr17	63052437	A	T	c.275T>A	p.V92E	61.02 (620/1016)	14.97 (69/461)
ID3	Training	*KMT2D*	NM_003482.3	chr12	49425896	G	A	c.12592C>T	p.R4198*	39.68 (873/2200)	3.77 (128/3395)
ID4	Training	*HIST1H1E*	NM_005321.2	chr6	26156937	C	T	c.319C>T	p.L107F	17.69 (222/1255)	-
ID4	Training	*ITPKB*	NM_002221.3	chr1	226923964	C	T	c.1196G>A	p.S399N	13.86 (154/1111)	-
ID4	Training	*ITPKB*	NM_002221.3	chr1	226924421	C	G	c.739G>C	p.E247Q	5.51 (72/1307)	-
ID4	Training	*ITPKB*	NM_002221.3	chr1	226924604	G	A	c.556C>T	p.P186S	5.93 (55/928)	-
ID4	Training	*ITPKB*	NM_002221.3	chr1	226923733	A	GCATTCT	c.1420_1426del7bp	p.R474fs*13	7.33 (89/1214)	-
ID4	Training	*PCLO*	NM_033026.5	chr7	82784833	T	+30bp	c.1123_1124ins30bp	p.Q375Ins10AA	2.73 (34/1244)	-
ID4	Training	*PCLO*	NM_033026.5	chr7	82584246	G	C	c.6023C>G	p.T2008R	32.30 (686/2124)	-
ID4	Training	*TBL1XR1*	NM_024665.4	chr3	176756172	T	-ACA	c.973_975del3bp	p.C325DelC	19.59 (269/1104)	-
ID5	Training	*CCND3*	NM_001760.4	chr6	41903745	C	+G	c.811_812insC	p.P271fs*	45.79 (1115/2435)	6.99 (84/1201)
ID5	Training	*EP300*	NM_001429.3	chr22	41573446	C	T	c.5731C>T	p.P1911S	0.89 (33/3716)	-
ID5	Training	*FBXW7*	NM_033632.3	chr4	153244092	G	A	c.2065C>T	p.R689W	46.40 (1979/4265)	6.11 (205/3357)
ID6	Training	*KMT2D*	NM_003482.3	chr12	49434090	G	A	c.7463C>T	p.S2488L	4.21 (40/951)	3.34 (19/569)
ID6	Training	*MYD88*	NM_002468.4	chr3	38182292	G	A	c.728G>A	p.S243N	10.49 (284/2708)	28.19 (1074/3810)
ID6	Training	*PIM1*	NM_001243186.1	chr6	37138769	C	G	c.475C>G	p.H159D	-	1.43 (33/2303)
ID6	Training	*PIM1*	NM_001243186.1	chr6	37139237	C	T	c.850C>T	p.L284F	-	4.29 (91/2121)
ID6	Training	*TNFAIP3*	NM_001270508.1	chr6	138200167	A	+C	c.1585_1586insC	p.T529fs*5	6.08 (200/3291)	12.01 (412/3431)
ID6	Training	*TP53*	NM_000546.5	chr17	7577094	G	A	c.844C>T	p.R282W	7.97 (207/2597)	15.72 (469/2984)
ID7	Training	*B2M*	NM_004048.2	chr15	45003780	A	-CT	c.37_38delCT	p.L13fs*43	60.92 (410/673)	0.87 (6/692)
ID7	Training	*BCL2*	NM_000633.2	chr18	60985757	A	C	c.143T>G	p.I48S	34.37 (221/643)	0.62 (9/1451)
ID7	Training	*CREBBP*	NM_004380.2	chr16	3795276	A	T	c.3914+2T>A		37.69 (798/2117)	1.99 (51/2559)
ID7	Training	*EZH2*	NM_004456.4	chr7	148508728	A	T	c.1936T>A	p.Y646N	31.12 (539/1732)	0.95 (15/1577)
ID7	Training	*KMT2D*	NM_003482.3	chr12	49445616	A	G	c.1850T>C	p.L617P	4.61 (64/1389)	-
ID7	Training	*KMT2D*	NM_003482.3	chr12	49441770	T	C	c.4214A>G	p.H1405R	34.33 (505/1471)	1.7 (38/2230)

ID7	Training	*LRP1B*	NM_018557.2	chr2	141201923	C	T	c.10270G>A	p.D3424N	-	1.32 (23/1739)
ID7	Training	*SPEN*	NM_015001.2	chr1	16256910	C	G	c.4175C>G	p.S1392*	39.81 (1207/3032)	1.2 (46/3829)
ID7	Training	*TBL1XR1*	NM_024665.4	chr3	176767845	A	T	c.642T>A	p.C214*	38.50 (954/2478)	2.69 (44/1637)
ID7	Training	*TNFRSF14*	NM_003820.2	chr1	2491336	T	G	c.379T>G	p.C127G	54 (27/50)	-
ID8	Training	*NOTCH2*	NM_024408.2	chr1	120458053	T	C	c.7292A>G	p.H2431R	2.79 (19/681)	-
ID8	Training	*NOTCH2*	NM_024408.2	chr1	120458048	C	-AGAGT	c.7292_7296del5bp	p.H2431fs*3	36.7 (407/1109)	11.47 (233/2031)
ID8	Training	*TBL1XR1*	NM_024665.4	chr3	176756174	C	-AAG	c.971_973del3bp	p.S324DelS	35.62 (285/800)	5.62 (65/1157)
ID8	Training	*TP53*	NM_000546.5	chr17	7577559	G	A	c.722C>T	p.S241F	14.72 (29/197)	0.72 (9/1251)
ID8	Training	*XPO1*	NM_003400.3	chr2	61711077	T	C	c.2672A>G	p.D891G	32.13 (187/582)	6.3 (52/825)
ID9	Training	*CCND3*	NM_001760.4	chr6	41903692	C	G	c.865G>C	p.A289P	30.40 (642/2112)	6.22 (84/1351)
ID9	Training	*FBXW7*	NM_033632.3	chr4	153244155	T	C	c.2002A>G	p.S668G	-	1.38 (75/5429)
ID9	Training	*PCLO*	NM_033026.5	chr7	82579618	T	A	c.10286A>T	p.N3429I	4.29 (164/3827)	0.47 (26/5511)
ID9	Training	*SPEN*	NM_015001.2	chr1	16261296	C	T	c.8561C>T	p.S2854F	26.49 (648/2446)	1.79 (99/5535)
ID9	Training	*SPEN*	NM_015001.2	chr1	16255679	G	T	c.2944G>T	p.E982*	65.82 (4167/6331)	8.59 (638/7425)
ID10	Training	*ATM*	NM_000051.3	chr11	108236105	A	G	c.9041A>G	p.Q3014R	-	0.47 (23/4853)
ID10	Training	*BCL2*	NM_000633.2	chr18	60985405	C	G	c.495G>C	p.E165D	34.30 (3363/9804)	1.61 (62/3842)
ID10	Training	*BCOR*	NM_017745.5	chrX	39932419	G	A	c.2180C>T	p.P727L	-	1.31 (31/2358)
ID10	Training	*CHD2*	NM_001271.3	chr15	93521519	C	T	c.2633C>T	p.S878L	-	0.62 (31/5007)
ID10	Training	*CREBBP*	NM_004380.2	chr16	3786036	C	T	c.4728+1G>A		28.38 (1242/4376)	1.25 (45/3590)
ID10	Training	*EZH2*	NM_004456.4	chr7	148508727	T	A	c.1937A>T	p.Y646F	52.94 (3331/6292)	1.67 (50/2999)
ID10	Training	*KLHL6*	NM_130446.2	chr3	183273174	G	A	c.268C>T	p.L90F	47.86 (3025/6321)	2.4 (69/2881)
ID10	Training	*MYC*	NM_002467.4	chr8	128750803	C	G	c.340C>G	p.L114V	1.42 (94/6620)	-
ID10	Training	*MYC*	NM_002467.4	chr8	128750684	C	T	c.221C>T	p.P74L	10.24 (308/3008)	-
ID10	Training	*MYC*	NM_002467.4	chr8	128750625	G	C	c.162G>C	p.E54D	34.99 (1978/5653)	0.17 (4/2334)
ID10	Training	*MYC*	NM_002467.4	chr8	128750843	A	G	c.380A>G	p.N127S	35.20 (2777/7890)	1.15 (56/4883)
ID10	Training	*PCLO*	NM_033026.5	chr7	82582514	C	G	c.7755G>C	p.L2585F	-	0.66 (27/4111)
ID10	Training	*PCLO*	NM_033026.5	chr7	82579795	C	A	c.10109G>T	p.S3370I	-	0.66 (33/4990)
ID10	Training	*PCLO*	NM_033026.5	chr7	82580115	C	T	c.9789G>A	p.M3263I	-	0.72 (36/4977)
ID11	Training	*TP53*	NM_000546.5	chr17	7577594	A	-CAGT	c.683_686del4	p.D228fs*18	15.81 (163/1031)	-
ID12	Training	*BRAF*	NM_004333.4	chr7	140453193	T	C	c.1742A>G	p.N581S	62.65 (3573/5703)	3.99 (87/2181)
ID12	Training	*NOTCH1*	NM_017617.3	chr9	139391369	G	-GA	c.6820_6821delTC	p.S2274fs*79	4.16 (83/1993)	-
ID12	Training	*TP53*	NM_000546.5	chr17	7579484	T	+CTGG	c.202_203ins4	p.E68fs*82	81.91 (1200/1465)	0.51 (10/1950)
ID12	Training	*ZMYM3*	NM_005096.3	chrX	70460810	T	A	c.4069A>T	p.I1357F	86.68 (742/856)	1.35 (17/1255)
ID13	Training	*EP300*	NM_001429.3	chr22	41560092	A	G	c.3764A>G	p.H1255R	-	27.65 (885/3201)
ID13	Training	*ID3*	NM_002167.4	chr1	23885728	G	A	c.190C>T	p.L64F	40.28 (1705/4233)	19.17 (696/3631)
ID13	Training	*KMT2D*	NM_003482.3	chr12	49422609	A	C	c.14382+2T>G		17.54 (455/2594)	10.2 (332/3254)
ID13	Training	*KMT2D*	NM_003482.3	chr12	49426961	G	A	c.11527C>T	p.Q3843*	37.13 (874/2354)	19.97 (543/2719)
ID13	Training	*PIM1*	NM_001243186.1	chr6	37138901	C	T	c.514C>T	p.P172S	-	0.83 (28/3378)

ID13	Training	*PIM1*	NM_001243186.1	chr6	37138908	G	A	c.521G>A	p.G174D	-	1.15 (38/3305)
ID13	Training	*TBL1XR1*	NM_024665.4	chr3	176756174	C	A	c.974G>T	p.C325F	40.93 (562/1373)	13.59 (238/1751)
ID13	Training	*TNFAIP3*	NM_001270508.1	chr6	138199594	G	T	c.1012G>T	p.E338*	67.75 (1714/2530)	35.23 (856/2430)
ID13	Training	*TP53*	NM_000546.5	chr17	7577538	C	T	c.743G>A	p.R248Q	68.37 (1314/1922)	32.01 (894/2793)
ID15	Training	*ATM*	NM_000051.3	chr11	108168100	G	T	c.4996G>T	p.E1666*	16.51 (389/2356)	41.25 (905/2194)
ID15	Training	*B2M*	NM_004048.2	chr15	45007727	C	-TT	c.175_176delTT	p.L59fs*8	22.29 (820/3678)	33.5 (671/2003)
ID15	Training	*PCLO*	NM_033026.5	chr7	82580704	C	T	c.9200G>A	p.R3067K	4.08 (147/3599)	13.49 (400/2966)
ID15	Training	*POT1*	NM_015450.2	chr7	124467278	A	G	c.1676T>C	p.L559P	17.39 (358/2059)	43.19 (827/1915)
ID15	Training	*STAT6*	NM_003153.4	chr12	57496255	C	T	c.1330G>A	p.E444K	27.42 (811/2958)	36.75 (781/2125)
ID15	Training	*TP53*	NM_000546.5	chr17	7579415	C	T	c.272G>A	p.W91*	9.27 (243/2621)	13.91 (324/2330)
ID16	Training	*CD58*	NM_001779.2	chr1	117086941	T	-AAAG	c.352_355del4bp	p.Y117fs*12	25.51 (504/1976)	33.43 (1365/4083)
ID16	Training	*CD58*	NM_001779.2	chr1	117064573	T	-G	c.660delC	p.P220fs*4	26.94 (284/1054)	40.48 (1144/2826)
ID16	Training	*CD79B*	NM_001039933.1	chr17	62006798	T	G	c.590A>C	p.Y197S	28.2 (238/844)	36.97 (505/1366)
ID16	Training	*CREBBP*	NM_004380.2	chr16	3817738	G	T	c.3233C>A	p.S1078*	31.03 (650/2095)	37.3 (1593/4271)
ID16	Training	*KLHL6*	NM_130446.2	chr3	183273248	A	G	c.194T>C	p.L65P	1.39 (201/1446)	13.71 (295/2151)
ID16	Training	*KLHL6*	NM_130446.2	chr3	183273170	G	C	c.272C>G	p.A91G	9.83 (76/773)	15.77 (194/1230)
ID16	Training	*KLHL6*	NM_130446.2	chr3	183273223	A	C	c.219T>G	p.D73E	32.08 (502/1565)	32.09 (750/2337)
ID16	Training	*KMT2D*	NM_003482.3	chr12	49443617	G	A	c.3754C>T	p.R1252*	56.68 (721/1272)	80.41 (2812/3497)
ID16	Training	*NOTCH2*	NM_024408.2	chr1	120458372	G	A	c.6973C>T	p.Q2325*	29.08 (301/1035)	40.27 (1119/2779)
ID16	Training	*PIM1*	NM_001243186.1	chr6	37139167	G	C	c.780G>C	p.K260N	-	1.35 (26/1926)
ID16	Training	*PIM1*	NM_001243186.1	chr6	37138651	T	A	c.458T>A	p.L153*	-	3.36 (35/1041)
ID16	Training	*PIM1*	NM_001243186.1	chr6	37138808	G	A	c.513+1G>A		26.74 (368/1376)	33.78 (629/1862)
ID16	Training	*PIM1*	NM_001243186.1	chr6	37140894	C	A	c.1003C>A	p.H335N	28.79 (463/1608)	39.52 (1180/2986)
ID17	Training	*BCL2*	NM_000633.2	chr18	60985644	G	A	c.256C>T	p.L86F	26.14 (177/677)	4.44 (16/360)
ID17	Training	*CREBBP*	NM_004380.2	chr16	3788618	G	A	c.4336C>T	p.R1446C	21.8 (801/3674)	5.6 (110/1965)
ID17	Training	*EZH2*	NM_004456.4	chr7	148508727	T	G	c.1937A>C	p.Y646S	36.43 (1242/3409)	5.97 (139/2328)
ID17	Training	*KMT2D*	NM_003482.3	chr12	49444250	G	A	c.3121C>T	p.Q1041*	20.92 (627/2997)	3.27 (25/765)
ID17	Training	*KMT2D*	NM_003482.3	chr12	49433620	G	A	c.7933C>T	p.R2645*	22.05 (741/3360)	5.68 (97/1709)
ID17	Training	*MAP2K1*	NM_002755.3	chr15	66774131	G	A	c.607G>A	p.E203K	3.15 (75/2380)	-
ID17	Training	*MEF2B*	NM_001145785.1	chr19	19260128	G	-AA	c.163_164delTT	p.F55fs*28	22.36 (588/2630)	8.07 (125/1548)
ID17	Training	*STAT6*	NM_003153.4	chr12	57498348	A	G	c.1111T>C	p.C371R	20.27 (238/1174)	5.02 (95/1892)
ID17	Training	*TNFRSF14*	NM_003820.2	chr1	2488123	G	A	c.20G>A	p.W7*	13.26 (127/958)	1.77 (18/1016)
ID19	Training	*B2M*	NM_004048.2	chr15	45003773	T	C	c.29T>C	p.L10P	NA	8.43 (43/510)
ID19	Training	*CARD11*	NM_032415.4	chr7	2979559	C	T	c.688G>A	p.D230N	NA	2.88 (55/1907)
ID19	Training	*HIST1H1E*	NM_005321.2	chr6	26157009	C	T	c.391C>T	p.P131S	NA	1.71 (29/1696)
ID19	Training	*LRP1B*	NM_018557.2	chr2	141751573	C	G	c.2635G>C	p.V879L	NA	1.49 (21/1409)
ID19	Training	*LRP1B*	NM_018557.2	chr2	141232880	T	C	c.9452A>G	p.E3151G	NA	4.52 (78/1726)
ID21	Training	*CREBBP*	NM_004380.2	chr16	3788618	G	A	c.4336C>T	p.R1446C	21.71 (345/1589)	5.79 (120/2074)

ID21	Training	*KMT2D*	NM_003482.3	chr12	49443844	GCTGT	G	c.3523_3526del4bp	p.T1175fs*36	-	7.55 (126/1669)
ID21	Training	*KMT2D*	NM_003482.3	chr12	49416115	G	A	c.16360C>T	p.R5454*	22.35 (76/340)	9.89 (91/920)
ID21	Training	*PCLO*	NM_033026.5	chr7	82546096	T	C	c.11206A>G	p.T3736A	NA	0.85 (21/2474)
ID22	Training	*KMT2D*	NM_003482.3	chr12	49447293	G	A	c.805C>T	p.Q269*	NA	1.89 (32/1697)
ID22	Training	*TNFAIP3*	NM_001270508.1	chr6	138198245	C	-G	c.839_839delG	p.R280fs*7	NA	2.17 (55/2537)
ID23	Training	*ARID1A*	NM_006015.4	chr1	27102199	G	C	c.5124+1G>C		36.88 (648/1757)	28.94 (722/2495)
ID23	Training	*CD79B*	NM_001039933.1	chr17	62007129	C	G	c.552+1G>C		42.16 (390/925)	23.53 (368/1564)
ID23	Training	*MYC*	NM_002467.4	chr8	128751037	T	A	c.574T>A	p.Y192N	43.66 (303/694)	10.14 (73/720)
ID23	Training	*MYC*	NM_002467.4	chr8	128751056	C	T	c.593C>T	p.A198V	44.75 (290/648)	11.11 (86/774)
ID23	Training	*MYC*	NM_002467.4	chr8	128750938	C	G	c.475C>G	p.L159V	48.29 (989/2048)	18.17 (325/1789)
ID23	Training	*MYC*	NM_002467.4	chr8	128750750	C	G	c.287C>G	p.S96C	51.21 (1232/2406)	18.89 (221/1170)
ID23	Training	*MYC*	NM_002467.4	chr8	128750602	T	G	c.139T>G	p.Y47D	48.84 (609/1247)	19.38 (289/1491)
ID23	Training	*MYC*	NM_002467.4	chr8	128750849	G	A	c.386G>A	p.S129N	-	3.66 (96/2623)
ID23	Training	*MYC*	NM_002467.4	chr8	128750686	C	T	c.223C>T	p.P75S	49.70 (750/1509)	8.12 (65/800)
ID23	Training	*PIM1*	NM_001243186.1	chr6	37139210	C	G	c.823C>G	p.L275V	-	12.45 (162/1301)
ID23	Training	*PIM1*	NM_001243186.1	chr6	37139114	C	G	c.727C>G	p.L243V	84.27 (1698/2015)	39.6 (531/1341)
ID26	Training	*CARD11*	NM_032415.4	chr7	2984030	G	A	c.500C>T	p.T167M	NA	2.32 (29/1248)
ID26	Training	*DDX3X*	NM_001356.3	chrX	41204667	G	A	c.1181G>A	p.R394H	NA	3.36 (14/417)
ID26	Training	*FBXW7*	NM_033632.3	chr4	153245397	A	+T	c.1793_1794insA	p.N598fs*7	NA	0.75 (17/2255)
ID26	Training	*LRP1B*	NM_018557.2	chr2	141812805	C	A	c.1432G>T	p.D478Y	NA	1.08 (15/1395)
ID26	Training	*TNFAIP3*	NM_001270508.1	chr6	138199911	T	-G	c.1330_1330delG	p.E444fs*33	NA	1.14 (19/1671)
ID26	Training	*ZMYM3*	NM_005096.3	chrX	70464194	C	A	c.3238G>T	p.A1080S	NA	3 (26/866)
ID28	Training	*ARID1A*	NM_006015.4	chr1	27105554	G	T	c.5165G>T	p.R1722L	14.29 (254/1778)	17.39 (483/2777)
ID28	Training	*BRAF*	NM_004333.4	chr7	140481403	C	G	c.1405G>C	p.G469R	-	1.69 (46/2729)
ID28	Training	*CCND3*	NM_001760.4	chr6	41903719	G	A	c.838C>T	p.Q280*	-	2.89 (49/1696)
ID28	Training	*CCND3*	NM_001760.4	chr6	41903731	G	A	c.826C>T	p.Q276*	23.02 (399/1733)	31.31 (541/1728)
ID28	Training	*IKBKB*	NM_001556.2	chr8	42166429	T	A	c.578T>A	p.L193Q	18.09 (226/1249)	22.44 (438/1952)
ID28	Training	*KMT2D*	NM_003482.3	chr12	49431113	G	-C	c.10025_10025delG	p.R3342fs*16	12.78 (187/1463)	16.21 (406/2504)
ID28	Training	*KMT2D*	NM_003482.3	chr12	49425455	T	+G	c.13032_13033insC	p.K4345fs*27	30.53 (722/2365)	35.3 (1036/2935)
ID28	Training	*KRAS*	NM_033360.2	chr12	25398281	C	T	c.38G>A	p.G13D	29.53 (344/1165)	32.85 (837/2548)
ID28	Training	*LRP1B*	NM_018557.2	chr2	141597656	T	G	c.5115-2A>C		NA	21.2 (300/1415)
ID28	Training	*TBL1XR1*	NM_024665.4	chr3	176767919	C	T	c.568G>A	p.D190N	14.35 (248/1728)	18.8 (478/2543)
ID28	Training	*TNFAIP3*	NM_001270508.1	chr6	138199831	A	T	c.1249A>T	p.K417*	33.05 (1084/3280)	48.55 (1437/2960)
ID30	Training	*CD58*	NM_001779.2	chr1	117087037	GAC	G	c.258_259delGT	p.S87fs*2	NA	0.14 (4/2862)
ID30	Training	*NOTCH1*	NM_017617.3	chr9	139390846	TG	T	c.7344_7344delC	p.P2448fs*29	NA	43.3 (294/679)
ID30	Training	*PIM1*	NM_001243186.1	chr6	37138805	C	G	c.511C>G	p.L171V	NA	1.91 (33/1727)
ID30	Training	*STAT6*	NM_003153.4	chr12	57492812	C	G	c.1941G>C	p.K647N	NA	0.57 (18/3145)
ID30	Training	*TP53*	NM_000546.5	chr17	7577021	CG	C	c.917_917delG	p.R306fs*39	NA	16.93 (564/3332)

ID30	Training	*TP53*	NM_000546.5	chr17	7576887	TTC	T	c.957_958delGA	p.K320fs*16	NA	18.39 (582/3165)
ID31	Validation	*CD79B*	NM_001039933.1	chr17	62006804	T	C	c.584A>G	p.H195R	33.78 (586/1735)	18.65 (130/697)
ID31	Validation	*MEF2B*	NM_001145785.1	chr19	19260064	C	T	c.229G>A	p.E77K	39.89 (919/2304)	21.85 (248/1135)
ID31	Validation	*HIST1H1E*	NM_005321.2	chr6	26157118	C	A	c.500C>A	p.A167D	1.12 (25/2238)	1.49 (9/558)
ID32	Validation	*TNFRSF14*	NM_003820.2	chr1	2491292	C	T	c.335C>T	p.S112F	19.17 (158/824)	75.16 (230/306)
ID32	Validation	*KRAS*	NM_033360.2	chr12	25378562	C	T	c.436G>A	p.A146T	28.69 (2662/9278)	55.39 (4907/8859)
ID32	Validation	*MYD88*	NM_001172567.1	chr3	38182025	G	T	c.649G>T	p.V217F	15.14 (885/5846)	28.82 (1591/5521)
ID32	Validation	*CD58*	NM_001779.2	chr1	117087122	G	A	c.175C>T	p.Q59*	18.47 (1987/10757)	24.90 (1953/7842)
ID32	Validation	*CD58*	NM_001779.2	chr1	117086941	T	-AAAG	c.352_355del4bp	p.L118fs*11	13.92 (1003/7208)	18.60 (928/4990)
ID32	Validation	*NOTCH2*	NM_024408.2	chr1	120458147	G	A	c.7198C>T	p.R2400*	15.93 (439/2756)	37.99 (827/2177)
ID32	Validation	*PRDM1*	NM_182907.2	chr6	106552886	G	+A	c.449_450insA	p.R150fs*140	-	4.72 (461/9759)
ID32	Validation	*CD58*	NM_001779.2	chr1	117078656	A	G	c.559T>C	p.C187R	-	5.42 (403/7431)
ID32	Validation	*CD58*	NM_001779.2	chr1	117087208	G	-A	c.88_88delT	p.S30fs*12	-	16.44 (1199/7291)
ID34	Validation	GNA13	NM_006572.4	chr17	63010482	A	C	c.1027T>G	p.Y343D	31.11 (1455/4677)	1.61 (44/2740)
ID34	Validation	CREBBP	NM_004380.2	chr16	3788657	A	C	c.4297T>G	p.Y1433D	42.74 (1107/2590)	0.88 (11/1257)
ID34	Validation	KMT2D	NM_003482.3	chr12	49431399	A	+G	c.9739_9740insC	p.L3247fs*4	41.26 (861/2087)	0.62 (8/1286)
ID34	Validation	EP300	NM_001429.3	chr22	41568590	G	A	c.4540G>A	p.E1514K	-	1.56 (38/2431)
ID35	Validation	*NOTCH1*	NM_017617.3	chr9	139391902	G	A	c.6289C>T	p.P2097S	1.83 (22/1203)	-
ID36	Validation	*EP300*	NM_001429.3	chr22	41536154	A	G	c.1771A>G	p.I591V	20.26 (1298/6406)	-
ID37	Validation	*TNFRSF14*	NM_003820.2	chr1	2489255	T	A	c.160T>A	p.C54S	64.57 (277/429)	10.46 (16/153)
ID37	Validation	*TNFRSF14*	NM_003820.2	chr1	2491399	C	T	c.442C>T	p.Q148*	12.59 (51/405)	1.32 (2/151)
ID37	Validation	*TBL1XR1*	NM_024665.4	chr3	176750838	T	G	c.1337A>C	Y446S	10.35 (300/2898)	-
ID37	Validation	*TBL1XR1*	NM_024665.4	chr3	176750808	T	A	c.1367A>T	p.Y456F	10.56 (300/2840)	-
ID37	Validation	*CARD11*	NM_032415.4	chr7	2977614	T	A	c.1070A>T	p.D357V	21.15 (246/1163)	6.33 (25/395)
ID37	Validation	*CARD11*	NM_032415.4	chr7	2979499	A	G	c.748T>C	p.S250P	19.79 (977/4936)	0.71 (6/846)
ID37	Validation	*EZH2*	NM_004456.4	chr7	148508728	A	T	c.1936T>A	p.Y646N	9.17 (131/1429)	-
ID37	Validation	*EZH2*	NM_004456.4	chr7	148508727	T	A	c.1937A>T	p.Y646F	23.47 (334/1423)	2.27 (8/353)
ID37	Validation	*KMT2D*	NM_003482.3	chr12	49420661	G	A	c.15088C>T	p.R5030C	35.80 (600/1676)	5.74 (50/331)
ID37	Validation	*KMT2D*	NM_003482.3	chr12	49425836	G	A	c.12652C>T	p.Q4218*	37.02 (810/2188)	11.18 (52/465)
ID37	Validation	*BCL2*	NM_000633.2	chr18	60985529	A	G	c.371T>C	p.F124S	21.05(498/2366)	0.87 (6/687)
ID37	Validation	*BCL2*	NM_000633.2	chr18	60985839	A	G	c.61T>C	p.Y21H	22.04 (410/1860)	6.35 (25/394)
ID37	Validation	*BCL2*	NM_000633.2	chr18	60985798	A	C	c.102T>G	p.D34E	4.71 (74/1572)	-
ID37	Validation	*BCL2*	NM_000633.2	chr18	60985874	T	A	c.26A>T	P.Y9F	3.26 (63/1931)	4.77 (21/440)
ID37	Validation	*CREBBP*	NM_004380.2	chr16	3786703	T	A	c.4508A>T	p.Y1503F	36.15 (1624/4492)	5.71 (45/788)
ID37	Validation	*CREBBP*	NM_004380.2	chr16	3786801	G	T	c.4296C>A	p.H1432Q	35.80 (1142/3190)	11.42 (50/438)
ID38	Validation	*HIST1H1E*	NM_005321.2	chr6	26156976	G	A	c.358G>A	p.A120T	20.39 (300/1472)	10.66 (214/2007)
ID38	Validation	*HIST1H1E*	NM_005321.2	chr6	26157154	C	T	c.536C>T	p.A179V	25.74 (323/1255)	12.80 (182/1422)
ID38	Validation	*ZMYM3*	NM_005096.3	chrX	70465518	C	T	c.2824G>A	p.D956N	63.23 (724/1145)	14.68 (143/974)

ID38	Validation	*MYC*	NM_002467.4	chr8	128750638	G	A	c.130G>A	p.A44T	15.50 (93/600)	3.88 (36/929)
ID38	Validation	*MYC*	NM_002467.4	chr8	128750677	C	T	c.169C>T	P57S	14.54 (115/791)	4.06 (47/1159)
ID38	Validation	*B2M*	NM_004048.2	chr15	45003780	A	-CT	c.37_38delCT	p.L13fs*43	-	12.28 (57/464)
ID38	Validation	*B2M*	NM_004048.2	chr15	45007851	T	C	c.298T>C	p.C100R	-	14.89 (391/2626)
ID39	Validation	*CCND3*	NM_001760.4	chr6	41903752	C	+T	c.588_589insA	p.S142fs*36	35.61 (365/1025)	11.33 (176/1553)
ID39	Validation	*STAT6*	NM_003153.4	chr12	57496662	C	T	c.1255G>A	p.D419N	61.8 (1016/1644)	28.09 (704/2506)
ID41	Validation	*CD58*	NM_001779.2	chr1	117078586	C	T	c.628+1G>A		-	11.07 (113/1021)
ID41	Validation	*CD58*	NM_001779.2	chr1	117087028	A	-GG	c.267_268delCC	p.L90Hfs*10	-	8.86 (143/1614)
ID41	Validation	*KMT2D*	NM_003482.3	chr12	49418670	G	A	c.15844C>T	p.R5282*	21.96 (412/1876)	23.69 (245/1034)
ID41	Validation	*KMT2D*	NM_003482.3	chr12	49421105	C	A	c.14644G>T	p.E4882*	48.25 (690/1430)	40.55 (221/545)
ID41	Validation	*STAT6*	NM_003153.4	chr12	57496661	T	C	c.1256A>G	p.D419G	62.82 (2491/3965)	56.29 (1253/2226)
ID41	Validation	*CREBBP*	NM_004380.2	chr16	3900780	G	A	c.316C>T	p.Q106*	27.33 (473/1731)	28.52 (223/782)
ID42	Validation	*TNFRSF14*	NM_003820.2	chr1	2489256	G	T	c.161G>T	p.C54F	11.76(16/136)	-
ID42	Validation	*EZH2*	NM_004456.4	chr7	148506467	G	C	c.2045C>G	p.A682G	25.65 (198/772)	1.95 (40/2047)
ID42	Validation	*CREBBP*	NM_004380.2	chr16	3781322	C	-AAG	c.5040_5042delCTT	p.L1643-	20.67 (37/179)	1.34 (10/746)
ID42	Validation	*BCL2*	NM_000633.2	chr18	60985760	C	T	c.140G>A	p.G47D	31.84 (149/468)	-
ID43	Validation	*BRAF*	NM_004333.4	chr7	140476786	A	T	c.1620T>A	p.H540Q	20.37 (552/2710)	18.61 (470/2526)
ID43	Validation	*KMT2D*	NM_003482.3	chr12	49431895	G	A	c.9244C>T	p.R3082W	21.6 (604/2796)	16.72 (298/1782)
ID43	Validation	*CD36*	NM_001779.2	chr7	80303421	T	A	c.1377T>A	p.F459L	-	0.68 (20/2923)
ID44	Validation	*EZH2*	NM_004456.4	chr7	148508728	A	T	c.1936T>A	p.Y646N	NA	4.16 (58/1395)
ID44	Validation	*KMT2D*	NM_003482.3	chr12	49427403	A	-G	c.11084_11084delC	p.P3695fs*54	NA	7.2 (208/2893)
ID44	Validation	*KMT2D*	NM_003482.3	chr12	49436416	A	T	c.5795T>A	p.L1932*	NA	4.05 (159/3926)
ID44	Validation	*GNA13*	NM_006572.4	chr17	63010628	T	A	c.596A>T	p.D294V	NA	4.8 (272/5667)
ID44	Validation	*GNA13*	NM_006572.4	chr17	63010818	T	C	c.691A>G	p.K231E	NA	4.36 (223/5112)
ID44	Validation	*EP300*	NM_001429.3	chr22	41574806	T	G	c.7091T>G	p.F2364C	NA	2.07 (70/3388)
ID46	Validation	*KMT2D*	NM_003482.3	chr12	49432396	G	A	c.8743C>T	p.R2915*	NA	8.69 (269/3096)
ID47	Validation	*ITPKB*	NM_002221.3	chr1	226822466	C	T	c.2747G>A	p.W916*	2.76 (66/2393)	-
ID47	Validation	*MYC*	NM_002467.4	chr8	128753192	C	G	c.1353C>G	p.N451K	22.99 (1624/7065)	-
ID48	Validation	*SPEN*	NM_015001.2	chr1	16258541	C	T	c.5806C>T	p.R1936*	40.23 (4199/10437)	15.39 (591/3839)
ID48	Validation	*TNFAIP3*	NM_001270508.1	chr6	138199805	C	+A	c.1223_1224insA	p.E409Rfs*20	19.71 (2169/11003)	7.27 (286/3933)
ID48	Validation	*EZH2*	NM_004456.4	chr7	148508728	A	T	c.1936T>A	p.Y646N	22.02 (522/2371)	5.27 (53/1004)
ID48	Validation	*B2M*	NM_004048.2	chr15	45003773	T	G	c.29T>G	p.L10R	17.14 (817/4767)	7.8 (38/486)
ID48	Validation	*B2M*	NM_004048.2	chr15	45003779	T	C	c.35T>C	p.L12P	16.54 (830/5019)	8.59 (44/512)
ID48	Validation	*GNA13*	NM_006572.4	chr17	63052591	C	A	c.121G>T	p.E41*	39.83 (2297/5767)	11.81 (132/1118)
ID49	Validation	TP53	NM_000546.5	chr17	7577509	C	T	c.772G>A	p.E258K	81.38 (647/795)	8.64 (114/1319)
ID49	Validation	GNA13	NM_006572.4	chr17	63010412	A	G	c.812T>C	p.I366T	77.15 (1087/1409)	8.14 (310/3808)
ID49	Validation	BCL2	NM_000633.2	chr18	60985689	G	A	c.211C>T	p.P71S	65.98 (192/291)	-
ID50	Validation	TNFRSF14	NM_003820.2	chr1	2488174	T	A	c.69+2T>A		NA	12.33 (82/665)

ID50	Validation	MYD88	NM_001172567.1	chr3	38182032	C	G	c.656C>G	p.S219C	NA	13.03 (265/2034)
ID50	Validation	*KLHL6*	NM_130446.2	chr3	183209942	C	T	c.1639G>A	p.E547K	NA	12.41 (259/2087)
ID50	Validation	*FBXW7*	NM_033632.3	chr4	153247247	A	C	c.1027T>G	p.Y519D	NA	15.12 (608/4020)
ID51	Validation	*SPEN*	NM_015001.2	chr1	16203031	T	A	c.739T>A	p.S247T	-	1.44 (27/1879)
ID51	Validation	*SPEN*	NM_015001.2	chr1	16261464	A	T	c.8729A>T	p.E2910V	-	1.22 (29/2381)
ID51	Validation	*MYD88*	NM_001172567.1	chr3	38182641	T	C	c.818T>C	p.L265P	20.91 (413/1975)	8.86 (114/1286)
ID51	Validation	*HIST1H1E*	NM_005321.2	chr6	26157144	G	A	c.526G>A	p.A176T	-	1.58 (31/1965)
ID51	Validation	*PIM1*	NM_001243186.1	chr6	37139063	G	A	c.403G>A	p.E135K	21.72 (425/1957)	9.81 (116/1183)
ID51	Validation	*PIM1*	NM_001243186.1	chr6	37138549	G	A	c.356G>A	p.G119D	16.87 (122/723)	-
ID51	Validation	*CD79B*	NM_001039933.1	chr17	62006647	A	C	c.632T>G	p.I211R	24.05 (416/1730)	4.06 (50/1232)
ID51	Validation	*CD79B*	NM_001039933.1	chr17	62006674	A	G	c.605T>C	p.I202T	22.69 (435/1917)	4.76 (69/1451)
ID51	Validation	*CD79B*	NM_001039933.1	chr17	62006799	A	C	c.589T>G	p.Y197D	24.05 (404/1680)	7.49 (78/1042)
ID52	Validation	*ARID1A*	NM_006015.4	chr1	27101054	C	T	c.4336C>T	p.R1446*	12.13 (151/1245)	9.85 (91/924)
ID52	Validation	*ITPKB*	NM_002221.3	chr1	226924223	T	C	c.936_936delC	p.T313fs*86	12 (303/2524)	4.87 (75/1539)
ID52	Validation	*TNFAIP3*	NM_001270508.1	chr6	138196027	G	A	c.341G>A	p.G114D	24.35 (809/3322)	16.67 (231/1386)
ID52	Validation	*HIST1H1E*	NM_005321.2	chr6	26156852	C	A	c.234C>A	p.S78R	-	2.35 (34/1449)
ID52	Validation	*TNFAIP3*	NM_001270508.1	chr6	138200387	C	-G	c.1806_1806delG	p.T604fs*93	25.18 (1634/6489)	16.55 (455/2749)
ID52	Validation	*STAT6*	NM_003153.4	chr12	57496654	A	T	c.1263T>A	p.N421K	24.34 (572/2350)	11.61 (119/1025)
ID52	Validation	*TP53*	NM_000546.5	chr17	7578406	C	T	c.524G>A	p.R175H	28.71 (552/1923)	11.17 (86/770)
ID53	Validation	*PRDM1*	NM_182907.2	chr6	106553111	C	T	c.968C>T	p.T359M	NA	1.54 (42/2736)

FIGURES LEGENDS

Figure 1. Prevalence and molecular spectrum of non-synonymous somatic mutations discovered in plasma cfDNA. Percentage of training **(A)** and validation **(D)** DLBCL cases harboring non-synonymous somatic mutations of genes investigated by targeted-resequencing of plasma cfDNA. Color codes indicate the distribution of mutations among germinal center (GC) and non-GC DLBCL. The final number and prevalence of mutated cases is given for each gene. Molecular spectrum of non-synonymous somatic mutations identified in plasma cfDNA of the training **(B)** and validation **(E)** DLBCL cases compared to the molecular spectrum of non-synonymous somatic mutations that have been detected in tumor gDNA of published DLBCL series and reported in the COSMIC database (v76). Genes mutated in >10% cases of the DLBCL series are reported. Position and type of non-synonymous somatic mutations identified in plasma cfDNA of the training **(C)** and validation **(F)** DLBCL cases. Genes mutated in >10% cases are reported. Mutation maps were obtained through MutationMapper v1.0.1.

Figure 2. Overview of the DLBCL genotype discovered in tumor gDNA. Case-level mutational profiles of 20 training **(A)** and 16 validation **(B)** DLBCL tumors genotyped on gDNA from the tissue biopsy. Each column represents one tumor sample, each row represents one gene. The fraction of tumors with

mutations in each gene is plotted on the right. The number of mutations in a given tumor is plotted above the heat map. Cases are clustered according to the cell of origin.

Figure 3. Concordance between plasma cfDNA and tumor gDNA genotyping. Number of mutations in a given tumor discovered in plasma cfDNA and/or tumor gDNA in the training **(A)** and validation **(D)** DLBCL (cases provided of paired plasma cfDNA and tumor biopsy gDNA are represented). Mutations are color coded if they were identified in both plasma cfDNA and tumor biopsy gDNA (red), only in plasma cfDNA (blue), or only in tumor biopsy gDNA (grey). Venn diagram summarizing the overall number of mutations discovered in both plasma cfDNA and tumor biopsy gDNA (red), only in plasma cfDNA (blue), or only in tumor biopsy gDNA (grey) from the training **(B)** and validation **(E)** DLBCL. The corresponding overall sensitivity of plasma cfDNA genotyping in discovering biopsy-confirmed mutations is shown. Fraction of tumor biopsy-confirmed mutations that were detected in plasma cfDNA in the training **(C)** and validation **(F)** DLBCL. Patients are ordered by decreasing detection rates. The training cohort included two patients (ID2 and ID14) who were not informative because devoid of mutations within the target region. The black portion of the bars marks the prevalence of tumor biopsy-confirmed mutations that were detected in plasma cfDNA. The gray portion of the bars marks the prevalence of tumor biopsy-confirmed mutations that were not detected in plasma cfDNA.

Figure 4. Accuracy of plasma cfDNA genotyping vs tumor gDNA genotyping. The mutation abundance in plasma cfDNA vs the mutation abundance in tumor gDNA is comparatively represented in the scatterplot for each variant identified in the training **(A)** and validation **(C)** DLBCL cases. Receiver operating characteristic (ROC) illustrating the performance of plasma cfDNA genotyping in detecting biopsy confirmed tumor variants according to the variant allele frequency of mutations in tumor gDNA in the training **(B)** and validation **(D)** DLBCL cases. The bar graph shows the allele frequency in tumor gDNA of the variants that were discovered in plasma cfDNA (red bars) or missed in plasma cfDNA (grey bars). The dash line tracks the 20% variant allele frequency threshold.

Figure 5. Detection of a *FBXW7* mutation in genomic tumor DNA from DLBCL cells of the cerebrospinal fluid. The p.S688G *FBXW7* somatic mutation was detected in the tumor gDNA from DLBCL cells of the cerebrospinal fluid at 0.42% frequency above the background, passing the identification cut-offs established by Bonferroni-corrected probabilities.

Figure 6. Longitudinal assessment of mutation abundance in plasma cfDNA upon R-CHOP treatment. The graphs represent the variant allele frequency of non-synonymous somatic mutations in plasma cfDNA at baseline before treatment start, before each R-CHOP course, and at the end of treatment. Responding **(A)** and non-responding **(B)** patients are pooled in the graphs. Each dot represents one

single individual mutation. Among responding patients **(A)**, baseline mutations (n=102; mean allele frequency 12.6%, range 0.17%-80.4%) disappeared from plasma cfDNA. Among non-responding patients **(B)**, baseline mutations (n=27; mean allele frequency 14.3%, range 0.14-53.9%) remain detectable in plasma cfDNA before R-CHOP cycle 2 (mean allele frequency 20.2%, range 0.65%-65.3%), before R-CHOP cycle 3 (mean allele frequency 18.8%, range 1.1-61.1%), and at the end of treatment (mean allele frequency 8.2%, range 0.1-34.2%).

Figure 7. Non-invasive real-time monitoring of the DLBCL clonal evolution on plasma cfDNA. Sequential analysis of plasma cfDNA collected at presentation, upon treatment and at the end of treatment. Disease response assessment by CT/PET is also shown at the corresponding timepoints. Representative patients are shown. Among non-responding patients (ID13 and ID12), baseline mutations remain detectable in plasma cfDNA. Among relapsing patients, baseline mutations cleared from plasma cfDNA and appeared again at the time of relapse (ID1) or persisted despite clinical remission (ID9). Color codes indicate different variants. Gene mutations selected by treatment are highlighted.

Figure 1. Prevalence and molecular spectrum of non-synonymous somatic mutations discovered in plasma cfDNA.

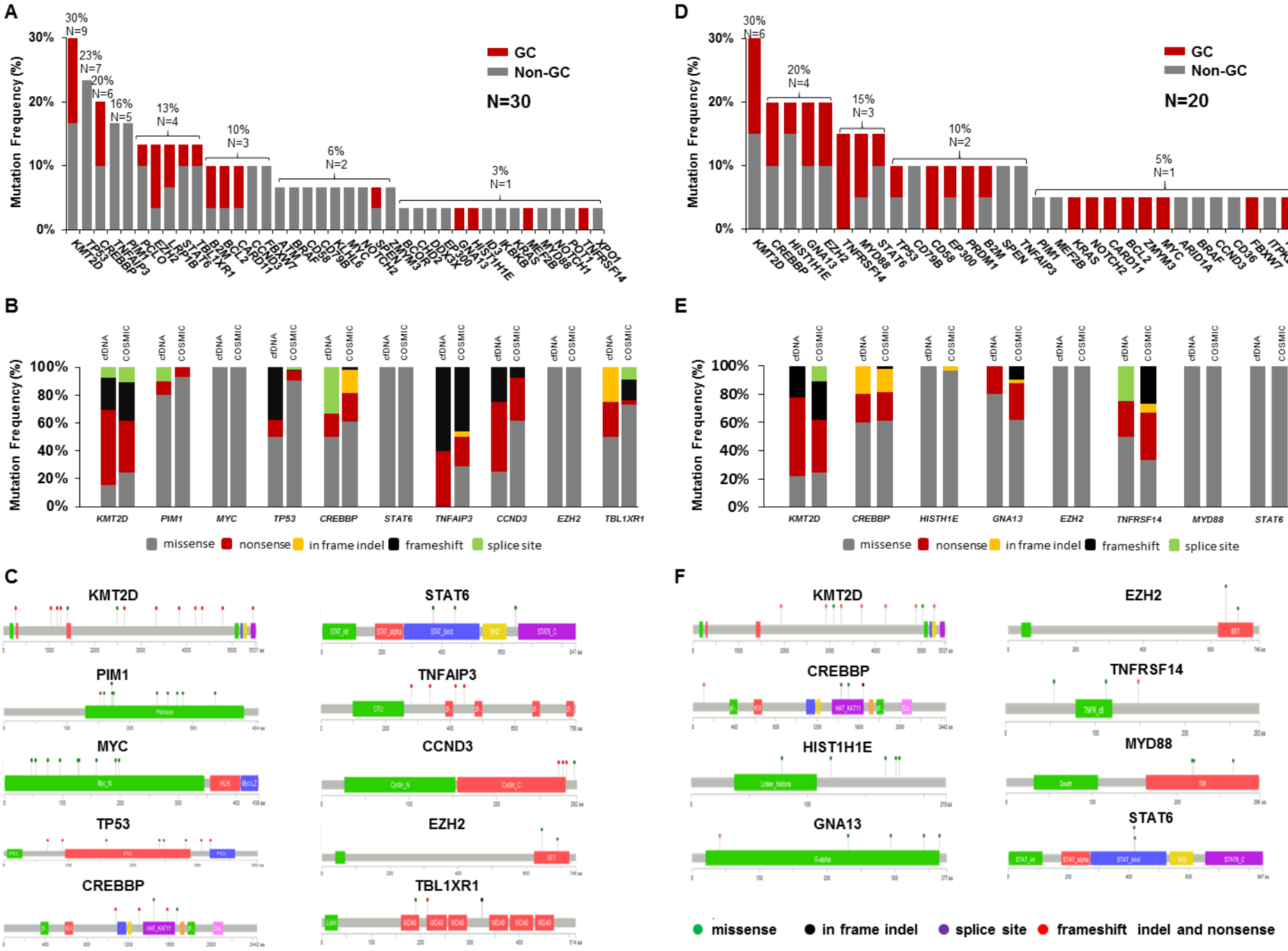


Figure 2. Overview of the DLBCL genotype discovered in tumor gDNA.

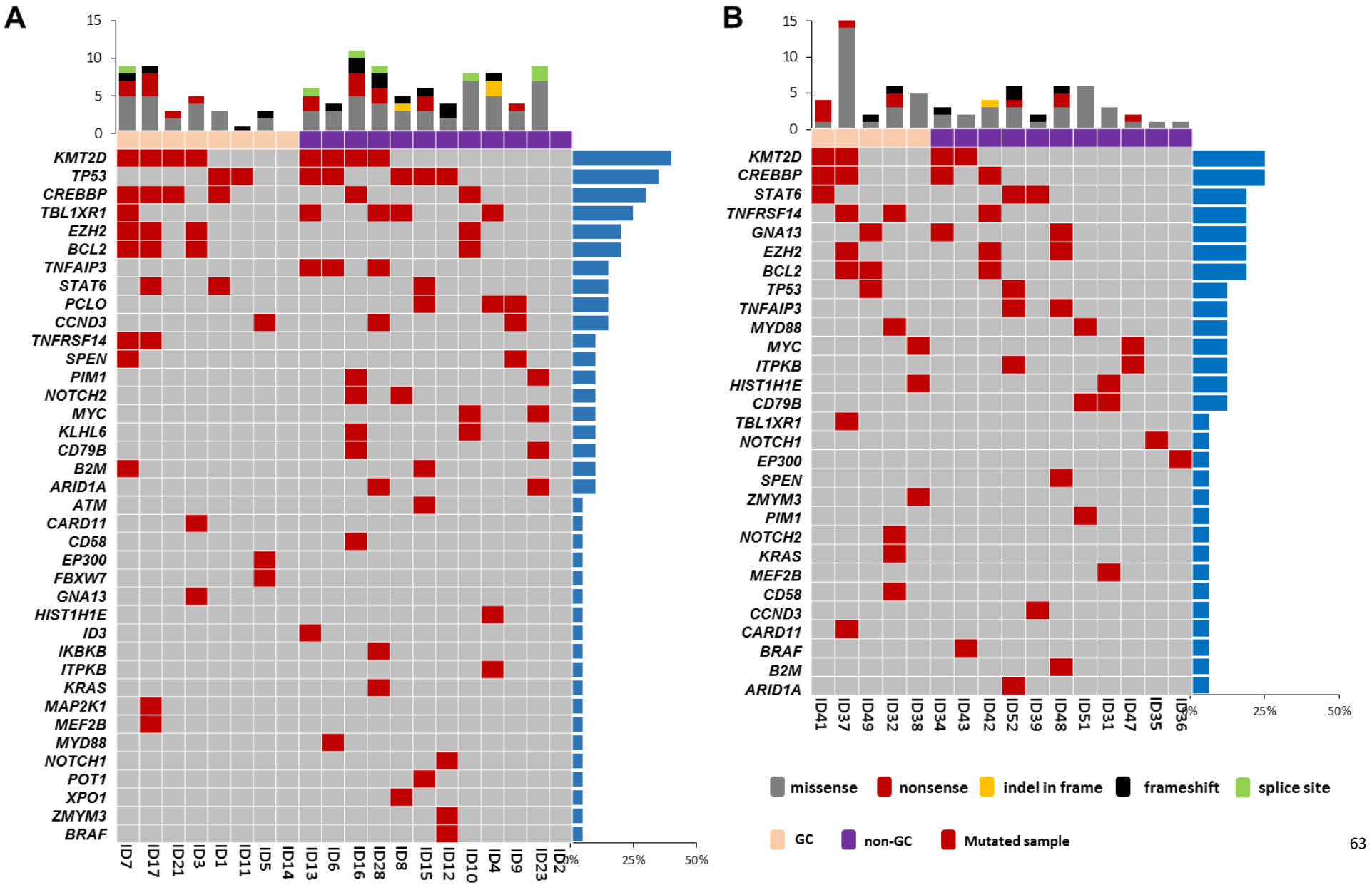


Figure 3. Concordance between plasma cfDNA and tumor gDNA genotyping.

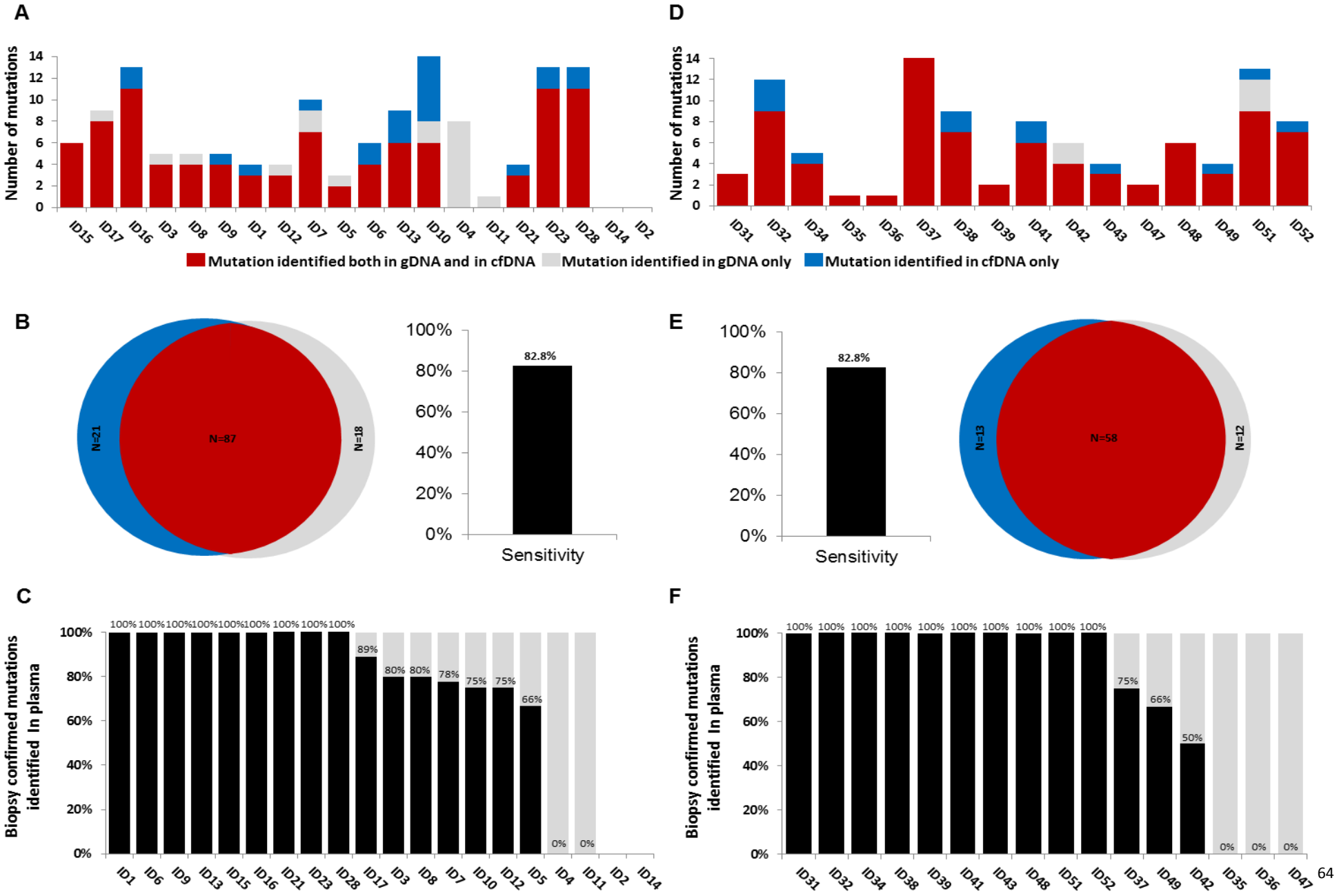


Figure 4. Accuracy of plasma cfDNA genotyping vs tumor gDNA genotyping.

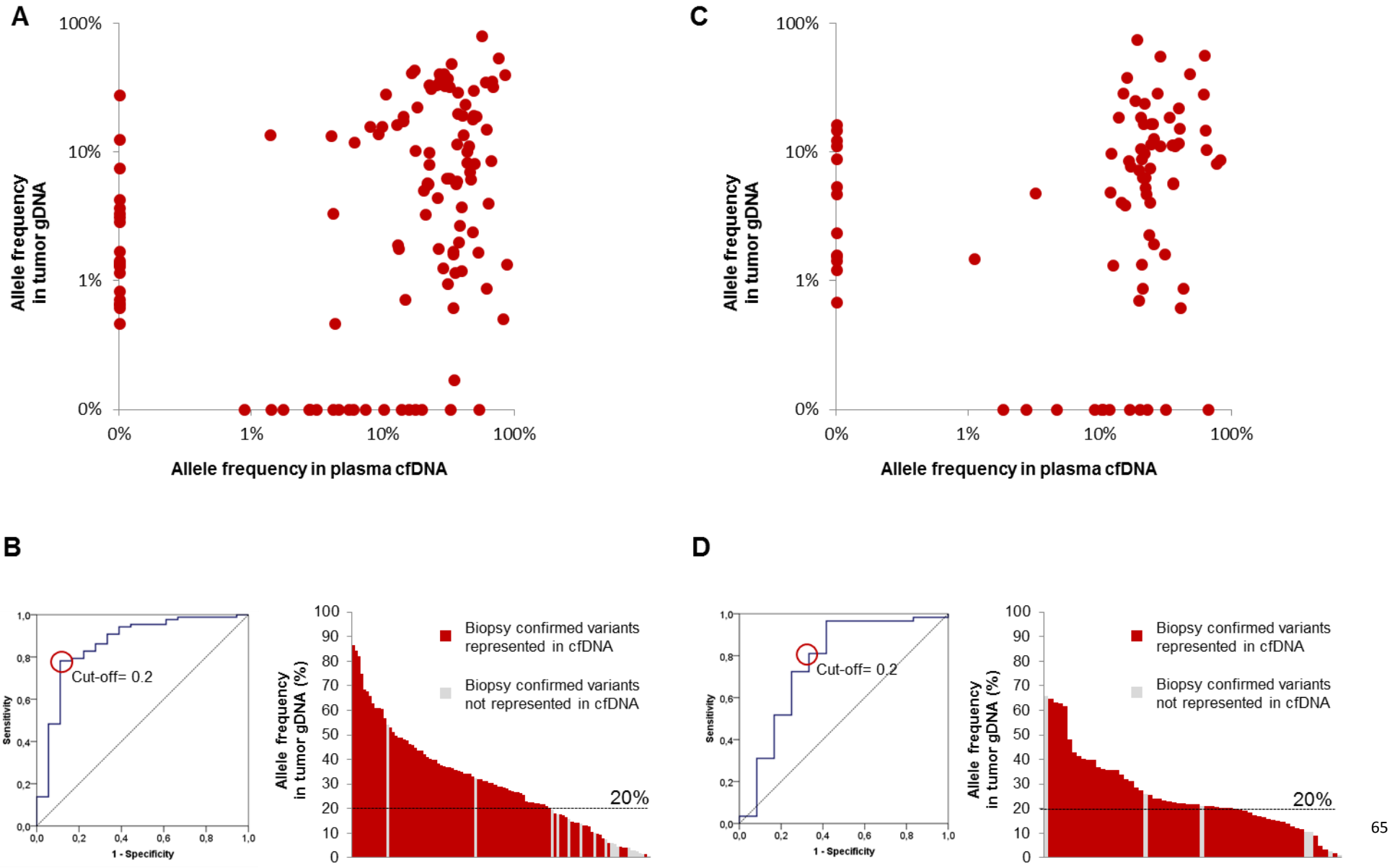


Figure 5. Detection of a *FBXW7* mutation in genomic tumor DNA from DLBCL cells of the cerebrospinal fluid.

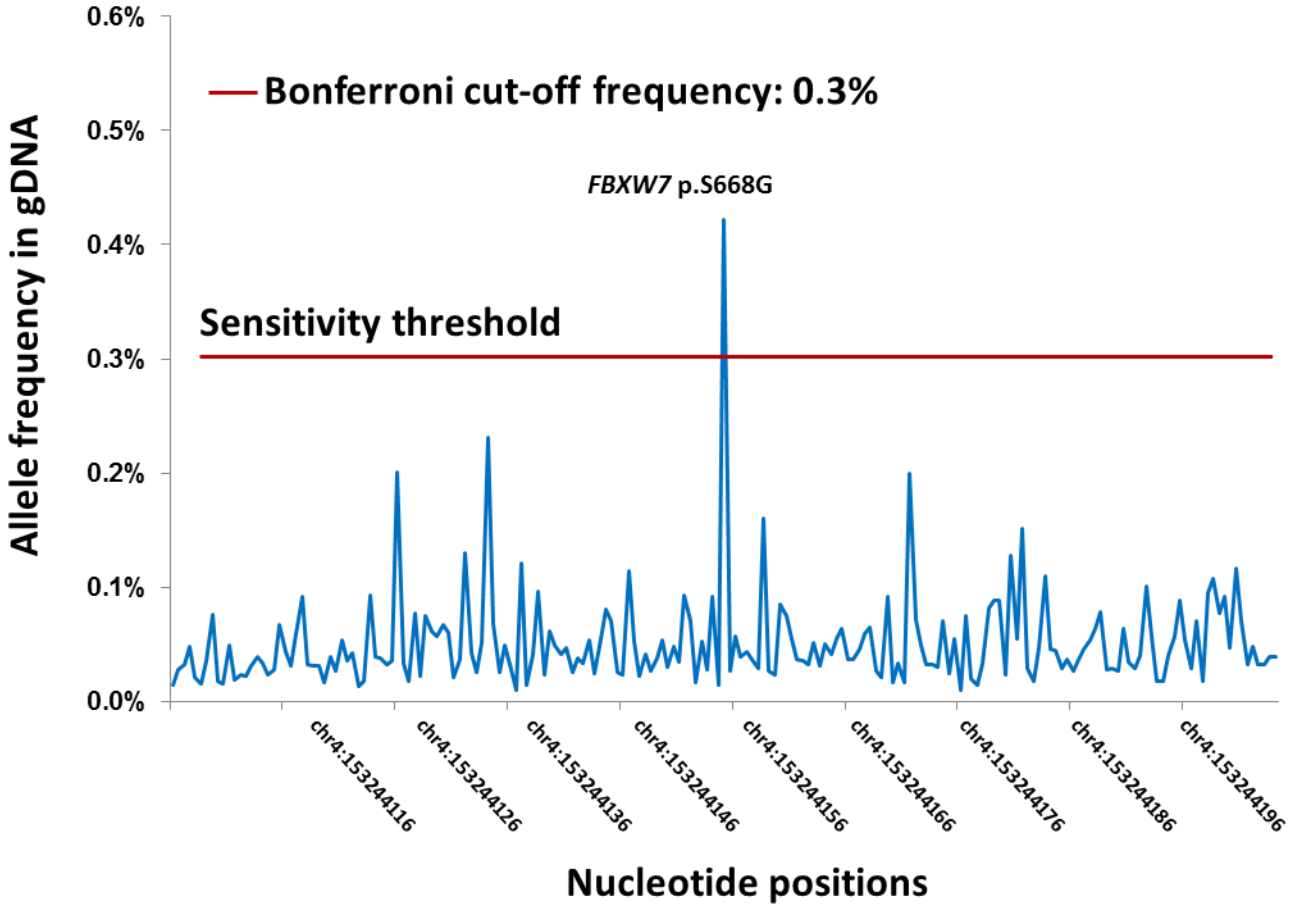


Figure 6. Longitudinal assessment of mutation abundance in plasma cfDNA upon R-CHOP treatment.

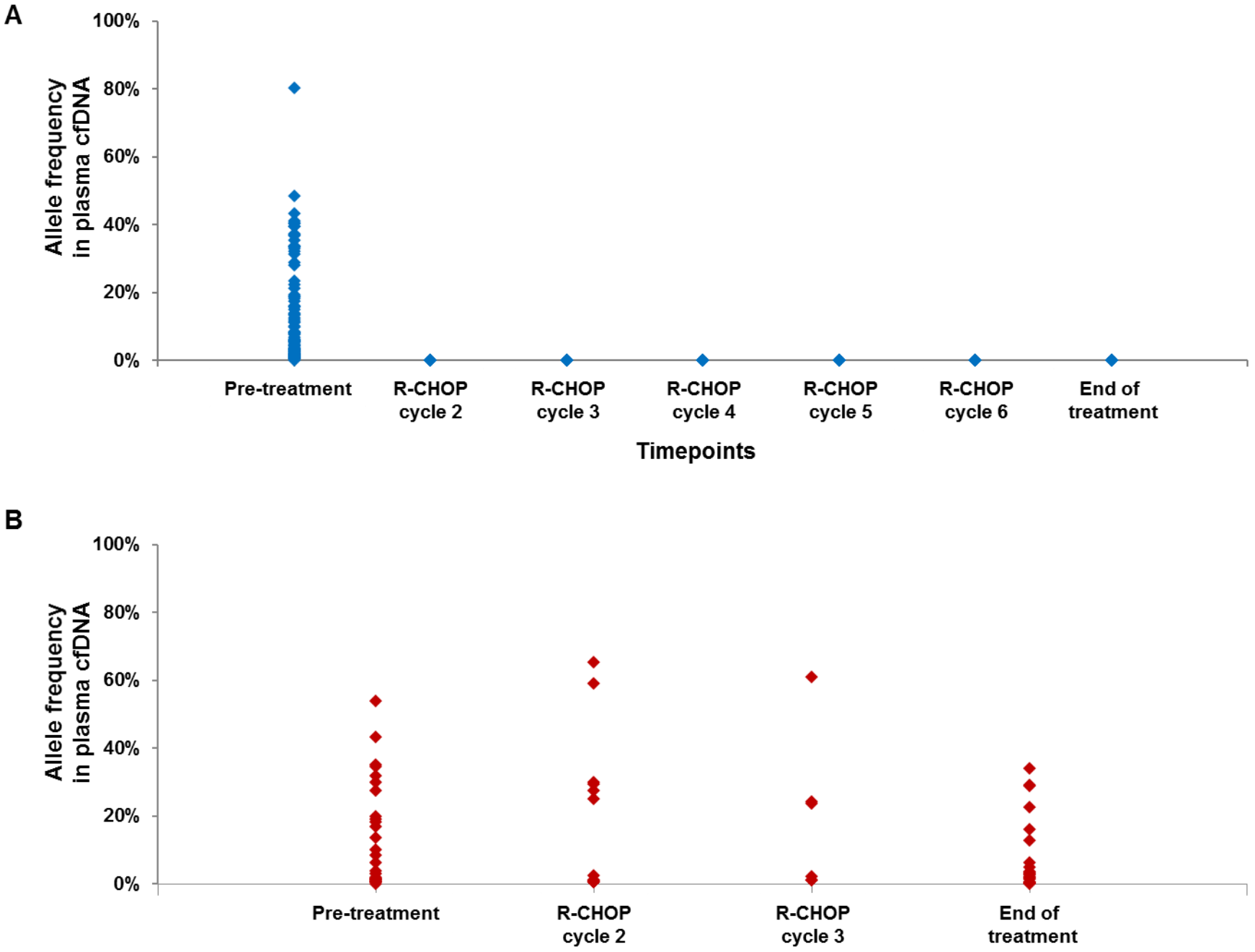
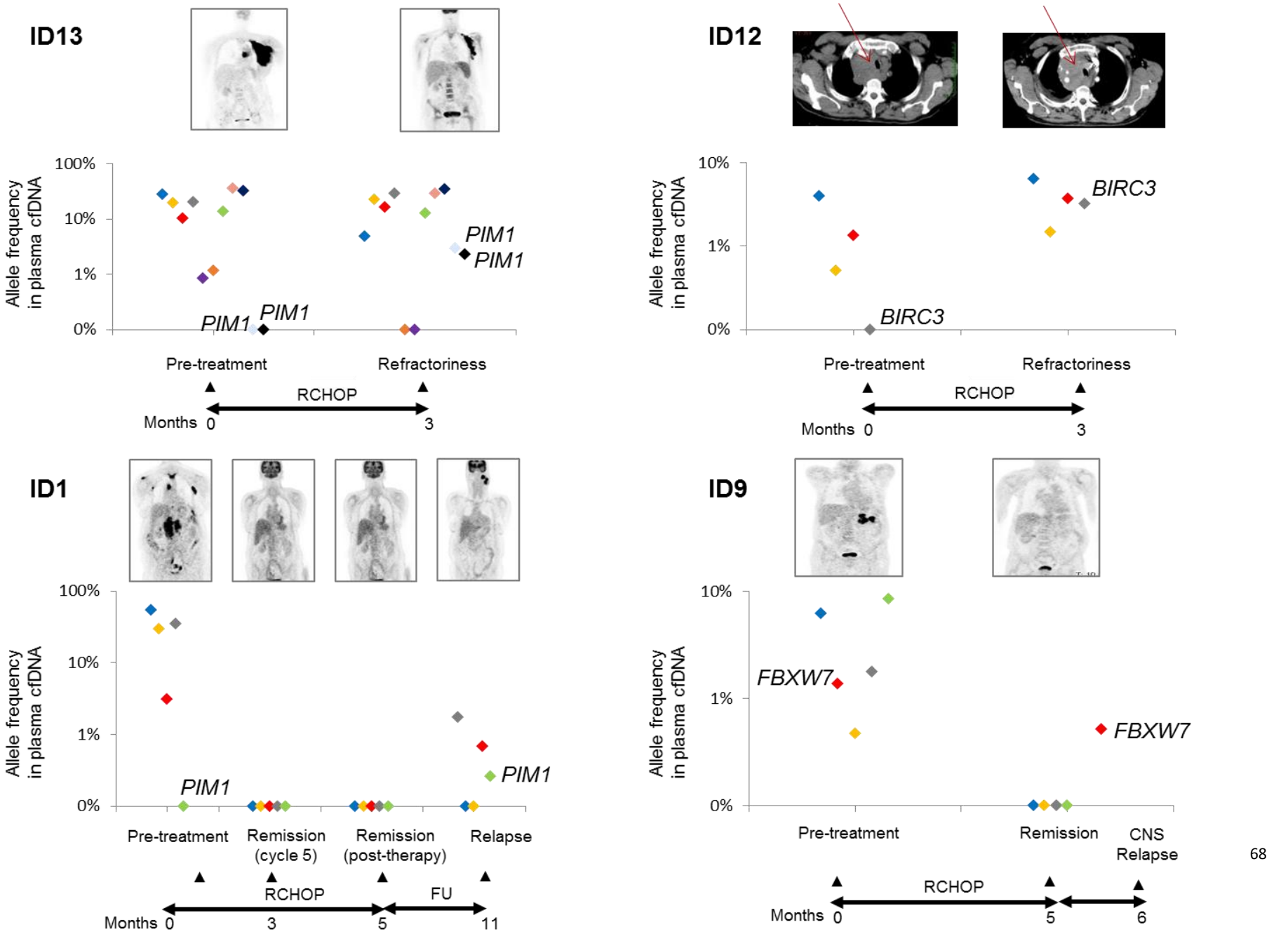


Figure 7. Non-invasive real-time monitoring of the DLBCL clonal evolution on plasma cfDNA.



SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Experimental validation of plasma cfDNA mutations identified by CAPP-seq. The x axis shows the variant allele frequency of the non-synonymous somatic mutations by the discovery CAPP-seq experiment. The y axis shows the variant allele frequency of the non-synonymous somatic mutations by the validation CAPP-seq experiment. Each dot represents one single individual mutation. The coefficient of correlation (R2) between the two experiments is shown. Panel A shows the biopsy confirmed variants. Panel B shows the variants discovered in plasma cfDNA but not in tumor gDNA.

Supplementary Figure 2. Sanger sequencing validation of mutations discovered in plasma cfDNA. Sanger sequencing was used to validate the most abundant plasma cfDNA mutations detected by CAPP-seq. Exemplificative results are shown. The electropherograms show plasma cfDNA mutations that have a variant allele frequency >8%.

Supplementary Figure 3. Analysis of plasma cfDNA from healthy donors by CAPP-seq. (A) The three peaks represent kernel distribution density plots of the frequencies of allele fractions in plasma cfDNA from DLBCL patients. By their low concentrations, somatic mutations (far left-hand small peak) can be generally distinguished from heterozygous (middle peak) and homozygous (right-hand peak) germline SNV listed in the

dbSNP138. The bottom dot plot represents the allele frequency of variants (in red somatic mutations, in grey germline SNV) **(B)** The two peaks represent kernel distribution density plots of the allele fractions in healthy donors. Heterozygous (middle peak) and homozygous (right-hand peak) germline SNV listed in the dbSNP138 are shown. Conversely, no far left-hand small peaks are visible indicating that the CAPP-seq approach does not pick up biological or analytical background noise in cfDNA. Plots are derived by deflagging dbSNP138 polymorphisms from filters applied to the final list of variants called by the algorithm. The bottom dot plot represents the allele frequency of variants (in grey germline SNV).

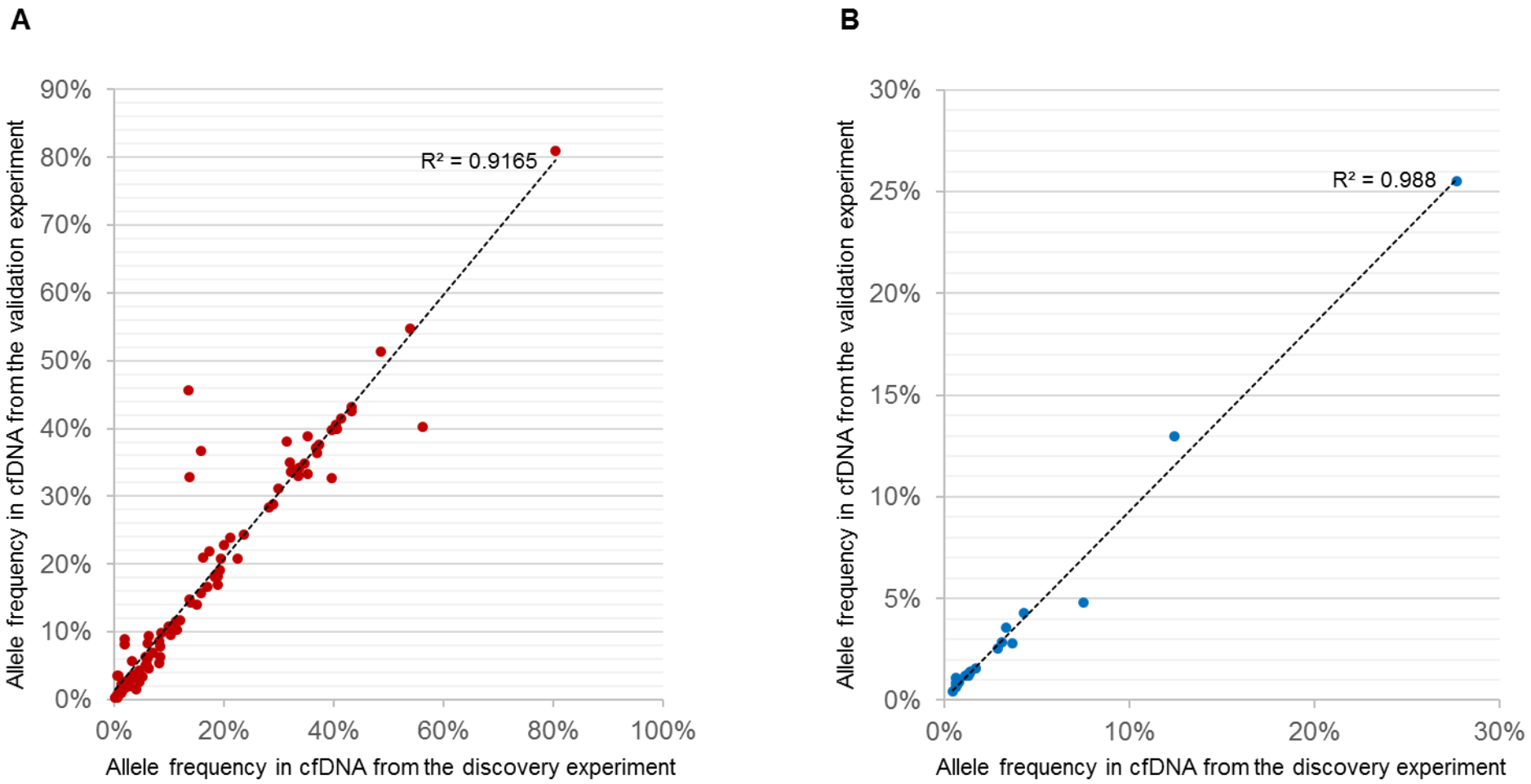
Supplementary Figure 4. Pairwise CAPP-seq of samples provided with both fresh and FFPE tissue biopsies. The graph comparatively shows the allele frequency of mutations in gDNA from paired fresh/FFPE samples processed from the same DLBCL biopsy.

Supplementary Figure 5. Coverage across the target region. Depth of coverage (y axis) across the target region (x axis) by CAPP-seq of plasma cfDNA from the training DLBCL (n=30) **(A)**, of tumor gDNA from fresh biopsies of the training DLBCL (n=17) **(B)**, of tumor gDNA from FFPE biopsies of the training DLBCL (n=3) **(C)**, of plasma cfDNA from the validation DLBCL (n=20) **(D)**, of tumor gDNA from fresh biopsies of the validation DLBCL (n=8) **(E)**, of tumor gDNA from FFPE biopsies of the validation DLBCL (n=8) **(F)**. Each dot represents the

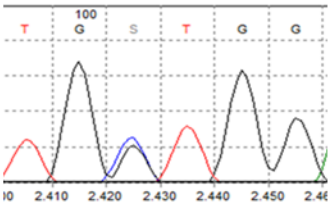
sequencing depth on that specific position of the target region of one single individual sample. The solid blue line shows the median depth of coverage, while the dash blue lines show the interquartile range.

Supplementary Figure 6. Correlation between tumor cfDNA load in plasma and clinical indices. Panel A shows disease stage. Panel B shows LDH levels. The band inside the box is the median tumor cfDNA load in plasma across patients, the bottom and top of the box are the first and third quartiles, and the ends of the whiskers represent the range. For each patient, the variant showing the larges allele frequency in plasma was considered as the marker of the tumor cfDNA load in plasma.

Supplementary Figure 1. Experimental validation of plasma cfDNA mutations identified by CAPP-seq.

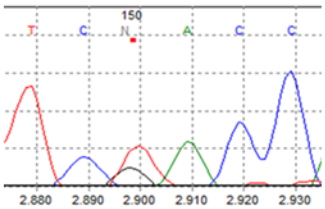


Supplementary Figure 2. Sanger sequencing validation of mutations discovered in plasma cfDNA.



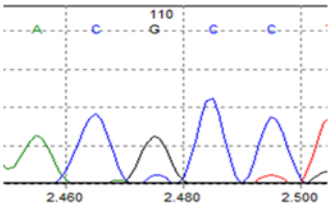
39.6%

PIM1 c.727 C>G p.L243V

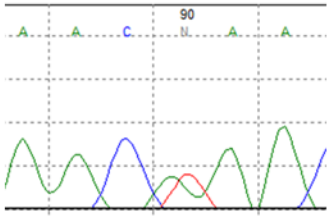


19.38%

MYC c.193 T>G p.Y47D

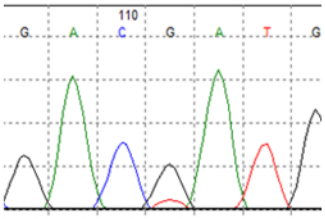


16.21%



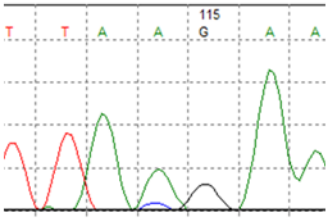
48.55%

TNFAIP3 c.1250 A>T p.K417*



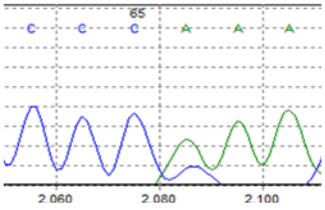
17.39%

ARID1A C.5165 G>T p.R1722L



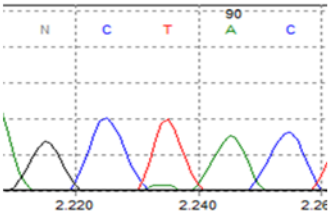
21.2%

LRP1B c.5115-2 A>C



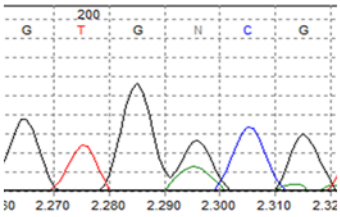
35.3%

KMT2D c.13032insC p.K4345fs*27

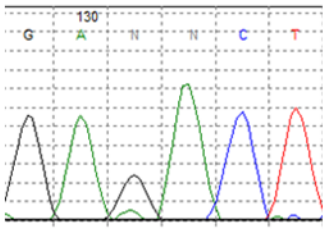


22.44%

IKBKB c.578 T>A p.L193Q

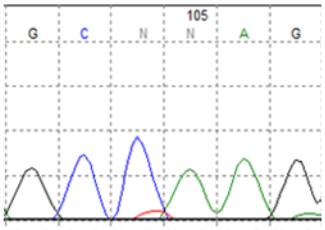


32.85%



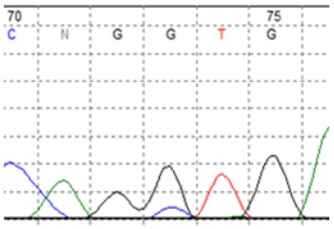
18.80%

TBL1XR1 c.567 G>A p.D190N



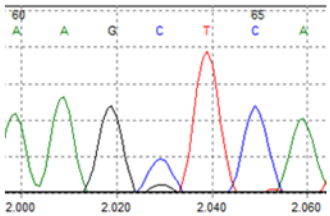
31.31%

CCND3 c.825 C>T p.Q276*



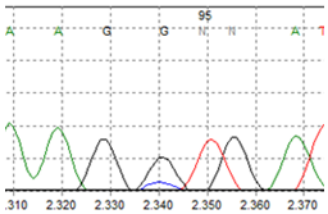
28.94%

ARID1A c.5124+1 G>C



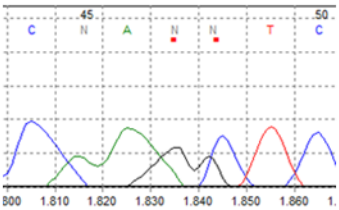
12.45%

PIM1 c.823 C>G p.L275V



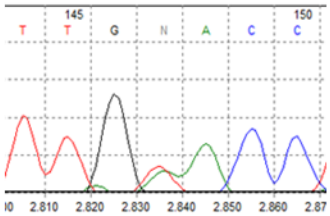
23.53%

CD79B c.549+1 G>C



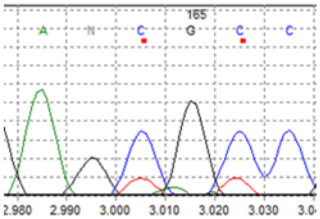
18.17%

MYC c.475 C>G p.L159V



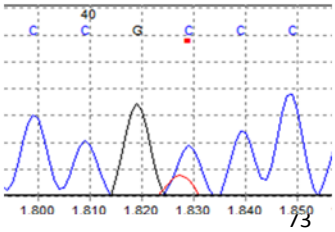
10.14%

MYC c.575 T>A p.Y192N



11.11%

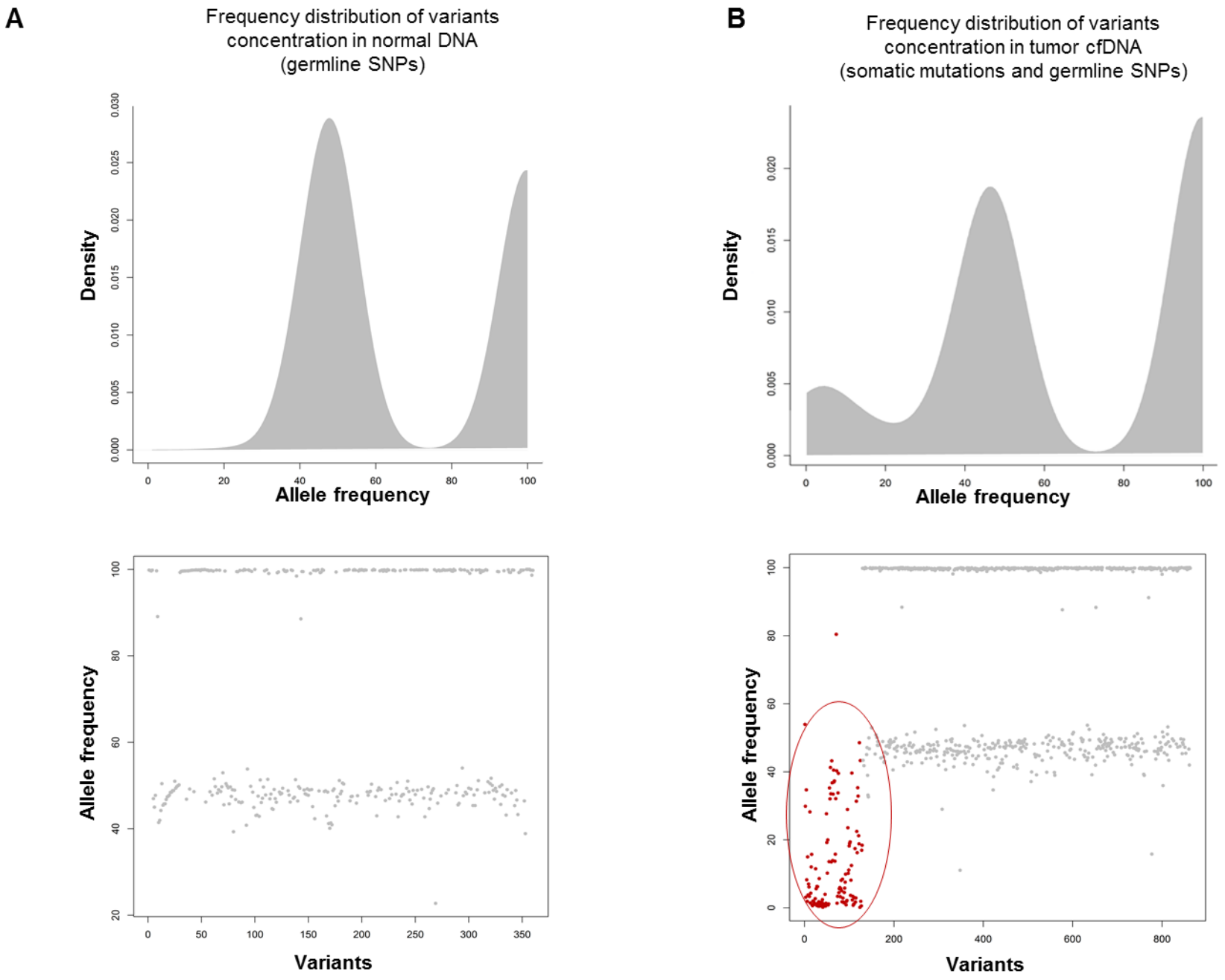
MYC c.593 C>T p.A198V



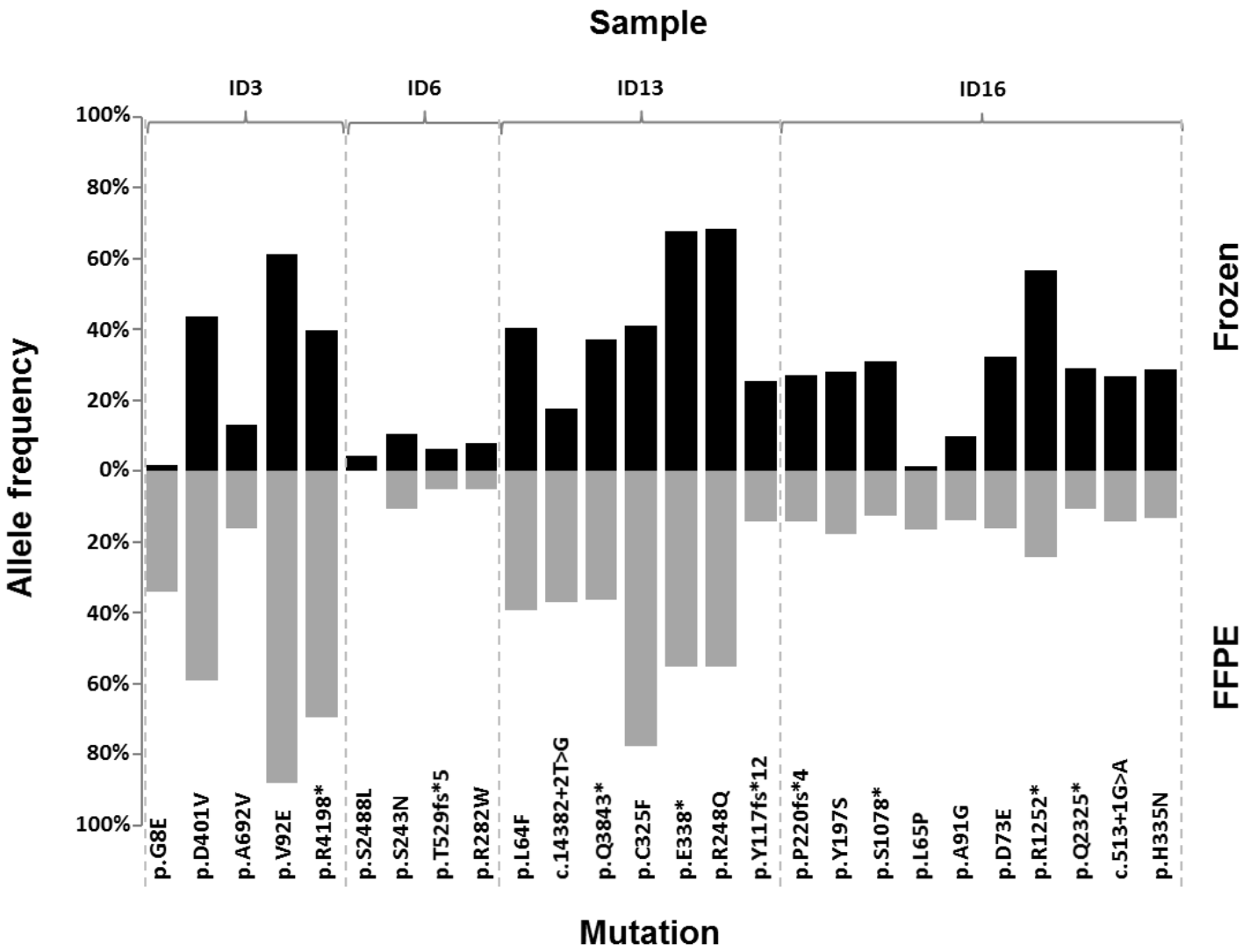
8.12%

MYC c.223C>T p.P75S

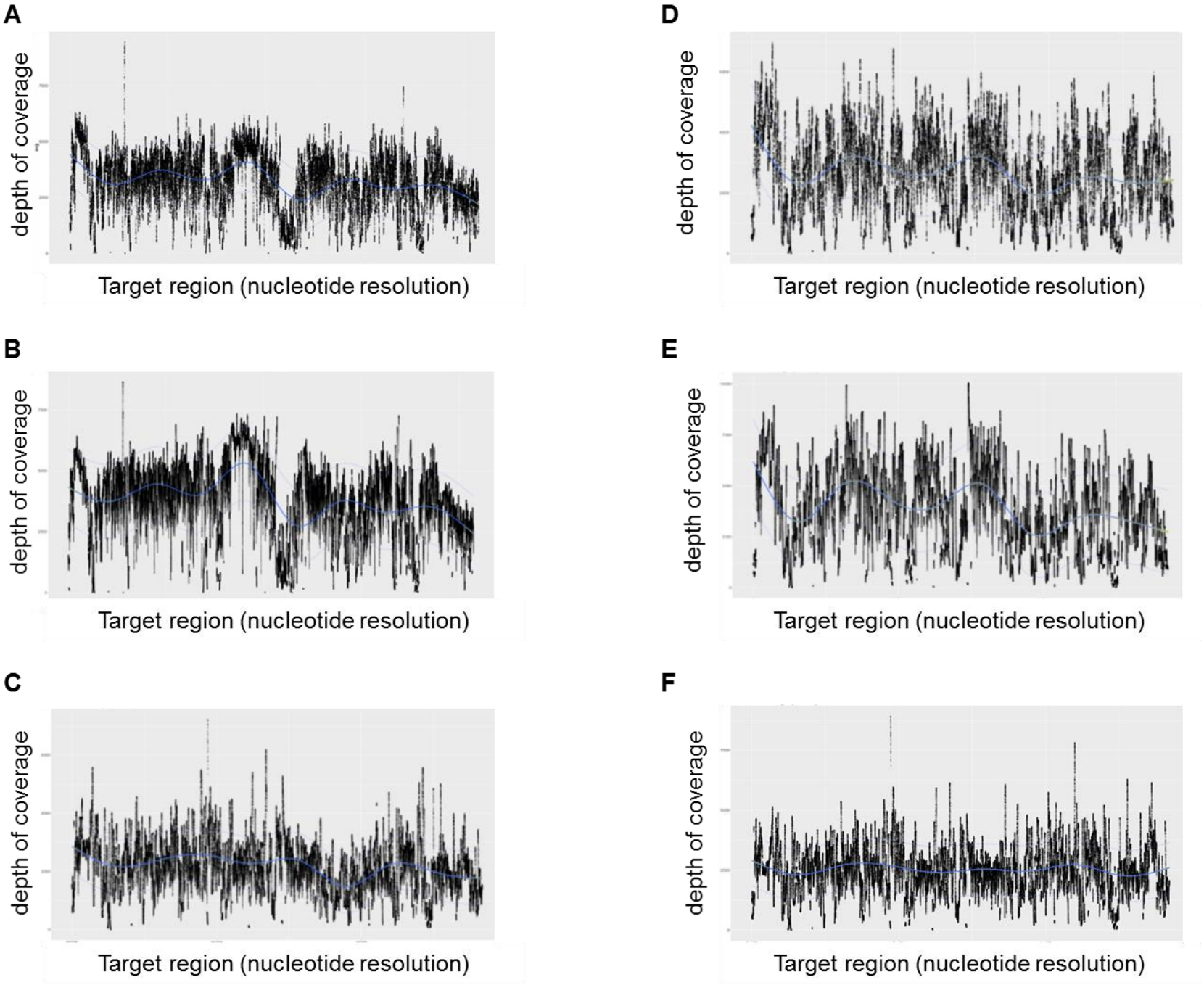
Supplementary Figure 3. Analysis of plasma cfDNA from healthy donors by CAPP-seq.



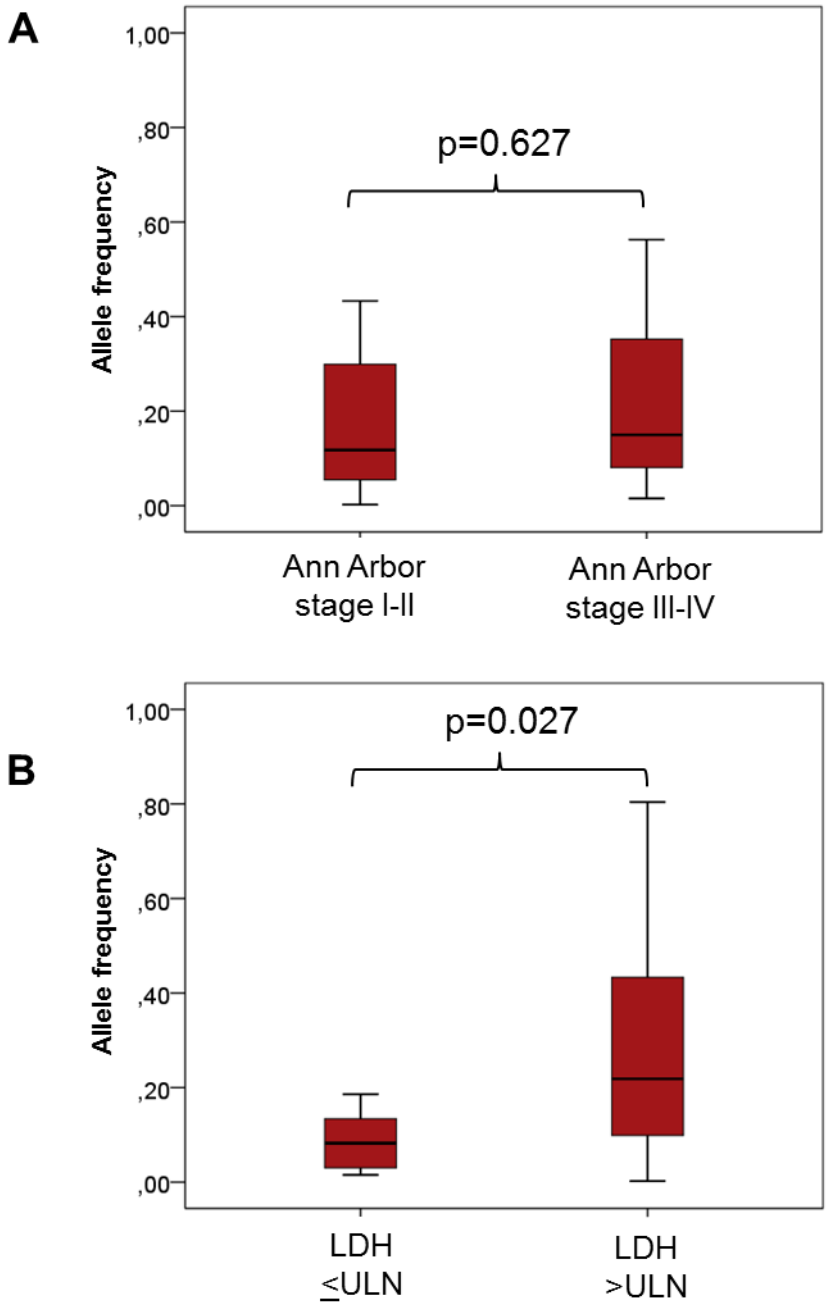
Supplementary Figure 4. Pairwise CAPP-seq of samples provided with both fresh and FFPE tissue biopsies.



Supplementary Figure 5. Coverage across the target region.



Supplementary Figure 6. Correlation between tumor cfDNA load in plasma and clinical indices.



RINGRAZIAMENTI

Dopo tre lunghi e intensi anni, finalmente termino il dottorato di ricerca. È stato un periodo di profondo apprendimento, non solo a livello scientifico, ma anche personale. Vorrei spendere due parole di ringraziamento nei confronti di tutte le persone che mi hanno sostenuto, sopportato e aiutato durante questo periodo.

Al Professor Gianluca Gaidano, Direttore del Laboratorio di Ematologia presso il Dipartimento di Medicina Traslazionale dell'Università del Piemonte Orientale, relatore di questa tesi di laurea e mio mentore, per l'aiuto, il sostegno e la disponibilità fornitomi in tutti questi anni e la grande conoscenza che mi ha donato.

A tutti i miei colleghi/amici: Chiara Favini, Simone Favini, Denise Peroni, Silvia Rasi, Sruthi Sagiraju, Riccardo Moia e Clara Deambrogi. Ci siamo sempre sostenuti a vicenda, nella buona e nella cattiva sorte, sia durante le fatiche e lo sconforto che hanno caratterizzato il mio percorso, sia nei momenti di gioia e soddisfazione. Quindi grazie di esserci stati fino all'ultimo. Avete avuto un peso determinante nel conseguimento di questo risultato, grazie infinite per la pazienza, la disponibilità, l'aiuto e tutti i preziosi consigli che mi avete elargito. Grazie per avermi aiutato a sopravvivere a questi tre anni. Vi voglio bene.

A tutti i collaboratori del laboratorio di Ematologia, Ramesh Adhinaveni, Ahad Kodipad, Davide Rossi, Alessio Bruscaggin, Valeria Spina, tutti i medici del reparto di ematologia dell'Ospedale maggiore della carità di Novara.

Alla mia famiglia per il sostegno durante questi anni. Un grazie particolare a mia madre per non aver mai smesso di credere in me e per non avermi permesso di smettere di credere in me stessa.

A tutti gli amministrativi e agli informatici, in particolare, Ingrid Cappa, Maurizio Pietroni, Ferdinando Maffia, Anna Rapa, Letizia Iadanza e Roberto Serra.

All'agenzia viaggi Novarseti, in particolare a Giusi Ferraris per tutto l'aiuto e la pazienza.

All'Associazione Italiana contro le Leucemie, Linfomi e Mieloma (A.I.L.) per l'aiuto e il sostegno che mi ha dato nella mia crescita professionale.

Per ultimi ma non meno importanti, i miei amici (tutti quanti!!). Grazie per la vostra costante ma discreta presenza.