

Shallow-depth sequencing of cell-free DNA for Hodgkin and diffuse large B-cell lymphoma (differential) diagnosis: a standardized approach with underappreciated potential



Lennart Raman,^{1,2,*} Malaika Van der Linden,^{1,2,*} Ciel De Vriendt,^{3,*} Bliede Van den Broeck,⁴ Kristoff Muylle,⁵ Dries Deeren,⁶ Franceska Dedeurwaerdere,⁷ Sofie Verbeke,¹ Amélie Dendooven,^{1,8} Katrien De Grove,³ Saskia Baert,³ Kathleen Claes,² Björn Menten,² Fritz Offner^{3#} and Jo Van Dorpe^{1#}

¹Department of Pathology, Ghent University, Ghent University Hospital, Ghent; ²Center for Medical Genetics, Department of Biomolecular Medicine, Ghent University, Ghent University Hospital, Ghent; ³Department of Clinical Hematology, Ghent University, Ghent University Hospital, Ghent; ⁴Department of Nuclear Medicine, Ghent University, Ghent University Hospital, Ghent; ⁵Department of Nuclear Medicine, AZ Delta, Roeselare; ⁶Department of Hematology, AZ Delta, Roeselare; ⁷Department of Pathology, AZ Delta, Roeselare and ⁸Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium

*LR, MVDL and CDV contributed equally as co-first authors.

#JVD and FO contributed equally as co-senior authors.

Haematologica 2022
Volume 107(1):211-220

ABSTRACT

Shallow-depth sequencing of cell-free DNA, an inexpensive and standardized approach to obtain molecular information on tumors non-invasively, has been insufficiently explored for the diagnosis of lymphoma and disease follow-up. This study collected 318 samples, including longitudinal liquid and paired solid biopsies, from a prospectively-recruited cohort of 38 Hodgkin lymphoma (HL) and 85 aggressive B-cell non-HL patients, represented by 81 diffuse large B-cell lymphoma (DLBCL) cases. Following sequencing, copy number alterations and viral read fractions were derived and analyzed. At diagnosis, liquid biopsies showed detectable copy number alterations in 84.2% of HL patients (88.6% for classical HL) and 74.1% of DLBCL patients. Of the DLBCL patients, copy number profiles between liquid-solid pairs were highly concordant ($r=0.815\pm 0.043$); and, compared to tissue, HL liquid biopsies had abnormalities with higher amplitudes ($P=0.010$). This implies that tumor DNA is more abundant in plasma. Additionally, 39.5% of HL and 13.6% of DLBCL cases had a significantly elevated number of plasma Epstein-Barr virus DNA fragments, achieving a sensitivity of 100% compared to the current standard. A longitudinal analysis determined that, when detectable, copy number patterns were similar across (re)staging moments in refractory or relapsed patients. Further, the overall profile anomaly correlated highly with the total metabolic tumor volume ($P<0.001$). To conclude, as a proof of principle, we demonstrate that liquid biopsy-derived copy numbers can aid diagnosis: e.g., by differentiating HL from DLBCL, random forest modeling is represented by an area under the receiver operating characteristic curve of 0.967. This application is potentially useful when tissue is difficult to obtain or when biopsies are small and inconclusive.

Introduction

B-cell-derived lymphomas can develop at different stages along the B-lymphocyte differentiation pathways. This generates a range of tumor entities¹ including Hodgkin lymphoma (HL) as well as diverse non-HL (nHL) varieties, with diffuse large B-cell lymphoma (DLBCL) the most common nHL.

Correspondence:

JO VAN DORPE
Jo.VanDorpe@uzgent.be

Received: August 3, 2020.

Accepted: November 25, 2020.

Pre-published: December 10, 2020.

<https://doi.org/10.3324/haematol.2020.268813>

©2022 Ferrata Storti Foundation

Material published in *Haematologica* is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>.

Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



Both HL and DLBCL typically display distinct morphological, immunohistochemical and (epi)genetic features.²⁻⁵ Their high cell turnover makes them excellent candidates for 'liquid biopsy'-based research. Necrotic and apoptotic tumor cells have been shown to shed DNA⁶ into the peripheral circulation and this is informative with regard to diagnosis, prognosis and therapy.⁷ In theory, this would enable real-time (serial) sampling of the entire (often heterogeneous)⁸ genomic lymphoma architecture in a convenient and non-invasive manner through traditional blood sampling.

Recent studies⁹⁻¹¹ have focused on plasma cell-free DNA (cfDNA) analysis using ultra-deep targeted sequencing¹² for single nucleotide variant (SNV) and translocation detection. Overall, these studies have provided two main insights: firstly, that variant allele frequencies often correlate with disease status; and, secondly, that surveilling circulating tumor DNA (ctDNA) might outperform ¹⁸F-fluoro-2-deoxyglucose positron emission tomography/computed tomography (PET/CT) scans in terms of sensitivity. This latter finding holds a great deal of potential for relapse risk assessment by quantifying minimal residual disease.

Despite their potential, ultra-deep sequencing techniques are still expensive and require targeted panels. In contrast, shallow (coverage ~0.25x) whole-genome sequencing (sWGS) for primarily copy number detection is cheaper and fully operative in hospitals that offer non-invasive prenatal testing.¹³ One of the hallmarks of sWGS is its high specificity in identifying malignant cells: in comparison to SNV, large (i.e., >5 Mb) somatic copy number alterations are rarely detected in unaffected subjects¹⁴ whereas SNV accumulate over time. Worth mentioning is that this also occurs in hematopoietic stem cells, a phenomenon referred to as clonal hematopoiesis,¹⁵ potentially resulting in incorrect driver identifications.

sWGS of cfDNA may therefore have clinical potential, especially for patients with lesions that are difficult to biopsy (e.g. those in brain; deep lymph nodes in the thoracic cavity or abdomen); or when dealing with ambiguous PET/CT scans, as both imaging and clinical symptoms of lymphoma are often non-specific. Nevertheless, this approach has only been superficially investigated.¹⁶ We, therefore, evaluated whether sWGS could serve as a molecular test in addition to established diagnostic methods, using a diverse set of 123 prospectively recruited lymphoma patients, comprising baseline and longitudinal blood samples, supplemented with paired formalin-fixed paraffin-embedded (FFPE) biopsies.

Methods

Ethics statement

This study was approved by the institutional ethics committee at Ghent University Hospital (EC/2016/0307). Written informed consent was obtained from all patients.

Study population

Between January 2016 and November 2019, 123 lymphoma patients were recruited at Ghent University Hospital and AZ Delta Roeselare. The subtypes studied include 38 HL (3 nodular lymphocyte predominant HL; 23 nodular sclerosis classical HL [cHL]; 5 mixed cellularity cHL; 5 lymphocyte-rich cHL; 1 lym-

phocyte-depleted cHL; and 1 not otherwise specified [NOS]); 81 DLBCL (61 NOS; 4 Epstein-Barr virus [EBV]-positive; 10 primary mediastinal large B-cell lymphomas [PMBCL]; 2 high-grade B-cell lymphomas; 2 T-cell/histiocyte-rich large B-cell lymphomas; 1 intravascular large B-cell lymphoma; and 1 plasmablastic lymphoma) and four grey-zone lymphomas (GZL) (*Online Supplementary File S2: Tables S1 and S2*). Patients were included at the start, 'baseline', of a new line of therapy (i.e., baseline represents initial diagnosis for 115 cases; after ineffective treatment in 6 other cases; and at relapse for the 2 remaining cases). Although patient recruitment was random, the final cohort was enriched for HL patients.

Sample series

Liquid biopsies were collected at baseline for all patients (n=123) and a random selection of paired FFPE tissue was made (n=33); this was obtained from 15 HL and 18 DLBCL patients. Longitudinal liquid biopsies were taken following milestones in treatment with 93 sequenced for 31 patients. These were intentionally enriched for refractory or relapsed disease: interim evaluations after two and four cycles of ABVD in HL; interim evaluations after four cycles of R-CHOP in nHL; at restaging after ineffective treatment or relapse; following successful treatment; and every 6 months for patients in maintained complete remission (CR). For HL (n=9), six patients had obtained CR within a year of inclusion whereas three were refractory. For nHL (n=22), 11 patients reached CR within a year; six were refractory; and five relapsed.

Negative liquid biopsies were included from non-invasive prenatal assays (n=60), supplemented with benign FFPE tissue (n=9) as a control for the solid biopsies. In total, 318 samples were thus analyzed.

Clinical and pathological characteristics

Patient demographics and clinical variables were available through routine clinical practice. The total metabolic tumor volume (MTV) was measured by PET/CT using AW workstation semi-automated segmentation software (GE medical systems, Waukesha, WI, USA), with a standard uptake value cutoff of 2.5.

Pathology diagnosis included standard histological examination of FFPE biopsy material in combination with immunohistochemistry (IHC). Double/triple hit (*MYC*, *BCL2* and/or *BCL6*) and double expressor (*MYC* and *BCL2*) samples were detected using fluorescence *in situ* hybridization and IHC, respectively. The germinal center B-cell (GCB) cell of origin (COO) status was assigned through the Hans algorithm.¹⁷

When using chromogenic *in situ* hybridization (CISH), the presence of EBV-encoded RNA (EBER) was evaluated via INFORM EBER probes (Ventana Medical Systems, Tucson, AZ, USA). Selected liquid biopsies were additionally tested using the quantitative EBV-polymerase chain reaction (PCR) method, as established by Bordon *et al.*¹⁸

Sequencing and bioinformatic analysis

Each laboratory step and subsequent computational analysis can be found in *Online Supplementary File S1: Supplemental methods*. These include: copy number profiling;¹⁹ defining the tumor burden using the copy number profile abnormality (CPA) score²⁰ and estimated tumor fraction;²¹ methods used to derive viral read fractions of HIV-1, HIV-2, EBV, JC polyomavirus, human T-lymphotropic virus 1, human herpesvirus 8 and hepatitis C virus; random forest modeling²² of copy number profiles to predict tumor subtype; detection of copy number driver peaks;²³ and general statistical testing.

Results

Shallow-depth sequencing of cell-free DNA enables accurate copy number profiling in lymphoma

For healthy individuals, copy number profiles are assumed to be approximately 'flat', implying that large alterations are absent. In a previous study,²⁰ in order to statistically distinguish cancerous from control profiles, we developed a score that increases with rising deviance from the naturally-occurring healthy background variance: the CPA score. Based on this variable, 95/123 baseline liquid biopsies were detected as cancerous (Figure 1A; *Online Supplementary File S2: Table S3*). As expected, the score reached significantly higher ($P < 0.001$; Welch test) values in the lymphoma group compared to the control group. There was no significant difference ($P = 0.250$; Fisher exact test) in the fraction of abnormal observations between HL (32/38; 84.2%) and DLBCL (60/81; 74.1%), nor when specifically considering cHL (31/35; 88.6%; $P = 0.091$). At Ann Arbor stage I, only one of six liquid biopsies showed deviations whereas stage II to IV presented comparable detection rates (*Online Supplementary File S1: Figure S1*).

In order to confirm whether the copy number profiles derived from blood samples represent those from tissue biopsies, we executed a concordance analysis using 33 liquid-solid pairs (*Online Supplementary File 1: Liquid-solid pairs*). One representative DLBCL case, patient 59, demonstrated very similar deviations in both samples (Figure 1B), and the Pearson correlation between this patient's smoothed profiles amounted to 0.916. We can note, from left to right, four clusters distinguished in a corresponding scatter plot (Figure 1C): deletions, copy neutral bins, duplications and amplifications. It should be noted that the amplitudes of alterations, a concept defined as the absolute value of a segment's log₂ ratio,¹⁹ are more extreme in liquid than in tissue biopsies (Figure 1B), caused by a greater ctDNA fraction in patient 59's liquid biopsy. This is also indicated by a steeper identity line than least squares fit in the scatter plot (Figure 1C). Overall, when the CPA score in the liquid biopsy was abnormal, DLBCL pairs showed high concordance ($r = 0.815 \pm 0.043$; 95% confidence interval). However, this was not the case for HL ($r = 0.260 \pm 0.055$; 95% confidence interval) patients (Figure 1D).

Next, the amplitudes were compared between paired

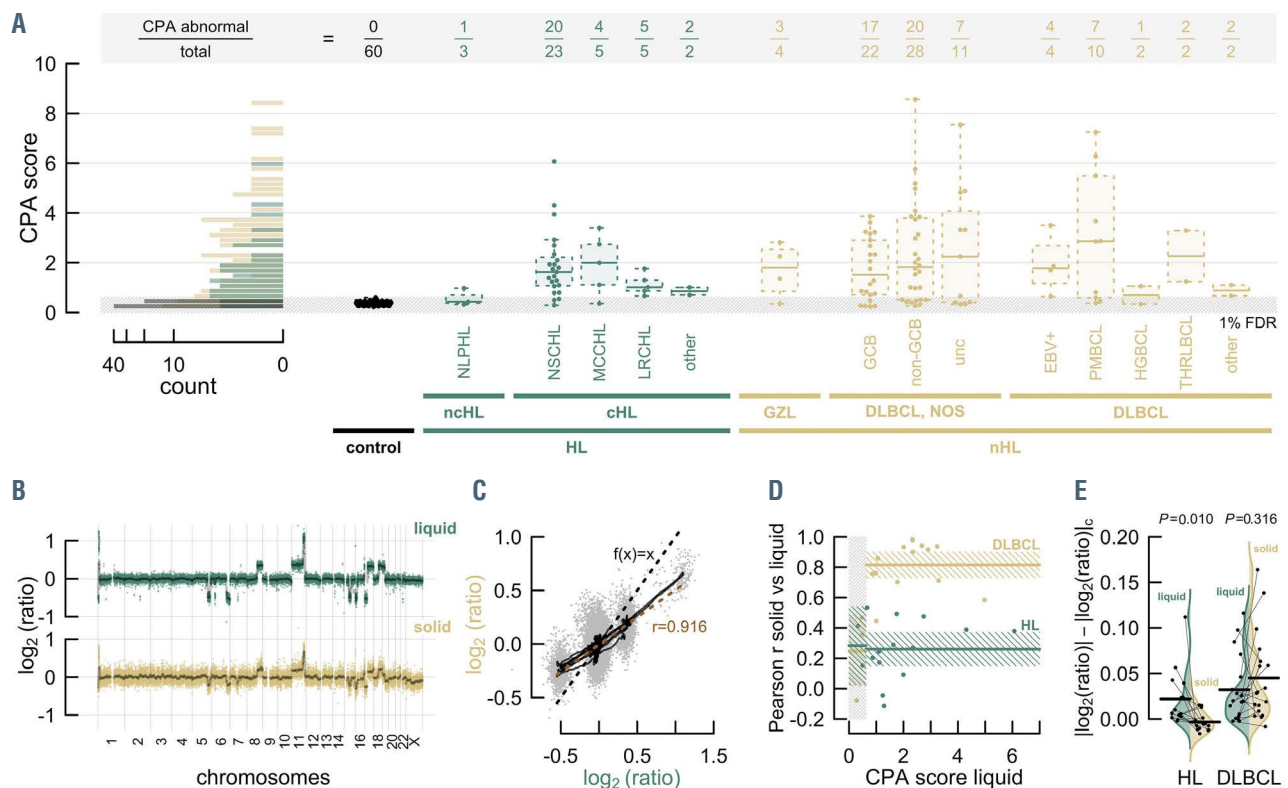


Figure 1. Feasibility of copy number profiling through sWGS of cfDNA. (A) CPA score comparison between common HL and DLBCL subtypes. Dots represent blood samples at baseline; default box plots indicate underlying distributions. The bottom grey box delineates the abnormality cutoff defined by the controls. According to this limit, the numbers at the top define which samples have abnormal copy number profiles. The overlapping histograms on the left display the overall CPA distribution. (B) Copy number profile comparison between the liquid and solid biopsy of patient 59. Dots are bins for which copy number is inferred whereas graph lines represent smoothed profiles. The 'ratio' equals the observed over the expected number of reads. (C) Scatter plot comparison of bins (grey dots) and smoothed values (solid black graph line) between the liquid and solid biopsy of patient 59. Identity line (black dotted) and least squares fit (red dotted), with corresponding Pearson correlation coefficient (r), are shown. (D) Scatter plot of all Pearson correlations between solid and liquid biopsy pairs, defined as in C, in relation to the CPA of the liquid biopsy. The grey box on the left matches the abnormality cutoff from A. Horizontal colored lines embedded in colored boxes represent means and their uncertainty (95% confidence interval), respectively. (E) Control-normalized amplitude comparison between the liquid and solid biopsies of HL and DLBCL patients. Samples are represented by dots; same-patient biopsies are connected; the underlying distribution is clarified by violin plots. P -values result from paired t -tests. sWGS: shallow whole-genome sequencing; cfDNA: cell-free DNA; CPA: copy number profile abnormality; HL: Hodgkin lymphoma; nHL: non-HL; DLBCL: diffuse large B-cell lymphoma; cHL: classical HL; nCHL: non-cHL; NLPHL: nodular lymphocyte predominant HL; NSCHL: nodular sclerosis cHL; MCCHL: mixed cellularity cHL; LRCHL: lymphocyte-rich cHL; GZL: grey zone lymphoma; NOS: not otherwise specified; GCB: germinal center B-cell; unc: unclassified; EBV: Epstein-Barr virus; PMBCL: primary mediastinal large B-cell lymphoma; HGBCL: high-grade B-cell lymphoma; THRLBCL: T-cell/histiocyte-rich large B-cell lymphoma; c, control.

liquid and solid biopsies (Figure 1E). These numbers were averaged per sample and additionally normalized by subtracting the mean across the control samples (0.005 for liquid; 0.026 for solid), which results in a fairer comparison, as FFPE material tends to produce profiles with higher levels of noise,²⁴ naturally increasing the overall deviance. This variable was found to be higher in liquid than in solid biopsies for HL ($P=0.010$; paired t -test), but not for DLBCL ($P=0.316$; paired t -test). Worthy of note is that, within HL, the mean of the solid biopsies barely varied from zero, meaning that most of these profiles are hard to differentiate from normal controls, and, therefore, the low liquid-solid correlation observed in HL patients (Figure 1D), seems to be caused by a scarcity of tumor cells in the FFPE samples rather than a low ctDNA fraction. From a biological perspective, this is reasonable: neoplastic Hodgkin and Reed-Sternberg cells are always embedded in an inflammatory background of non-affected cells,²⁵ causing resemblance to healthy tissue.

Shallow-depth sequencing of cell-free DNA enables sensitive Epstein-Barr virus detection in lymphoma

Sequencing reads were mapped to seven different lymphoma-associated viruses. Of the lymphoma liquid biopsies, 39/216 had abnormally-elevated EBV read fractions compared to the negative controls (*Online Supplementary File S2: Table S3*); whereas one patient tested positive for JC polyomavirus (13 reads) at surveillance (patient 79). In accordance with clinical records, no other viruses were detected. Within control samples, no or exclusively low amounts of EBV reads were noted. The latter was the case for 15/60 liquid (read fraction range 0.032-0.059 parts per million [ppm]) and 6/9 solid controls (0.029-0.08 ppm).

The EBV read fraction was evaluated for all baseline liquid biopsies, with 28/123 patients testing positive (Figure 2A). For GZL, 2/4 patients were positive, and, as expected, more HL samples were EBV-positive (15/38; 39.5%) than DLBCL samples (11/81; 13.6%) ($P=0.004$; Fisher exact test). All but one EBV-positive sample had an abnormal CPA

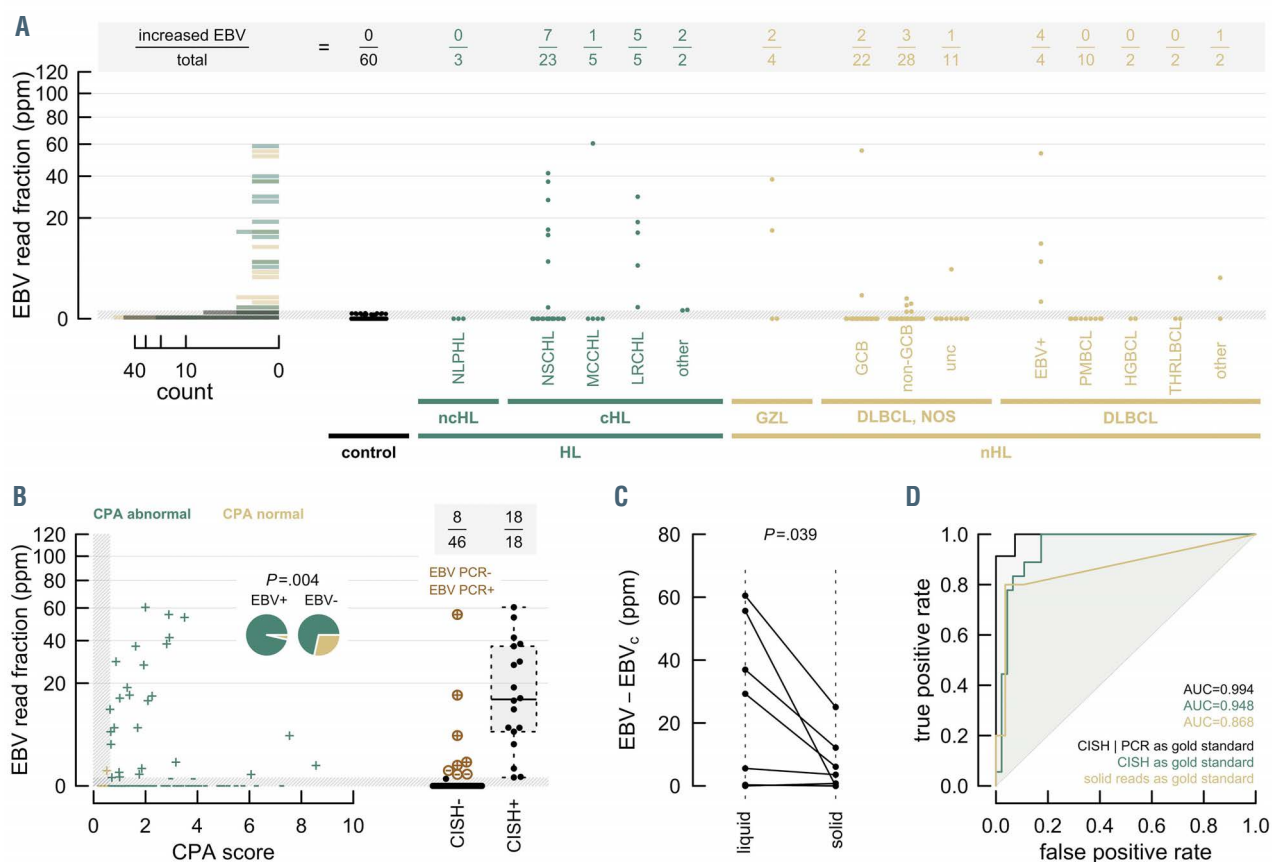


Figure 2. Feasibility of EBV detection through SWGS of cfDNA. (A) EBV read fraction comparison between common HL and DLBCL subtypes. Dots represent blood samples at baseline. The bottom grey box delineates the abnormality cutoff defined by the controls. According to this limit, the numbers at the top define which samples have abnormally-elevated EBV fractions. The overlapping histograms on the left display the overall EBV read fraction distribution. (B) Scatter plot between the CPA score and the EBV read fraction. Symbols represent blood samples at baseline; color shows whether CPA is abnormal; the + or - symbol is assigned in accordance to A, thereby specifying EBV positivity or negativity, respectively. Pie charts visualize contingency table; the P -value results from Fisher's exact test. Swarm and box plots on the right-hand side indicate the EBV read fraction, grouped by CISH for EBER detection outcome. Dots represent blood samples; numbers at the top are assigned as in A. Colored dots are additionally tested by PCR in blood samples. (C) Control-normalized EBV comparison between the liquid and solid biopsies. Samples are represented by dots; same-patient biopsies are connected. Patients without elevated an EBV level in both their solid and liquid biopsies were omitted. The P -value results from a paired Wilcoxon rank-sum test. (D) ROC analyses executed using ranked viral-read fractions and three different gold standards: CISH, overridden by PCR; CISH only; solid biopsy EBV read fraction. EBV: Epstein-Barr virus; SWGS: shallow whole-genome sequencing; cfDNA: cell-free DNA; HL: Hodgkin lymphoma; nHL: non-HL; DLBCL: diffuse large B-cell lymphoma; CPA: copy number profile abnormality; CISH: chromogenic *in situ* hybridization; EBER: EBV-encoded RNA; PCR: polymerase chain reaction; cHL: classical HL; nCH: non-cHL; ROC: receiver operating characteristic; AUC: area under the curve; NLPHL: nodular lymphocyte predominant HL; NSCHL: nodular sclerosis cHL; MCCHL: mixed cellularity cHL; LRCHL: lymphocyte-rich cHL; GZL: grey zone lymphoma; NOS: not otherwise specified; GCB: germinal center B-cell; unc: unclassified; PMBCL: primary mediastinal large B-cell lymphoma; HGBCL: high-grade B-cell lymphoma; THRLBCL: T-cell/histiocyte-rich large B-cell lymphoma; c: control.

score in addition (Figure 2B), which significantly associates the latter categorical variables ($P=0.004$; Fisher exact test).

The sensitivity and specificity of EBV detection through sWGS of cfDNA was established by a comparison with routine CISH for EBER detection results, performed on FFPE tissue. All 18 samples that tested positive for EBER, as well as 8/46 EBER-negative samples, had abnormally-increased EBV levels in plasma, achieving a sensitivity and specificity of 100% and 82.6%, respectively (Figure 2B; *Online Supplementary File S2: Tables S3 and S4*). The liquid biopsies with the highest viral read fractions of the EBER-negative cases did, however, accommodate increased levels of EBV fragments in blood, as confirmed by a subsequent PCR analysis (Figure 2B). This indicates that routine CISH might have inadequate sensitivity or that viral fragments may have been absent in the tumor cells studied. When considering the latter five as true positives, specificity increases to 93.5%.

As for the copy number alteration amplitudes, liquid biopsies had higher ($P=0.039$; paired Wilcoxon signed-rank test) EBV read fractions than solid biopsies (Figure 2C). To conclude, three receiver operating characteristic (ROC) analyses (Figure 2D) were included to further demonstrate

the performance of this EBV detection approach. Ranked plasma EBV viral read fractions and three different gold standards were used: CISH overridden by PCR (area under the curve [AUC]=0.994); CISH only (AUC=0.948); and solid biopsies reads (AUC=0.868).

Copy numbers derived from liquid biopsies can aid histological classification

When visualizing fractions of aberrant samples across genomic loci per histological subtype, both general (e.g., gains at 2p, holding *REL*) and subtype-specific fingerprints can be distinguished (Figure 3). Unsupervised clustering²⁶ applied to these summarizing profiles results in the anticipated histological hierarchy. Indeed, nHL and HL are separated as two definite entities, with the exception of GZL and PMBCL, which cluster alongside the HL group. Notwithstanding their non-Hodgkin histology, both have been described as being molecularly related to HL,^{27,28} hence the structure of the dendrogram. GZL in particular demonstrates overlapping features with cHL and PMBCL.²⁸

To mimic how this translates as a potential clinical application for samples that are difficult to classify morphologically, or in cases in which obtaining tumor tissue is compli-

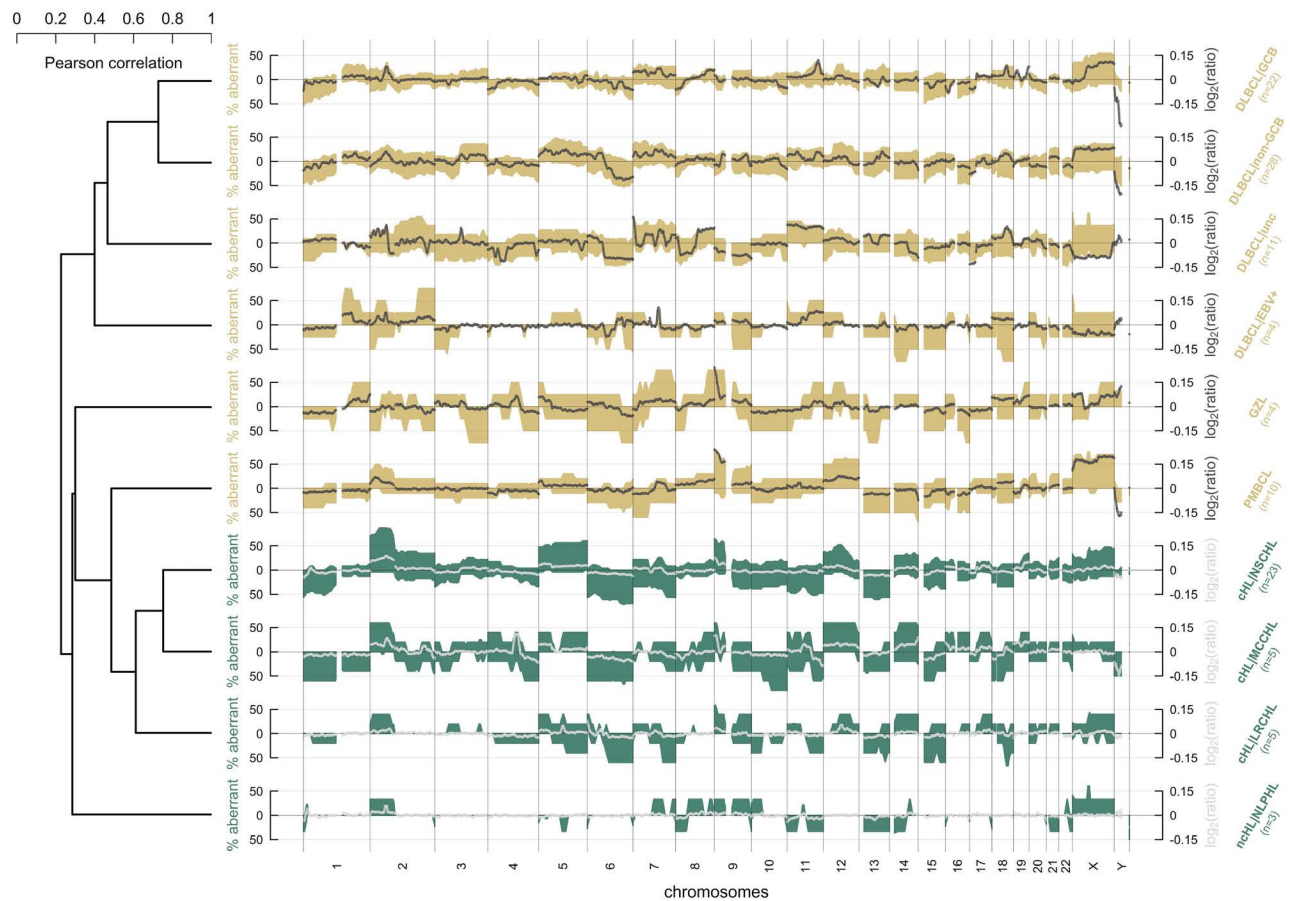


Figure 3. Overview of aberrations detected using liquid biopsies across lymphoma subtypes. Colored waves represent fractions of aberrant samples across genomic loci. Patterns above the $y=0$ lines indicate gains whilst opposite contours represent losses (e.g., when the top of the wave is at 50%, 50% of samples have a gain at the corresponding locus; when the bottom of the wave is at 10%, 10% of samples have a loss at the corresponding locus). The dendrogram (left) results from hierarchical clustering applied to the Pearson distances (i.e., $d = [1 - r]/2$) between the means of these waves. Graph lines represent smoothened mean \log_2 ratios. The number of used baseline samples (n) is indicated on the right-hand side. Subtypes represented by fewer than three patients were excluded. DLBCL: diffuse large B-cell lymphoma; GCB: germinal center B-cell; unc: unclassified; EBV: Epstein-Barr virus; GZL: grey zone lymphoma; PMBCL: primary mediastinal large B-cell lymphoma; HL: Hodgkin lymphoma; cHL: classical HL; nCHL: non-cHL; NSCHL: nodular sclerosis cHL; MCCHL: mixed cellularity cHL; LRCHL: lymphocyte-rich cHL; NLPHL: nodular lymphocyte predominant HL.

cated, random forest predictive modeling, in combination with cross-validation, was applied to four differential diagnoses (Figure 4A): HL versus DLBCL; cHL versus PMBCL; DLBCL, NOS versus PMBCL; and within DLBCL, NOS, GCB versus non-GCB. For these analyses, obviously only baseline liquid biopsies were employed. Two ROC AUC can be discussed here: one for all samples and one disregarding flat copy number profiles. The latter statistic is considered in the next paragraph as it represents the performance of a rational application: profiles without aberrations will never be subjected to machine learning algorithms since predictions would be random.

The differential diagnosis of HL and DLBCL, using computationally-processed copy number profiles, was found to be accurate (AUC of 0.967). Out of bag principles (i.e., cross-validation for bootstrapping methods, to overcome train/test bias) were applied to a set of 119 samples, meaning that, at each iteration, sufficient data were available for model training; and, biologically, HL and DLBCL express divergent profiles (Figure 3). PMBCL has features between those of cHL and DLBCL.²⁷ Nonetheless, copy number profiles can aid differential diagnoses, indicated by AUC values of 0.931 and 0.981, respectively. Finally, the GCB versus non-GCB COO status does not seem to trigger specific copy number alterations, as shown by the resulting AUC of 0.450. We could not test whether the activated B-cell versus GCB distinction by gene expression profiling²⁹ would perform better. This should be explored in future work.

During training, learning algorithms assign higher weights to features that improve classification. For the HL

versus DLBCL problem, the top three discriminative features - or loci, in our case - are located at 2p, 9p and 11q (Figure 4B, bottom). In addition, the GISTIC software aims to define driver genes in pattern peaks across a set of samples. Of note, there were three interesting peaks in these regions that were significantly enriched (Figure 4B, top), matching several anticipated genes: in both HL and DLBCL, 2p16.1 (*REL*) gains;^{30,31} 9p24.1 (*JAK2* and *PD-L1*) gains in HL;³² 9p21.3 (*CDKN2A*) losses in DLBCL;³³ and to conclude, 11q23.3 (*CEP164*) gains in DLBCL³⁴ (Online Supplementary File S2: Table S5). Although 2p16.1 was found to be amplified, or duplicated, in both HL and DLBCL, the model gives importance to the 2p arm, as these aberrations are more frequently detected in HL compared to DLBCL (Figure 4B, top).³⁵

Finally, locus importance was examined for the two other well-performing models: to a certain extent, 5p, 9q, and chromosomes 14 and 19 allow separation of cHL and PMBCL, while more specific 7p and, once more, 9p24.1 (*JAK2* and *PD-L1*)³² cytobands are decisive in differentiating DLBCL from PMBCL (Figure 4C).

Longitudinal liquid biopsies correlate with overall disease status

Serial blood samples were analyzed to observe copy number alteration dynamics over time and to study longitudinal changes in CPA and EBV read fractions. One example of an intra-patient evolution is demonstrated by patient 32, who had refractory DLBCL (Figure 5A). In this case, alterations disappeared during treatment (MTV on PET/CT

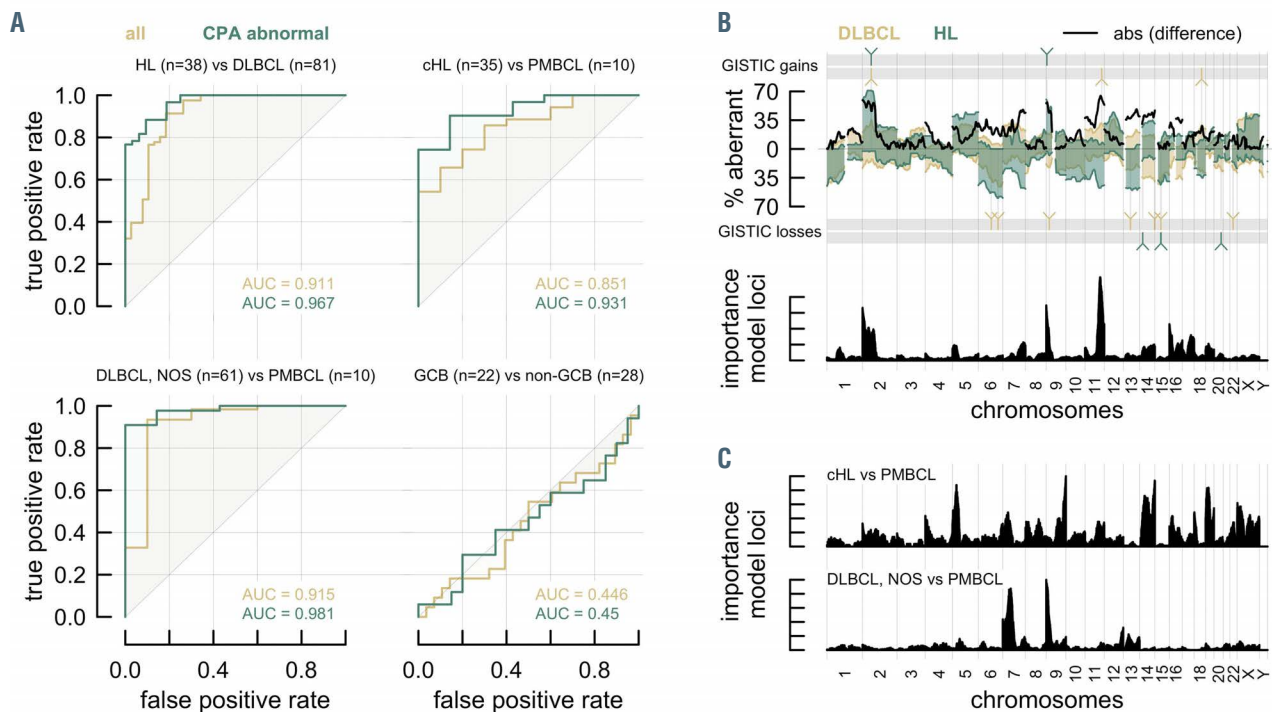


Figure 4. Feasibility of copy number profiles as a clinical lymphoma subtyping tool. (A) ROC analysis following out of bag random forest modeling to differentially diagnose HL versus DLBCL; cHL versus PMBCL; DLBCL, NOS versus PMBCL; and within DLBCL, NOS, GCB versus non-GCB. Exclusively, blood samples at baseline were included. ROC curves when omitting flat profiles (i.e., only CPA abnormal cases) are shown in addition. (B) The top plot is generated as in Figure 3. The black graph line shows the mean absolute difference between these profiles. The four horizontal grey bars are colored at significantly-enriched aberration peaks, assumed to hold driver genes, detected by GISTIC (gains on top, losses below; yellow for DLBCL, green for HL; emphasized by (inverted) 'V' symbols). The plot below visualizes the importance of regions for the differential diagnosis according to a HL vs. DLBCL random forest. (C) Similar to (B) but for the cHL versus PMBCL and DLBCL, NOS versus PMBCL classification problems. ROC: receiver operating characteristic; AUC: area under the curve; CPA: copy number profile abnormality; HL: Hodgkin lymphoma; cHL: classical HL; DLBCL: diffuse large B-cell lymphoma; PMBCL: primary mediastinal large B-cell lymphoma; NOS: not otherwise specified; GCB: germinal center B-cell.

decreased from 915 cm³ at baseline to 4 cm³ between both interim moments), and reappeared when progression was diagnosed after 212 days (MTV increased to 79 cm³). The smoothed profiles at initial staging and subsequent restaging at progression were very similar ($r=0.946$; Figure 5B), notwithstanding the long time interval, suggesting that

a portion of ‘remnant’ tumor had never been completely eliminated and that minimal residual disease had caused recidivism. Despite the significant correlation, evidence of tumor evolution was present, as indicated by several changes along the copy number alteration pattern, e.g., on chromosome arms 2q and 6p; and on chromosome 14

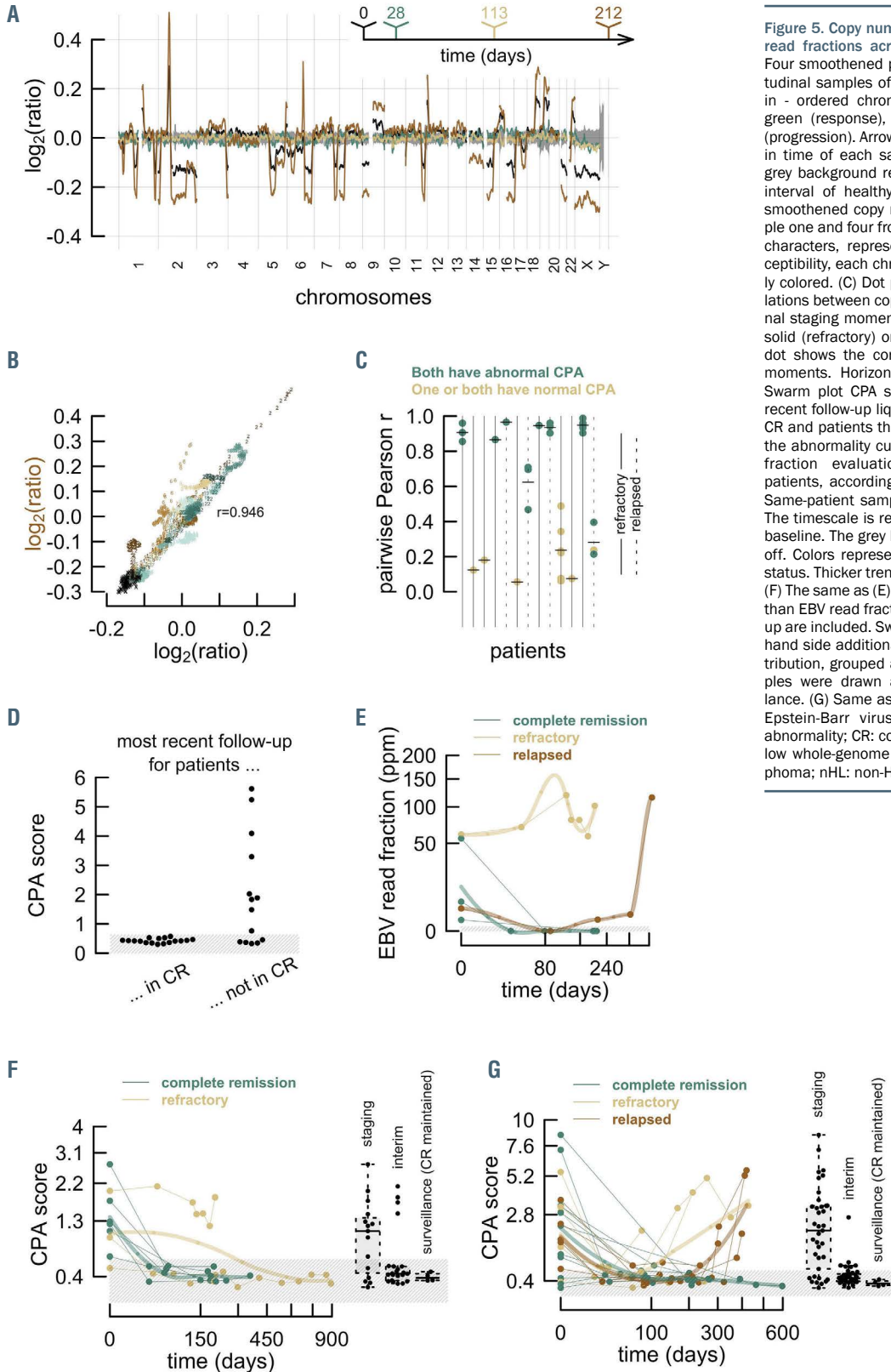


Figure 5. Copy number profile dynamics and EBV read fractions across serial liquid biopsies. (A) Four smoothed profiles corresponding to longitudinal samples of refractory patient 32, colored in - ordered chronologically - black (diagnosis), green (response), yellow (response) and brown (progression). Arrow on top indicates relative point in time of each sample collection moment. The grey background represents the 95% confidence interval of healthy controls. (B) Scatter plot of smoothed copy number profiles between sample one and four from (A). Scatter symbols are tiny characters, representing chromosomes; for perceptibility, each chromosome character is uniquely colored. (C) Dot plot of pairwise Pearson correlations between copy number profiles at longitudinal staging moment. Patients are represented by solid (refractory) or dotted lines (relapsed). Each dot shows the correlation between two staging moments. Horizontal lines indicate means. (D) Swarm plot CPA score comparison of the most recent follow-up liquid biopsy between patients in CR and patients that are not. The grey box shows the abnormality cutoff. (E) Longitudinal EBV read fraction evaluation. Exclusively EBV-positive patients, according to sWGS data, are included. Same-patient samples are connected with lines. The timescale is relative, where 0 corresponds to baseline. The grey box shows the abnormality cutoff. Colors represent a patient's current disease status. Thicker trend lines result from loess fitting. (F) The same as (E) but evaluates CPA score rather than EBV read fraction. All HL patients with follow-up are included. Swarm and box plots on the right-hand side additionally indicate the CPA score distribution, grouped according to whether the samples were drawn at staging, interim or surveillance. (G) Same as (F), but for nHL patients. EBV: Epstein-Barr virus; CPA: copy number profile abnormality; CR: complete remission; sWGS: shallow whole-genome sequencing; HL: Hodgkin lymphoma; nHL: non-HL.

(Figure 5A). These zones are represented by irregularities in the scatter plot trend between the profiles (Figure 5B).

Overall, observations were consistent for refractory and relapsed patients: when detectable, copy number patterns were very similar across different longitudinal staging moments (Figure 5C). One exception was found in patient 73 (rightmost patient in Figure 5C), who appeared to relapse with a divergent copy number profile (*Online Supplementary File S1: Longitudinal samples*). This case concerned an aggressive EBV-positive plasmablastic lymphoma that is known to present complex karyotypic abnormalities and, presumably, an unstable genome.³⁶

The most recent follow-up of all 17 patients in CR showed that none had remaining copy number alterations; in contrast, this was the case in only 5/14 patients in whom CR was thought not to be reached (Figure 5D). In total, five EBV-positive patients, according to sWGS of cfDNA, were detected at diagnosis of the 31 patients with follow-up. Of these, three went into CR and, consequently, no longer had abnormally elevated EBV levels; one refractory patient (patient 6) was noted to have recurrently high EBV levels and one relapsed patient (patient 73) was detected with an unusually high EBV read fraction, following an earlier drop to 0 ppm at first response evaluation (Figure 5E). It should be noted that, for this patient, no tumor was detected under PET/CT at the fourth time point (*Online Supplementary File 2: Table S4*), however, the steadily increasing EBV fraction anticipated actual relapse at the fifth time point, after which the MTV increased rapidly along with the plasma concentration of viral fragments (*Online Supplementary File S1: Longitudinal samples*).

A similar analysis, concerning the CPA score, was performed for all patients with HL (Figure 5F) and nHL (Figure 5G) with follow-up. Similar conclusions were made according to the trend lines, which largely summarize the longitudinal analyses: a prompt drop for patients in CR; varying values for refractory patients; and a drop, followed by a newly-triggered rise, for relapsed patients. It is significant how the CPA decreased rapidly when treatment was initiated: interim samples had normal CPA scores unless they represented a refractory patient. Finally, 6/31 patients never transcended the CPA abnormality cutoff, meaning that these, probably, have insufficient ctDNA. In two cases (patients 66 and 71), however, the MTV reached extreme levels (>1000 cm³) while the CPA score stayed low (*Online Supplementary File S1: Longitudinal samples*). For these patients, it is more likely that copy number alterations really were absent.

Of the above cases, additional clinical information from four interesting patients further illustrates the potential of shallow-depth sequencing for disease monitoring - e.g., when dealing with ambiguous PET/CT images or tumor and viral entities beyond initial suspicion. These are discussed as short case-reports in *Online Supplementary File S1: Case reports*.

Copy number anomalies detected in liquid biopsies correlate with clinical parameters

Five lymphoma-related clinical variables were included in an extensive concordance analysis of the baseline and longitudinal liquid biopsies: Ann Arbor stage; International Prognostic Index/Score; lactate dehydrogenase concentration; $\beta 2$ microglobulin concentration; and MTV (*Online Supplementary File S1: Figure S2*). The estimated tumor fraction, according to copy number profiles, positively and sig-

nificantly correlates with all of these clinical variables, rendering the former estimate relevant. The CPA score is intrinsically correlated to the tumor fraction, so similar associations as for the tumor fraction arise. Since all these variables are naturally cross-related, the derived relations are obviously not claimed to be independently significant.

Discussion

To date, molecular profiling of lymphoma has largely been focused on the characterization of actionable targets, such as point mutations and translocations. This profiling requires panel sequencing, currently using tissue biopsies as the DNA source and targeting a substantial number of genes. However, until now, the availability of this precise information has contributed little to actual progress in patient care, despite its tempting rationale.³⁷ In addition, applying these concepts in a minimally-invasive manner requires ultra-deep sequencing (~2000x),³⁸ of which the implementation in routine practice remains challenging.

In this study, we evaluated sWGS of cfDNA for the diagnosis, differential diagnosis, and disease monitoring of HL and DLBCL. In contrast to targeted sequencing, this approach provides genomic insight by copy number profiling in an accessible and standardized manner. Indeed, non-invasive prenatal testing, which involves the same laboratory and computational steps, has evolved into an application performed daily in molecular diagnostic laboratories. The short turnaround time (approximately, 4 days) and low-price tag (around \$200, including processing costs) renders sWGS *a fortiori* feasible.

Copy number profiling is not a surrogate for targeted sequencing, nor does it enable the characterization of phenotypic traits, such as CD20 expression. However, to a limited extent, it does enable the prediction of targetable genes, such as *JAK2*, *PD-L1* and *REL*. Further, regarding tumor fraction and specificity, it has the advantage of easier and more stable interpretation: once a copy number alteration is visible, the presence of sufficient ctDNA is most likely ascertained given that large somatic copy number alterations are rarely detected in unaffected subjects.¹⁴

For DLBCL, the sensitivity of copy number profiling is a shortcoming: only 74.1% of baseline liquid biopsies were found to have detectable aberrations, which is lower than that reported from targeted ultra-deep duplex sequencing studies (e.g., 98% in Kurtz *et al.*³⁹). On the other hand, the sensitivity was surprisingly high for HL (84.2% for HL overall and 88.6% for cHL in this study; 88.9% in a smaller proof-of-concept study¹⁰) and compares to that of SNV-based studies (e.g., 81.2% in Spina *et al.*¹¹). It ought to be noted that the lower limit of detection for CNA is often considered to be 3% tumor fraction.²¹ It is worth mentioning that this limit could be lowered as paired-end sequencing at greater depths (1x-1.5x) is expected to become available for routine diagnosis at a similar cost. This would empower computational tumor-read enrichment by insert-size filtering, as shorter cfDNA fragments are more likely to be tumor-derived, thus increasing sensitivity.^{20,40,41}

The signs and symptoms of aggressive lymphoma are often non-specific, having much in common with a broad range of different disorders, including autoimmune, immunodeficiency and systemic diseases, infections and other malignancies. Therefore, currently, accurate lymphoma diagnosis is essentially only possible through inva-

sive tissue biopsy. However, affected sites are sometimes difficult to reach and tissue-based subtyping may be complicated when dealing with inconclusive or small biopsies, or partially-involved lymph nodes. Since several cases of differential diagnosis by computationally-processed copy number profiles have been found to be accurate, copy number profiling of cfDNA might enter clinical practice as a test for patients with: suspected lymphoma with a mediastinal mass (to guide the need for mediastinoscopy); deep-seated pathologic lymph nodes in the abdomen, especially following surgery; a remaining suspicion of lymphoma after a negative biopsy, caused by ambiguous PET/CT images; probable central nervous system lymphoma (to avoid radical surgery); probable intravascular localization;⁴² and, generally, lymphoma in children (to avoid unnecessary surgery). For these indications, the approach presented here could be easily integrated with other tests in order to decide whether invasive procedures are still necessary.

sWGS of cfDNA was found to be 100% sensitive for the detection of EBV. However, one fact that could preclude its application in lymphoma diagnosis is that all possible causes of elevated EBV levels (e.g., ongoing infectious mononucleosis) are detected simultaneously through blood samples irrespectively of whether these are related to lymphoma. CISH does not have this shortcoming and can be used to assess a neoplastic association. Nonetheless, the combination of an increased EBV level and typical HL or DLBCL alterations strongly indicates an EBV-driven lymphoma. This is especially relevant for the diagnosis of post-transplant or immunodeficiency-associated lymphoma. sWGS could be employed, in part, to detect EBV prior to performing a CISH test (around \$140), when required. The latter test should only be performed when EBV is positive.

Future in-house research will focus on the inclusion of other (rarer) lymphoma subtypes, including T-cell lymphomas, to obtain a broader image of the alterations' speci-

ficity compared to subtype. Studies will also concentrate on novel methods to study cfDNA (epi)genetics, including cell-free reduced-representation bisulfite sequencing for methylation detection, which are likely to improve specificity. Since this study was concluded, new liquid biopsies have been analyzed by sWGS at our institutions, mostly in cases of diagnostic ambiguity. While negative results currently never exclude the presence of malignancies, positive results ascertain disease.

Disclosures

No conflicts of interest to disclose.

Contributions

Study supervision: FO and JVD; patient recruitment: CDV, DD, FO, KDG and SB; clinical/pathological data acquisition/generation: CDV, BVDB, KM, DD, FD, SV, AD, FO and JVD; data processing: MVDL; study design: LR, MVDL, CDV, BM, FO and JVD; data analysis: LR; manuscript writing: LR and JVD; all authors read, revised and approved the final manuscript.

Acknowledgments

The authors would like to thank all collaborators at AZ Delta Roeselare and Ghent University Hospital for the active cohort recruitment, sample collection and provision of individual patient data.

Funding

This study was supported by Bijzonder Onderzoeksfonds (BOF), Ghent University, in the form of a doctoral research grant (BOF.STA.2017.0002.01 to LR).

Data-sharing statement

Copy number profiles and viral reads counts of all sequenced samples are included in Online Supplementary file: Table S6 and S7, respectively.

References

- Shaffer AL, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol.* 2012;30:565-610.
- Re D, Küppers R, Diehl V. Molecular pathogenesis of Hodgkin's lymphoma. *J Clin Oncol.* 2005;23(26):6379-6386.
- Steidl C, Connors JM, Gascoyne RD. Molecular pathogenesis of Hodgkin's lymphoma: increasing evidence of the importance of the microenvironment. *J Clin Oncol.* 2011;29(14):1812-1826.
- Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med.* 2018;378(15):1396-1407.
- Steidl C, Gascoyne RD. The molecular pathogenesis of primary mediastinal large B-cell lymphoma. *Blood.* 2011;118(10):2659-2669.
- Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer.* 2011;11(6):426-437.
- Dominguez-Vigil IG, Moreno-Martinez AK, Wang JY, Roehrl MHA, Barrera-Saldaña HA. The dawn of the liquid biopsy in the fight against cancer. *Oncotarget.* 2017;9(2):2912-2922.
- Araf S, Wang J, Korfi K, et al. Genomic profiling reveals spatial intra-tumor heterogeneity in follicular lymphoma. *Leukemia.* 2018;32(5):1261-1265.
- Rossi D, Diop F, Spaccarotella E, et al. Diffuse large B-cell lymphoma genotyping on the liquid biopsy. *Blood.* 2017;129(14):1947-1957.
- Bohers E, Vially P-J, Becker S, et al. Non-invasive monitoring of diffuse large B-cell lymphoma by cell-free DNA high-throughput targeted sequencing: analysis of a prospective cohort. *Blood. Cancer J.* 2018;8(8):74.
- Spina V, Brusca A, Cuccaro A, et al. Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma. *Blood.* 2018;131(22):2413-2425.
- Newman AM, Bratman SV, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med.* 2014;20(5):548-554.
- Raman L, Baetens M, De Smet M, Dheedene A, Van Dorpe J, Menten B. PREFACE: In silico pipeline for accurate cell-free fetal DNA fraction prediction. *Prenat Diagn.* 2019;39(10):925-933.
- Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature.* 2006;444(7118):444-454.
- Busque L, Patel JP, Figueroa M, et al. Recurrent Somatic TET2 Mutations in Normal Elderly Individuals With Clonal Hematopoiesis. *Nat Genet.* 2012;44(11):1179-1181.
- Vandenberghe P, Wlodarska I, Tousseyn T, et al. Non-invasive detection of genomic imbalances in Hodgkin/Reed-Sternberg cells in early and advanced stage Hodgkin's lymphoma by sequencing of circulating cell-free DNA: a technical proof-of-principle study. *Lancet Haematol.* 2015;2(2):e55-65.
- Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood.* 2004;103(1):275-282.
- Bordon V, Padalko E, Benoit Y, Dhooge C, Laureys G. Incidence, kinetics, and risk factors of Epstein-Barr virus viremia in pediatric patients after allogeneic stem cell transplantation. *Pediatr Transplant.* 2012;16(2):144-150.
- Raman L, Dheedene A, De Smet M, Van Dorpe J, Menten B. WisecondorX: improved copy number detection for routine shallow whole-genome sequencing. *Nucleic Acids Res.* 2019;47(4):1605-1614.
- Raman L, Van der Linden M, Van der Eecken K, et al. Shallow whole-genome sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung carcinoma. *Genome Med.* 2020;12(1):35.

21. Adalsteinsson VA, Ha G, Freeman SS, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun.* 2017; 8(1):1324.
22. Liaw A, Wiener M. Classification and Regression by random Forest. *R News.* 2002;2(3):18-22.
23. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
24. Van der Linden M, Raman L, Vander Trappen A, et al. Detection of copy number alterations by shallow whole-genome sequencing of formalin-fixed, paraffin-embedded tumor tissue. *Arch Pathol Lab Med.* 2020;144(8):974-981.
25. Mathas S, Hartmann S, Küppers R. Hodgkin lymphoma: Pathology and biology. *Semin Hematol.* 2016;53(3):139-147.
26. Ward JH Jr. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 1963;58(301):236-244.
27. Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med.* 2003;198(6):851-862.
28. Eberle FC, Rodriguez-Canales J, Wei L, et al. Methylation profiling of mediastinal gray zone lymphoma reveals a distinctive signature with elements shared by classical Hodgkin's lymphoma and primary mediastinal large B-cell lymphoma. *Haematologica.* 2011;96(4):558-566.
29. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 2000;403(6769):503-511.
30. Joos S, Menz CK, Wrobel G, et al. Classical Hodgkin lymphoma is characterized by recurrent copy number gains of the short arm of chromosome 2. *Blood.* 2002;99(4):1381-1387.
31. Houldsworth J, Mathew S, Rao PH, et al. REL proto-oncogene is frequently amplified in extranodal diffuse large cell lymphoma. *Blood.* 1996;87(1):25-29.
32. Green MR, Monti S, Rodig SJ, et al. Integrative analysis reveals selective 9p24.1 amplification, increased PD-1 ligand expression, and further induction via JAK2 in nodular sclerosing Hodgkin lymphoma and primary mediastinal large B-cell lymphoma. *Blood.* 2010;116(17):3268-3277.
33. Lenz G, Wright GW, Emre NCT, et al. 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.
34. Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet.* 2011;43(9):830-837.
35. Kober-Hasslacher M, Schmidt-Suppran M. The unsolved puzzle of c-Rel in B cell lymphoma. *Cancers.* 2019;11(7):941.
36. Castillo JJ, Reagan JL. Plasmablastic lymphoma: a systematic review. *Scientific World Journal.* 2011;11:687-696.
37. Iacoboni G, Zucca E, Ghielmini M, Stathis A. Methodology of clinical trials evaluating the incorporation of new drugs in the first-line treatment of patients with diffuse large B-cell lymphoma (DLBCL): a critical review. *Ann Oncol.* 2018;29(5):1120-1129.
38. Scherer F, Kurtz DM, Newman AM, et al. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci Transl Med.* 2016;8(364):364ra155.
39. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol.* 2018;36(28):2845-2853.
40. Mouliere F, Mair R, Chandrananda D, et al. Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. *EMBO Mol Med.* 2018;10(12):e9323.
41. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature.* 2019;570(7761):385-389.
42. Deeren D, Van Der Linden M, Dedeurwaerdere F, et al. Circulating cell-free DNA for response evaluation of intravascular lymphoma. *Ann Hematol.* 2019;98(8):2021-2023.
43. Rinaldi A, Kwee I, Poretti G, et al. Comparative genome-wide profiling of post-transplant lymphoproliferative disorders and diffuse large B-cell lymphomas. *Br J Haematol.* 2006;134(1):27-36.
44. Morscio J, Dierickx D, Tousseyn T. Molecular pathogenesis of B-cell posttransplant lymphoproliferative disorder: what do we know so far? *Clin Dev Immunol.* 2013;2013:150835.