

Whole-genome Sequencing of Reed-sternberg Cells Identifies Outcome Determinants in Hodgkin Lymphoma

Jesús Velasco-Suelto

University Hospital Virgen de la Victoria

Laura Gálvez-Carvajal

University Hospital Virgen de la Victoria

Esperanza López-López

University Hospital Virgen de la Victoria

Antonio Cantarero-Cuenca

University Hospital Virgen de la Victoria

Silvia Sequero-López

University Hospital Clínico San Cecilio

Natividad Martínez-Banaclocha

Hospital General Universitario de Alicante Doctor Balmis

Marcos Melián-Sosa

Instituto Valenciano de Oncología (IVO)

María Casanova-Espinosa

University Hospital Costa del Sol

Luis Cruz-Merino

Biomedicine Institute of Seville (IBIS), CSIC, University Hospital Virgen Macarena, University of Seville

Natalia Palazón-Carrión

Biomedicine Institute of Seville (IBIS), CSIC, University Hospital Virgen Macarena, University of Seville

Francisco Ramón García-Arroyo

University Hospital of Pontevedra

Daniel Prieto

Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND

Martina Álvarez

Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND

Javier Pascual

University Hospital Virgen de la Victoria

Mariano Provencio-Pulla

University Hospital Puerta de Hierro-Majadahonda

Emilio Alba

University Hospital Virgen de la Victoria

Rocío Lavado-Valenzuela

University Hospital Virgen de la Victoria

Iñaki Comino-Méndez

inaki.comino@ibima.eu

University Hospital Virgen de la Victoria

Antonio Rueda-Domínguez

University Hospital Virgen de la Victoria

Article

Keywords: Classical Hodgkin lymphoma (cHL), Whole-Genome Sequencing (WGS), Single-nucleotide variants (SNVs), Copy number variations (CNVs), Structural Variants (SVs), COSMIC mutational signatures.

Posted Date: June 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9880329/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 4th, 2026. See the published version at <https://doi.org/10.1038/s41598-026-69457-8>.

Abstract

Classical Hodgkin lymphoma (cHL) is a B-cell-derived malignancy characterized by rare Hodgkin and Reed–Sternberg (HRS) cells within an extensive inflammatory microenvironment. Although most patients achieve durable remission with first-line therapy, a subset relapse or experience disease progression. We performed whole-genome sequencing (WGS) of purified HRS-cell pools isolated from diagnostic biopsies of 76 cHL patients using DEPAArray™ technology. Somatic copy number variations (CNVs), structural variants (SVs), and COSMIC mutational signatures were correlated with clinical characteristics, treatment response, and progression-free survival (PFS). HRS genomes displayed marked chromosomal instability with recurrent CNVs. Several CNV peaks were associated with baseline features and treatment response, while 2p11.2 amplification and 11q25 and 2p13.2 deletions were linked to inferior PFS. Mutational spectra were dominated by SBS5, SBS8, SBS84, SBS9 and SBS1, with an inverse SBS5–SBS9 axis and an AID-like SBS9-high/SBS5-low profile enriched in nodular sclerosis cHL. SV patterns were largely represented by SV2, with specific signatures associated with disease stage and treatment-related toxicity. TRACER-HL, integrating a Genomic Instability Score with relapse-associated CNVs, achieved an AUC of 0.81 and an NPV of 95%. These findings identify candidate genomic features associated with outcome and support further evaluation of genomically informed risk stratification strategies in cHL.

INTRODUCTION

Classical Hodgkin lymphoma (cHL) is a B-cell-derived lymphoid malignancy characterized by the presence of malignant Hodgkin Reed-Sternberg (HRS) cells surrounded by a dense but ineffective immune infiltrate¹. Although most patients achieve long-term remission with standard first-line therapy, a significant proportion, particularly those with advanced-stage disease, experience relapse or refractory progression. This clinical challenge underscores the need for improved strategies to identify patients at higher risk of treatment failure. Current prognostic tools, including interim positron emission tomography (PET) imaging, fail to capture the molecular complexity and heterogeneity of the disease, and lack sensitivity for detecting minimal residual disease, among other limitations^{2–4}.

HRS cells originate from germinal center B cells but lose their lineage identity through profound epigenetic reprogramming, leading to the silencing of B-cell receptor genes and aberrant activation of oncogenic pathways such as NF- κ B and JAK/STAT^{5–9}. Despite representing only a small fraction of the total tumor cellularity, these cells play a central role in disease pathogenesis and progression, sustained by a combination of genetic alterations, immune escape mechanisms, and dynamic interactions with the tumor microenvironment^{5,10–13}. While early studies employed laser-capture microdissection and single-cell PCR, more recent advances in flow cytometry and low-input sequencing have enabled genomic profiling of purified HRS cells^{11,13}.

Characterizing the full spectrum of genomic alterations in HRS cells is essential to understand the biological mechanisms driving tumor evolution and immune evasion in cHL. Whole-genome approaches

offer the opportunity to investigate global features such as copy number imbalance, ploidy state, and structural complexity, which may be associated with clinical outcome in this disease¹³. However, both the prognostic value of these global genomic features and their relationship with clinical presentation remain poorly defined. While previous genomic studies have profiled purified HRS cells using exome or whole-genome approaches, the clinical implications of global CNV burden, SV signatures and baseline relapse-oriented genomic modelling remain insufficiently defined in cHL^{11,13,14}.

In this study, we used single-cell sorting to prospectively isolate highly pure HRS cells from diagnostic cHL biopsies and subjected them to whole-genome profiling. We systematically characterized somatic Copy Number Variations (CNVs) and mutational signatures, and integrated these data with clinical variables and longitudinal outcomes to identify genomic correlates of presentation and treatment failure. Finally, we developed TRACER-HL, a composite risk model combining focal CNVs and chromosomal instability to stratify relapse risk at baseline and support risk-adapted management in cHL.

METHODS

Patient cohort and sample collection

Patients were prospectively enrolled at the following institutions under institutional-approved protocols: Hospital Universitario Virgen de la Victoria, Instituto Valenciano de Oncología, Hospital Universitario San Cecilio, Hospital General Universitario Dr. Balmis, Hospital Universitario Virgen de la Macarena, Hospital Universitario Costa del Sol, Complejo Hospitalario de Pontevedra, and Hospital Universitario Puerta de Hierro Majadahonda. The study was conducted in accordance with the Declaration of Helsinki and under protocols approved by the Regional Ethics Committee of Málaga. Written informed consent was obtained from all participants.

Eligible patients had a confirmed diagnosis of cHL and underwent staging according to international guidelines¹⁵. Formalin-fixed paraffin-embedded (FFPE) tumor samples obtained at diagnosis, prior to treatment initiation, were used for tissue dissociation and downstream single-cell analyses. In total, 76 FFPE samples were processed. A detailed overview of sample inclusion and distribution across the different genomic analyses is provided in (**Supplementary Fig. S1**).

Tissue dissociation from FFPE sections

FFPE tissue sections (60 µm) were cut from diagnostic blocks, deparaffinized, rehydrated through graded ethanol and subjected to heat-induced antigen retrieval in Tris/EDTA buffer (pH 8.0). After cooling and washing, samples were enzymatically and mechanically dissociated using collagenase-based digestion and GentleMACS™ programs to obtain single-cell suspensions, which were filtered, centrifuged and washed prior to immunofluorescent staining (**Supplementary Methods**).

Immunofluorescent staining

Single-cell suspensions were subjected to immunofluorescent staining for CD30 and nuclear counterstaining with DAPI, including peroxidase quenching, serum blocking and tyramide signal amplification to enhance CD30 detection. Stained cells were finally resuspended in SB115 buffer and incubated at 4°C before DEPArray™ PLUS sorting (**Supplementary Methods**).

DEPArray™ PLUS-based cell sorting and isolation

Stained cells were loaded onto DEPArray™ PLUS cartridges in SB115 buffer according to the manufacturer's instructions. Single cells were imaged using DAPI, FITC (Alexa Fluor 488) and brightfield channels, and HRS cells were identified based on CD30 expression and large-cell morphology as previously described¹⁶ (**Supplementary Fig. S2**). For this study, bulk HRS-cell pools were recovered per sample (mean 24.5 cells, range 8–62) and pelleted by centrifugation; supernatants were removed and dry pellets were stored at – 20°C until whole-genome amplification (**Supplementary Methods**).

Whole-genome sequencing analysis

Purified HRS-cell pellets were lysed and subjected to whole-genome amplification using the PicoPLEX® Gold Single Cell DNA-Seq kit, followed by dual-index barcoding (DNA HT Dual Index Kit) according to the manufacturer's instructions. Libraries were pooled, purified and sequenced on an Illumina NovaSeq™ X platform with 2 × 150 bp paired-end reads, achieving a mean coverage of ~ 15× in tumor samples and ~ 4× in normal controls (**Supplementary Methods**).

Whole-genome sequencing data processing

Raw sequencing reads underwent quality control with fastQC, adapter trimming and quality filtering with Trim Galore, and alignment to the GRCh38 reference genome using BWA-MEM. Downstream processing followed GATK best practices for duplicate marking, base quality score recalibration and correction of technical artefacts. Alignments were managed with Samtools and final quality control was performed with Qualimap. Samples with a mean sequencing depth below 5× were excluded from further analyses (**Supplementary Methods, Supplementary Fig. S3; Supplementary Table S1**).

Copy-number and structural variant events

CNVs were inferred using the GATK CNV pipeline with a panel of normals comprising 16 control samples. Segmented CNVs profiles were analyzed with GISTIC2.0 to identify recurrent focal amplifications and deletions ($q < 0.05$), using $\log_2$ copy-number ratio thresholds of 0.2 and – 0.25 to define gains and losses. SVs were called with svABA, which detects deletions, insertions, duplications, inversions and translocations by leveraging local assembly and discordant read-pair information. Low-

quality events were filtered using internal quality metrics, and somatic SVs were distinguished from germline using GaTSV. Cohort-level SVs integration was performed with svtools and SURVIVOR, and functionally annotated using AnnotSV (**Supplementary Methods**).

Single nucleotide variants detection

High-confidence somatic single nucleotide variants (SNVs) were called in a tumor-only setting using Mutect2 (GATK) against the GRCh38 reference genome, restricting analyses to high-mappability regions. A panel of normals and gnomAD population data were used to suppress germline variants and sequencing artefacts. Variants were filtered with FilterMutectCalls, annotated with ANNOVAR and retained for downstream analyses if they had a population allele frequency < 0.5% in gnomAD, sequencing depth ≥ 8 reads, ≥ 3 mutant-supporting reads and a variant allele frequency $\geq 20\%$ (**Supplementary Methods**).

Mutational COSMIC Signatures

Mutational signature analysis was performed using the SigProfiler suite with COSMIC v3.3 reference catalogues (GRCh38) for Single Base Substitution (SBS) and SVs signatures. Signatures attributable to sequencing or library-preparation artefacts, or deemed biologically implausible in this context, were excluded. SBS signatures were refitted using the mmsig R package, and used for downstream correlative analyses (**Supplementary Methods**)¹⁷.

Development of genomic predictors and risk modelling

To quantify chromosomal instability, we derived a Genomic Instability Score (GIS) from WGS-based CNVs profiles. For each sample, global tumour ploidy was computed from GATK CNV segmentation as the length-weighted mean copy number:

$$Ploidy = \frac{\sum_i (CN_i \times L_i)}{\sum_i L_i}$$

where CN_i and L_i denote the absolute copy number and length of segment i , respectively. The fraction of genome altered (GA) was calculated as the cumulative length of all CNVs segments in the sample divided by the size of the human genome (3,200 Mb):

$$\%GA = \frac{\sum_i L_i}{3200 \text{ Mb}}$$

Ploidy and %GA were independently z-transformed across the cohort. The GIS was then defined as the difference between z-transformed %GA and z-transformed ploidy, so that higher GIS values reflect a high CNV burden not explained by compensatory genome doubling.

In parallel, three CNVs peaks associated with adverse outcome (2p11.2 amplification and 11q25 and 2p13.2 deletions) were combined into a CNVs risk score by assigning one point per lesion; patients harboring ≥ 2 of these events were considered CNVs-positive.

A multivariable logistic regression model integrating GIS (continuous) and CNVs status (binary) was developed to predict relapse and termed TRACER-HL (Tumor Relapse And Copy-number Evaluation in Reed–Sternberg cells – Hodgkin Lymphoma). Internal robustness was assessed by bootstrap resampling with 1,000 iterations and out-of-bag evaluation to derive confidence intervals for the area under the curve (AUC) and classification performance (**Supplementary Methods**).

Statistical analysis

Descriptive statistics were used to summarize clinical and molecular variables. Categorical data were compared using Fisher's exact or Chi-squared tests, and continuous variables using Mann–Whitney U or Kruskal–Wallis tests, as appropriate. The impact of CNVs on progression-free survival (PFS) was analyzed with Kaplan–Meier curves and log-rank tests. Associations between mutational signature exposures and clinical variables were assessed with non-parametric tests; in particular, the relationship between AID-like SBS9 and clock-like SBS5 activity was examined using Spearman's rank correlation for continuous exposures and Wilcoxon rank-sum tests for group comparisons. Selected genomic features were evaluated in multivariable logistic regression models for relapse prediction, with model discrimination quantified by the AUC and optimal cut-offs defined by the Youden index. Two-sided p-values < 0.05 were considered statistically significant for prespecified analyses. All statistical analyses were performed using R, SPSS and Python (**Supplementary Methods**).

RESULTS

Clinical and pathological characteristics of the patient cohort

A total of 76 patients diagnosed with cHL were included for integrative clinicogenomic analyses (Table 1, **Supplementary Fig. S1**). Figure 1A provides an overview of the study design, from patient recruitment and HRS-cell purification to genome-wide characterization and development of the relapse-risk model. The cohort comprised 40 females (52.6%) and 36 males (47.4%). According to age at diagnosis, 49 patients (64.5%) were under 45 years of age. The most frequent histological subtype was nodular sclerosis, observed in 52 patients (68.4%), followed by mixed cellularity in 15 patients (19.7%). The remaining 9 patients (11.8%) were classified as other or unknown subtypes. Based on German Hodgkin Study Group (GHSG) criteria, 34 patients (44.7%) presented with advanced-stage disease, 24 (31.6%) with intermediate-stage, and 18 (23.7%) with early-stage disease. Bulky disease was reported in 13 patients (17.1%), and absent in 63 (82.9%).

Table 1

Baseline clinical characteristics of the study cohort. Summary of demographic and disease-related features for the 76 patients with classical Hodgkin lymphoma (cHL) included in the study. Data are presented as number of patients and percentage (%). GHSG, German Hodgkin Study Group; Lugano Response, metabolic response at the end of first-line treatment according to Lugano 2014 criteria; Bulky disease, defined by Ann Arbor/Cotswold classification.

Clinical characteristics	n (%)
Sex	
Female	40 (52.63%)
Male	36 (47.37%)
Age at diagnosis	
< 45 years	49 (64.5%)
≥ 45 years	27 (35.5%)
GHSG Stage Group	
Advanced stage	34 (44.74%)
Intermediate stage	24 (31.58%)
Early stage	18 (23.68%)
Histological Subtype	
Nodular sclerosis	52 (68.42%)
Mixed cellularity	15 (19.74%)
Lymphocyte rich	5 (6.58%)
Lymphocyte depleted	1 (1.31%)
Unknown	3 (3.95%)
Bulky Disease (Ann Arbor/Cotswold)	
No	63 (82.89%)
Yes	13 (17.11%)
Lugano Response	
Complete response	64 (84.21%)

Clinical characteristics	n (%)
Sex	
Partial response	4 (5.26%)
Progression	4 (5.26%)
Unknown	4 (5.26%)
Progression	
No	64 (84.21%)
Yes	12 (15.79%)
Death	
No	72 (94.74%)
Yes	3 (3.95%)
Unknown	1 (1.31%)

All patients received standard first-line chemotherapy with ABVD (doxorubicin, bleomycin, vinblastine, and dacarbazine). At the end of treatment, radiological response was evaluated according to Lugano 2014 criteria¹⁸, showing complete response in 64 patients (84.21%), partial response in 4 patients (5.26%), progression in 4 patients (5.26%) and unknown or unevaluable status in 4 patients (5.26%). At the time of last follow-up, 64 patients (84.2%) remained progression-free, while 12 patients (15.8%) had experienced disease relapse. After a median follow-up of 40.3 months (range: 16.5–73.2), 3 patients (3.95%) had died, 72 (94.74%) were alive, and the vital status of 1 patient (1.31%) was unknown at the time of last follow-up.

These data define a well-characterized and uniformly treated cHL cohort, allowing robust correlative analyses between genomic alterations in HRS cells and treatment outcomes.

Genomic landscape of copy-number alterations in HRS cells

To characterise the genomic architecture of HRS cells, we performed WGS in 76 cHL cases. Of these, WGS data were successfully obtained for 67 cases, which were included in downstream genomic analyses (**Supplementary Fig. S1**). A genome-wide CNVs map revealed marked chromosomal complexity and substantial inter-patient heterogeneity (Fig. 1B). Using GISTIC2.0, we identified 31 recurrent amplification peaks and 62 deletion peaks at $q < 0.05$ (**Supplementary Fig. S4; Supplementary Table S2**). This discovery-level peak set was used for downstream association testing with clinical features and relapse. Within the subset of highly prevalent events ($\geq 30\%$ of tumors; Fig. 1C), recurrent copy-number gains were dominated by alterations at 19q13.32 (*BCL3/SPIB*) and 19p13.12, together with recurrent amplifications at 2p23.3 (*DNMT3A*), 2p16.1 (*REL/XPO1*) and 9p24.1 (*JAK2/PD-L1/PD-L2*). This

pattern is consistent with the established biology of cHL, linking frequent CNVs to core pathways such as NF- κ B signaling (*REL*, *CARD11*, *BCL3*), immune evasion and JAK/STAT activation (9p24.1), and epigenetic/chromatin regulation (*DNMT3A*). On the deletion side, recurrent losses included 1p36.32 (*TNFRSF14*) and multiple additional loci recurrently affected across the cohort (e.g., 13q, 10q22.1/10q26.3 and 6q25-27), supporting a model in which widespread chromosomal instability coexists with preferential targeting of immune-regulatory and signaling nodes.

Clinical associations and prognostic impact of CNVs

Several CNVs showed significant associations with baseline features and PET response (**Supplementary Fig. S5**). Deletions at 1q21.1 and 10q11.22 were more frequent in patients > 45 years, amplification at 7p22.2 was enriched in early-stage disease, and amplifications at 8q24.3 together with deletions at 9p11.2 and 9q13 were more common in intermediate/advanced stages. Deletion of 17p11.2 associated with bulky disease, whereas amplifications at 2p23.3, 2p22.3, 2p16.1 and 2p11.2 correlated with extranodal involvement. For treatment response, amplification at 10q22.2 and deletion at 11q25 were enriched in patients with partial metabolic response (PMR) at interim PET, deletion of 1p36.32 occurred only in complete metabolic responders, and deletions at 8q24.3, 10q26.3 and 11q25 associated with PMR or progression at end of treatment. Together, these patterns support the clinical relevance of CNVs profiling for stratifying disease burden and response.

We next assessed the impact of individual CNVs peaks on PFS. Amplification at 2p11.2 and deletions at 11q25 and 16q22.2 were each associated with inferior PFS (log-rank $p < 0.05$), and deletion at 2p13.2 showed a similar trend at $p < 0.1$ (**Supplementary Fig. S6**). To refine prognostic candidates, we trained a Random Forest classifier on binary CNVs calls; 11q25, 2p11.2 and 2p13.2 emerged as the top-ranked features, enriched in relapsed cases and independent of clinical stage or baseline tumor burden. We therefore defined a simple composite classifier in which the presence of ≥ 2 of these CNVs identified a high-risk genomic subgroup, which showed markedly shorter PFS than the remaining patients (log-rank $p = 0.00033$; Fig. 2).

Mutational signature landscape in HRS cells

In the 67 isolated HRS-cell samples, the SBS mutational landscape was dominated by a small set of COSMIC signatures. Across the cohort, five signatures accounted for virtually all substitutions, namely SBS5, SBS8, SBS84, SBS1 and SBS9 (Fig. 3A). When aggregating exposures across samples, SBS5 (clock-like) represented approximately 40% of all SBS mutations, followed by SBS8 (~ 30%), SBS84 (~ 15%), SBS9 (~ 10%) and SBS1 (~ 3%) and other signatures explaining < 1% of the total burden. Per-sample stacked barplots showed that this restricted palette of processes was consistently observed across tumors, with variable relative contributions.

As an additional quality check for potential FFPE-related distortion, inspection of the SBS96 substitution profiles showed that, although C > T substitutions were frequent, the spectra were not dominated exclusively by this change and preserved substantial contributions from other substitution types across

cases (**Supplementary Fig. S7**). The relative proportion of C > T substitutions in our cohort was within the range reported in cHL and other B-cell lymphoid malignancies, supporting that the substitution spectrum was not dominated by an extreme FFPE-like pattern^{19,20}.

In addition, analysis of SVs signatures revealed a relatively simple SVs mutational landscape in this cHL cohort. Per-sample composition plots showed that most cases were dominated by two signatures, SV2 and SV8, with only minor contributions from additional patterns in a subset of tumors (Fig. 3B).

Association of mutational signatures with clinical features and outcome

To explore global patterns of signature activity, we performed unsupervised hierarchical clustering based on the four most prevalent SBS signatures (SBS5, SBS8, SBS84 and SBS9) (**Supplementary Fig. S8**). The clustering did not reveal multiple sharply separated genomic subgroups; instead, the dendrogram was dominated by a major bifurcation that separated the cohort into two broad arms differing primarily by the presence versus absence of SBS9. Visual inspection of the heatmap showed that tumors in the SBS9-positive arm systematically exhibited lower SBS5 exposure than SBS9-negative cases, and this inverse relationship was supported by a highly significant difference in SBS5 contribution between the two branches (Wilcoxon rank-sum, $p = 2.6 \times 10^{-17}$) (**Supplementary Fig. S9**). To further examine the relationship between SBS9 and SBS5, we analysed per-sample signature-assigned counts. This analysis showed that SBS9 was detectable in a subset of tumours, whereas SBS5 was broadly distributed across the cohort. Among cases with detectable contributions from both signatures, SBS5 and SBS9 assigned counts were strongly positively correlated (Spearman's $\rho = 0.95$, $p = 7.5 \times 10^{-18}$), arguing against a simple reciprocal depletion of SBS5 in SBS9-positive tumours (**Supplementary Fig. S9**). Rather, these findings indicate that the main source of variation reflects heterogeneous acquisition of SBS9 across cases on a background of broadly distributed SBS5 activity. Together, these analyses define an SBS5–SBS9 axis as a major source of mutational variation in our HRS genomes.

On this basis, we operationally defined an AID-like group comprising cases with concomitant SBS9 positivity and relatively low SBS5 exposure and compared it with the remaining tumors (Fig. 3C). We then explored its association with baseline clinical and treatment variables. This composite mutational profile was significantly more frequent in nodular sclerosis than in other histological subtypes (Fisher's exact, $p = 0.0111$). Consistently, nodular sclerosis cases cluster were predominantly present in the SBS9-high/SBS5-low region, whereas mixed cellularity and other subtypes are over-represented among SBS5-enriched, SBS9-negative profiles (Spearman $\rho = -0.802$, $p = 5.8 \times 10^{-17}$). In additional analyses of other signatures, SBS1 exposure was significantly lower in bulky compared with non-bulky disease (Wilcoxon rank-sum $p = 0.0216$) (**Supplementary Fig. S9**).

Unsupervised hierarchical clustering based on SVs signature exposures did not segregate patients into clearly distinct genomic subgroups, and SVs patterns were not obviously related to histological subtype or relapse status (**Supplementary Fig. S10**). However, specific SVs signatures showed clinical

associations, where SV2 exposure was higher in patients who developed grade 3–4 treatment-related toxicity (Wilcoxon rank-sum test, $p = 0.0079$) and was increased in GHSG early-stage compared with advanced-stage disease (Wilcoxon rank-sum test, $p = 0.0085$). Conversely, SV8 levels were higher in patients without severe toxicity (Wilcoxon rank-sum test, $p = 0.037$) and SV8 exposure varied across GHSG stages but was lower in early than in advanced disease (Wilcoxon rank-sum test, $p = 0.045$) (**Supplementary Fig. S11**). Taken together, these findings suggest that, although SVs mutational processes appear comparatively homogeneous in cHL, subtle differences in the activity of selected SVs signatures may reflect underlying disease biology or treatment vulnerability.

Development of a composite genomic risk score (TRACER-HL)

To enhance the predictive power of individual genomic features, we developed a composite biomarker termed TRACER-HL, which integrates two independent layers of tumor genomic complexity, global chromosomal instability and the presence of relapse-associated CNVs peaks (Fig. 4A).

The GIS was calculated for each patient as the difference between the z-score of the percentage of GA and the z-score of tumor ploidy, such that higher values reflect high chromosomal disruption in the absence of compensatory genome doubling. Independently, we generated a structural CNVs risk signature by assigning one point per peak among three specific CNVs previously associated with shorter progression-free survival (amplification of 2p11.2, deletions of 11q25 and 2p13.2). Patients harboring two or more of these peaks were classified as CNVs-high (**Supplementary Table S3**).

In multivariable logistic regression analysis, CNVs-high status was significantly associated with relapse ($p = 0.0046$), while GIS exhibited a positive but non-significant association ($p = 0.117$). The combined model integrating both variables achieved an Area Under the Curve (AUC of 0.81, outperforming either predictor alone (GIS only: AUC = 0.69; CNVs only: AUC = 0.72) (**Supplementary Fig. S12**). To evaluate the clinical performance of the TRACER-HL score, an optimized probability threshold of 0.15 was selected by maximizing the Youden index ($J = 0.564$) across all evaluated cutoffs (**Supplementary Fig. S13**). At this threshold, the model achieved a sensitivity of 83.3%, correctly identifying 10 out of 12 patients who experienced relapse, and a specificity of 73.1%, with 38 out of 52 non-relapsing patients accurately classified as low risk. The positive predictive value (PPV) was 41.7%, while the negative predictive value (NPV) reached 95.0%, highlighting the model's strong capacity to rule out relapse in patients deemed low risk (**Supplementary Fig. S13**). These metrics correspond to a contingency distribution of 10 true positives, 38 true negatives, 14 false positives, and 2 false negatives.

To further assess the robustness and generalizability of the TRACER-HL model, we performed an in-silico bootstrap validation using 1000 iterations of resampling with replacement. This approach allowed us to evaluate the stability of model performance metrics under simulated cohort variability. The mean AUC across all bootstrap replicates was 0.80 (95% CI: 0.63–0.82) (Fig. 4B), supporting the strong discriminative capacity of the model in distinguishing patients who would eventually relapse. The distribution of Youden index values was also consistent, with a mean of 0.56 (95% CI: 0.29–0.78), further

validating the selection of 0.15 as the optimal probability threshold (**Supplementary Fig. S14**). These distributions illustrate the model's robustness across simulated subsamples and reinforce the reliability of its prognostic utility.

DISCUSSION

In this study, we performed genome-wide profiling of HRS cells from patients with cHL to define genomic correlates of clinical outcome. Using low-pass WGS, we uncovered a highly complex CNVs landscape, consistent with the marked genomic complexity of HRS cells^{13,14}. Many recurrent gains and losses affected loci were previously implicated in cHL pathogenesis^{21–29}, reinforcing the contribution of immune escape, NF-κB and JAK/STAT signalling, and epigenetic deregulation to HRS-cell survival and disease biology^{11,14}.

This recurrent CNV landscape is further supported by its strong overlap with alterations previously described in purified HRS cells. In particular, amplification of the 9p24.1 locus, encompassing *PD-L1*, *PD-L2* and *JAK2*, is a defining genomic feature of cHL and a major driver of immune evasion and JAK–STAT pathway activation. The identification of this lesion in our cohort is therefore consistent with prior studies and further supports the biological relevance of our CNV findings^{24,30,31}. Likewise, Wienand et al. reported recurrent deletions involving 1p36.32 and 6q12, both of which were also detected in our series, with frequencies of 30.30% and 25.76%, respectively. The deletion at 1p36.32 includes *TNFRSF14*, a gene recurrently altered in germinal centre-derived B-cell lymphomas and implicated in immune escape mechanisms¹¹. Similarly, Maura et al. identified recurrent amplifications of 9p24 and 2p16 (*REL–XPO1*), as well as multiple recurrent deletions, several of which were also represented in our cohort, including 10q11 (31.82%), 8p23 (22.73%), 14q32 (18.18%), 16p11 (13.64%), 6p21 (15.15%) and 2p11 (15.15%). Taken together, these overlapping events reinforce the view that disruption of loci involved in immune regulation and B-cell signalling is a central feature of cHL biology¹³.

Beyond this descriptive landscape, we identified a small set of focal CNVs with prognostic relevance. Amplification at 2p11.2 and deletions at 2p13.2 and 11q25 were associated with shorter PFS, and 2p11.2 and 2p13.2 remained enriched in relapsed cases in machine-learning analyses. Although 2p gains are frequent in cHL²², these peaks do not define canonical cHL drivers, they affect regions containing genes plausibly linked to lymphoid signalling, transcriptional control, genome maintenance, and treatment resistance^{32,33}.

The SBS landscape of purified HRS cells was relatively simple, with most substitutions explained by five COSMIC signatures. Beyond ubiquitous clock-like processes³⁴, the prominence of SBS9 and SBS84 is consistent with the germinal-center B-cell origin of HRS cells and with AID-related mutagenesis^{35–37}. Our data further refine prior cHL WGS studies¹³ by identifying an inverse SBS5–SBS9 axis, in which an SBS9-high/SBS5-low AID-like profile was enriched in nodular sclerosis cHL and SBS1 exposure was modestly

lower in bulky disease. This pattern may reflect intertumoral differences in the relative contribution of germinal-centre AID activity versus background mutagenesis.

By contrast, SVs signatures outlined a comparatively homogeneous landscape dominated by SV2, with smaller contributions from additional patterns. The predominance of SV2, a COSMIC structural variant signature mainly characterized by non-clustered translocations, suggests that a relatively common background process of structural genome rearrangement underlies cHL rather than multiple sharply distinct SV programmes. As SV2 has been reported across diverse tumour types, its presence in our cohort is consistent with a broad pattern of chromosomal instability rather than a cHL-specific rearrangement mechanism.

Building on these observations, we developed TRACER-HL, a composite genomic risk model integrating global chromosomal instability and a focal CNV-based signature. This framework was designed to capture complementary dimensions of tumor-intrinsic relapse risk and may help identify, at diagnosis, patients less likely to respond to standard therapy. In this setting, baseline genomic information could complement clinical and imaging-based assessment to refine risk stratification and support more individualized treatment approaches.

This study has several limitations. It is based on a single prospective cohort with few relapse events and therefore requires external validation in larger independent series. Another limitation is the relatively small number of HRS cells recovered per case, which may not fully capture intratumoral heterogeneity or rare subclonal populations. Therefore, our genomic profiles should be interpreted as representative of the dominant tumour compartment rather than a comprehensive reconstruction of subclonal diversity. In addition, FFPE single-cell WGS and tumour-only analyses may limit the precision of mutation calling and somatic–germline discrimination. Functional studies will also be needed to clarify how relapse-associated CNVs and mutational processes contribute to HRS biology and treatment resistance. Nevertheless, our findings provide a genomic framework for baseline risk stratification in cHL.

Our study shows that cHL is characterized by a complex genomic landscape in HRS cells, dominated by recurrent CNVs affecting pathways involved in immune evasion, NF- κ B and JAK/STAT signalling. Beyond describing these alterations, we identified focal CNVs and mutational patterns associated with adverse clinical features and shorter progression-free survival, supporting their biological and potential clinical relevance. Importantly, integrating chromosomal instability with relapse-associated CNVs into the TRACER-HL model improved baseline risk stratification and identified a subgroup of patients with a very low probability of relapse. Overall, these findings support genome-wide characterization of purified HRS cells as a complementary approach to current clinical and imaging-based assessment and provide a rationale for future validation of genomically informed, risk-adapted strategies in cHL.

Abbreviations

cHL

Classical Hodgkin lymphoma

HRS
Hodgkin Reed-Sternberg
PET
Positron emission tomography
CNVs
Copy Number Variations
WGS
Whole-genome sequencing
SVs
Structural Variants
FFPE
Formalin-fixed paraffin-embedded
SNVs
Single nucleotide variants
SBS
Single base substitution
GIS
Genomic instability score
GA
Genomic altered
TRACER-HL
Tumor Relapse And Copy-number Evaluation in Reed–Sternberg cells - Hodgkin Lymphoma
AUC
Area under the curve
PFS
Progression-free survival
GHSG
German Hodgkin Study Group
ABVD
Doxorubicin,bleomycin,vinblastine,and dacarbazine
PMR
Partial metabolic response
PPV
Positive predictive value
NPV
Negative predictive value.

Declarations

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Written informed consent was obtained from all participants prior to sample collection. The study was approved by the Comité de Ética de la Investigación Provincial de Málaga and conducted in accordance with the Declaration of Helsinki and all applicable national and institutional regulations.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests

FUNDING

This study was partially funded by a research grant from the Sociedad Andaluza de Oncología Médica (SAOM) awarded to Laura Gálvez-Carvajal, the Grupo Oncológico para el Tratamiento y Estudio de los Linfomas (GOTEL) to Antonio Rueda-Domínguez, and institutional support from the Biomedical Research Institute of Málaga (IBIMA Plataforma BIONAND) awarded to Iñaki Comino-Méndez.

Author Contribution

J.V.S. and I.C.M. designed and performed experiments. L.G.C. coordinated patient recruitment and sample logistics, in collaboration with N.M.B., M.M.S., M.C.E., L.C.M., F.R.G.A., N.P.C., M.P.P. and A.R.D. E.L.L. and A.C.C. conducted bioinformatic analyses and generated genomic visualizations. M.A. and D.P. contributed to FFPE tissue processing. R.L.V., J.P., and E.A. supervised patient inclusion, clinical data curation, and sample collection. I.C.M. and A.R.D. jointly supervised the study as senior authors. All authors contributed to data interpretation, critically revised the manuscript, approved the final version, and consented to its publication.

Acknowledgement

We are deeply grateful to the patients and their families for their participation and trust. We thank the clinical staff and research teams at Hospital Universitario Virgen de la Victoria for their continuous support, and particularly the Medical Oncology Unit at Hospital Universitario Virgen de la Victoria for facilitating the development of this project. We also acknowledge the technical assistance provided by the Laboratorio de Biología Molecular del Cáncer (LBMC)/CIMES/University of Málaga. Funding for contract personnel was provided by the Spanish Government through the Ramón y Cajal Grant RYC2023-

Data Availability

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available in the NCBI Sequence Read Archive (SRA) under accession number PRJNA1302745 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA1302745>).

References

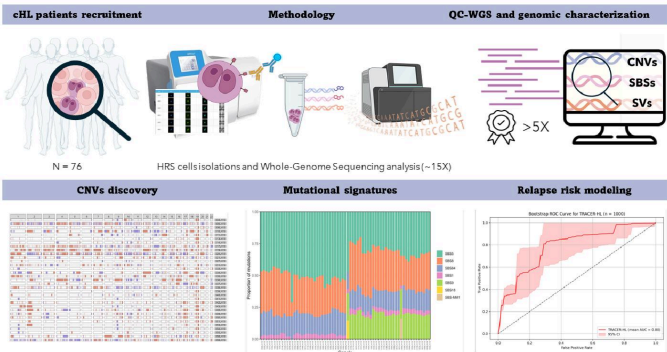
1. Mathas, S., Hartmann, S. & Kuppers, R. Hodgkin lymphoma: Pathology and biology. *Semin Hematol.* **53**, 139–147. 10.1053/j.seminhematol.2016.05.007 (2016).
2. Velasco-Suelto, J., Galvez-Carvajal, L., Comino-Mendez, I. & Rueda-Dominguez, A. Hodgkin lymphoma and liquid biopsy: a story to be told. *J. Exp. Clin. Cancer Res.* **43**, 184. 10.1186/s13046-024-03108-6 (2024).
3. Keraliya, A. R., Tirumani, S. H., Shinagare, A. B. & Ramaiya, N. H. Beyond PET/CT in Hodgkin lymphoma: a comprehensive review of the role of imaging at initial presentation, during follow-up and for assessment of treatment-related complications. *Insights Imaging.* **6**, 381–392. 10.1007/s13244-015-0407-z (2015).
4. El-Galaly, T. C. et al. Positron emission tomography/computed tomography surveillance in patients with Hodgkin lymphoma in first remission has a low positive predictive value and high costs. *Haematologica* **97**, 931–936. 10.3324/haematol.2011.056010 (2012).
5. Alaggio, R. et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* **36**, 1720–1748. 10.1038/s41375-022-01620-2 (2022).
6. Campo, E. et al. The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. *Blood* **140**, 1229–1253. 10.1182/blood.2022015851 (2022).
7. Mottok, A. & Steidl, C. Biology of classical Hodgkin lymphoma: implications for prognosis and novel therapies. *Blood* **131**, 1654–1665. 10.1182/blood-2017-09-772632 (2018).
8. Weniger, M. A. & Kuppers, R. NF-kappaB deregulation in Hodgkin lymphoma. *Semin Cancer Biol.* **39**, 32–39. 10.1016/j.semcancer.2016.05.001 (2016).
9. Tiacci, E. et al. Pervasive mutations of JAK-STAT pathway genes in classical Hodgkin lymphoma. *Blood* **131**, 2454–2465. 10.1182/blood-2017-11-814913 (2018).
10. Masel, R., Roche, M. E. & Martinez-Outschoorn, U. Hodgkin Lymphoma: A disease shaped by the tumor micro- and macroenvironment. *Best Pract. Res. Clin. Haematol.* **36**, 101514. 10.1016/j.beha.2023.101514 (2023).

11. Wienand, K. et al. Genomic analyses of flow-sorted Hodgkin Reed-Sternberg cells reveal complementary mechanisms of immune evasion. *Blood Adv.* **3**, 4065–4080. 10.1182/bloodadvances.2019001012 (2019).
12. Aoki, T. & Steidl, C. Novel insights into Hodgkin lymphoma biology by single-cell analysis. *Blood* **141**, 1791–1801. 10.1182/blood.2022017147 (2023).
13. Maura, F. et al. Molecular Evolution of Classic Hodgkin Lymphoma Revealed Through Whole-Genome Sequencing of Hodgkin and Reed Sternberg Cells. *Blood cancer discovery.* **4**, 208–227. 10.1158/2643-3230.Bcd-22-0128 (2023).
14. Reichel, J. et al. Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. *Blood* **125**, 1061–1072. 10.1182/blood-2014-11-610436 (2015).
15. Ansell, S. M. & Hodgkin lymphoma A 2020 update on diagnosis, risk-stratification, and management. *Am. J. Hematol.* **95**, 978–989. 10.1002/ajh.25856 (2020).
16. Mangano, C. et al. Precise detection of genomic imbalances at single-cell resolution reveals intra-patient heterogeneity in Hodgkin's lymphoma. *Blood cancer J.* **9** 10.1038/s41408-019-0256-y (2019).
17. Rustad, E. H. et al. mmsig: a fitting approach to accurately identify somatic mutational signatures in hematological malignancies. *Commun. Biol.* **4**, 424. 10.1038/s42003-021-01938-0 (2021).
18. Cheson, B. D. et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J. Clin. Oncol.* **32**, 3059–3068. 10.1200/JCO.2013.54.8800 (2014).
19. He, X. et al. Analysis of genomic alterations in primary central nervous system lymphoma. *Med. (Baltim).* **102**, e34931. 10.1097/MD.00000000000034931 (2023).
20. Mata, E. et al. Analysis of the mutational landscape of classic Hodgkin lymphoma identifies disease heterogeneity and potential therapeutic targets. *Oncotarget* **8**, 111386–111395. 10.18632/oncotarget.22799 (2017).
21. Steidl, C. et al. Genome-wide copy number analysis of Hodgkin Reed-Sternberg cells identifies recurrent imbalances with correlations to treatment outcome. *Blood* **116**, 418–427. 10.1182/blood-2009-12-257345 (2010).
22. Joos, S. et al. Classical Hodgkin lymphoma is characterized by recurrent copy number gains of the short arm of chromosome 2. *Blood* **99**, 1381–1387, (2002). 10.1182/blood.v99.4.1381
23. Martin-Subero, J. I. et al. Recurrent involvement of the REL and BCL11A loci in classical Hodgkin lymphoma. *Blood* **99**, 1474–1477. 10.1182/blood.v99.4.1474 (2002).
24. Green, M. R. 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* **116**, 3268–3277. 10.1182/blood-2010-05-282780 (2010).
25. Joos, S. et al. Genomic imbalances including amplification of the tyrosine kinase gene JAK2 in CD30 + Hodgkin cells. *Cancer Res.* **60**, 549–552 (2000).

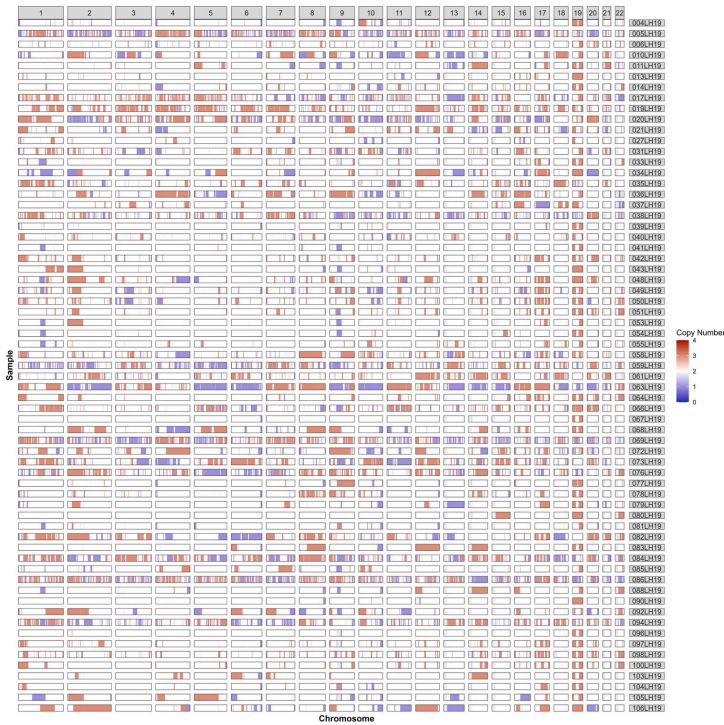
26. Rui, L. et al. Cooperative epigenetic modulation by cancer amplicon genes. *Cancer Cell*. **18**, 590–605. 10.1016/j.ccr.2010.11.013 (2010).
27. Kuppers, R., Engert, A. & Hansmann, M. L. Hodgkin lymphoma. *J. Clin. Invest.* **122**, 3439–3447. 10.1172/JCI61245 (2012).
28. Camus, V. et al. Detection and prognostic value of recurrent exportin 1 mutations in tumor and cell-free circulating DNA of patients with classical Hodgkin lymphoma. *Haematologica* **101**, 1094–1101. 10.3324/haematol.2016.145102 (2016).
29. Salipante, S. J. et al. Recurrent somatic loss of TNFRSF14 in classical Hodgkin lymphoma. *Genes Chromosomes Cancer*. **55**, 278–287. 10.1002/gcc.22331 (2016).
30. Vranic, S. et al. PD-L1 Status in Refractory Lymphomas. *PloS one*. **11**, e0166266. 10.1371/journal.pone.0166266 (2016).
31. Roemer, M. G. et al. PD-L1 and PD-L2 Genetic Alterations Define Classical Hodgkin Lymphoma and Predict Outcome. *J. Clin. oncology: official J. Am. Soc. Clin. Oncol.* **34**, 2690–2697. 10.1200/jco.2016.66.4482 (2016).
32. Salaverria, I. et al. A recurrent 11q aberration pattern characterizes a subset of MYC-negative high-grade B-cell lymphomas resembling Burkitt lymphoma. *Blood* **123**, 1187–1198. 10.1182/blood-2013-06-507996 (2014).
33. Buedts, L. et al. The landscape of copy number variations in classical Hodgkin lymphoma: a joint KU Leuven and LYSA study on cell-free DNA. *Blood Adv.* **5**, 1991–2002. 10.1182/bloodadvances.2020003039 (2021).
34. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101. 10.1038/s41586-020-1943-3 (2020).
35. Weniger, M. A. & Küppers, R. Molecular biology of Hodgkin lymphoma. *Leukemia* **35**, 968–981. 10.1038/s41375-021-01204-6 (2021).
36. Hernandez-Verdin, I. et al. Pan-cancer landscape of AID-related mutations, composite mutations, and their potential role in the ICI response. *NPJ Precis Oncol.* **6** 10.1038/s41698-022-00331-2 (2022).
37. Wong, J. K. L. et al. Association of mutation signature effectuating processes with mutation hotspots in driver genes and non-coding regions. *Nat. Commun.* **13**, 178. 10.1038/s41467-021-27792-6 (2022).

Figures

Fig. 1 A



B



C

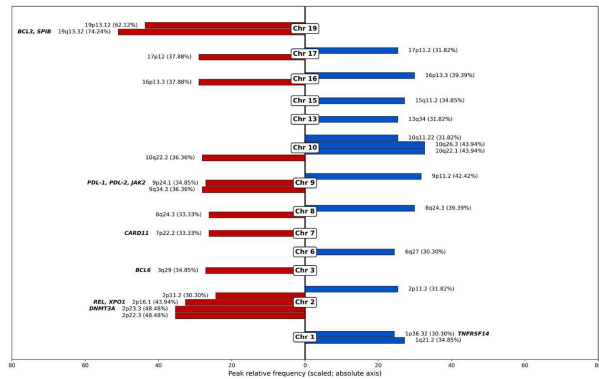


Figure 1

Genomic landscape of CNVs in HRS cells. (A) Overview of the study workflow, including patient recruitment, HRS-cell isolation, whole-genome sequencing, genomic characterization and relapse-risk modelling. **(B)** Genome-wide distribution of CNVs across the cohort. Gains (red) and losses (blue) are displayed across all autosomes chromosomes. **(C)** Representation of recurrent CNVs identified in the discovery analysis and filtered to retain events present in $\geq 30\%$ of patients. Amplification peaks are

shown on the left (red) and deletion peaks on the right (blue). Each label indicates the cytoband region and its frequency, with selected driver/biologically relevant genes annotated in bold italics.

Fig. 2

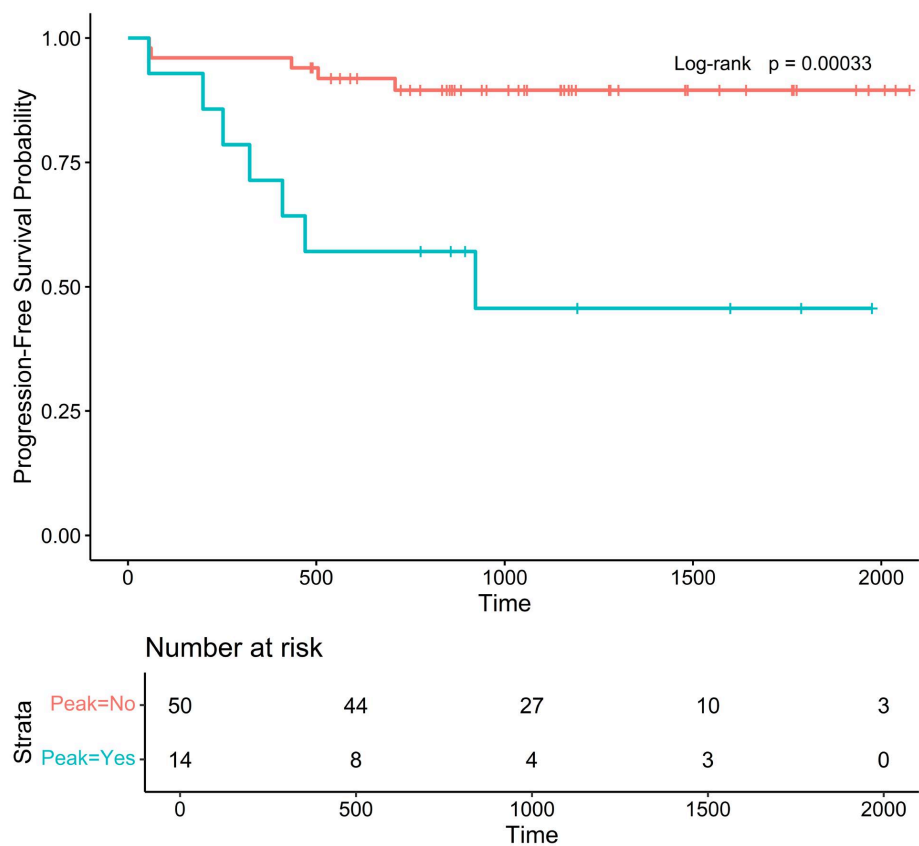


Figure 2

Kaplan-Meier curve showing PFS according to the presence of the three specific CVNs identified as associated with relapse. Patients were stratified based on the detection of more than two of these

peaks (Peak = Yes, n = 14) versus their absence (Peak = No, n = 50). The presence of these peaks was significantly associated with inferior PFS (log-rank p = 0.00033). Time is represented in days from diagnosis.

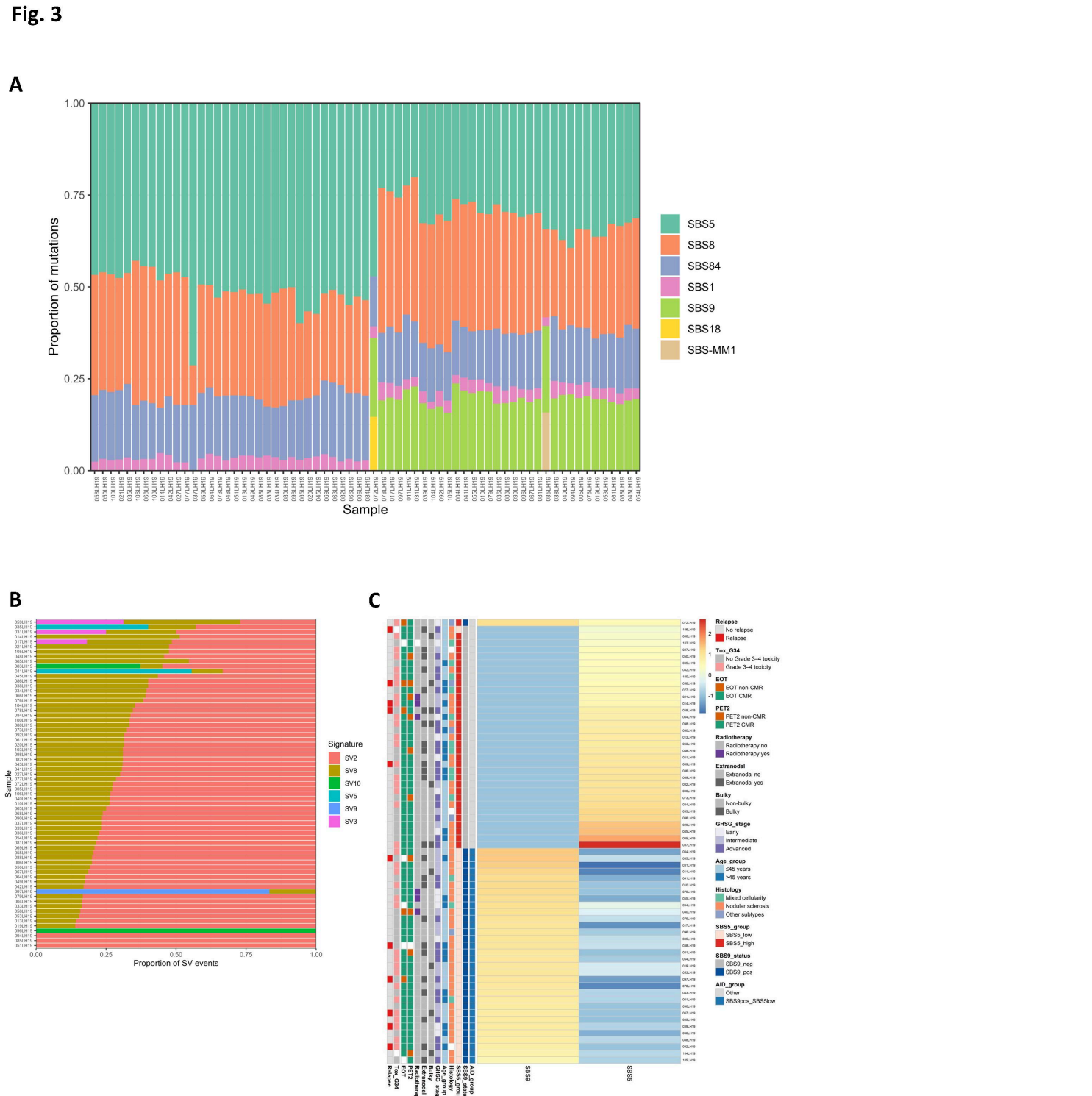


Figure 3

SBS/SVs mutational landscape and AID-like SBS9/SBS5 pattern in cHL. **(A)** Per-sample composition of COSMIC SBS signatures showing the six most prevalent signatures; all remaining signatures are grouped as “Other”. Bars represent the proportion of total SBS mutations attributed to each signature in each tumour. **(B)** Stacked barplot showing the relative contribution of SVs mutational signatures in each case. Each horizontal bar represents one tumour sample, and coloured segments indicate the fraction of SV events attributed to each detected SVs signature. Signatures with non-zero contribution in at least one sample are displayed. **(C)** Heatmap of SBS5 and SBS9 exposure across the cohort, highlighting the emergence of an AID-like subgroup. Signature exposures are displayed as z-scores (blue = lower, red = higher than the cohort mean). Rows correspond to individual samples. The annotation tracks summarize clinical and pathological variables, as well as derived mutational categories (SBS5_group, SBS9_status and AID_group, defined as SBS9_pos/SBS5_low versus all other cases). Unavailable clinical data are shown in white in the plot.

Fig. 4



Figure 4

In-silico validation of the TRACER-HL model. (A) Summary plot of genomic and clinical features for each patient included in the model. Each row represents a patient, and each column corresponds to a clinical or genomic variable. Patients on the left remained relapse-free, whereas those on the right relapsed. **(B)** Receiver operating characteristic (ROC) curve of the TRACER-HL model applied to the full

cohort. The model achieved an AUC of 0.80, demonstrating good discriminative performance in relapse prediction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPLEMENTARYTABLES13.xlsx](#)
- [SUPPLEMENTARYMETHODS.docx](#)
- [SUPPLEMENTARYFIGURES114.pdf](#)