Asynchronous Transitions from Hepatoblastoma to Carcinoma in High-Risk Pediatric Tumors

Xinjian Yu[1,*], Stephen Sarabia[1,2,*], Martin Urbicain[1,2,*], Sonal Somvanshi[1], Roma Patel[1,3], Tuan M Tran[4], Yen-Ping Yeh[5], Keng-Shih Chang[5], Yi-Tzu Lo[5], Jessica Epps[1], Kathleen A. Scorsone[1], Hua-Sheng Chiu[1], Emporia Faith Hollingsworth[1,2], Cintia R. Perez[1,2], Mohammad Javad Najaf Panah[1], Barry Zorman[1], Milton Finegold[1,2], John A. Goss[1,2], Rita Alaggio[6], Angshumoy Roy[1,2], Kevin E. Fisher[1,2], Andras Heczey[1], Sarah Woodfield[1,3], Sanjeev Vasudevan[1,3], Kalyani Patel[1,2], Ting-Wen Chen[5,#], Dolores Lopez-Terrada[1,2,#], Pavel Sumazin[1,#]

1. Department of Pediatrics, Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX, USA.
2. Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, USA.
3. Divisions of Pediatric Surgery and Surgical Research, Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, USA.
4. Department of Systems Biology, Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX, USA.
5. Biological Science and Technology, Center for Intelligent Drug Systems and Smart Bio-Devices, and Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.
6. Department of Pathology, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

*Equal contribution.

#Correspondence to dodochen@nctu.edu.tw, dhterrad@texaschildrens.org, sumazin@bcm.edu

ABSTRACT

Most malignant hepatocellular tumors in children are classified as either hepatoblastoma (HB) or hepatocellular carcinoma (HCC), but some tumors demonstrate features of both HB and HCC[1-3]. These tumors have been recognized under a provisional diagnostic category by the World Health Organization and are distinguished from HB and HCC by a combination of histological, immunohistochemical, and molecular features[4-6]. Their outcomes and cellular composition remain an open question[7-9]. The heterogeneous histological and molecular profiles of hepatoblastomas with carcinoma features (HBCs)[4] may result from cells with combined HB and HCC characteristics (HBC cells) or from mixtures of cells displaying either HB or HCC signatures. We used multiomics profiling to show that HBCs are mixtures of HB, HBC, and HCC cell types. HBC cells are more chemoresistant than HB cells, and their chemoresistance—a driver of poor outcomes[10-12]—is determined by their cell types, genetic alterations, and embryonic differentiation stages. We showed that the prognosis of HBCs is significantly worse than that of HBs. We also showed that HBC cells are derived from HB cells at early hepatoblast differentiation stages, that aberrant activation of WNT-signaling initiates HBC transformation, and that WNT inhibition promotes differentiation and increases sensitivity to chemotherapy. Furthermore, our analysis revealed that each HBC is the product of multiple HB-to-HBC and HBC-to-HCC transitions. Thus, multiomics profiling of HBCs provided key insights into their biology and resolved major questions regarding the etiology of these childhood liver tumors.

INTRODUCTION

Hepatoblastomas (HBs) and hepatocellular carcinomas (HCCs) are the most frequent primary hepatocellular tumors diagnosed in children[13]. While most cases can be definitively diagnosed as either HBs or HCCs, some hepatocellular tumors demonstrate combined HB-HCC molecular and histological features. Such tumors were formally recognized in 2014 under the provisional hepatocellular neoplasm–not otherwise specified (HCN NOS) diagnostic category[14]. Initially, the diagnosis and classification of HCN NOS were based solely on histopathology[1,2]. However, recent studies have characterized their molecular features and proposed integrated diagnostic approaches that combine immunohistochemistry and molecular biomarkers to accurately diagnose patients with HCN NOS and related tumors under the broader hepatoblastoma with carcinoma features (HBC) category[4]. Consequently, using a combination

of immunohistochemistry and molecular biomarkers, we distinguish between three types of childhood hepatocellular tumors: HB, HBC, and HCC[4,5].

Histological evaluations of HBCs suggested classification into three subtypes: biphasic, which contain homogenous HB and HCC regions; equivocal, which are composed of intermingled cells with HB and HCC features; and focal, which, before and after chemotherapy, contain small areas displaying high-grade HB-HCC features—including atypia and pleomorphism—enveloped by HB cells. The molecular profiles of all HBC subtypes revealed recurrent genetic alterations alongside the activation of marker genes and pathways associated with HB and HCC[4]. A key unresolved question is whether HBCs are solely composed of a mixture of HB and HCC cells or if they also include cells with unique HBC-specific DNA alterations and combined HB-HCC immunohistochemical and gene expression profiles (HBC cells). Furthermore, due to limited data, the incidence of HBC remains unclear, and the outcomes have been reported to be comparable to either those of HB or HCC, with respective survival rates exceeding 70% and falling below 40%, respectively[15,16]. In summary, HBC biology and etiology remain poorly understood and strategies for their treatment are being investigated[17].

To address critical questions surrounding HBC biology and outcomes, we clinically characterized 42 HBC cases and used high-resolution multiomics approaches to profile a subset of these cases. Our findings confirm that HBCs are linked to poor outcomes, with overall survival rates below 50%. Our analyses addressed unresolved questions regarding the cellular composition of HBCs and the phylogenetic relationships among the subclones that constitute them. Ultimately, our research offers valuable insights into the etiology of HBC, their responses to therapies, and the conditions necessary for HBC evolution.

RESULTS

We identified molecular signatures that distinguish between histologically classified HBs and HCCs at single-cell resolutions and showed that these features can be used to distinguish HB from HBC and HCC cells and tumors. Namely, tumors that were definitively diagnosed as HBs or HCCs and all tumor cells that comprise them were accurately classified as HBs or HCCs, respectively. In contrast, tumors falling under the HBC category, including all tumors diagnosed as HCN NOS, were composed of neoplastic cells with distinct HB molecular signatures, HCC molecular signatures, or their combination (HBC cells). Analogous to HBC tumors, HBC cells were characterized by specific DNA alterations and the activation of both HB and HCC marker genes and pathways. Our results support the hypothesis that HBCs are derived from HB precursor cells[2,4] and can transition to cells with HCC histology and molecular features.

Lower survival rates for HCCs are linked to their higher resistance to chemotherapy[18]. We showed that HBC cells exhibit greater chemoresistance than HB cells, resulting in poor patient outcomes and an overall survival rate below 50%. We observed that each HBC tumor contains multiple genetic subclones, and each genetic subclone comprises cells from multiple tumor types—HB, HBC, and HCC—at early embryonal differentiation stages. Thus, we concluded that each HBC tumor arises from multiple asynchronous transitions from HB to HBC and from HBC to HCC at early embryonic differentiation stages. Our results suggest that both HBCs and their precursor cells are stochastically halted at early hepatoblast differentiation stages by HB-initiating genetic alterations that activate the WNT-signaling pathway[19-21]. Moreover, we show that WNT-signaling inhibition can promote HBC differentiation and enhance their sensitivity to standard-of-care chemotherapy. Note that our results on initiating DNA alterations in WNT-signaling pathway genes are consistent with previous findings in HBs and HCCs[22-24].

HBC outcomes are poor

Studies that clinically characterized HBs at scale include the Children's Hepatic Tumors International Collaboration with 1605 HBs with at least 2 years of follow up post-diagnosis[23]; Cairo et al. (2008 and 2020) with 85 and 174 HBs, respectively[25,26]; Sumazin et al. (2017) with 69 HBs[27]; Nagae et al. (2021) with 139 HBs[28]; and Hirsch et al. (2021) with 86 HBs[29]. Here, we collected outcome data for 42 HBCs, including 41 with recorded events for at least 2 years post-diagnosis (Table S1). A comparison of event and outcomes data suggests that HBCs have significantly worse outcomes than HBs, including worse event-free survival (Figure 1A) and overall survival (Figure 1B). The overall survival for HBs across these studies was 79% on average, compared with 43% for HBCs. Pre-treatment HBC biopsies were highly heterogeneous, while post-treatment HBC resections often revealed regions of chemoresistant cancer

cells that differed from those observed in HBs; representative pre- and post-treatment HBC regions are illustrated in Figures S1A-B.

HBC resections contain HBC cells

To molecularly distinguish between the profiles of HB and HCC cells, we developed a gene expression-based classification criterion using RNA expression profiles of 4 HBs, 3 HCCs, and 1 tumor-adjacent liver sample at single-cell resolutions (Figure S2A). Our analysis revealed 45 genes specifically upregulated in HB cells and 265 genes uniquely elevated in HCC cells (Table S2), both of which were also upregulated compared to tumor-adjacent normal cells. The HB-HCC classifier evaluates whether either HB or HCC biomarker sets are significantly elevated relative to other cells (see Methods), and it effectively identified cells and samples that were profiled at single-cell resolutions (Figure S2B and Table S3) and at bulk by RNA sequencing (RNA-Seq) and NanoString (Figure S2C and Table S4). Given that HBC tumors and HBC cells are expected to exhibit upregulation of both the HB and HCC biomarker sets, they could not be classified as HBs or HCCs and were instead identified as HBCs.

We used the HB-HCC classifier to assess single-nuclei RNA-Seq (snRNA-Seq) profiles of resection samples from four biphasic HBCs (TLTX01, TLTX23, TLTX25, and TLTX33) and one equivocal HBC (TLTX18). As supplementary controls, we also analyzed profiles from a resection of a high-risk HB in an older child and biopsies from two HBCs involved in the AGAR trial[30]. The HB sample was selected as a control due to its comparable risk and age characteristics to HBCs, despite its homogeneous HB histology, and we expected its composition to consist entirely of HB cells. In contrast, the AGAR trial represents a phase 1 salvage clinical study aimed at treating pediatric patients with relapsed and refractory tumors using T cells that have been genetically engineered with a GPC3-CAR and IL15 (NCT04377932). HCC cells are more resistant to therapy[18], and patients with HBC participating in this trial have undergone intensive treatment involving a combination of chemotherapies and immunotherapies. Consequently, these samples were expected to contain few remaining HB cells. Our analysis revealed that all HBC resections, including biphasic HBCs that were expected to contain homogeneous regions that are solely composed of HB or HCC cells, contained HBC cells (Figures 2A-B and S2D and Table S5). As expected, the high-risk HB sample was exclusively composed of HB cells, and AGAR HBCs were predominantly composed of HCC cells (93%). Each HBC sample contained at least 14% HBC cells, with HBC cells constituting more than half (51%) of the total cells across the profiled HBC resections.

Both HB and HCC signature genes are upregulated in HBC cells

Hierarchical clustering of bulk RNA expression profiles of fetal liver and hepatocellular tumor samples is known to produce distinct clusters of fetal liver, HB, and HCC profiles[27]. In contrast, bulk RNA expression profiles of HBC resections are known to encompass a transcriptomic spectrum ranging from HB to HCC. Some HBC profiles align with HBs and others with HCCs; however, the majority form an intermediate cluster characterized by the upregulation of both HB- and HCC-specific genes and pathways[4]. These observations indicate that HBCs are intermediate HB-HCC tumors[4], yet they do not fully address the intratumoral heterogeneity of HBCs. Given that HBCs consist of a mixture of HB, HBC, and HCC cells, we hypothesized that single-cell resolution profiles will yield valuable insights into the transcriptomic similarities among the various cell types within HBC tumors.

Our analysis of single-cell RNA (scRNA) profiles of HBCs indicated that HBC cells exist in an intermediate state between HB and HCC. This finding is illustrated in a uniform manifold approximation and projection (UMAP) plot (Figure 2C) and supported by hierarchical clustering (Figure 2D and Table S6). Namely, HB and HCC cells from each HBC tumor clustered with their respective counterparts, while the HBC cells exhibited a transcriptomic continuum from HB to HCC. For instance, 71% of TLTX18 cells were HBCs and were predominantly clustered in an intermediate HB-HCC cluster, while 23% and 6% of TLTX18 cells were classified as HCCs and HBs, respectively, and these cells clustered alongside HCC and HB cells from other tumors. Moreover, we identified genes and pathways that were differentially expressed across tumor types, and as expected, these genes and pathways had intermediate HB-HCC expression patterns in HBC samples and clusters (Figures 2E-F and Table S7). For example, while Cairo's HB signature[26]

was upregulated in HB cells, their high-risk signature was more upregulated in HBC and HCC cells, whereas Hoshida's HCC gene set[31] was downregulated in HB cells.

High-risk HB, HBC, and HCC cells are less differentiated

We developed a model to infer the embryonic differentiation stage of liver cells based on single-cell RNA sequencing (scRNA-Seq) profiles of human embryonic liver cells at 5 and 19 weeks post-conception (Figure S3A)[32]. This model allows for the scoring of cells or samples according to their RNA expression profiles; detailed methodology can be found in the Methods section. To assess the model's performance, we scored human embryonic liver cells 6 to 16 weeks post-conception (see Figure S3B and Table S8) and mouse embryonic liver cells at 11 to 17.5 days post-conception (refer to Figure S3C and Table S8). Both datasets demonstrated accurate classification of cells, effectively reproducing the embryonic age of the samples in each dataset, with statistical significance ($p=8E-3$ and $p=7E-3$, respectively, using hypergeometric distribution).

An analysis of the embryonic differentiation stages of cells and tumors revealed that pre-treatment HBs are predominantly composed of cells at later embryonic differentiation stages, while HCCs are enriched with cells at early embryonic differentiation stages. In contrast, post-treatment HBs and HCCs are the most embryonically undifferentiated (refer to Figure S3F and Table S9). Furthermore, a comparative study of the composition of scRNA-Seq profiled HBs[33] suggested that low-risk tumors are associated with a higher prevalence of cells at later embryonic differentiation stages (Figure S3G, Table S9) and that HBCs are less differentiated than low-risk HBs (Figures S3F-H and Table S9, $p=0.02$ by t-test). Similarly, HB differentiation scores are closely aligned with risk estimates for HB tumors (Figures 3A-B, Table S10).

Comparative analysis of embryonic differentiation scores between pediatric and adult non-cancerous and cancerous samples demonstrated that adult HCCs[34] and fetal liver samples exhibit a significant enrichment of cells at early embryonic differentiation stages when compared to adult liver samples ($p<2E-16$ and $p=7E-4$, respectively, by t-test). Notably, pediatric liver cancer cells were significantly more undifferentiated than other cancer and non-cancer liver samples (Figure 3C and Table S10, $p=7E-10$ by t-test). Moreover, the inferred proliferative activity of tumor cells indicates that undifferentiated HBC cells tend to be more quiescent (Figure S3I), and quiescence is linked to chemoresistance[35]. Overall, these findings indicate that enrichment with cancer cells at early embryonic differentiation stages is associated with higher chemoresistance in both HBs and HBCs.

LIN28B is a recognized biomarker for both embryonic differentiation[27,36] and HB risk (Figure 3B). We showed that LIN28B expression is correlated with embryonic differentiation in the developing liver, HBs, and HBCs (Figures 3D and S3J). LIN28B regulates[37-39] and its expression is inversely correlated with *let-7* microRNA (miRNA) activity in HBC cells (Figure 3D). By suppressing *let-7* miRNA biogenesis, LIN28B helps preserve the stem cell properties of hepatoblasts during liver development[40]. Furthermore, information-theoretical analysis using the data processing inequality revealed that the association between cell differentiation and *let-7* miRNA activity is dependent on LIN28B expression, suggesting that LIN28B abundance modulates *let-7* miRNA activity in HBCs (Figure 3D and Table S11).

In addition to LIN28B, other prognostic HB biomarkers exhibit expression profiles that correlate with the embryonic differentiation stages of cancer cells. Notably, HMGA2 is upregulated while IGF2[27] is downregulated in undifferentiated cancer cells (Figures S3K and S3L). Pathways associated with mature-liver expression and activity, including Hsiao Liver Specific Genes and Lipid Metabolism, were activated in differentiated cells. In contrast, pathways associated with HCC and high-risk factors were enriched in undifferentiated cells (Figure 3E and Table S12). We note that both lipid metabolism and its associated gene *IGF2* demonstrated a correlation with embryonic cell differentiation in both liver cancer and non-cancerous cells (Figures 3E and S3L).

WNT-signaling activation regulates liver differentiation

DNA alterations in WNT-signaling genes drive aberrant WNT-signaling activation in nearly all HBs[41-43], and DKK1—an indicator of WNT-signaling activity—is a prognostic biomarker for HB (Figure 4A)[27]. WNT-signaling activity is closely related to embryonal liver differentiation, with DKK1 expression declining

exponentially from 5 to 9 weeks post-conception (Figure S4A). This decline is particularly evident in HB cells, where DKK1 levels are closely associated with the stages of embryonic differentiation in HB (Figure 4B). Although DKK1 downregulation also correlated with embryonic differentiation in HBC cells, its expression remains significantly elevated even in more differentiated HBC cells (Figure 4C).

We targeted the patient-derived undifferentiated HBC cell line HB17[17] with the WNT-inhibitor ICG-001[44] to evaluate the effect of WNT signaling on cancer cell differentiation. Although HB17 cells exhibited slow differentiation in culture, treatment with ICG-001 at concentrations of 5 or 10 μM for durations of 5 or 10 days led to a significant decrease in the expression of both DKK1 and LIN28B. This finding indicates a reduction in WNT-signaling activity and increased differentiation (Figures S4B-C and Table S13). To evaluate the effects of WNT inhibition at the single-cell level, we treated HB17 cells with ICG-001 (5 μM) for either (i) 10 days or (ii) 5 days followed by 5 days of incubation without the inhibitor. Comparative analysis of single-cell RNA expression profiles of treated cells to those cultured with a placebo revealed a significant shift in the transcriptomes of HB17 cells after treatment (Figures S5A and 4D). Remarkably, treated HB17 cells differentiated at a significantly faster rate than their untreated counterparts (Figure 4E and Table S14), resulting in clusters characterized by diverse DKK1 and LIN28B expression profiles.

Embryonic differentiation scores closely aligned with LIN28B expression levels (refer to Figures 4D and F). After 10 days, 80% of the placebo-treated cells maintained high DKK1 and LIN28B expression and remained in early embryonic differentiation stages. In contrast, over 70% of the cells treated with ICG-001 for 10 days exhibited low DKK1 and low LIN28B expression and were inferred to be more differentiated—with differentiation scores on par with fetal liver cells (Figure S3E). Interestingly, over 70% of the cells treated for 5 days and then rested for 5 days exhibited high DKK1 and low LIN28B expression and were predicted to be more differentiated. This result suggests that WNT-signaling inhibition with ICG-001 promoted differentiation, and that WNT-signaling activity recovered in the absence of the inhibitor, but cell differentiation was not reversed (Figures 4F and S5B-C and Table S15). As expected, very few cells with high DKK1 and high LIN28B expression were inferred to be more differentiated (Figure 4F); we note that to avoid circularity, LIN28B expression was not used for embryonic cell differentiation inference.

WNT-signaling inhibition sensitizes HBC cells to cisplatin by promoting cell differentiation

Adjuvant therapy with cisplatin followed by surgical resection is the standard of care for pediatric patients with liver cancer[45,46]. We developed a model for inferring cisplatin resistance based on bulk RNA expression profiling and resistance evaluation of pediatric liver cancer cell lines[29] (Figure S6A). This model effectively assessed resistance to cisplatin from scRNA-Seq profiles of treated patient-derived tumor spheroids[33] (Figure S6B). Furthermore, it was predictive of HB risk stratification (Figure S6C and Table S16) and identified HCC cells as more resistant to cisplatin than HB cells (Figure S6D). Notably, the model suggested an increase in mean cisplatin resistance following treatment in a partially chemoresistant patient-derived xenograft (PDX) model, which was profiled using scRNA-Seq both before and after cisplatin administration (Figure S6D). Importantly, LIN28B pseudobulk expression in our HB17 cell clusters (Figure 4G and Table S17) was significantly predictive of the model's cisplatin-resistance score, suggesting that HBC differentiation increases sensitivity to cisplatin.

We assessed the resistance of cells and samples to cisplatin using our predictive model. Following treatment with the WNT-inhibitor ICG-001, HB17 cells exhibited a marked decrease in cisplatin resistance (Figure 4H and Table S18). To explicitly evaluate the effectiveness of WNT inhibition and chemotherapy in combination, we administered ICG-001 at a concentration of 5 μM to HB17 cells, employing a 5-day-on and 5-day-off regimen. This approach resulted in a 5% reduction in cell viability (IC5), after which we applied cisplatin at the maximum dosage that did not affect cell viability (IC0) and at a dosage that induced a 10% reduction in viability (IC10). Treatment with ICG-001 at IC5—over 5 days, with or without a subsequent 5-day rest—followed by cisplatin at IC0 and IC10 led to significant decreases in cell viability compared to cisplatin alone, with p-values of 0.07 and 0.007, respectively (Figure 4I). Specifically, ICG-001 at IC5 followed by cisplatin at IC0 resulted in a 15% reduction in cell viability, while the combination of ICG-001 at IC5 and cisplatin at IC10 caused a 28% reduction. We conclude that targeting the WNT-signaling pathway may enhance the sensitivity of undifferentiated HB and HBC cells to chemotherapy.

Furthermore, our model indicates that HBC clusters exhibit greater resistance to cisplatin than their HB counterparts, regardless of differentiation stage (Figure 4J and Table S17, p=0.002 by t-test).

HBC cells are derived from HBs at early embryonic differentiation stages and transition to HCC

To reconstruct the phylogeny and map the cell subtypes and differentiation stages of HBCs, we used a combination of assays, including multi-region longitudinal whole exome sequencing (WES)[22], Affymetrix OncoScan profiles[4], single-nuclei DNA-Seq (snDNA-Seq)[47,48], Texas Children's Pediatric Solid Tumor Cancer Mutation Panels[49,50], NanoString bulk RNA expression profiles[4], and snRNA-Seq[51]. DNA alterations, including mutations and copy number variations (CNVs), were identified based on evidence from multiple assays. Note that the samples and patients described in Table S1 and regions from TLTX36 before and after treatment were profiled at bulk resolutions, while multiple regions from TLTX18, TLTX23, TLTX25, and TLTX33 were profiled at bulk, snDNA-Seq, and snRNA-Seq.

Because only WES profiles of one pre-treatment biopsy and five post-treatment resection regions were available to support the discovery of most mutations in TLTX36, we mandated that each mutation be identified in at least two regions. Inferred TLTX36 phylogeny indicated that alterations at *CTNNB1* and *KIF14* were present in each identified TLTX36 genetic subclone and that TLTX36 HBC cells had HB cell progenitors. Notably, while the pre-treatment TLTX36 biopsy primarily consisted of embryonically differentiated (fetal) HB cells, post-treatment resections exhibited a higher prevalence of embryonically undifferentiated HB and HBC cells. Profiled HB cells demonstrated loss of heterozygosity in chromosomes 9q and 11q, along with mutations in *NACA2*, *ARMC7*, *WDR87*, and *TSN* (Figure 5A and Tables S19 and S20). In contrast, HBC cells acquired additional copies of chromosomes 1, 8, and 17, along with mutations in HELZ and BIRC6. HBC cells continued to accumulate DNA alterations, exhibited a reduced likelihood of further differentiation, and displayed genetic divergence across spatial regions (see Figure 5A). To assess the accuracy of our inference of embryonic differentiation stages, slides were annotated for fetal and embryonal compositions, as well as for pleomorphic, teratoid, and mesenchymal histology by a team of clinical pathologists (MJF, DLT, and KP). Their qualitative assessments based on histology aligned with our quantitative embryonic differentiation scores (Table S19).

Evidence for mutations and copy number alterations in HBC cells were derived from OncoScan, snDNA-Seq (Figure S7), snRNA-Seq (Figure S8), and WES (Figure S9) assays; see Methods for details. We first categorized cells into genetic subclones based on their common genetic alterations, and then, each genetic subclone was further classified according to cancer cell types (HB, HBC, and HCC) and embryonic differentiation stages (early, intermediate, and late). The resulting phylogenies and cellular classifications are illustrated in Figures 5D-E. TLTX18 and TLTX33 profiles indicated the presence of 4 and 5 distinct genetic subclones, respectively, as determined by WES and snDNA-Seq profiles. However, only 3 out of 4 and 2 out of 5 of these subclones, respectively, corresponded with snRNA-Seq profiles in adequate cell abundance. Figures 5B-C present the inferred phylogenies and composition of TLTX18 and TLTX33, demonstrating that each genetic subclone encompassed cells from all cancer cell types, with each subclone subpopulation containing cells at early embryonic differentiation stages. Few HB cells from TLTX23 and HCC cells from TLTX25 were effectively profiled by snRNA-Seq, but similar to TLTX18 and TLTX33, each genetic subclone included undifferentiated cells from each tumor type (Figures 5D-E). Given the extremely low likelihood that, in each tumor, cells of each tumor type and each differentiation stage independently gained the same genetic alterations, these phylogenies revealed that transitions from HB to HBC and HBC to HCC take place during the early stages of differentiation among genetically distinct cancer cells within each profiled HBC tumor.

DISCUSSION

The provisional HCN NOS diagnostic category was included in the International Pediatric Liver Tumors Consensus during the International Pathology Symposium[14] in 2014 and is recognized by the World Health Organization. A detailed examination of HCN NOS and related tumors, including HBs that demonstrate focal atypia or pleomorphism, indicated that these tumors have common clinical, histological, and molecular characteristics[4]. Consequently, these tumor types were collectively referred to as HBs with HCC features (HBCs)[4]. Sumazin et al. (2022) distinguished between HBCs with histologically distinct HB or HCC tumor regions (biphasic), those exhibiting intermediate histological traits,

characterized by cells that display overlapping HB and HCC histological features (equivocal), and tumors with focal high-grade HB-HCC features (focal). Bulk molecular profiles of these tumors revealed that both HB- and HCC-specific gene expression and pathway activity levels are heightened in HBCs. Nevertheless, analysis of these profiles did not resolve a critical question regarding the cellular composition of HBCs: Are HB- and HCC-specific biomarkers elevated in HBC tumors because they comprise both HB and HCC cells, or do HBCs contain cells that express both HB- and HCC-specific genes?

We showed that most of the cells in HBC resections express both HB- and HCC-specific genes and pathways (HBC cells). Notably, this observation encompassed analyses of biphasic tumors, which were hypothesized to consist of homogeneous HB and HCC regions based on histological assessments. Furthermore, while previous reports based on bulk RNA-Seq profiles of HBC tumors indicated that HBCs possess intermediate HB-HCC expression programs[4], our findings revealed that each profiled HBC contained both HB and HCC cells with expression patterns akin to those found in HB and HCC tumors, respectively. Notably, most cells in HBC tumors displayed expression profiles that ranged from HB-like to HCC-like (Figure 2D), indicating that insights about the intermediate HB-HCC characteristics of HBC tumors extend to HBC cells. We note that HBC resections were obtained from post-treatment samples, where a diminished presence of HB cells was expected due to their lower chemoresistance.

We investigated the clonal evolution of HBCs using tumor maps that illustrated their clonal composition, categorized by tumor type and embryonic differentiation stages, as well as the phylogenetic relationships among genetic subclones (Figure 5). These maps revealed that HBC tumors emerge from multiple asynchronous transitions between HB and HBC, as well as between HBC and HCC. Specifically, we deduced that each genetic subclone consists of a blend of molecular cancer cell types—HB, HBC, and HCC—and that each clonal subpopulation encompasses cells at early stages of hepatoblast differentiation. Given the improbability of genetic convergence and embryonic differentiation reversal at this magnitude across all profiled tumors and genetic subclones, we concluded that each tumor is a product of multiple transitions originating from HB cells at early differentiation stages. This evidence negates the single-progenitor model for HBC genesis, which posits that all HBC cells stem from a singular HB-to-HBC transition event. Instead, it demonstrated that HBCs occupy an intermediate state between HBs and HCCs, indicating that HBCs derive from cells exhibiting HB molecular and histological characteristics, and that HBC cells have the potential to transition into cells displaying HCC features.

HBs have an average of fewer than 0.2 somatic mutations per Mb of sequenced DNA[27], while HBC mutation rates are estimated to be 10-20 times higher. Despite these low mutation rates, mutations in WNT-signaling pathway (WNT mutations) are hallmarks of HBs, HBCs, and pediatric non-fibrolamellar HCCs, with over 90% of HBs[27] and 100% of HBCs[4] harboring WNT mutations. WNT signaling plays a key role in regulating liver differentiation[43,52], and WNT mutations initiate liver tumor transformation[22], including in HBCs (Figure 5). Moreover, the extent of WNT-signaling pathway activation correlates with hepatoblast differentiation and is predictive of patient outcomes (Figures 4A-C and S4A). Our findings indicate that WNT-signaling pathway inhibition promotes liver cancer cell differentiation (Figures 4D-F and S5) and increases their sensitivity to chemotherapy (Figures 4H-I). This evidence suggests that initiating WNT mutations drive embryonic cell differentiation arrest in pediatric liver cancers and resolves key questions about the observed phenotypic heterogeneity of liver cancers. Specifically, heterogeneous HBs may undergo transformation during early stages of hepatoblast embryonic differentiation due to changes in WNT-signaling activity, which slows but does not halt the differentiation process of childhood hepatocellular tumor cells. This finding also implies that childhood hepatocellular tumors are composed of cells at multiple embryonic differentiation stages, including cells that may transition from HBs to HBCs and HCCs. Thus, HBs and HBCs can exhibit phenotypic heterogeneity even in the absence of genetic diversity.

HBCs exhibit a high prevalence of mutations commonly associated with HCC, with 84% of HBCs containing such mutations, even when excluding the frequently altered *CTNNB1* gene. HBCs are also enriched in alterations in pathways associated with stem cell pluripotency, PI3K-AKT, and mTOR signaling. Notably, mutations in the *TERT promoter* and *NFE2L2* have been recurrently identified in

HBCs, and HBCs are characterized by recurrent gains in chromosomes 1q, 2q, and 20, alongside losses in chromosomes 1p and 4q[4]. In our HBC cohort, we observed *TERT promoter* mutations in 33% of cases and *NFE2L2* mutations in 5% (Table S1). Tumors from patients with HBC that were profiled at single-cell resolutions (Figure 5) exhibited mutations in *NFE2L2*, *KEAP1*, *ARID1B* and *NRAS*; note that KEAP1 is a regulator of the protein coded by *NFE2L2*. These genetic alterations were recurrently found in HB cells across HBC genetic subclones, suggesting that these alterations may precede and facilitate transitions from HB to HBC. However, note that while we have established that HBC tumors are a product of transition events, specific molecular drivers that could be targeted in HBCs remain unidentified.

The study demonstrates that HBC composition, clonal genetic alterations, and differentiation stages play a role in tumor cell resistance to chemotherapy (Figure 4). Notably, HBC cells exhibit a higher resistance to chemotherapy than even high-risk HB cells (Figure 4J), and the 2-year survival of patients with HBC is below 50% (Figure 1). While chemotherapy is a prerequisite for complete surgical resection and successful treatment for most patients[10-12,25,53], its combination with surgery is primarily effective for lower-risk HBs. The overall survival rate for high-risk HBs and HCCs is below 50%[54,55]. The inherent heterogeneity of HBCs, which encompasses a wide transcriptomic range, from HB to HCC, poses significant challenges to their treatment. Although we have proposed strategies to target chemoresistant HBCs and other high-risk undifferentiated cancer cells, there is a pressing need for further high-resolution molecular investigations to uncover the underlying mechanisms driving their transition and associated chemoresistance. By leveraging the findings and conclusions of our study, these subsequent investigations will guide the development of combination therapies aimed at effectively targeting HBCs.

METHODS

snRNA-Seq protocol

To prepare nuclei, we added cold lysis buffer (NST-DAPI buffer with 0.1-0.2 U/μL murine RNase inhibitor) to frozen tumor tissues and cut the tissues into tiny pieces with scalpels. We used 1.5 mm beads to isolate nuclei by using a bead blaster at 4.0 m/s for three cycles of 20s on, with 10s pauses between cycles. We incubated the nuclear extract on ice for 5 min and passed it through a fluorescence-activated cell sorting (FACS) filter. We sorted the cells using FACS (BD FACSMelody) into 6 to 10 μL RNase inhibitor in 1.5 mL low-binding tubes coated with phosphate-buffered saline (PBS) + 1% bovine serum albumin (BSA). Suspensions were filtered through a 40-μm mesh, and single nuclei were flow-sorted. The DAPI intensity was used to set gates for both diploid and aneuploid cell populations. After sorting, we centrifuged the samples at 550 *g* for 5 min at 4°C and removed the supernatant (~20 μL left). Then, we slowly added resuspension buffer (PBS + 1% BSA + 0.2 U/μL RNase inhibitor) without resuspending the nuclei pellets and incubated the samples on ice for 5 to 10 min for buffer exchange. Finally, we centrifuged the samples at 550 *g* for 5 min at 4°C, removed the supernatant (~40 μL left), counted the cells, and performed snRNA-Seq[51].

Isolation of single nuclei by FACS

Frozen tumor tissue was lysed using an NST-DAPI lysis buffer as previously described[56]. Cell suspensions were filtered through a 40-μm mesh, and single nuclei were flow-sorted (BD FACSMelody). The DAPI intensity was used to set gates for both diploid and aneuploid cell populations. Single nuclei were sorted and then deposited into individual wells of 384-well plates (Eppendorf, 951020702). The sorting instrument alignment was assessed under a microscope before each experiment to ensure single nuclei were accurately deposited into the center of each well using a flat-bottom 384-well plate (Greiner, 781091).

Acoustic cell tagmentation procedure

FACS-sorted cells in 384-well plates were spun at 1500 *g* for greater than 4 min. The Echo 525 liquid handler (Labcyte) was used to dispense tagmentation reagents (Illumina, FC-131–1096) at nanoliter scales, with plate and liquid types detailed in the following steps. Thorough mixing and spinning of each plate after every dispense and incubation period is crucial to maximizing assay performance. Nuclei were lysed in 200 nL (384PP_SPHigh) of freshly prepared Tx lysis buffer (1.36 AU/mL protease diluted 1:9 in 5% Tween 20, 0.5% Triton X-100, and 30 mM Tris pH 8.0). Lysis thermocycler settings were programmed as 55°C 10 min, 75°C 15 min, 4°C ∞, lid = 80°C, vol = 1 μL. After lysis, 600 nL tagmentation reaction

mixture (TD: ATM 2:1, 384PP-Plus_GPSA) was dispensed per well. The ACT reaction settings on the thermocycler were 55°C 5 min, 4°C ∞, lid = 60°C, vol = 1 µL. The ACT reaction was neutralized with 200 nL (384PP_SPHigh) of NT buffer for 5 min at room temperature. The final polymerase chain reaction (PCR) included 1.11 µM N7XX (5′-CAAGCAGAAGACGGCATACGAGAT**XXXXXXXX**GTCTCGTGGGCTCGG-3′) and S5XX (5′-AATGATACGGCGACCACCGAGATCTACAC**XXXXXXXX**TCGTCGGCAGCGTC-3′) primers (384PP_AQBP) in 2× HiFi HotStart ReadyMix (Roche, KK2602, 6RES_GPSA). Dual barcode sequences in primers are denoted by "**XXXXXXXX**." Unique dual barcode combinations for each well of a 384-well plate were achieved by dispensing 16 unique N7XX barcodes across each row and 24 unique S5XX barcodes across each column. The PCR settings were 72°C 3 min, 98°C 30 s, then 15 to 18 cycles of 98°C 10 s, 63°C 30 s, 72°C 30 s, followed by 72°C 5 min, 4°C ∞, lid = 105°C, vol = 6 µL. ACT performance was evaluated by fluorometer (Qubit) and TapeStation (Agilent) from selected cell libraries. Final libraries were pooled and purified with 1.8× AMPure XP beads. The final libraries were sequenced at 50 single-read cycles with dual barcodes on the Illumina NextSeq 2000 P3.

Single nuclei 10x Genomics Chromium Single Cell 3′ protocols

We FACS-sorted nuclear suspensions (in PBS with 1% BSA and RNase inhibitor) and manually counted and concentrated the nuclei within the recommended range of 300 to 700 cells/µL for loading at a volume to recover 3000 to 8000 cells for snRNA-Seq. Single-nuclei capture, barcoding, and library preparation followed 10x Genomics Chromium Single Cell 3′ protocols (CG000183, v3.1). Four libraries were pooled to give a final concentration of 10 nM, and pooled samples were further tested by quantitative PCR for final concentration before submission for sequencing with the Illumina NextSeq 2000 sequencer using the P3 100 cycles flowcell. The samples were sequenced with 28 cycles for read 1, 8 cycles for i7 index, and 91 cycles for read 2 through the Advanced Technology Genomics Core at the University of Texas MD Anderson Cancer Center. Reads from single cells were demultiplexed and aligned using the human GRCh38-2020-A genome reference using the 10x Genomics Cell Ranger 6.1 pipeline for snRNA-Seq.

snRNA-Seq and scRNA-Seq preprocessing and clustering

The R package Seurat (v4.3.0)[57] was used for single-cell data preprocessing and clustering. For patient snRNA or scRNA-Seq data, cells with mitochondrial gene content above 15% and expressing fewer than 200 genes were excluded. The raw count matrix was normalized using the "LogNormalize" method with a scale factor of 10,000 and then scaled across cells. Principal component analysis was performed on all the genes. HBC samples profiled in two batches were integrated using canonical correlation analysis[58]. Clustering was based on the first 15 principal components and visualized by UMAP. Annotation of tumor and normal cell clusters was based on established pediatric liver cancer markers[26,27] and the automated annotation tool SingleR (v2.0.0)[59]. Because we focused on tumor composition, only high-confidence tumor cell clusters were included in further analyses. Cell clusters with ambiguous tumor or normal assignments were excluded. P17 cell clusters were integrated by sample using the standard Seurat package based on integration anchor identification (FindIntegrationAnchors function). When analyzing scRNA-Seq profiles of PDX samples, we classified the xenograft-derived sequence read data for further analyses using Xenome[60], which effectively handles a mix of reads from both the host and the graft. Cells from PDXs with mitochondrial gene content higher than 15% and those expressing fewer than 500 genes or more than 4000 genes were excluded to eliminate poorly or inconsistently profiled cells.

snDNA-Seq analysis

AneuFinder[61] R package (v1.24.0) was used for snDNA-Seq data preprocessing, quality control, and detecting aneuploidy. We used this tool to align reads to the GRCh38 reference genome, partition the genome into 1-Mb non-overlapping bins per cell, and apply a hidden Markov model to the binned data for unbiased detection of CNVs and aneuploidy.

Inferring CNVs from snRNA-Seq

We used the inferCNV R package (v1.8.1) to infer CNVs for each tumor cell from snRNA-Seq profiles of TLTX18, TLTX23, TLTX25, and TLTX33. A total of 3189 non-cancer hepatocytes from pediatric adjacent normal liver samples were used as the reference for CNV inference. To reduce false-positive detections,

only CNVs that were supported by at least one additional DNA profile of the same patient, either snDNA-Seq, WES, or Affymetrix OncoScan, were included in further analyses.

To define copy number states of each cell for detected CNV regions, we used two alternative criteria, in which the first assigned CNVs for more cells, while the second was more restrictive and aimed to eliminate potential contamination between genetic subtypes: (1) The inferCNV package applied a hidden Markov model to predict the CNV states. To reduce false-positive detections, CNVs supported by more than 7 of 10 runs of the hidden Markov model were used to define the CNV patterns for each cell. (2) For copy number gain, the average inferred copy number value across the chromosome regions of interest greater than an upper cutoff was defined as *copy number gain*, while less than a lower cutoff was defined as *copy number neutral*. The lower cutoff was set at the 50% quantile of the inferred copy number value of normal reference cells for each CNV region. For TLTX23, TLTX25, and TLTX33, the upper cutoff was set at the 95% quantile of the inferred copy number value of normal reference cells for each CNV region. For TLTX18, which had more tumor cells profiled, the upper cutoff was set at the 99% quantile. Similarly for copy number loss, the average inferred copy number value across the chromosome regions of interest less than a lower cutoff was defined as *copy number loss*, while greater than an upper cutoff was defined as *copy number neutral*. The upper cutoff was set at the 50% quantile of the inferred copy number value of normal reference cells for each CNV region. For TLTX23, TLTX25, and TLTX33, the lower cutoff was set at the 5% quantile of the inferred copy number value of normal reference cells for each CNV region. For TLTX18, the lower cutoff was set at the 1% quantile.

Whole exome sequencing analysis

A total of 40 tumor samples, including two biopsies of TLTX23 and TLTX36, and 12 matched normal samples were collected with consent from 12 patients from Texas Children's Hospital with approval from the institutional review board (H13999). The exon regions of these 52 DNA samples were captured using Agilent SureSelect and sequenced. Quality control was performed on sequence reads with FastQC[62], and sequence reads were trimmed with Trimmomatic[63]. Somatic variant discovery was performed with the Genome Analysis Toolkit (v4.1.9.0)[64]. The quality-controlled reads were aligned to the human GRCh38 reference genome[65] with the BWA-MEM algorithm in the Burrows-Wheeler Aligner[66] using options --TRAILING 20, --MINLEN 25, and PE, and the BAM files were indexed using Samtools[67].

To detect somatic variants, the two modes of Mutect2[68], single tumor sample and jointly multiple tumor samples, with the panel of 12 normal samples were used to call somatic variants. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were filtered with PASS tags by GATK FilterMutectCalls with default parameters and subsequently annotated by ANNOVAR (October 4, 2019, release) with refGene reference[69]. The results of these 2 modes were intersected to extract the final SNV and INDEL files for subsequent analysis. Mutations in select genes were also called using the Texas Children's Pediatric Solid Tumor Cancer Mutation Panel.

To detect CNVs in WES data, FACETS (v0.6.2)[70] was used to calculate the total copy number, minor copy number, and cancer purity based on SNV read counts from tumor-normal BAM files, with GRCh38p7_GATK_00-common_all as the reference. The cellular fraction (CF) of CNV (copy number alteration) estimated by FACETS can be transformed into a cancer cell fraction (CCF) using the formula CCF = CF/p, where p denotes tumor purity[71].

Inferring CNVs by Affymetrix OncoScan

Genome-wide copy numbers were estimated using DNA isolated from formalin-fixed paraffin-embedded (FFPE) tumor tissue using the Affymetrix OncoScan FFPE Assay Kit and OncoScan Console 1.3, and data were analyzed by Affymetrix ChAS 3.1 and OncoScan Nexus Copy Number 7.5 after Burrows-Wheeler Aligner alignment to GRCh38, as previously reported[4]. Affymetrix OncoScan detects copy number alterations at a 50-kb resolution and was extensively validated in our laboratory for clinical use.

RNA expression profiling by NanoString

We profiled the gene expression of each FFPE tissue region using a custom NanoString nCounter panel (Table S21), partially based on the PanCancer Pathways Panel and HB-specific genes (Table S22), as previously reported[4]. HB-specific genes were selected based on differential expression and risk-

predictive potential observed in previously reported HB expression profiling efforts and to represent key hepatocarcinogenesis pathways. Multiplexed measurements of gene expression through digital readouts of the relative abundance of mRNA transcripts were performed as follows: Hybridization of RNA to fluorescent reporter probes and capture probes, purification of the target/probe complexes using nCounter SPRINT cartridges, and imaging and analysis using nCounter SPRINT Profiler. Gene expression profiles across samples were normalized to equate positive-control probes, and expression estimates were z-score-normalized independently for each probe.

Tumor histopathology and patient characterization

With institutional review board approval, we clinically and molecularly characterized patients with HBC, including collecting information about age, diagnoses, observations in surgical pathology reports, histological reviews, and material available for molecular profiling. Histological reviews of glass slides from all cases, including diagnostic biopsies and resection, transplant, and metastasis samples, were performed by three pathologists (MJF, DLT, and KP), who confirmed the features in clinical pathology reports and selected representative areas of the tumor most suitable for molecular testing. Frozen samples and representative tumor areas were selected from resections based on displays of heterogeneity, atypia, or pleomorphism. Selected clinical information was collected for these patients, as allowed by the approved protocol.

HB17 Culturing and drug-response assays

The HB17 patient-derived cell line was cultured in media composed of 50% Minimum Essential Media (MEM) and 50% Hepatocyte Basal Media (HBM, Lonza, Walkerville, MD). The cells were incubated at 37°C in an atmosphere containing 5% CO2. Cisplatin and ICG-001 were administered at specified doses and time points. Cell viability was evaluated using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assays. IC values were determined by GraphPad Prism (GraphPad Software, Inc., La Jolla, CA) employing a nonlinear regression model of log(inhibitor) versus normalized response with a variable slope. RNA was extracted from the treated cells for qRT-PCR experiments to measure the expression levels of *DKK1* and *LIN28B*. Treated cells were also profiled by scRNA-Seq.

Estimation of cancer cell fractions

Variants that met the following criteria were analyzed using PyClone-VI (--num-clusters 20 --density beta-binomial --num-restarts 100)[22,71,72]: (1) The exonic function of mutations is neither a synonymous SNV nor unknown, following the annotations from the GRCh38 reference genome; (2) the coverage of mutations is at least 10 reads in all samples from a patient; and (3) the total copy number of the segment overlapping the mutations is not equal to zero. We ensured that variants with zero copy number were analyzed by adjusting the total copy number input to 2 if the copy number was zero and if the cancer cell fraction (CCF) of the CNV was less than 1.

SNVs with exonic function, i.e., non-synonymous SNV, startloss, stopgain, and stoploss, analyzed in PyClone-VI were converted to browser extensible data (BED) files to generate pileup files for scDNA and to variant call format (VCF) files for scRNA by Samtools mpileup. GRCh38.d1.vd1.fa and refdata-gex-GRCh38-2020-A (10x Genomics) were used in mpileup as references in scDNA and scRNA, respectively. The pileup files of each cell were used to extract reference counts and alternative counts. VCFs were used to check whether WES SNVs were identified in the scRNA BAM files before extracting a single-cell variant. The VarTrix coverage mode from 10x Genomics was used to extract single-cell reference counts and alternative counts for scDNA-Seq.

Reconstructing evolutionary trees from WES and single-cell resolution profiles

TLTX36 phylogeny was reconstructed based on the WES profiles of six tumor regions, a tumor-adjacent region, and a blood sample; TLTX36 A1, included in Figure 5A, is the diagnostic biopsy, while the rest are resection samples. Phylogeny inference by Chimaera followed previously described protocols[22]. For TLTX18 and TLTX25, the phylogenetic trees were reconstructed based on observations from their snDNA-Seq profiles (Figures S5A and S5C), assuming that tumor cells evolve continuously and each CNV arises only once. Tumor cells from the snRNA-Seq profiles with the corresponding inferred CNV patterns were then assigned to each node in the phylogenetic trees. We used criterion (1) for Inferring

CNVs from snRNA-Seq when defining CNV patterns. Any cells with contradictory node assignments based on CNVs from snRNA-Seq inference criterion (2) were excluded from further analysis. For TLTX23 and TLTX33, the phylogenetic trees were reconstructed based on inferred CNVs from their snRNA-Seq profiles. Using criterion (2) with a more stringent cutoff (upper cutoff at the 99% quantile for copy gain and lower cutoff at the 1% quantile for copy loss), we tested the sequence of CNV pairs of interest based on the number of cells supporting which CNV occurred first. If there were significantly more cells with CNV A and without CNV B than cells with CNV B and without CNV A, then CNV A was considered to occur before CNV B ($p<0.01$, hypergeometric). CNV pairs that could not be ordered were not distinguished and were included in the same node. Cells were assigned to the established nodes based on criterion (2). Finally, predicted ancestral relations between genetic subclones were tested to ensure criteria were significant and non-overlapping (Figure S8). Mutations inferred from snRNA-Seq data supported by at least one additional DNA profile from the same patient—either by snDNA-Seq or WES—were mapped to their comprising nodes in our phylogenetic trees.

HB-HCC subtype classification

We identified HB and HCC biomarker sets using a 2-step process. First, differential expression analysis comparing snRNA-Seq profiles of HB and HCC resections identified genes that were differentially expressed between HB and HCC cells. The samples included 2375 HB and 1468 HCC cells that passed quality control (Figure S2A). We identified 1427 differentially expressed genes, including 484 upregulated in HB and 943 upregulated in HCC. Then, to avoid sampling bias and ensure the generalizability of the gene set, we refined the gene set using two independent single-cell datasets. For HB, we compared 1918 HB cells with 1792 adjacent normal hepatocytes using a published scRNA-Seq dataset[33] and identified 45 of 484 of the HB genes as upregulated in HB cells relative to adjacent normal hepatocytes. Similarly, a comparison of 1487 HCC cells and 6170 adjacent normal cells from two pediatric AGAR-trial patients with HCC[30] found that 265 of 943 of the HCC genes were upregulated in HCC cells relative to adjacent normal hepatocytes. The resulting gene sets comprised the HB and HCC biomarker sets.

Cancer cell-type evaluations of profiled cells and nuclei samples were performed by comparing the mean normalized expression of the HB and HCC biomarker sets. For single-cell resolution profiles, raw count matrices were log-normalized with Seurat[57], and the expression values for HB- and HCC-specific genes were min-max normalized across cells. Raw HB and HCC scores were obtained by summing the normalized expression values of HB- and HCC-specific genes within cells. Finally, raw HB and HCC scores of tumor cells were transformed into z-scores using the means and standard deviations of raw HB and HCC scores of adjacent-normal cells. For the high-risk HB and 5 HBC resection samples, z-score transformations were based on adjacent normal cells in 5 HBCs. For the 2 AGAR-trial HBC samples, z-score transformations were based on adjacent normal cells within these 2 samples.

The cell-type classification criteria were as follows: Cells with an HB z-score greater than 2 and an HCC z-score less than 2 or a z-score difference of at least 2 between HB and HCC were classified as HB cells. Cells with an HB z-score less than 2 and an HCC z-score greater than 2 or a z-score difference of at least 2 between HB and HCC were classified as HCC cells. Cells with both HB and HCC z-scores greater than 2 were classified as HBC cells. Tumor cells that showed no significant upregulation (z-scores less than 2) of either HB or HCC genes were considered unassigned tumor cell types and were excluded from further analysis. The proportion of cells excluded in each sample was as follows. TLTX01: 25.7%; TLTX11: 4.3%; TLTX18: 2.6%; TLTX23: 4.0%; TLTX25: 0.5%; TLTX33: 5.9%; AGAR HBC: 4.1%. The visualization in Figure 2A shows evaluations of random samples (100 cells). When analyzing bulk profiles, evaluations were performed based on the available biomarkers—5 HB- and 8 HCC-specific genes by NanoString nCounter and 42 HB- and 245 HCC-specific genes by bulk RNA-Seq. The expression values for HB- and HCC-specific genes were min-max normalized across samples. Raw HB and HCC scores were obtained by summing the normalized expression values of HB- and HCC-specific genes within each sample. Like the methods described for single-cell resolution profiles, the raw scores were transformed into z-scores based on the means and standard deviations of the raw scores of the normal samples.

Liver cell embryonic differentiation inference

We trained a classifier to predict embryonic liver differentiation stages using scRNA-Seq profiles of human embryonic and fetal liver samples[32]. This dataset includes profiles of nine developing liver samples taken from embryos from 5 weeks (5W) to 19 weeks post-conception (19W) (Figure S3A). Cell-type annotation followed the outline provided by Wang et al[32]. We included 19,557 developing liver cells from the hepatoblast-hepatocyte lineage. We trained a random forest model to distinguish between 5W (early differentiation) and 19W (late differentiation) cells, selecting 20 tree-classifiers based on performance in training: Increasing the number of classifiers beyond 20 failed to result in significant training error improvement (Figure S3D). A total of 811 differentially expressed genes were used as predictive features. Here, given a profile, each classification tree produces a binary decision, <differentiated> or <undifferentiated>, and the differentiation score is the normalized count of <undifferentiated> calls, from 0 to 1. Namely, the predicted differentiation stage was quantified by a differentiation score as the proportion of <undifferentiated> votes among the 20 trees. To compare the proportion of undifferentiated cells in each tumor, we set a differentiation score cutoff based on its 95% quantile in profiles of 11,480 normal hepatocyte cells from a tumor-adjacent sample, producing a differentiation score cutoff of 0.4 (Figure S3E). To address modality differences and improve the generalizability of the trained model on bulk RNA profiles, each profile was transformed into ranks and the classifier evaluated gene ranks in each sample, pseudobulk, or cell profiles independently. Cells from intermediate time points (6W-16W) were used to evaluate classification quality, and independent testing was performed on an embryonic mouse liver scRNA-Seq dataset derived from samples taken from 11 (E11) to 17.5 (E17.5) embryonic days post-conception with 12,393 developing liver cells from the hepatoblast-hepatocyte lineage by Wang et al[32]. The results suggest significant quantitative predictive ability for human liver cells (Figure S3B, p=8E−3, hypergeometric distribution) and mouse liver cells (Figure S3C, p=7E−3, hypergeometric distribution).

Cisplatin-resistance prediction

Hirsh et al. profiled RNA and evaluated cisplatin resistance in nine pediatric liver cancer cell lines[29]. We identified 37 genes whose expression profiles were individually significantly predictive of cisplatin resistance in these lines (0.01% quantile by rank). This set included 20 genes whose expression correlated with cisplatin resistance (*AGPAT3*, *BRPF3*, *DENND5B*, *DTX2*, *FBXO6*, *FOXA1*, *GRTP1*, *IQCH-AS1*, *KLHL9*, *LINC00235*, *NFATC2*, *NFX1*, *PIK3C2A*, *RNPEP*, *SERPINB1*, *TAF8*, *THRB*, *UBAP1*, *UHRF2*, *ZNF417*), and 17 genes whose expression was negatively correlated with cisplatin resistance (*ANKRD34A*, *APEX2*, *ASB4*, *BMERB1*, *CA11*, *CAMTA1-IT1*, *CCDC74B*, *CERCAM*, *FAM216A*, *HMOX2*, *IL27RA*, *MAP4K1*, *MCTS1*, *METTL27*, *PRICKLE3*, *TBL3*, *UBE2A*). We computed the cisplatin-resistance score as the ratio between the mean expression profiles of genes in the two sets. When evaluating snRNA-Seq and scRNA-Seq data, the score was calculated based on pseudobulk transformation.

We evaluated our cisplatin-resistance score on the training data (Figure S6A, p=2E−4), on low- and high-risk HBs[27] (Figure S6B, p<0.01), on our reference HB and HCC samples (Figure S6C), on an HB PDX before and after treatment with cisplatin (Figure S6D), and on 4 HB organoids profiled and treated by Song et al.[33] (Figure S6E, p<0.04, hypergeometric distribution). When evaluating the cisplatin-resistance score for our HBC clusters, each cluster shown in Figure 2C for each patient was evaluated independently. The resistance scores were then normalized to ensure comparability by calculating the ratio of each resistance score to the resistance score of the corresponding patient's normal cells.

Evaluation of HBC proliferation vs. quiescence

We estimated the proliferative activity of HBC tumor cells across differentiation stages using cell cycle S-related genes identified by Whitfield et al[73]. To produce Figure S3I, HBC cells were categorized into early, medium, and late differentiation stages based on differentiation score cutoffs at 0.2 and 0.45. A proliferation score was calculated for each differentiation stage in the five patients with HBC by evaluating the mean of the pseudobulk expression values across the cell cycle S-related gene set and adjusting for the number of cells in each differentiation stage for each patient. To ensure comparability among the five patients with HBC, each proliferation score was converted to a ratio relative to the proliferation score of the early differentiation stage within each patient.

Differential gene expression analysis

When looking for differentially expressed genes—including during classifier construction—we used the "FindMarkers" function in Seurat, including genes that were detected in a minimum fraction of 25% of the cells in either of the two tested populations. Genes with a log2-fold change of greater than 0.5 and adjusted p-value of less than 0.01 (Bonferroni correction) were considered significantly expressed.

Gene set enrichment analysis

We performed gene set enrichment analysis (GSEA) to calculate pathway enrichment in single-cell clusters using the R package fgsea (v1.24.0). Gene sets were ranked based on fold change between the cell clusters so they could be compared, and the number of permutations was set at 10,000. Gene sets for pathways were obtained from the "C2" collection of the MSigDB database[74]. In Figure 2, GSEA was used to compare each HBC tumor cluster and adjacent normal cells. Figure 2D shows the $-\log_{10}$(adjusted p) values, with positive and negative signs determined by upregulation or downregulation relative to adjacent normal cells. For visualization, pathway scores were z-score transformed across the 10 HBC tumor clusters.

Estimating miRNA activity and data processing inequality from snRNA-Seq data

We used miReact[75] based on predicted miRNA targets[76,77] to estimate *let-7*-family[38] miRNA activity in cells with more than 1000 mapped unique molecular identifiers. Normalized mutual information (MI) was calculated between (i) estimated *let-7* activity and LIN28B expression, (ii) cell differentiation scores and LIN28B expression levels, and (iii) cell differentiation scores and estimated *let-7*-family miRNA activity. To evaluate MI significance and normalize MI, the expression values of LIN28B, estimated *let-7* activity, and cell differentiation scores were shuffled 1000 times, and the normalized MI calculated from the shuffled triplets was used to fit a generalized extreme value distribution. Furthermore, the differences were calculated between the smallest normalized MI values: One between *let-7* activity and LIN28B, and the other between cell differentiation scores and LIN28B expression levels, compared with the normalized MI between *let-7*activity and cell differentiation scores; p-values were estimated using a generalized extreme value distribution fit to the differences from shuffled data. The normalized MI was also calculated between (i) the expression levels of *let-7* targets and LIN28B, (ii) cell differentiation scores and LIN28B expression levels, and (iii) cell differentiation scores and the expression levels of let-7 targets. The same analysis was conducted for differentiation-related non-*let-7* target genes. A chi-squared test was used to examine whether the distribution of triplet pairs having the smallest normalized MI between cell differentiation scores and target genes differed significantly from that of differentiation-related non-*let-7* target genes. To evaluate the data processing inequality, the 3 MI values were compared across 223 verified *let-7*-family targets. For all but one target, the MI between the cell differentiation scores and the estimated *let-7*-family miRNA activity was the smallest of the three. Assuming that each comparison is as likely to be greater than the other, we can conclude that the relationship between cell differentiation scores and *let-7*-family miRNA activity is indirect at $p<1E-13$ (binomial distribution).

DATA AVAILABILITY

All data generated in this project, including raw sequencing and CELL files resulting from whole exome sequencing, targeted panel (Texas Children's Pediatric Solid Tumor Cancer Mutation Panel), Affymetrix OncoScan, and single-nuclei RNA sequencing, single-cell RNA sequencing, and single-nuclei DNA sequencing profiles, are freely available at the European Nucleotide Archive, accession PRJEB82511.

ACKNOWLEDGMENTS

This work was partially supported by the European Union Horizon 2020 award H2020-826121; the Cancer Prevention and Research Institute of Texas awards RP180674, RP190160, and RP230120; the National Institutes of Health awards R01CA258866, R01CA282467, R21CA286257, and UH3CA271227; and the Helis Research Program by the Helis Medical Research Foundation.

REFERENCES

1 Prokurat, A. *et al.* Transitional liver cell tumors (TLCT) in older children and adolescents: a novel group of aggressive hepatic tumors expressing beta-catenin. *Medical and pediatric oncology* **39**, 510-518 (2002).

2 Eichenmuller, M. *et al.* The genomic landscape of hepatoblastoma and their progenies with HCC-like features. *J Hepatol* **61**, 1312-1320, doi:10.1016/j.jhep.2014.08.009 (2014).

3 Seng, M. S. *et al.* Successful treatment of a metastatic hepatocellular malignant neoplasm, not otherwise specified with chemotherapy and liver transplantation. *Pediatric blood & cancer* **66**, e27603, doi:10.1002/pbc.27603 (2019).

4 Sumazin, P. *et al.* Hepatoblastomas with carcinoma features represent a biological spectrum of aggressive neoplasms in children and young adults. *Journal of Hepatology* **77**, 1026-1037 (2022).

5 Shrivastava, N. & O'Neill, A. F. in *Update in Pediatrics* 629-662 (Springer, 2024).

6 Chen Wongworawat, Y. *et al.* Molecular Profiling of a Hepatocellular Neoplasm Not Otherwise Specified (HCN-NOS) Demonstrates Distinct Molecular Features in Hepatoblastoma and HCC-Like Components. *Pediatric and Developmental Pathology* **27**, 169-175, doi:10.1177/10935266231204788 (2024).

7 Zhou, S. *et al.* Comparative Clinicopathologic and Genomic Analysis of Hepatocellular Neoplasm, Not Otherwise Specified, and Hepatoblastoma. *Modern Pathology* **37**, 100385, doi:https://doi.org/10.1016/j.modpat.2023.100385 (2024).

8 河原仁守 *et al.* HCN NOS (Hepatocellular neoplasm not otherwise specified) の 2 例. *北海道外科雑誌* **69**, 55-60 (2024).

9 Riehle, K. J. *et al.* Surgical management of liver tumors. *Pediatric blood & cancer*, e31155, doi:10.1002/pbc.31155 (2024).

10 Marin, J. J. G. *et al.* Mechanisms of Anticancer Drug Resistance in Hepatoblastoma. *Cancers* **11**, 407 (2019).

11 Hooks, K. B. *et al.* New insights into diagnosis and therapeutic options for proliferative hepatoblastoma. *Hepatology* **68**, 89-102 (2017).

12 Meyers, R. L. *et al.* Risk-stratified staging in paediatric hepatoblastoma: a unified analysis from the Children's Hepatic tumors International Collaboration. *The Lancet Oncology* **18**, 122-131 (2017).

13 Hadzic, N. & Finegold, M. J. Liver neoplasia in children. *Clinics in liver disease* **15**, 443-462 (2011).

14 López-Terrada, D. *et al.* Towards an international pediatric liver tumor consensus classification: proceedings of the Los Angeles COG liver tumors symposium. *Modern Pathology* **27**, 472-491 (2014).

15 Czauderna, P. *et al.* Hepatoblastoma state of the art: pathology, genetics, risk stratification, and chemotherapy. *Current opinion in pediatrics* **26**, 19-28 (2014).

16 Lau, C. S., Mahendraraj, K. & Chamberlain, R. S. Hepatocellular Carcinoma in the Pediatric Population: A Population Based Clinical Outcomes Study Involving 257 Patients from the Surveillance, Epidemiology, and End Result (SEER) Database (1973-2011). *HPB Surg* **2015**, 670728, doi:10.1155/2015/670728 (2015).

17 Espinoza, A. F. *et al.* A Novel Treatment Strategy Utilizing Panobinostat for High-Risk and Treatment- Refractory Hepatoblastoma. *Journal of Hepatology* **80**, 610-621 (2024).

18 Boster, J. M. *et al.* Predictors of survival following liver transplantation for pediatric hepatoblastoma and hepatocellular carcinoma: Experience from the Society of Pediatric Liver Transplantation (SPLIT). *Am J Transplant* **22**, 1396-1408, doi:10.1111/ajt.16945 (2022).

19 Klaus, A. & Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nature Reviews Cancer* **8**, 387-398, doi:10.1038/nrc2389 (2008).

20 Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**, 781-810, doi:10.1146/annurev.cellbio.20.010403.113126 (2004).

21 Perugorria, M. J. *et al.* Wnt–β-catenin signalling in liver development, health and disease. *Nature Reviews Gastroenterology & Hepatology* **16**, 121-136, doi:10.1038/s41575-018-0075-9 (2019).
22 Manica, M. *et al.* Inferring clonal composition from multiple tumor biopsies. *NPJ systems biology and applications* **6**, 27 (2020).
23 Czauderna, P. *et al.* The Children's Hepatic tumors International Collaboration (CHIC): Novel global rare tumor database yields new prognostic factors in hepatoblastoma and becomes a research model. *European journal of cancer* **52**, 92-101, doi:10.1016/j.ejca.2015.09.023 (2016).
24 Calderaro, J., Ziol, M., Paradis, V. & Zucman-Rossi, J. Molecular and histological correlations in liver cancer. *J Hepatol* **71**, 616-630, doi:10.1016/j.jhep.2019.06.001 (2019).
25 Cairo, S. *et al.* A combined clinical and biological risk classification improves prediction of outcome in hepatoblastoma patients. *European journal of cancer* **141**, 30-39 (2020).
26 Cairo, S. *et al.* Hepatic stem-like phenotype and interplay of Wnt/beta-catenin and Myc signaling in aggressive childhood liver cancer. *Cancer Cell* **14**, 471-484, doi:10.1016/j.ccr.2008.11.002 (2008).
27 Sumazin, P. *et al.* Genomic analysis of hepatoblastoma identifies distinct molecular and prognostic subgroups. *Hepatology* **65**, 104-121, doi:10.1002/hep.28888 (2017).
28 Nagae, G. *et al.* Genetic and epigenetic basis of hepatoblastoma diversity. *Nature Communications* **12**, 5423 (2021).
29 Hirsch, T. Z. *et al.* Integrated Genomic Analysis Identifies Driver Genes and Cisplatin-Resistant Progenitor Phenotype in Pediatric Liver Cancer. *Cancer discovery* **11**, 2524-2543, doi:10.1158/2159-8290.Cd-20-1809 (2021).
30 Steffin, D. *et al.* Interleukin-15-armoured GPC3 CAR T cells for patients with solid cancers. *Nature*, doi:10.1038/s41586-024-08261-8 (2024).
31 Hoshida, Y. *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res* **69**, 7385-7392, doi:10.1158/0008-5472.Can-09-1089 (2009).
32 Wang, X. *et al.* Comparative analysis of cell lineage differentiation during hepatogenesis in humans and mice at the single-cell transcriptome level. *Cell Research* **30**, 1109-1126 (2020).
33 Song, H. *et al.* Single-cell analysis of hepatoblastoma identifies tumor signatures that predict chemotherapy susceptibility using patient-specific tumor spheroids. *Nature communications* **13**, 4878 (2022).
34 TCGA. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e1323, doi:10.1016/j.cell.2017.05.046 (2017).
35 Talukdar, S. *et al.* Dormancy and cancer stem cells: An enigma for cancer therapeutic targeting. *Advances in cancer research* **141**, 43-84 (2019).
36 Nguyen, L. H. *et al.* Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. *Cancer Cell* **26**, 248-261, doi:10.1016/j.ccr.2014.06.018 (2014).
37 Piskounova, E. *et al.* Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147**, 1066-1079 (2011).
38 Ustianenko, D. *et al.* LIN28 selectively modulates a subclass of let-7 microRNAs. *Molecular cell* **71**, 271-283. e275 (2018).
39 Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science* **320**, 97-100, doi:10.1126/science.1154040 (2008).
40 Takashima, Y. *et al.* Suppression of lethal-7b and miR-125a/b Maturation by Lin28b Enables Maintenance of Stem Cell Properties in Hepatoblasts. *Hepatology* **64**, 245-260, doi:10.1002/hep.28548 (2016).
41 Armengol, C., Cairo, S., Fabre, M. & Buendia, M. Wnt signaling and hepatocarcinogenesis: the hepatoblastoma model. *The international journal of biochemistry & cell biology* **43**, 265-270 (2011).
42 Nelson, W. J. & Nusse, R. Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* **303**, 1483-1487, doi:10.1126/science.1094291 (2004).
43 Thompson, M. D. & Monga, S. P. WNT/beta-catenin signaling in liver health and disease. *Hepatology* **45**, 1298-1305, doi:10.1002/hep.21651 (2007).

44 Emami, K. H. *et al.* A small molecule inhibitor of beta-catenin/CREB-binding protein transcription [corrected]. *Proc Natl Acad Sci U S A* **101**, 12682-12687, doi:10.1073/pnas.0404875101 (2004).
45 Saeki, I. *et al.* Successful treatment of young childhood standard-risk hepatoblastoma with cisplatin monotherapy using a central review system. *Pediatric blood & cancer* **71**, e31255, doi:10.1002/pbc.31255 (2024).
46 Meyers, R. L. *et al.* Risk-stratified staging in paediatric hepatoblastoma: a unified analysis from the Children's Hepatic tumors International Collaboration. *The Lancet. Oncology* **18**, 122-131, doi:10.1016/s1470-2045(16)30598-8 (2017).
47 Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).
48 Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155-160, doi:10.1038/nature13600 (2014).
49 Parsons, D. W. *et al.* Actionable Tumor Alterations and Treatment Protocol Enrollment of Pediatric and Young Adult Patients With Refractory Cancers in the National Cancer Institute-Children's Oncology Group Pediatric MATCH Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **40**, 2224-2234, doi:10.1200/jco.21.02838 (2022).
50 David, M. P. *et al.* Multimodal molecular analysis of an atypical small cell carcinoma of the ovary, hypercalcemic type. *Cold Spring Harb Mol Case Stud* **4**, doi:10.1101/mcs.a002956 (2018).
51 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).
52 Wild, S. L. *et al.* The Canonical Wnt Pathway as a Key Regulator in Liver Development, Differentiation and Homeostatic Renewal. *Genes (Basel)* **11**, doi:10.3390/genes11101163 (2020).
53 Finn, R. S. *et al.* Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. *New England Journal of Medicine* **382**, 1894-1905 (2020).
54 Knight, K. R. *et al.* Group-Wide, Prospective Study of Ototoxicity Assessment in Children Receiving Cisplatin Chemotherapy (ACCL05C1): A Report From the Children's Oncology Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **35**, 440-445 (2017).
55 Freyer, D. R. *et al.* Effects of sodium thiosulfate versus observation on development of cisplatin-induced hearing loss in children with cancer (ACCL0431): a multicentre, randomised, controlled, open-label, phase 3 trial. *The Lancet Oncology* **18**, 63-74 (2017).
56 Minussi, D. C. *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302-308, doi:10.1038/s41586-021-03357-x (2021).
57 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529, doi:10.1016/j.cell.2021.04.048 (2021).
58 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).
59 Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172, doi:10.1038/s41590-018-0276-y (2019).
60 Conway, T. *et al.* Xenome--a tool for classifying reads from xenograft samples. *Bioinformatics* **28**, i172-178, doi:10.1093/bioinformatics/bts236 (2012).
61 Bakker, B. *et al.* Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol* **17**, 115, doi:10.1186/s13059-016-0971-7 (2016).
62 Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **7**, 1338, doi:10.12688/f1000research.15931.2 (2018).
63 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
64 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
65 Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **375**, 1109-1112, doi:10.1056/NEJMp1607591 (2016).

66 Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843-2851, doi:10.1093/bioinformatics/btu356 (2014).
67 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
68 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).
69 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
70 Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **44**, e131, doi:10.1093/nar/gkw520 (2016).
71 Dentro, S. C., Wedge, D. C. & Van Loo, P. Principles of Reconstructing the Subclonal Architecture of Cancers. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a026625 (2017).
72 Gillis, S. & Roth, A. PyClone-VI: scalable inference of clonal population structures using whole genome data. *BMC Bioinformatics* **21**, 571, doi:10.1186/s12859-020-03919-2 (2020).
73 Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**, 1977-2000, doi:10.1091/mbc.02-02-0030 (2002).
74 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550 (2005).
75 Nielsen, M. M. & Pedersen, J. S. miRNA activity inferred from single cell mRNA expression. *Scientific reports* **11**, 9170, doi:10.1038/s41598-021-88480-5 (2021).
76 Lorenzi, L. *et al.* The RNA Atlas expands the catalog of human non-coding RNAs. *Nature biotechnology* **39**, 1453-1465 (2021).
77 Chiu, H.-S. *et al.* Pan-cancer analysis of lncRNA regulation supports their targeting of cancer genes in each tumor context. *Cell reports* **23**, 297-312 (2018).

FIGURES

Figure 1. HBCs exhibit poor clinical outcomes. (**A**) Event-free survival analysis for our HBC cohort compared to three HB cohorts with recorded events; p-value by log-rank test. (**B**) Overall survival rates—based on at least two years of post-diagnosis surveillance— were compared between our HBC cohort ("HBC") and six HB cohorts with published patient outcome data; p-values were calculated using Fisher's exact test. CHIC refers to the Children's Hepatic Tumors International Collaboration.

Figure 2. HBC tumors comprise HB, HBC, and HCC cells. (**A**) The upregulation of HB and HCC biomarker sets (represented as z-scores) as observed in cells from a high-risk HB resection, five HBC resections, and AGAR-trial HBC biopsies, and (**B**) the cellular compositions estimates of each resection; a z-score exceeding 2 for either HB or HCC signature gene set or a z-score difference greater than 2 indicate significant evidence for HB or HCC; HBC cells demonstrated comparable z-scores greater than 2 for both signature gene sets. (**C**) UMAP analysis of 25,881 tumor cells from five HBC resections identified distinct cell clusters based on cancer-type classification. (**D**) Unsupervised hierarchical 2D clustering of cell profiles from our reference HB and HCC resections (left) and five HBC resections (right); cells are represented in columns, genes in rows, with genes ordered according to the reference clustering (left). (**E**) Expression levels of selected HB and HCC biomarkers in 15 HB, 16 HCC, 34 HBC, and 9 normal bulk samples profiled using the NanoString nCounter assay; statistical significance indicated as ***: $p<0.0001$, ***: $p<0.001$, **: $p<0.01$. (F) GSEA identified selected pathways enriched by cluster.

Figure 3. Tumor embryonic differentiation stages are predictive of risk. (**A**) Tumor embryonic differentiation scores and (**B**) the expression levels of the differentiation biomarker LIN28B were found to be predictive of HB risk; statistical significance indicated as ****: $p<0.0001$, ***: $p<0.001$. S.E.M. shown. (**C**) Distribution of embryonic differentiation scores in pediatric and adult liver cancer and non-cancer samples. (**D**) LIN28B expression was correlated with embryonic differentiation, while the activity of its miRNA biogenesis targets (let-7 miRNA family) showed an inverse correlation with embryonic differentiation. The data processing inequality, based on mutual information (MI) estimates between cell embryonic differentiation scores, LIN28B abundance, and let-7 activity (pairwise), suggests that the correlation between embryonic differentiation scores and let-7 activity is contingent upon LIN28B abundance. (**E**) Selected pathways associated with early (differentiation score≥0.45, red) and late (differentiation score≤0.2, green) embryonic differentiated cells in HBC resections by GSEA.

Figure 4. WNT-signaling activation regulates liver differentiation and chemoresistance. The expression of DKK1, a marker of WNT-signaling activation, is (**A**) predictive of HB risk and correlates with (**B**) HB and (**C**) HBC embryonic differentiation stages; DKK1 expression remains relatively elevated in HBCs across differentiation stages, with z-score normalized expression and S.E.M. shown. (**D**) UMAP of scRNA-Seq profiles from 14,194 patient-derived P17 cells before and after treatment with the WNT-inhibitor ICG-001. (**E**) Cumulative distributions of the inferred embryonic differentiation stages for wildtype untreated P17 cells and ICG-001-treated cells over 5 or 10 days; all cells were cultured for 10 days, with the 5-day treatment followed by 5-day inhibitor-free growth; p-values were calculated using the chi-squared test. (**F**) Composition of P17 cells exhibiting differential expression of LIN28B—a differentiation marker—and DKK1—a WNT-signaling activity marker—before and after ICG-001 treatments. (**G**) Correlation between inferred cisplatin resistance and LIN28B expression in P17 cell clusters, with LIN28B expression z-score transformed; circle sizes reflect cluster sizes. The Pearson correlation coefficient (r) and corresponding p-value are provided. (**H**) Inferred cisplatin resistance in wildtype P17 cells and post-ICG-001 treatments. (**I**) Relative P17 cell viability following low-dose treatment with ICG-001 (IC5), cisplatin (IC0 and IC10), or both; cells were treated with ICG-001 for 5 days and then rested for 5 days (IC5-A) and vice versa (IC5-B); S.E.M. is shown. Cis refers to cisplatin. (**J**) Inferred cisplatin resistance across cell clusters in five HBC resections (see Figure 2); p-values by t-test.

Figure 5. HBC phylogeny and composition reveal multiple transition events during early hepatoblast developmental stages. (**A**) The phylogeny of HBC TLTX36, including mutations and copy number alterations, was inferred from whole exome sequencing of a pre-treatment biopsy (A1) and five regions of a post-treatment resection (A14, A17, A18.B, A11.A, A18.A). Embryonic differentiation scores

for each region were derived from their NanoString RNA expression profiles. LOH denotes loss of heterozygosity. (**B-E**) The phylogeny, composition, and embryonic differentiation stages of cells in the HBC tumors (**B**) TLTX18, (**C**) TLTX33, (**D**) TLTX23, and (**E**) TLTX25 were inferred from single-cell resolution RNA and DNA profiles, as well as whole exome sequencing and SNV array profiling of multiple regions. The inferred phylogenies include mutations and copy number alterations; the sizes of the circles represent the number of cells, and colors indicate differentiation stages; relative tumor-type compositions—HB, HBC, and HCC—of each genetic subclone are presented for clarity.

SUPPLEMENTARY FIGURES

Figure S1. Representative histopathology of pre- and post-treatment HBC. Pre-treatment HBCs are composed of HB, HBC, and HCC cells. (**A**) Multiple regions of the pre-treatment biopsy of the equivocal HBC TLTX16 reveal cells with HB-like histology at variable embryonic differentiation stages (embryonal and fetal) and cells with HBC-like histology that are enveloped in regions that mainly comprise cells with HB-like histology. (**B**) The post-treatment resection (4 cycles of chemotherapy, including cisplatin, and vincristine-irinotecan) of TLTX16 reveals extensive necrotic regions adjacent to areas primarily containing viable cells with HBC-like histology.

Figure S2. Evaluation of HB and HCC biomarker sets. Our HB and HCC biomarker sets were derived from single-cell resolution profiles of HBs and HCCs, as reported by Song et al. (2022), samples from the AGAR trial by Steffin et al. (2024), and additional profiles, encompassing 2375 HB and 1468 HCC cells, respectively. We refer to these profiles—including snRNA-Seq profiles of 2375 HB and 1468 HCC cells—as our reference profiles. (**A**) The UMAP of our reference profiles demonstrates a perfect separation between HB (light blue) and HCC (red) cells using unbiased clustering. (**B**) HB and HCC biomarker z-scores of cells represented in our reference profiles, AGAR-trial HCCs, and adjacent normal cells (orange) within the reference HB and HCC and in 5 HBC samples. Z-scores greater than 2 indicated significant upregulation of HB or HCC biomarker gene sets. (**C**) HB and HCC biomarker gene set z-scores using bulk RNA-Seq profiles of 29 HBs, 4 HCCs, and 6 fetal samples from patients without cancer, alongside NanoString nCounter profiles of 15 HB, 16 HCC, and 9 fetal samples from patients without cancer. The NanoString panel includes only 5 of 45 HB and 8 of 265 HCC biomarker genes; p-values denote the significance of the separation between HB and HCC z-scores. (**D**) We classified HBC-profile clusters—in Figure 2C—into HBs, HB-HBCs, HBCs (purple), and HBC-HCCs. Random samples from each classification are displayed here. Namely, HB clusters predominantly contain cells with z-scores greater than 2 for HB biomarkers and less than 2 for HCC biomarkers. HBC clusters contain cells with z-scores greater than 2 for both HB and HCC biomarkers. HB-HBC and HBC-HCC clusters contain mixtures of cell types. Z-scores greater than 2 indicate significant upregulation of biomarker genes.

Figure S3. Liver cell differentiation inference. We used profiles of embryonic liver cells by Wang et al. (2020), spanning 5 to 19 weeks post-conception, to infer liver cell differentiation. Specifically, we developed a classifier based on profiles from cells at weeks 5 and 19 post-conception, while profiles of cells taken from embryos 6 to 16 weeks post-conception were used for evaluation. The classifier produces a differentiation score for each cell profile. (**A**) UMAP of 19,557 developing human liver cells from 5 to 19 weeks post-conception. (**B**) Differentiation scores for human embryonic liver cells from 5 to 19 weeks post-conception. (**C**) Differentiation scores of mouse embryonic liver cells from 11 to 17.5 days post-conception. (**D**) Differentiation inference was based on a random forest differentiation prediction classifier that could be trained with any given number of classification trees. We opted for a 20-tree classifier because increasing the number of trees to more than 20 did not significantly reduce the training error. Here, we show the training error as a function of the number of trees; OOB, out-of-bag error. (**E**) To compare the proportion of embryonically undifferentiated cells in each tumor, we set a differentiation score cutoff based on its 95% quantile in profiles of 11,480 normal hepatocyte cells from a tumor-adjacent sample, producing a differentiation score cutoff of 0.4; higher scores indicate undifferentiated embryonic cell stages, while lower scores indicate differentiated stages. (**F**) Inferred undifferentiated cell fractions of our HB reference sample, HCC reference sample, 2 AGAR-trial HBC samples, and 2 AGAR-trial HCC

samples. All differences are significant, and the results suggest that these pediatric HCC samples are less embryonically differentiated than HBs, while cancer cells remaining after AGAR-trial treatments are predominantly undifferentiated. (**G**) Inferred embryonically undifferentiated cell fractions of 7 HB samples by Song et al. Higher-risk samples are significantly less embryonically differentiated. (**H**) Inferred undifferentiated cell fractions of 5 HBC samples. TLTX23, characterized by predominant HCC and HBC composition, is less embryonically differentiated. (**I**) Normalized proliferation score of early (differentiation score ≤ 0.2), intermediate (0.2 < differentiation score < 0.45), and late (differentiation score ≥ 0.45) differentiated tumor cells in 5 HBC samples; p-values are based on the paired sample t-test. (**J**) Normalized expression of LIN28B in the human fetal liver during embryonic development (top) and in HBs across predicted embryonic differentiation stages (bottom). Normalized expression profiles of (**K**) HMGA2 and (**L**) IGF2 in human fetal liver development, HBs profiled by Song et al. (2022). and HBCs across predicted embryonic differentiation stages; S.E.M. is shown.

Figure S4. WNT-signaling inhibition decreases DKK1 and LIN28B expression. DKK1 is a widely used indicator of WNT-signaling activity, while LIN28 indicates liver cell differentiation (Sumazin et al., 2017). (**A**) Normalized expression of DKK1 in the human fetal liver during development suggests reduced WNT-signaling activity as a function of liver differentiation; S.E.M. is shown. (**B-C**) In vitro treatment with the WNT-inhibitor ICG-001 for 5 or 10 days at variable dosages reduced (**B)** DKK1 and (**C**) LIN28B expression in the patient-derived cell line P17; 5 biological replicates, S.E.M. is shown.

Figure S5. **WNT-signaling inhibition promotes cancer cell differentiation in vitro.** We treated the patient-derived cell line P17 with the WNT-inhibitor ICG-001 at 5 μM for 10 days, comparing WNT-signaling activity, cell differentiation, and resistance to cisplatin, the standard-of-care chemotherapy. All cells were cultured for 10 days: either untreated, treated for 5 days with the inhibitor and the following 5 days without the inhibitor (5+5), or treated for 10 days with the inhibitor. Cells were then profiled by scRNA-Seq. (**A**) UMAP of 14,194 P17 cells grouped by 5 μM ICG-001 treatment for 5+5 or 10 days. (**B**) The composition of cell clusters (from Figure 4D) in the integrated scRNA-Seq profiles identifies clusters with low and high DKK1 (WNT-signaling) and LIN28B (differentiation) expression. (**C**) Mean expression of DKK1 and LIN28B and the mean predicted differentiation score in each cell cluster after z-score transformation (mean and standard deviation based on cluster 0-11 only, due to limited cell counts in clusters 12 and 13); red and green indicate high and low values, respectively.

Figure S6. Gene expression-based scores predict resistance to cisplatin. (**A**) Correlation between predicted cisplatin-resistance scores and cisplatin drug test results, measured with the area under the curve (AUC) in nine pediatric cancer cell lines as reported by Hirsch et al (2021). Pearson correlation coefficient and associated p-value are shown. (**B**) Predicted cisplatin-resistance scores of 4 HB patient-derived organoids (PDOs) profiled and treated by Song et al. (2022), with their cell viability measurements after cisplatin treatment. Cisplatin-resistance inference was trained on bulk RNA-Seq profiles and evaluated on either bulk RNA-Seq or pseudobulk of scRNA-Seq profiles; S.E.M. is shown. (**C**) Predicted cisplatin-resistance scores in low-risk and high-risk HB cohorts profiled by bulk RNA-Seq; **: $p<0.01$. (**D**) Predicted cisplatin-resistance scores of our reference HB and HCC samples. (**E**) Predicted cisplatin-resistance scores of a PDX before and after cisplatin treatment profiled by scRNA-Seq.

Figure S7. Biclustering of cells and their copy number alterations. The results of the biclusterings of the snDNA-Seq profiles of (**A**) TLTX18, (**B**) TLTX23, (**C**) TLTX25, and (**D**) TLTX33.

Figure S8. Distinct evaluations of genetic alterations across HBC phylogeny nodes. (**A-D**) Distributions of the inferCNV scores—mean normalized expression estimates across all genes in the region—in the queried regions in cells from nodes in phylogenies presented in Figure 5. (**A**) TLTX18 nodes 2 and 3 inferCNV scores for chromosome 4q, which was identified as lost in node 3 cells. (**B**) TLTX23 nodes 2 and 3 inferCNV scores for chromosome 1q, which was identified as gained in node 3 cells. (**C**) TLTX25 nodes 2 and 5 inferCNV scores for chromosome 8, which was identified as gained in node 5 cells. (**D**) TLTX33 nodes 4 and 5 inferCNV scores for chromosome 19q, which was identified as lost in node 5 cells. (**E-L**) Cell-type composition of genetically distinct nodes. (**E**) TLTX18 node 2 and (**F**) node 3 were both composed of HB, HBC, and HCC cells. (**G**) TLTX23 node 2 and (**H**) node 3 were both

composed of HBC and HCC cells. (**I**) TLTX25 node 2 and (**J**) node 5 were both composed of HB and HBC cells. (**K**) TLTX33 node 4 and (**L**) node 5 were both composed of HB, HBC, and HCC cells.

Figure S9. CNV calls from multi-region WES data. CNV calls from WES profiles of 6 tumor regions of TLTX36, 1 tumor region of TLTX18, 3 tumor regions of TLTX33, 6 tumor regions of TLTX23, and 7 tumor regions of TLTX25.

SUPPLEMENTARY TABLES

Supplementary Table 1. Clinical information of our HBC cohort and the Sumazin et al. (2017) HB cohort with published patient outcomes data.

Supplementary Table 2. 45 HB- and 265 HCC-specific genes were identified from our reference HB and HCC snRNA-Seq sample. For each gene, the mean log2-fold changes, the proportion of HB and HCC cells expressing the gene, and the adjusted p-values are shown.

Supplementary Table 3. Evaluation of HB and HCC biomarker sets in reference samples—including HB, HCC, and HBC-adjacent samples—profiled at single-cell resolutions. HB and HCC biomarker z-scores of each cell are shown.

Supplementary Table 4. Evaluation of HB and HCC biomarker sets in bulk samples. HB and HCC biomarker z-scores of each bulk sample profiled by RNA-Seq or the NanoString nCounter assay.

Supplementary Table 5. HB and HCC biomarker z-scores of each cell and predicted cell-type composition for 1 high-risk HB resection, 5 HBC resections, and 2 AGAR-trial HBC biopsies. HB and HCC biomarker z-scores of each cell are shown.

Supplementary Table 6. Hierarchical clustering of the expression values of 45 HB- and 265 HCC-specific genes in our reference HB and HCC samples and 5 HBC resections.

Supplementary Table 7. Selected pathways enriched in the 10 HBC tumor clusters. Adjusted p-values and normalized enrichment scores (NES) by GSEA are shown.

Supplementary Table 8. Evaluation of liver cell differentiation prediction in human and mouse embryonic liver cells. Predicted differentiation scores of each cell are shown.

Supplementary Table 9. Predicted differentiation scores of each cell for the tumor-adjacent sample, reference HB and HCC samples, AGAR-trial HBC and HCC samples, 7 HB samples by Song et al., and 5 HBC resections.

Supplementary Table 10. Predicted differentiation scores of each sample for pediatric and adult liver cancer and non-cancer bulk samples.

Supplementary Table 11. Estimated MI between differentiation scores, the expression profile of LIN28B, and the expression profiles of *let-7*miRNA targets, for each verified target. This comparison included all cells with non-zero LIN28B mRNA counts. To evaluate data processing inequality (DPI), we counted the number of verified targets where MI (differentiationScore, *let-7*target expression) was smaller than both MI (differentiationScore, LIN28B expression) and MI (LIN28B expression, *let-7*target expression). This relationship was true for 222 of the 223 *let-7* targets.

Supplementary Table 12. Selected pathways enriched in differentiated and undifferentiated tumor cells. Adjusted p-values and normalized enrichment scores (NES) by GSEA are shown.

Supplementary Table 13. Expression of DKK1 and LIN28B by quantitative PCR in wildtype untreated P17 cells and in ICG-001-treated P17 cells for 5 or 10 days.

Supplementary Table 14. Predicted differentiation scores of each cell for wildtype untreated P17 cells and ICG-001-treated cells for 5+5 or 10 days.

Supplementary Table 15. The composition of cell clusters in the integrated scRNA-Seq profiles of P17 cells.

Supplementary Table 16. Predicted cisplatin-resistance scores of each patient in an HB cohort profiled by bulk RNA-Seq.

Supplementary Table 17. Predicted cisplatin-resistance scores based on pseudobulk profiles of P17 cell clusters and HBC tumor clusters.

Supplementary Table 18. Relative P17 cell viability after low-dose treatment with cisplatin, ICG-001, or both in three biological replicates per condition.

Supplementary Table 19. Predicted differentiation scores and histological examinations of each region for TLTX36. Predicted differentiation scores and cancer cell type regarding cancer maps for TLTX18, TLTX23, TLTX25, and TLTX33 as shown in Figure 5.

Supplementary Table 20. Mutation calls from WES, snDNA-Seq, and snRNA-Seq regarding cancer maps for TLTX36, TLTX18, TLTX23, TLTX25, and TLTX33 as shown in Figure 5.

Supplementary Table 21. The identities of probes used in NanoString assays.

Supplementary Table 22. NanoString expression profiles of 15 HB, 34 HBC, 16 HCC, and 9 normal bulk samples, in addition to NanoString expression profiles of six tumor regions and one adjacent normal region from TLTX36.

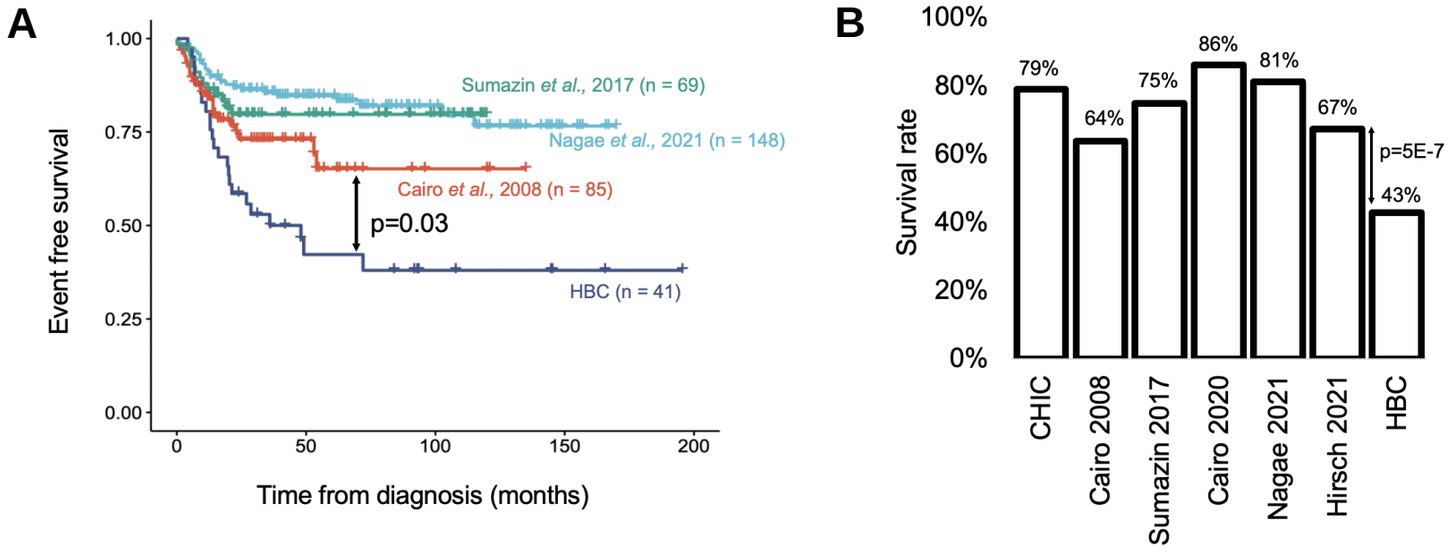

A
1.00
0.75
0.50
0.25
0.00
Event free survival
0
50
100
150
200
Time from diagnosis (months)
Sumazin et al., 2017 (n = 69)
Nagae et al., 2021 (n = 148)
Cairo et al., 2008 (n = 85)
p=0.03
HBC (n = 41)
B
100%
80%
60%
40%
20%
0%
Survival rate
79%
64%
75%
86%
81%
67%
43%
p=5E-7
CHIC
Cairo 2008
Sumazin 2017
Cairo 2020
Nagae 2021
Hirsch 2021
HBC

Figure 1

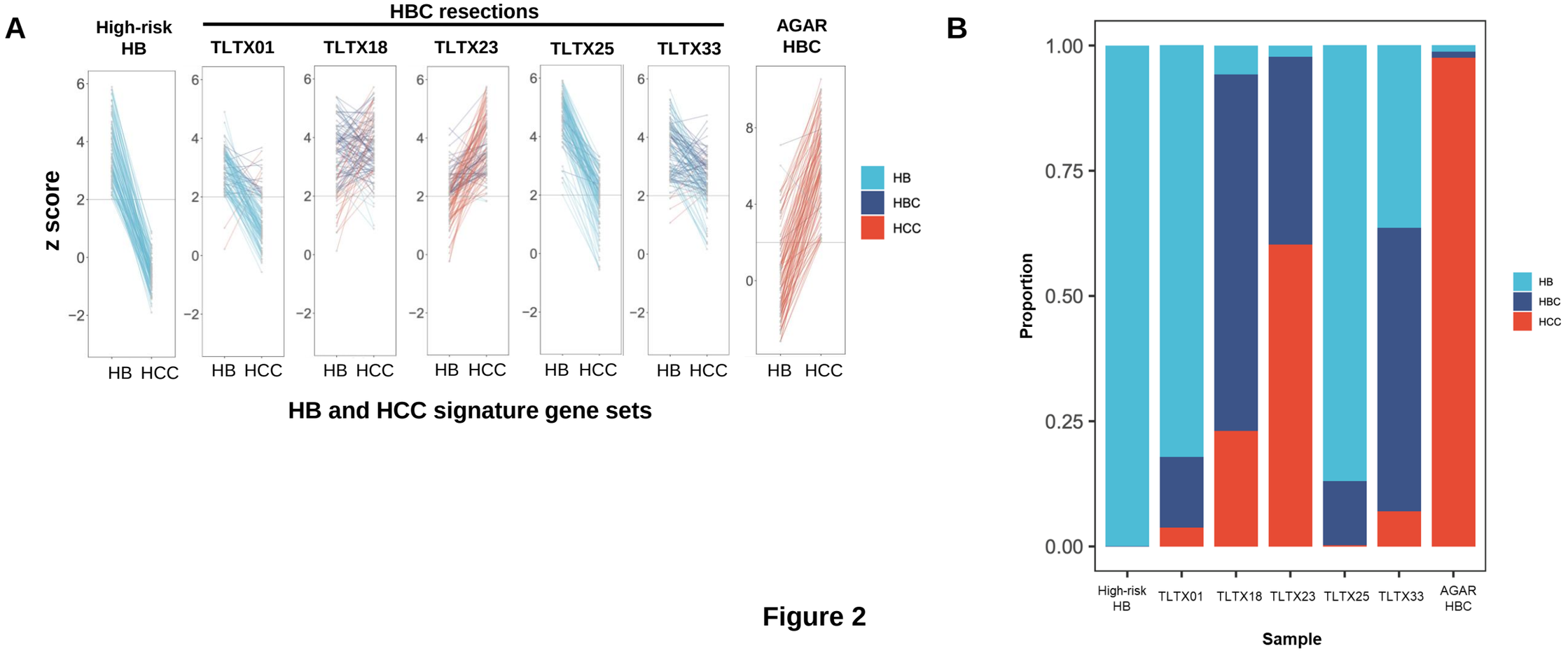

A
High-risk HB
HBC resections
TLTX01
TLTX18
TLTX23
TLTX25
TLTX33
AGAR HBC
z score
HB HCC
HB and HCC signature gene sets
HB
HBC
HCC
B
Proportion
1.00
0.75
0.50
0.25
0.00
High-risk HB
TLTX01
TLTX18
TLTX23
TLTX25
TLTX33
AGAR HBC
Sample

Figure 2

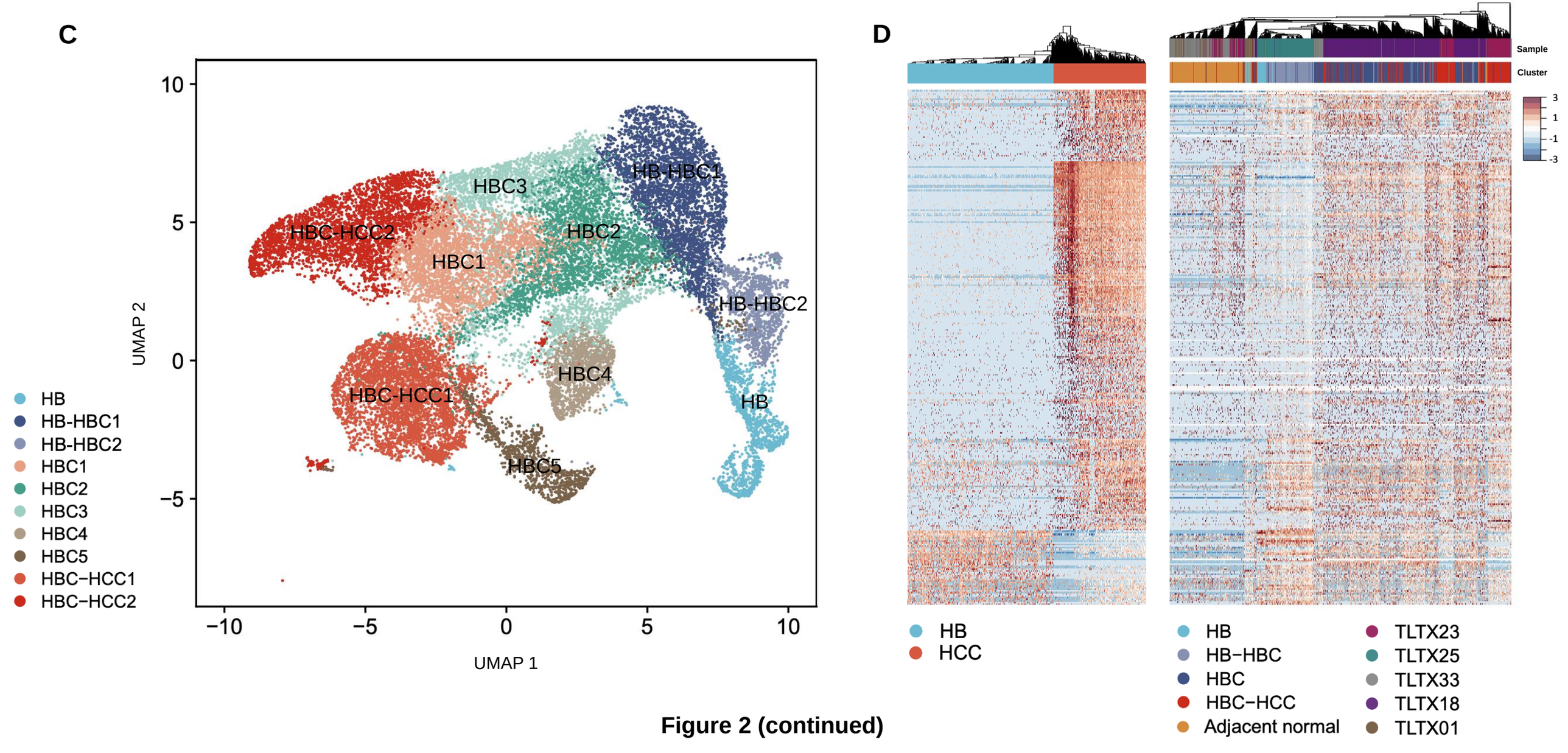
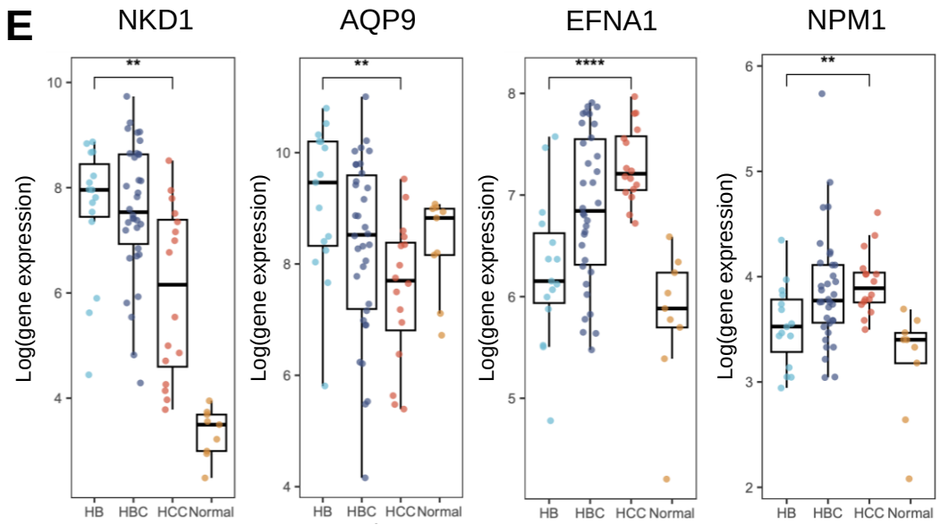


Figure 2 (continued)



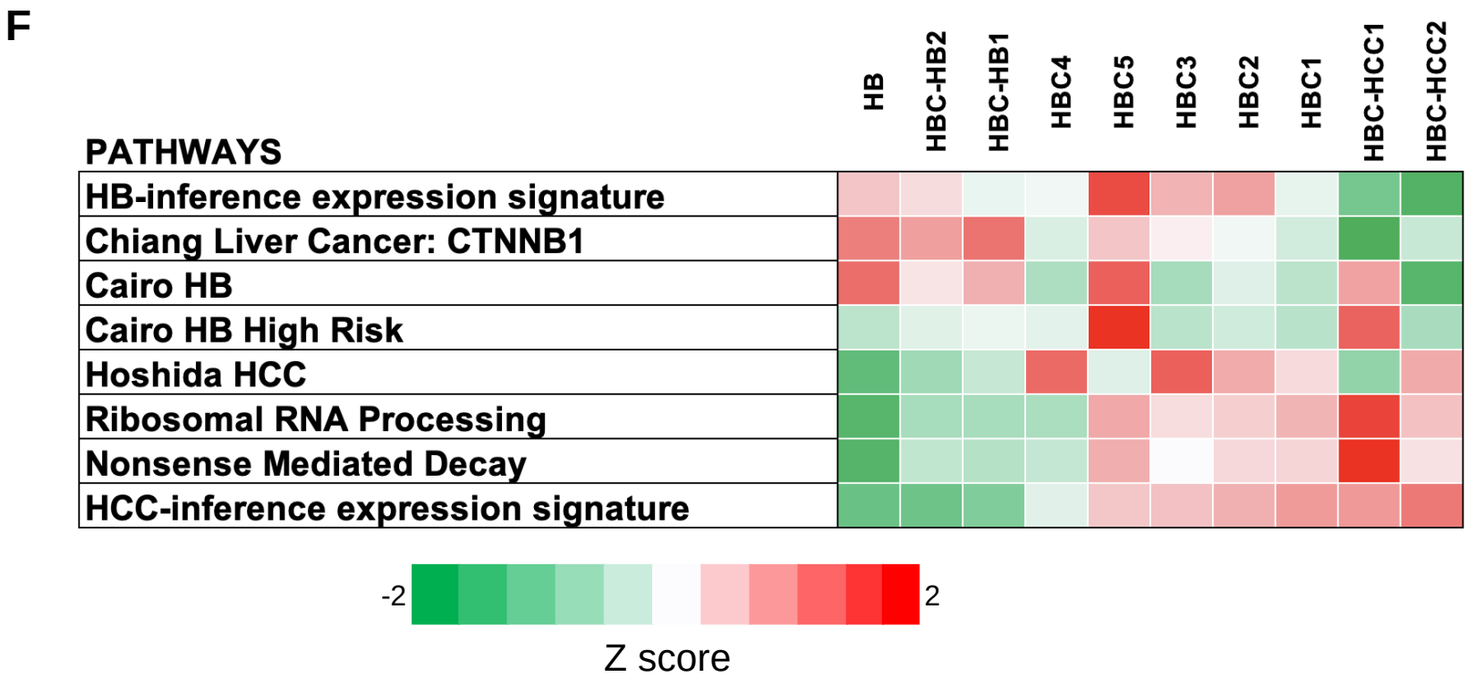


Figure 2 (continued)

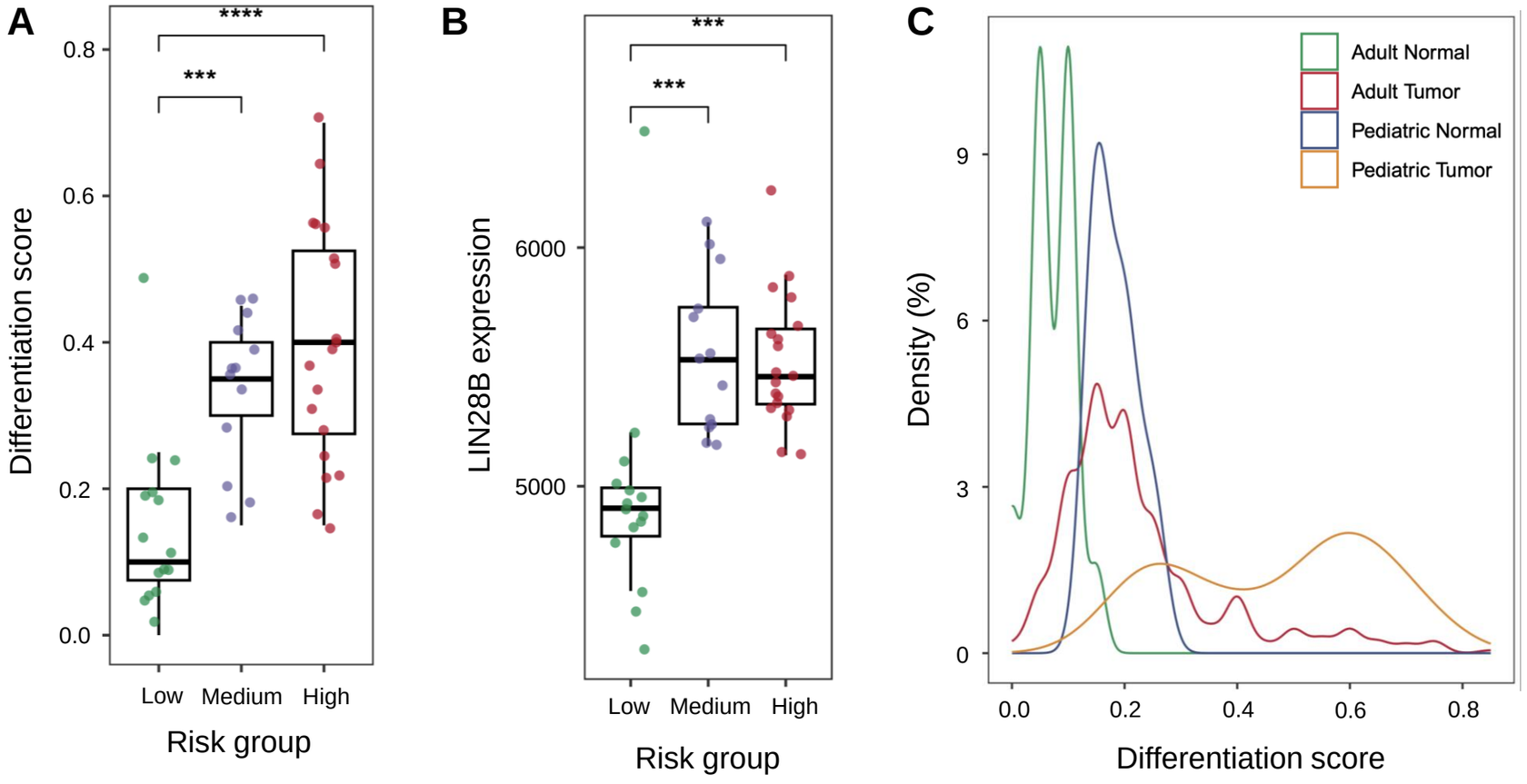
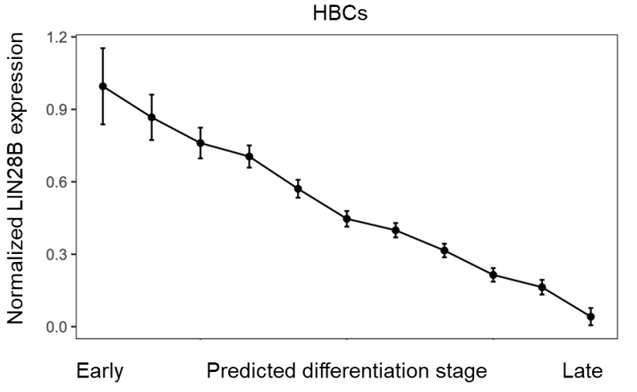
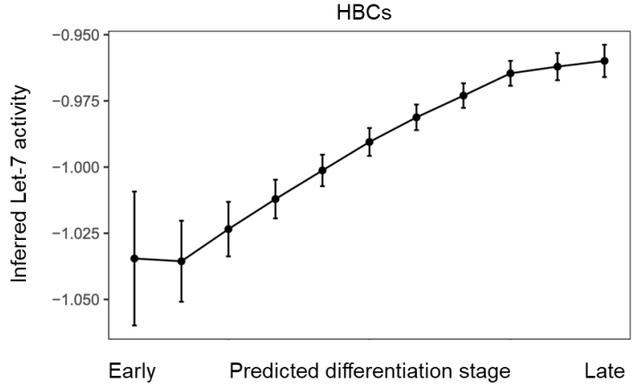
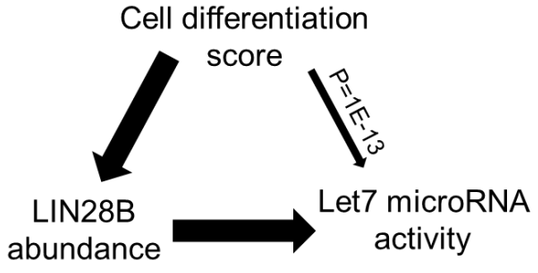

A
****
***
0.8
0.6
0.4
0.2
0.0
Differentiation score
Low
Medium
High
Risk group
B
***
***
6000
5000
LIN28B expression
Low
Medium
High
Risk group
C
Adult Normal
Adult Tumor
Pediatric Normal
Pediatric Tumor
9
6
3
0
Density (%)
0.0
0.2
0.4
0.6
0.8
Differentiation score

Figure 3

D







E

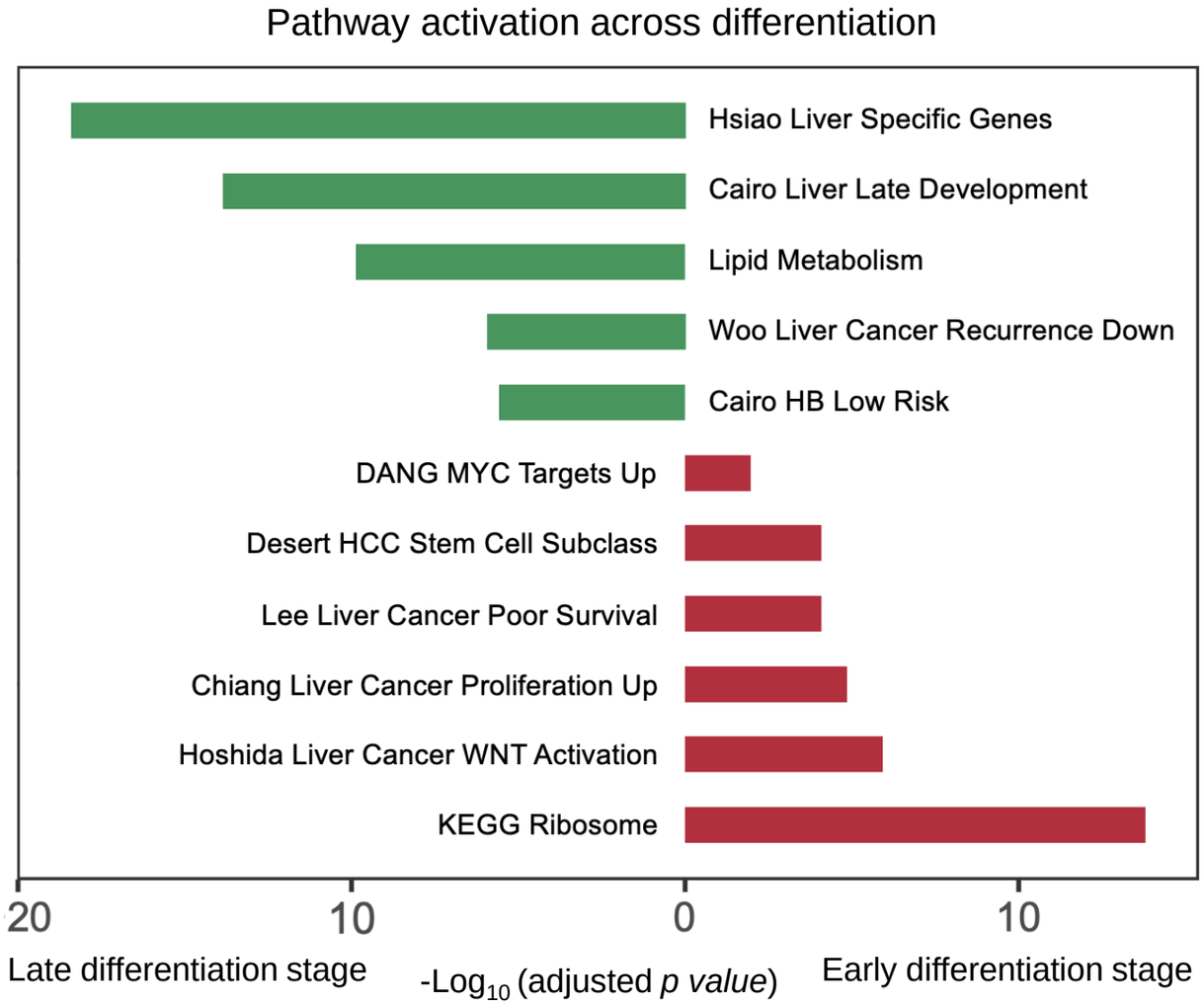


Figure 3 (continued)

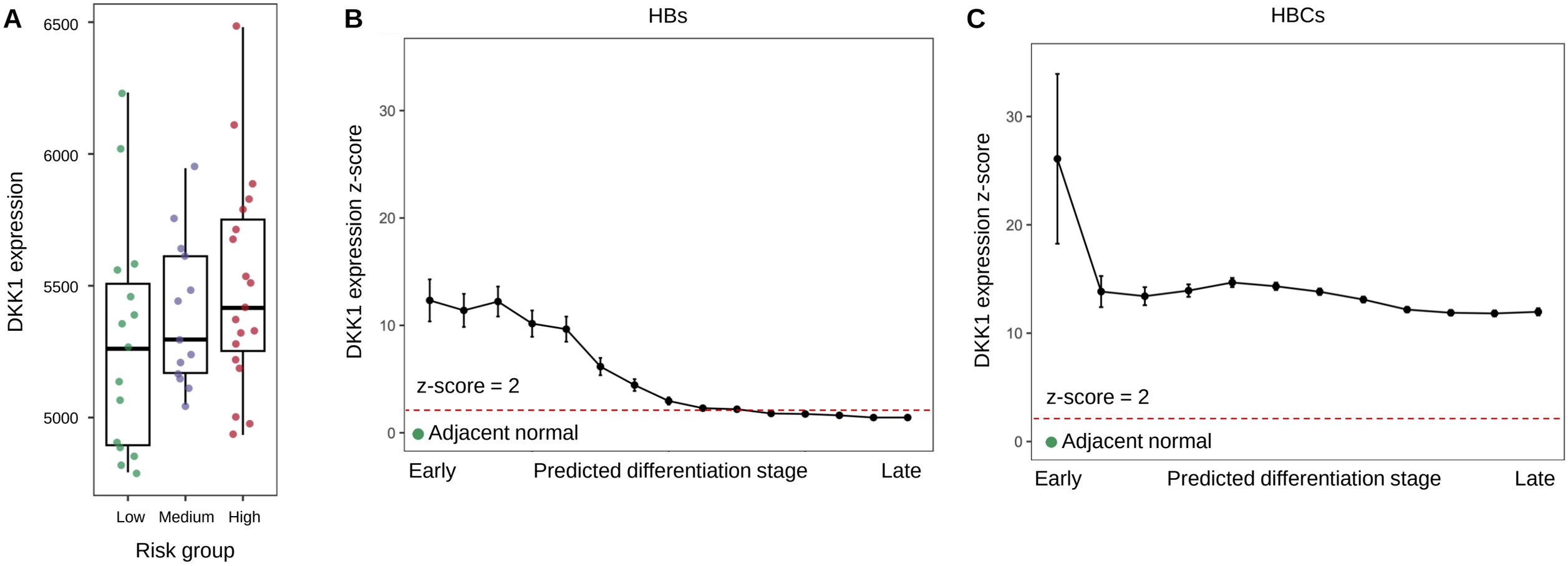

A
6500
6000
5500
5000
DKK1 expression
Low
Medium
High
Risk group
B
HBs
DKK1 expression z-score
30
20
10
0
z-score = 2
Adjacent normal
Early
Predicted differentiation stage
Late
C
HBCs
DKK1 expression z-score
30
20
10
0
z-score = 2
Adjacent normal
Early
Predicted differentiation stage
Late

Figure 4

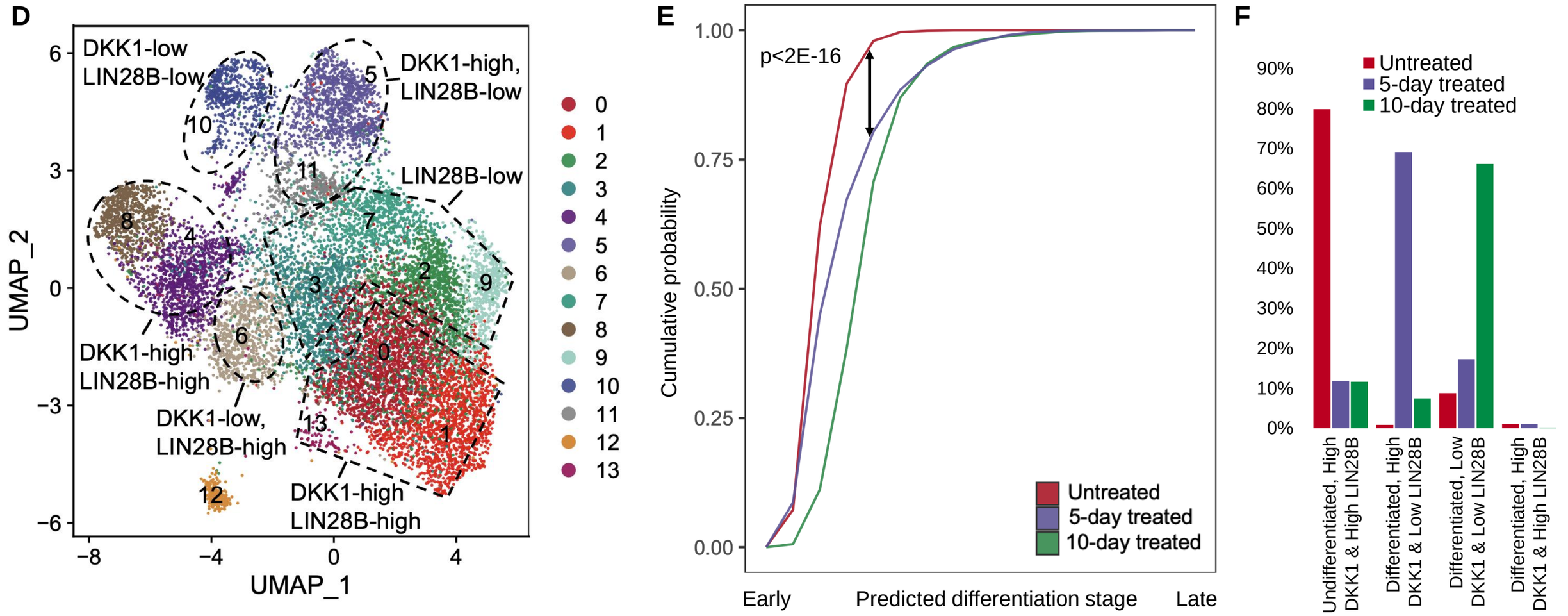


Figure 4 (continued)

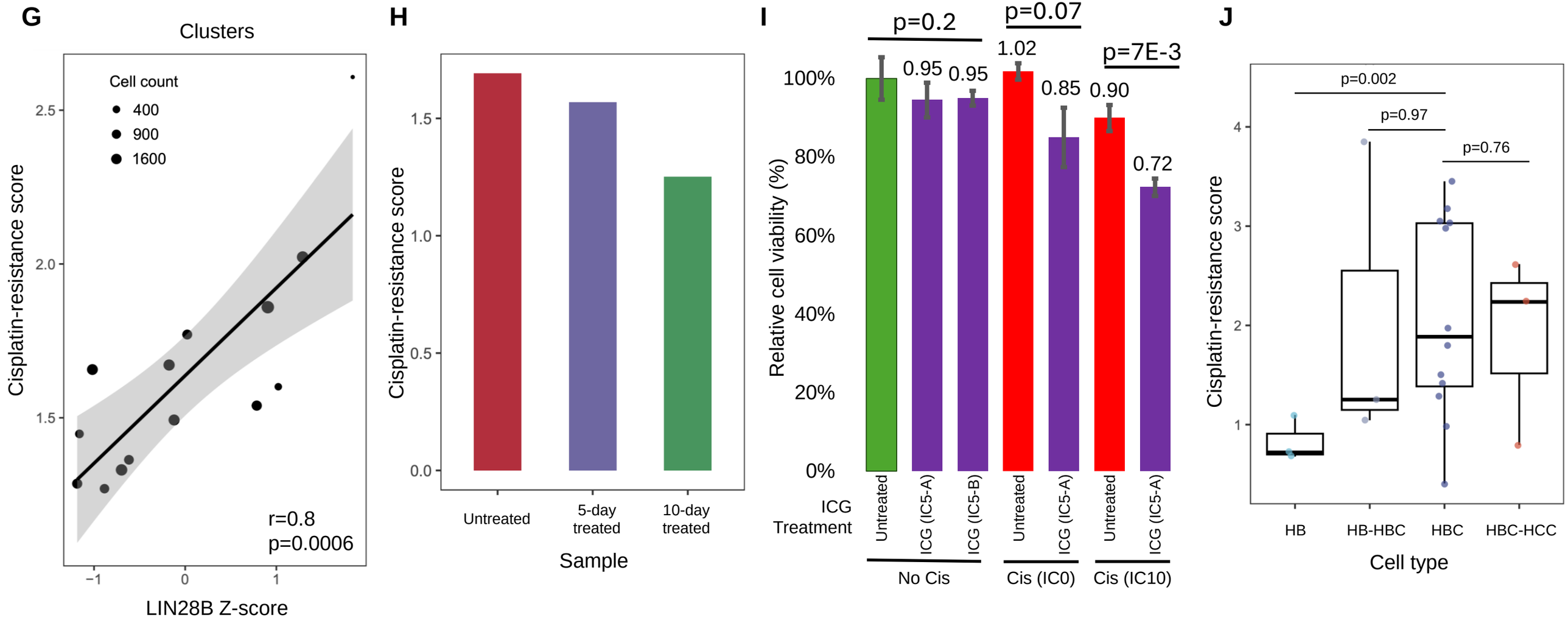


Figure 4 (continued)

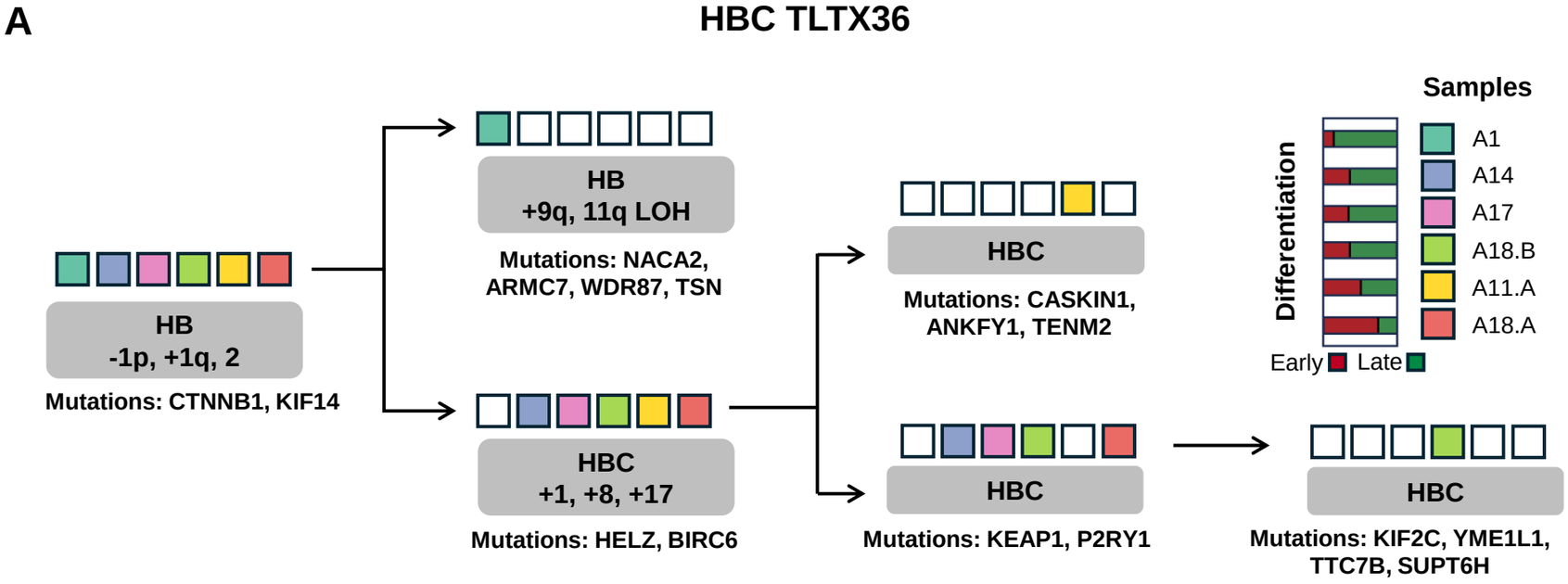

A
HBC TLTX36
HB
-1p, +1q, 2
Mutations: CTNNB1, KIF14
HB
+9q, 11q LOH
Mutations: NACA2, ARMC7, WDR87, TSN
HBC
+1, +8, +17
Mutations: HELZ, BIRC6
HBC
Mutations: CASKIN1, ANKFY1, TENM2
HBC
Mutations: KEAP1, P2RY1
HBC
Mutations: KIF2C, YME1L1, TTC7B, SUPT6H
Samples
Differentiation
Early
Late
A1
A14
A17
A18.B
A11.A
A18.A

Figure 5

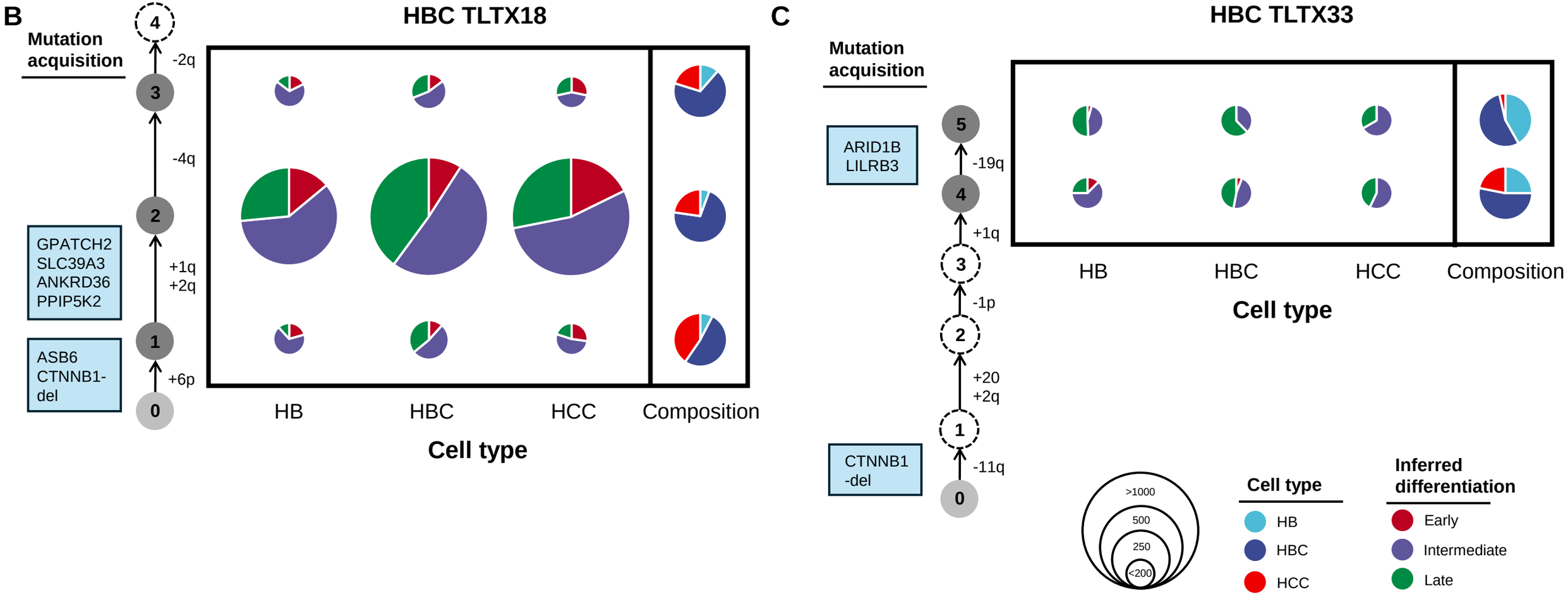


Figure 5 (continued)

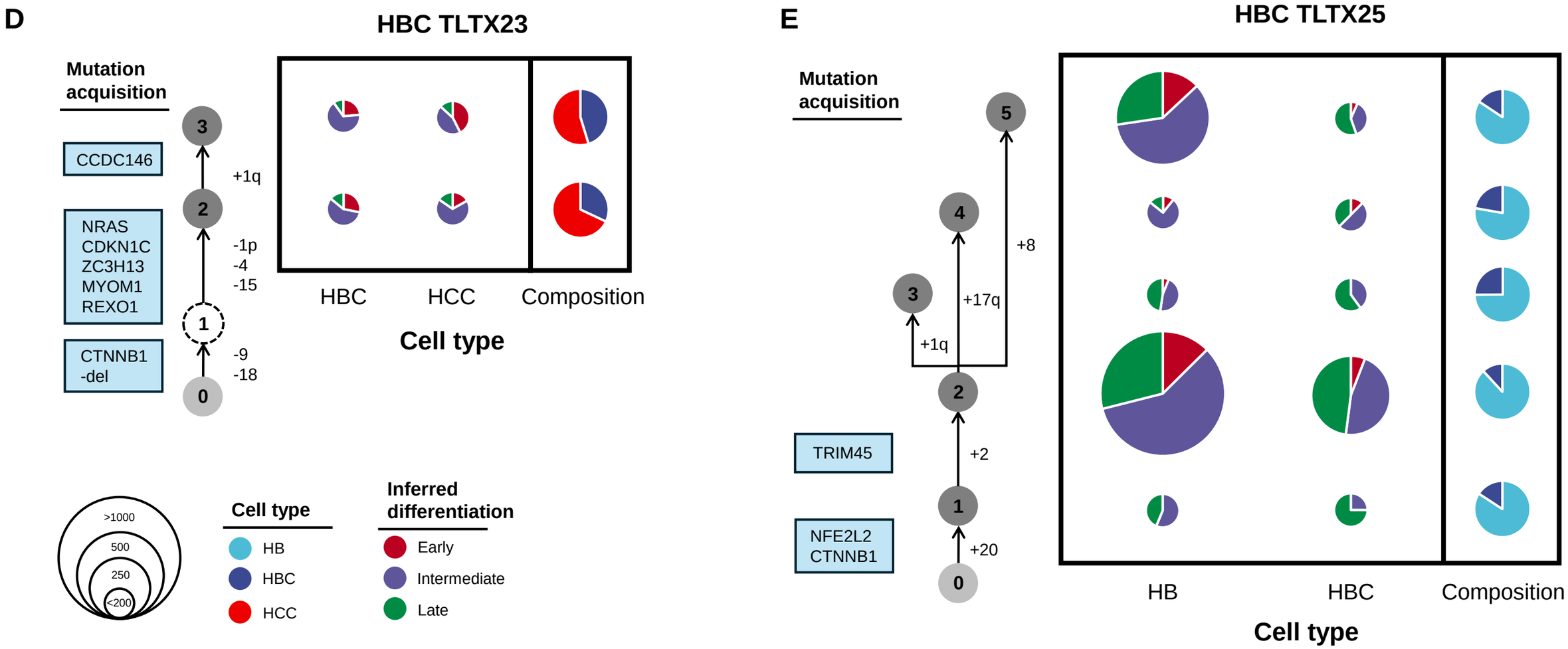


Figure 5 (continued)

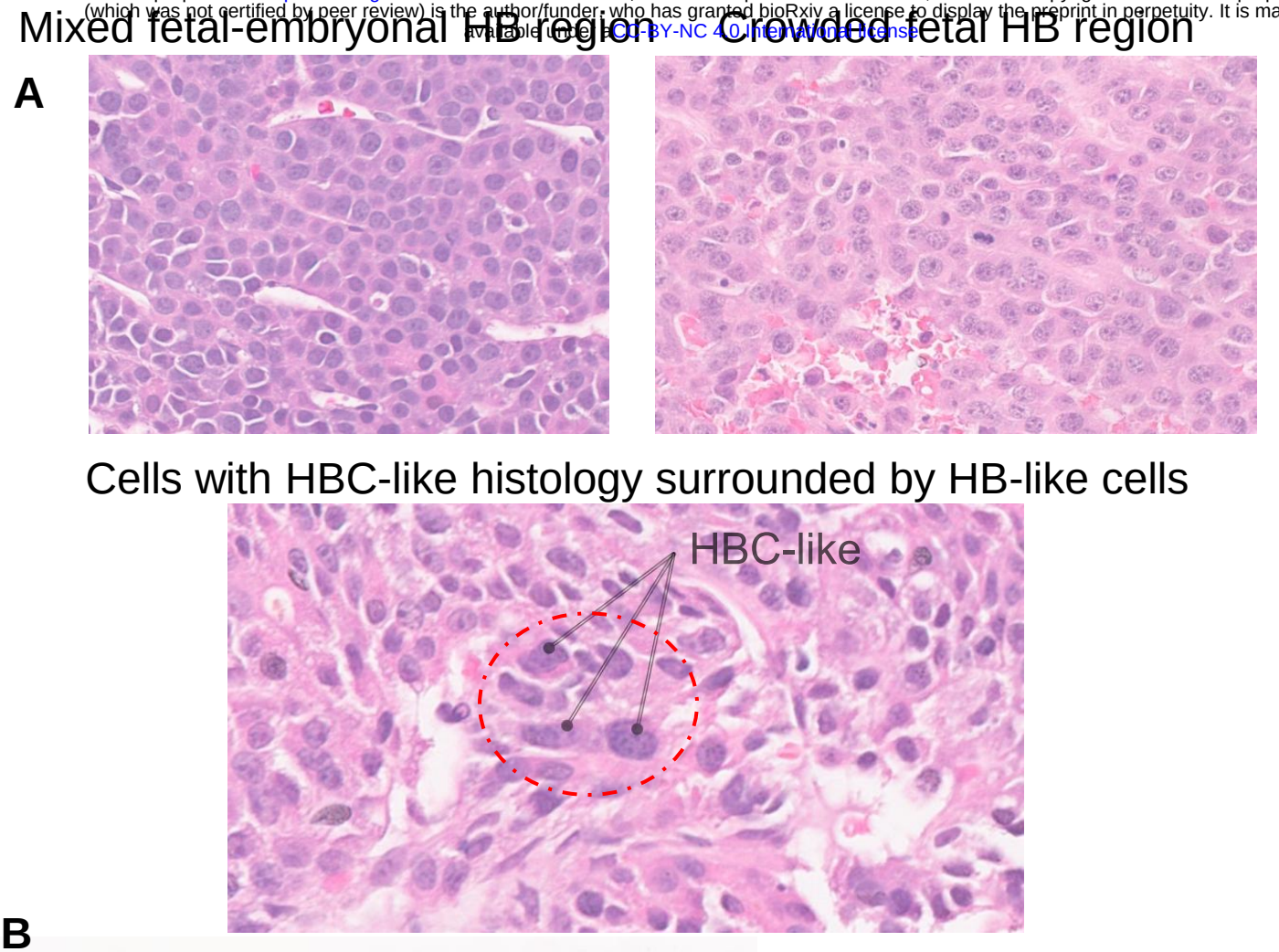
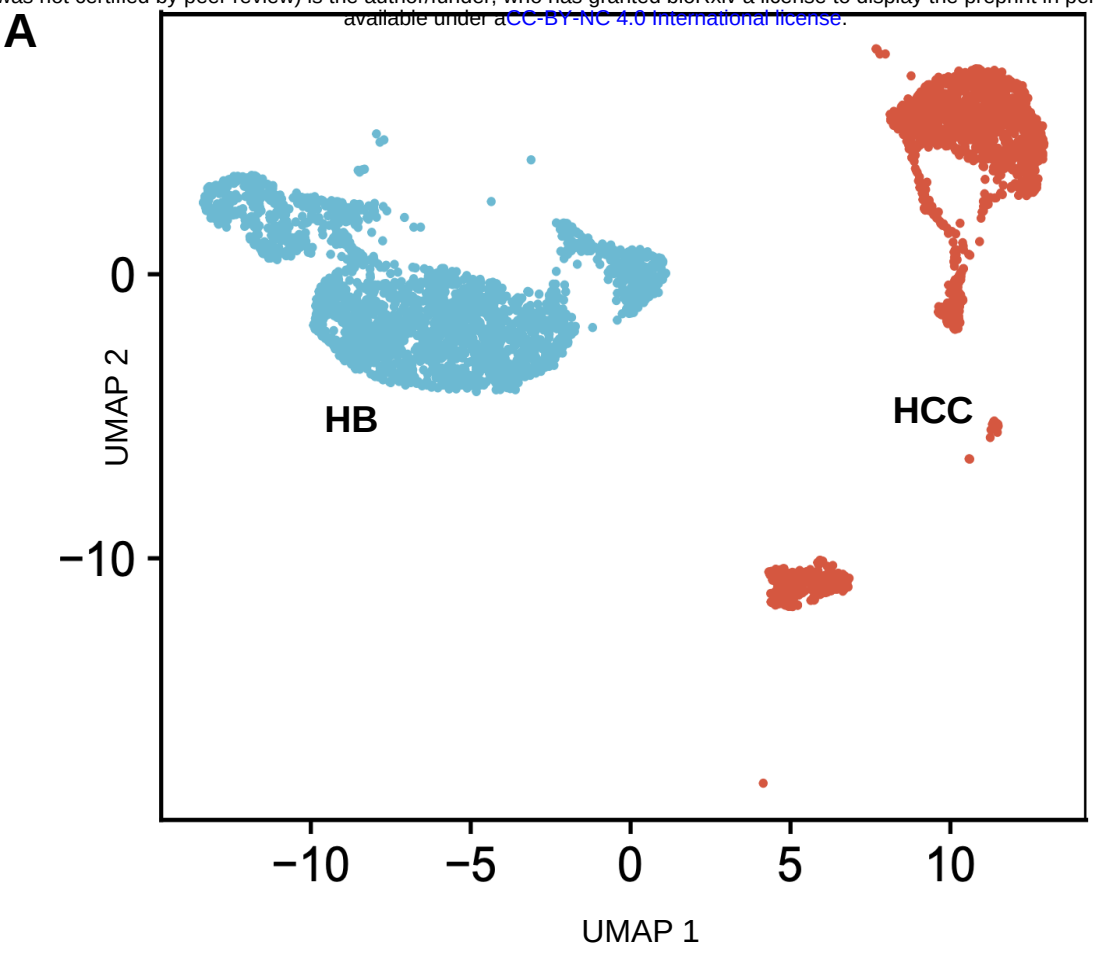


Figure S1



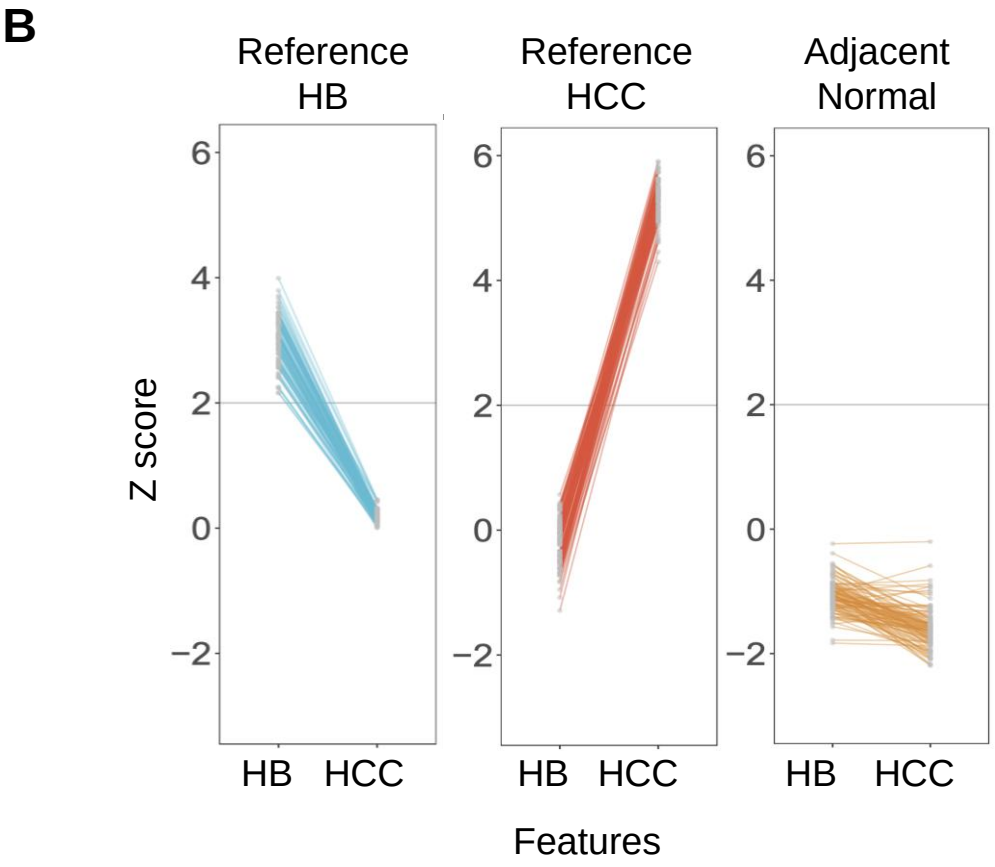
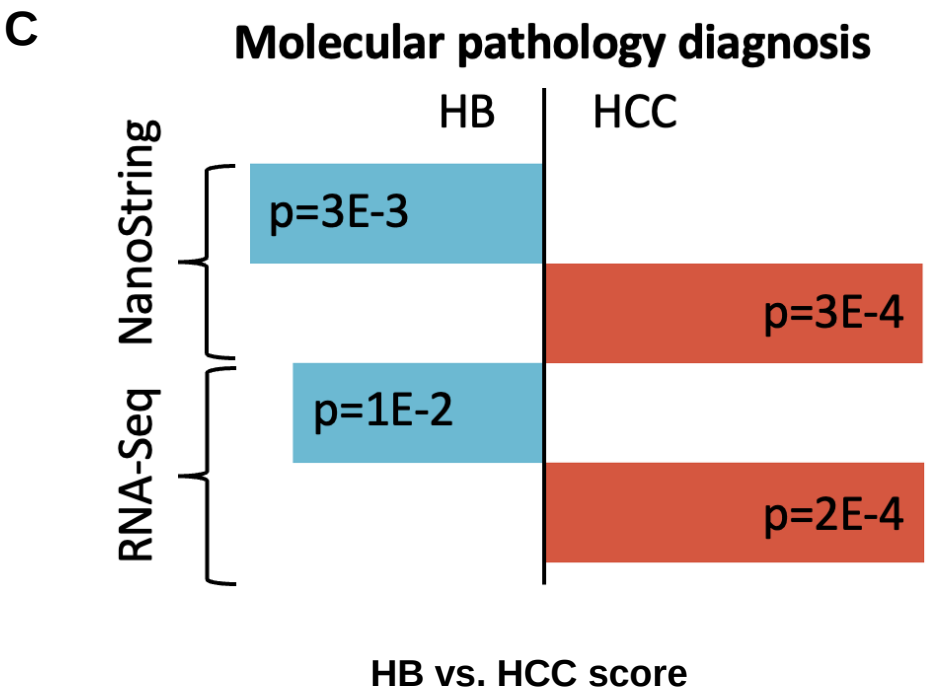


Figure S2



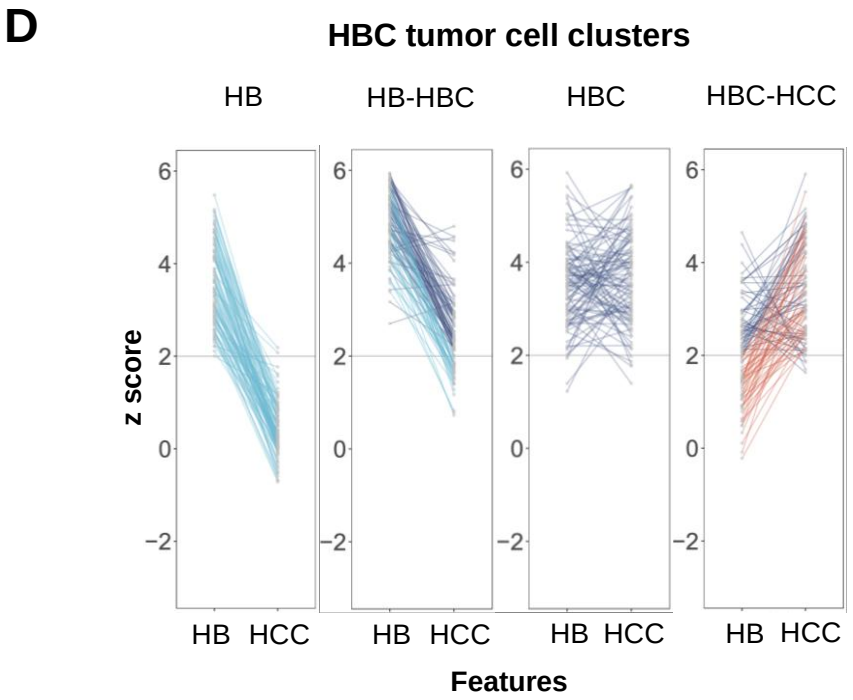


Figure S2

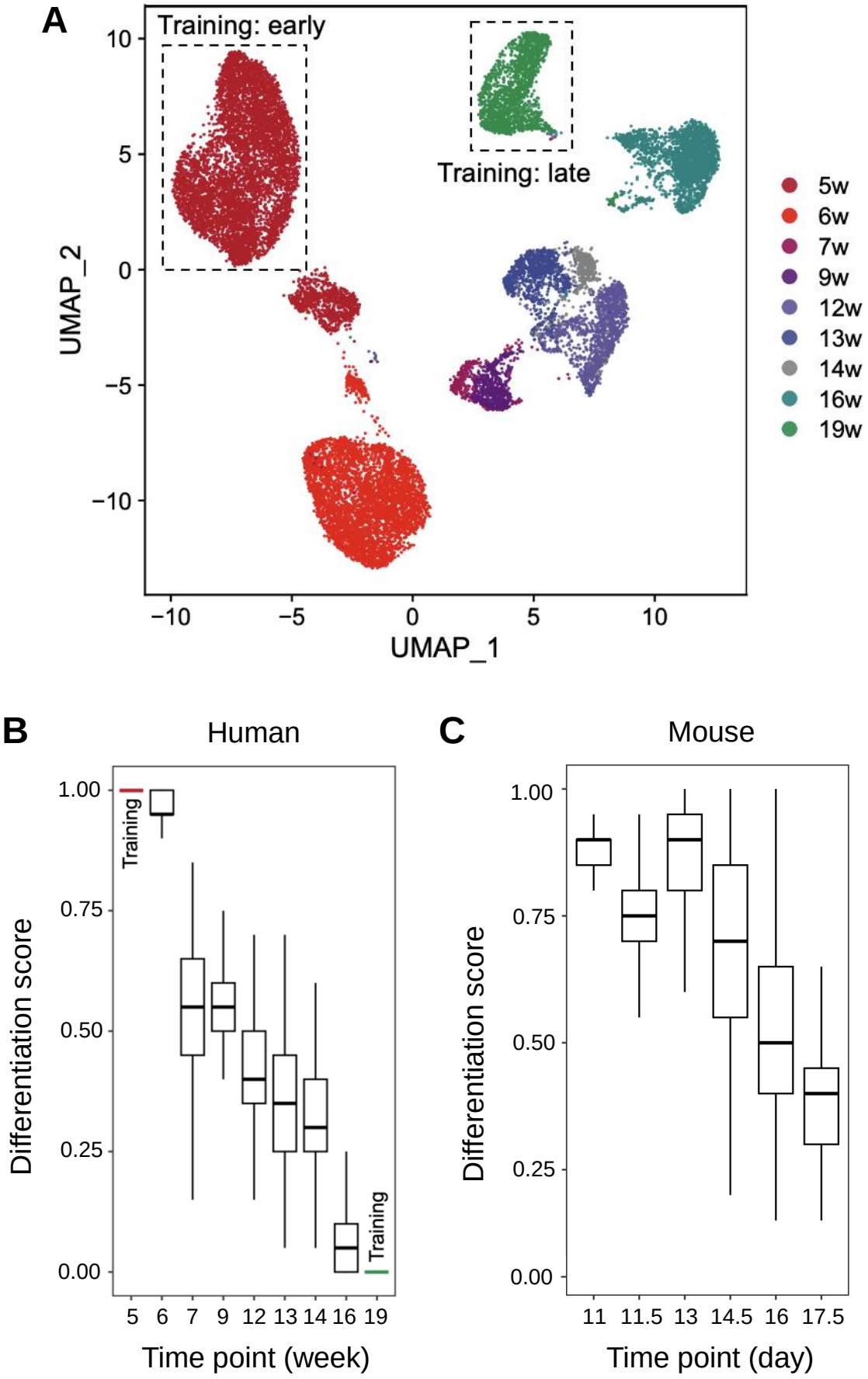


Figure S3

D

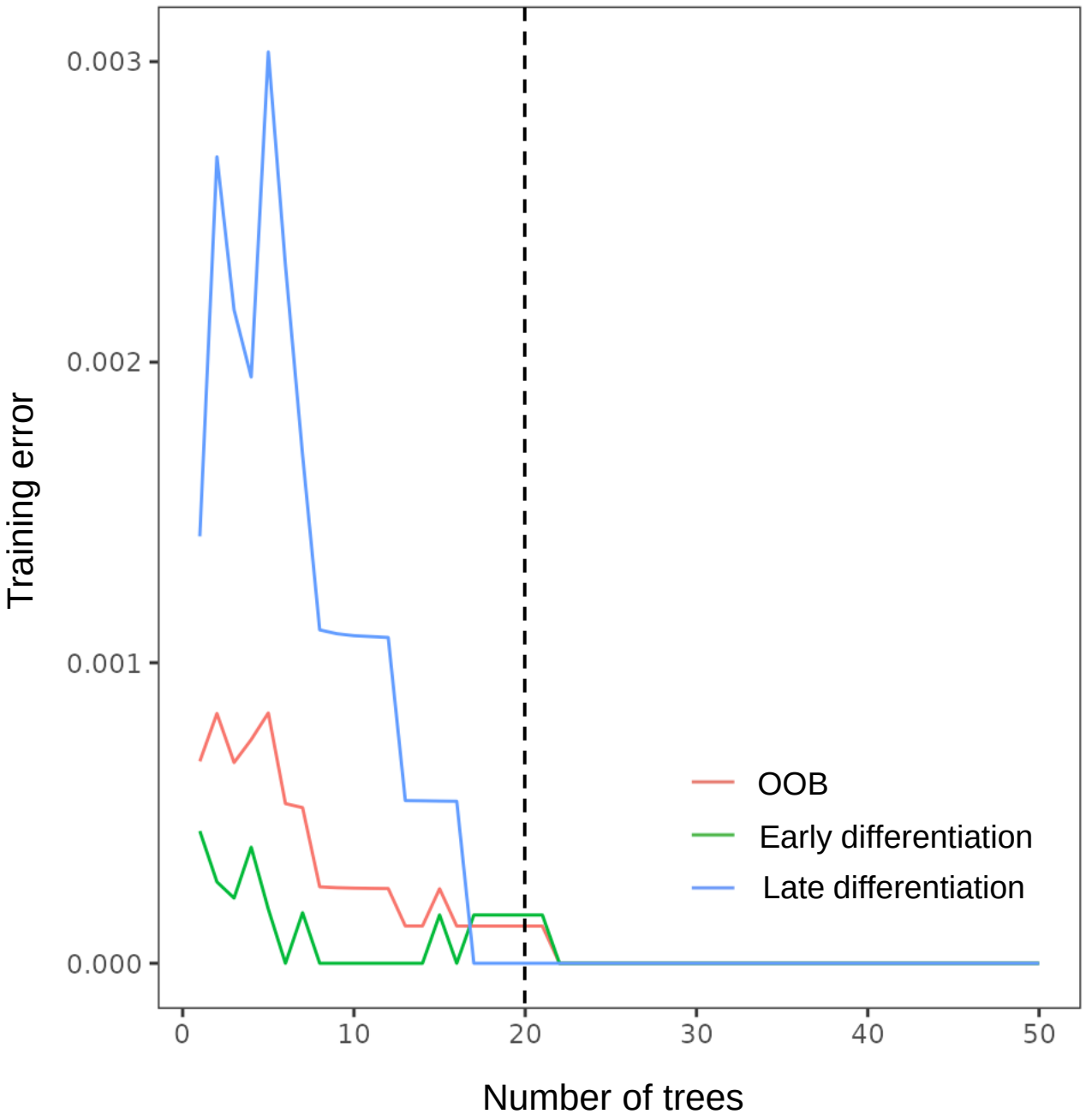


Figure S3

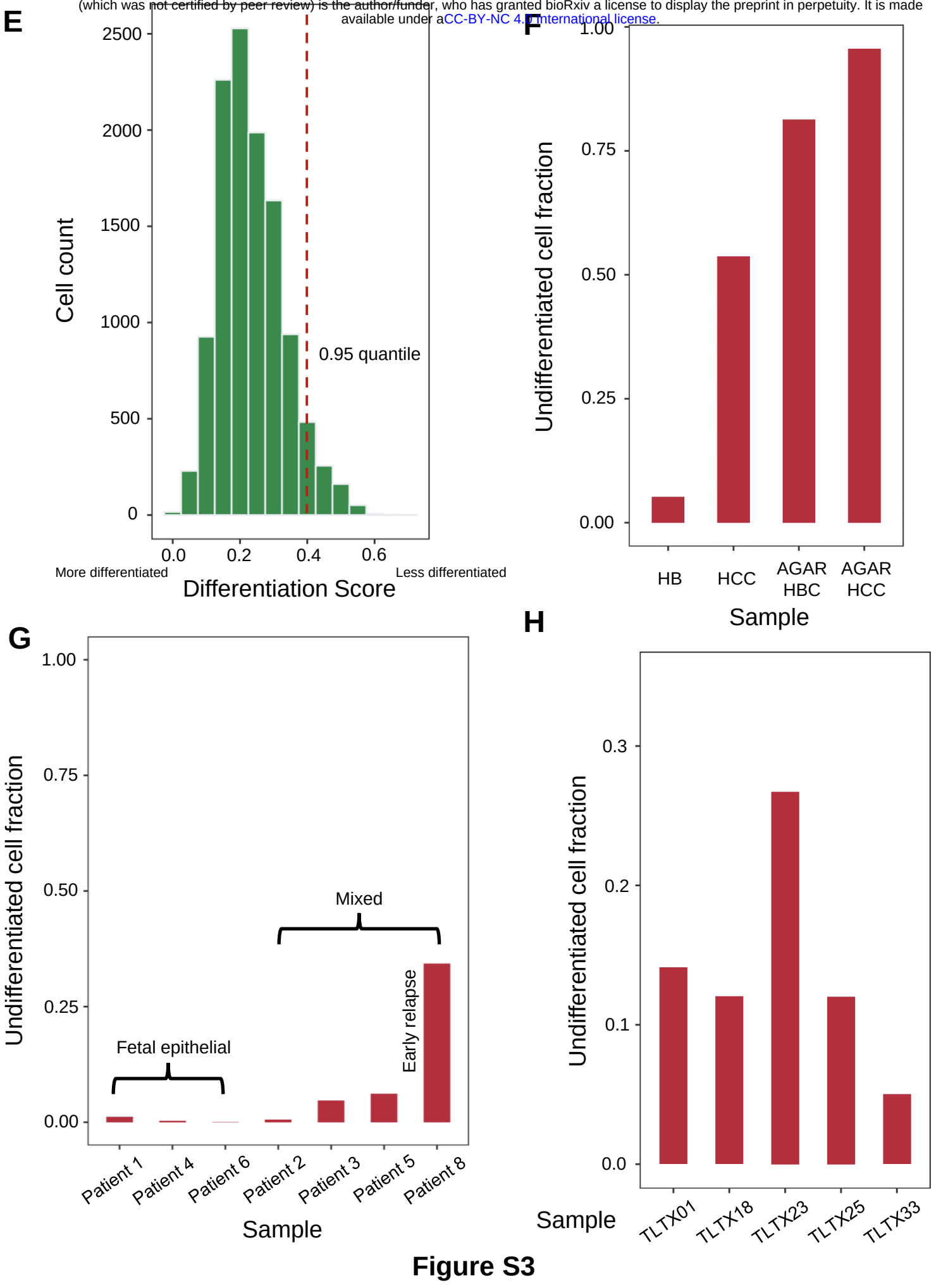


Figure S3

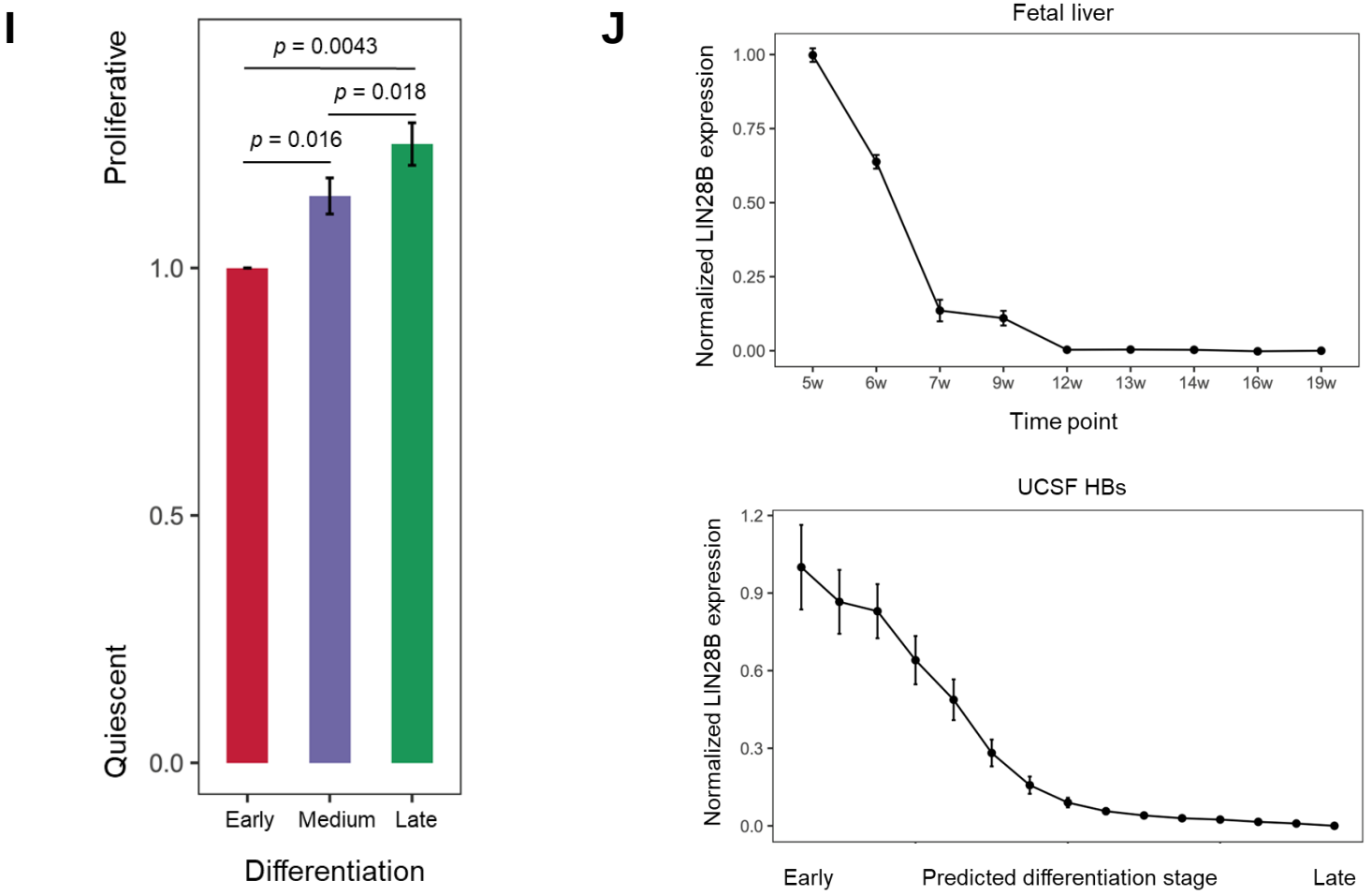


Figure S3

K

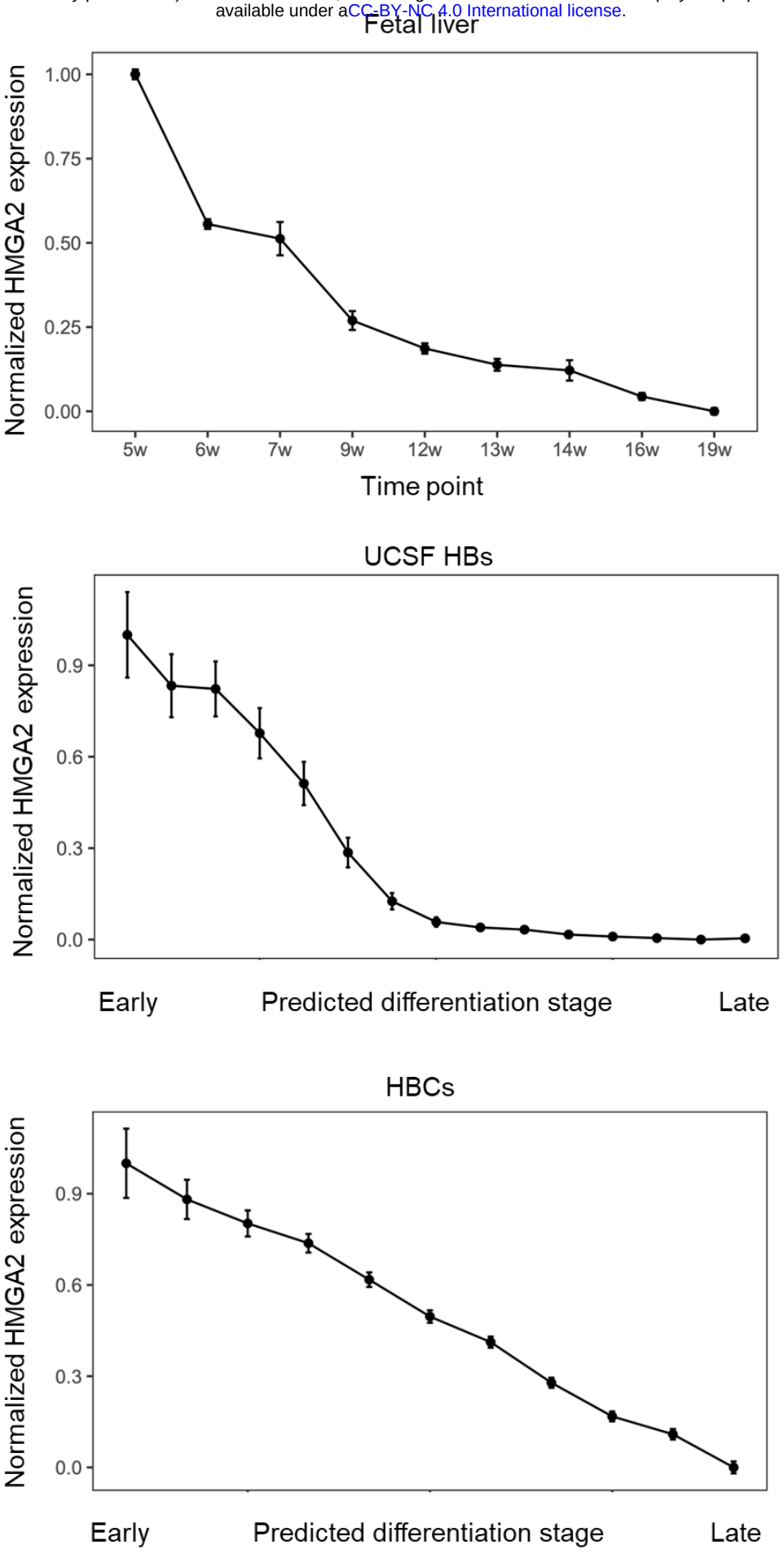


Figure S3

L

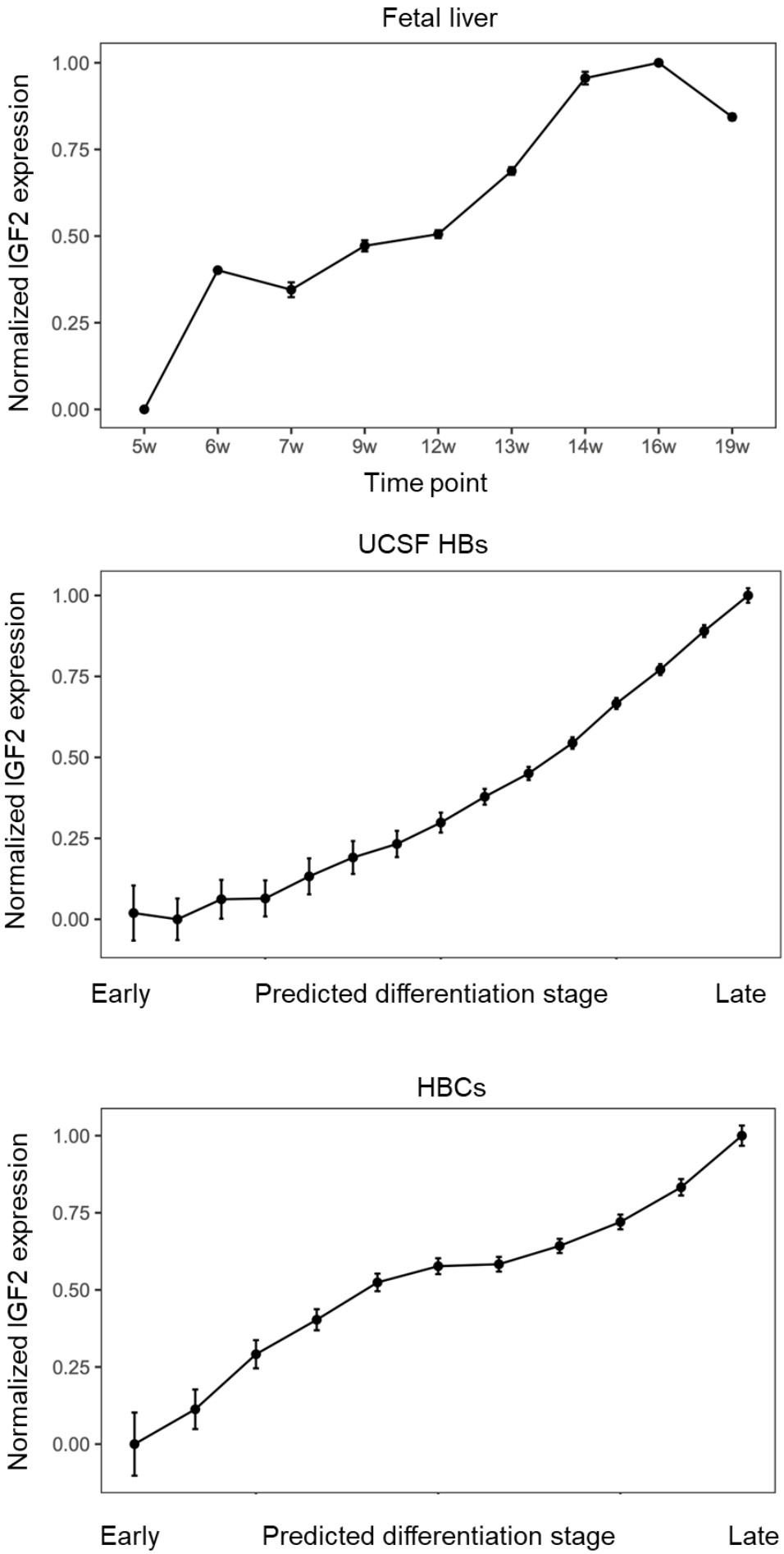
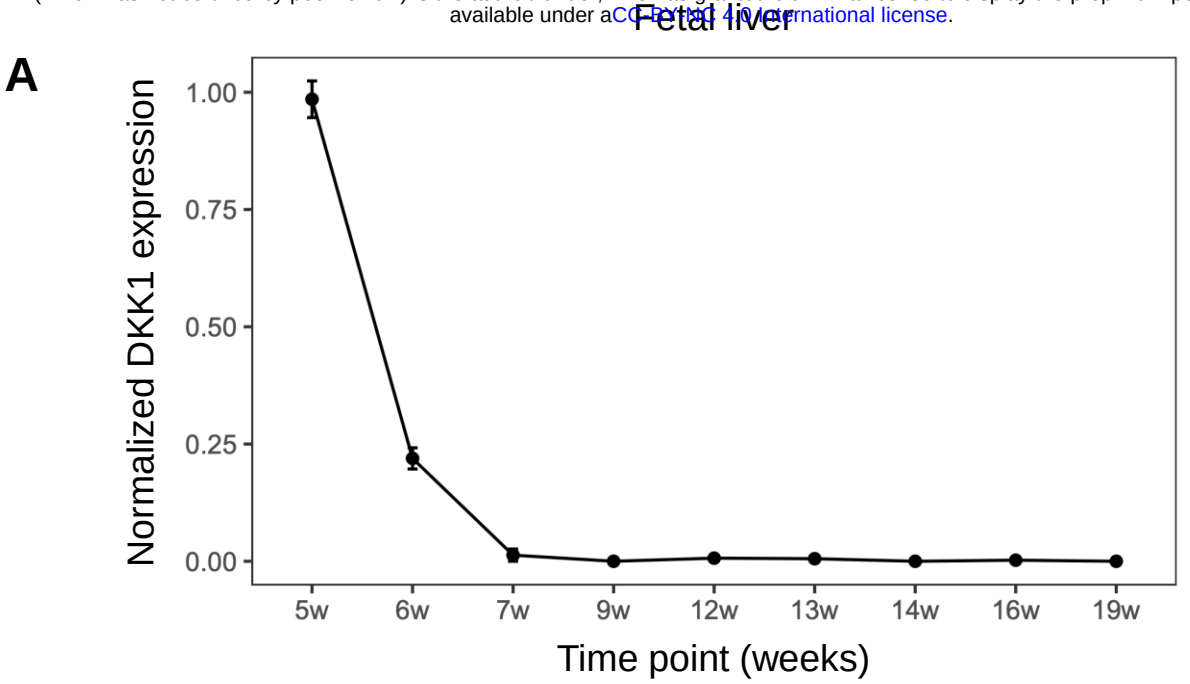
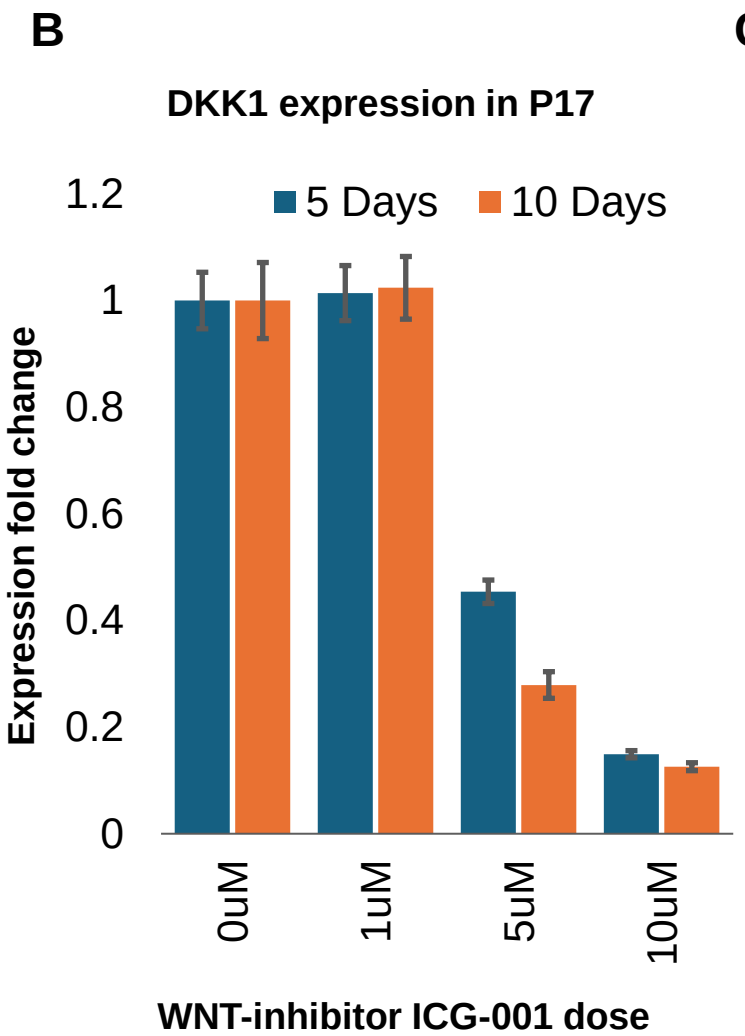


Figure S3





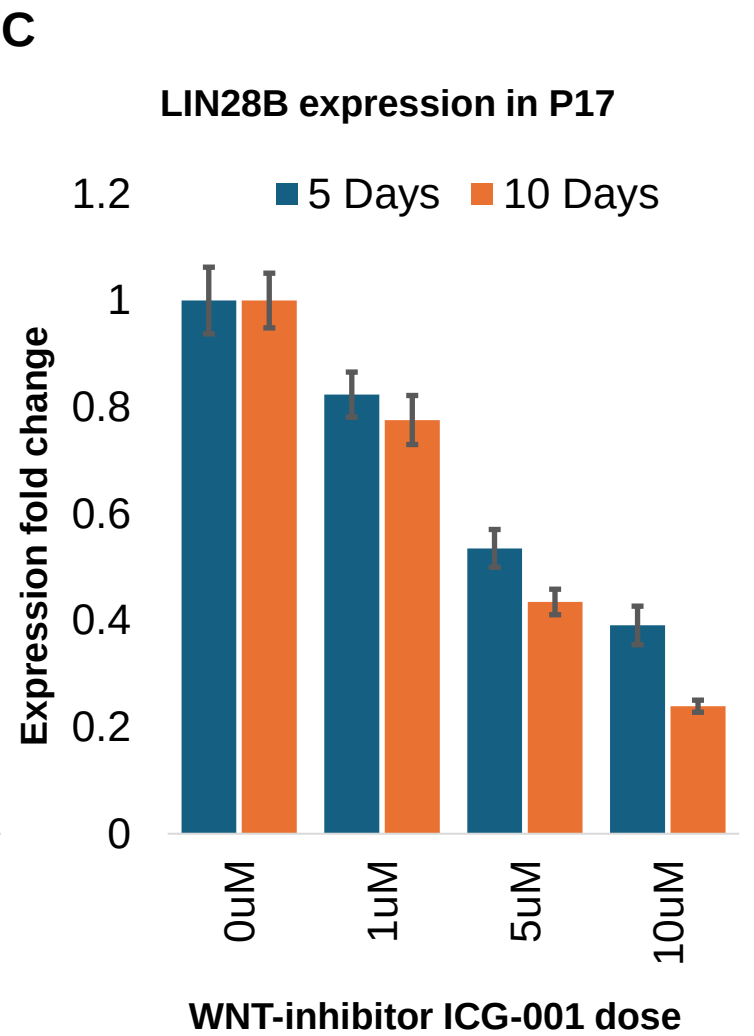


Figure S4

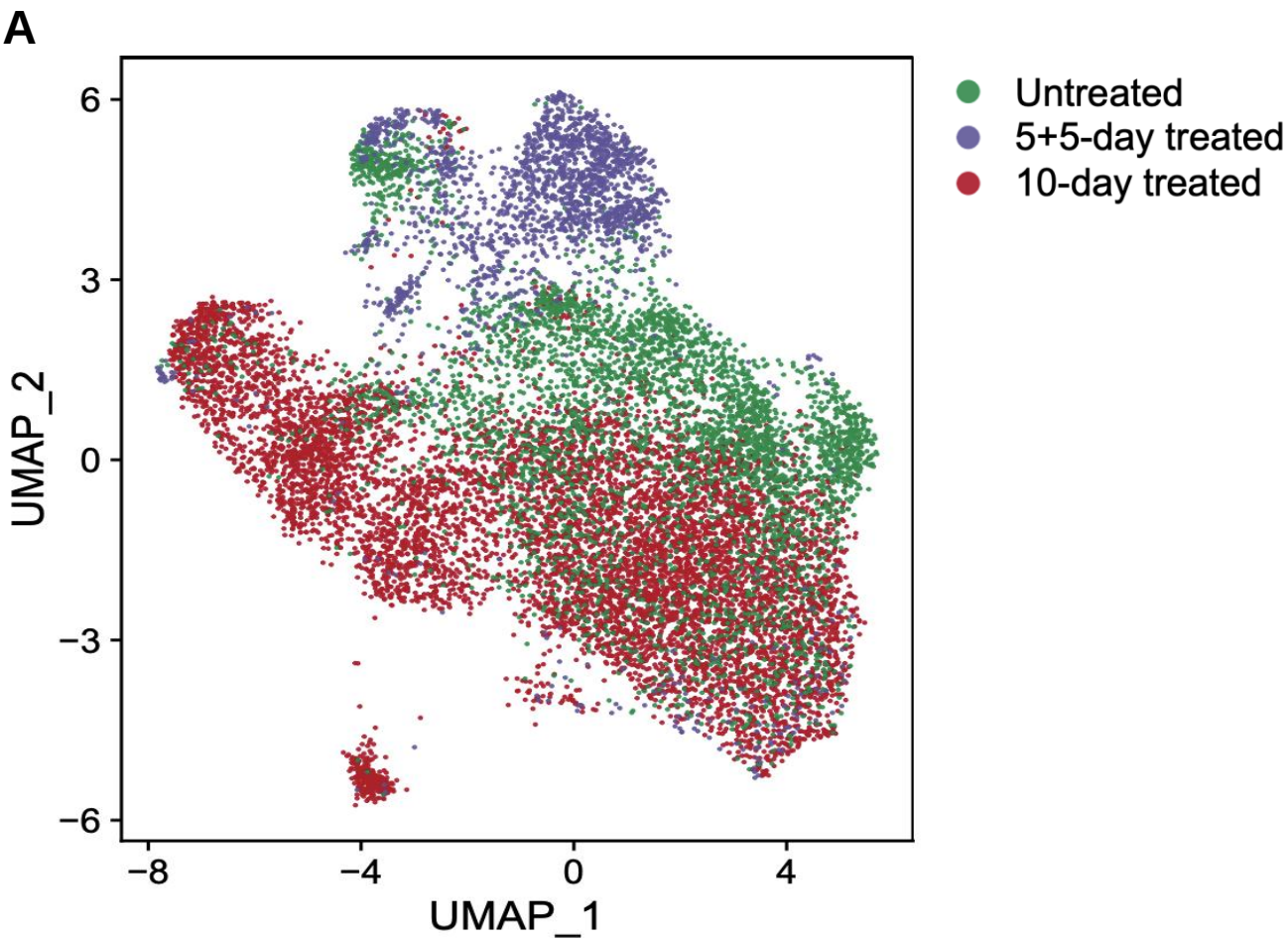
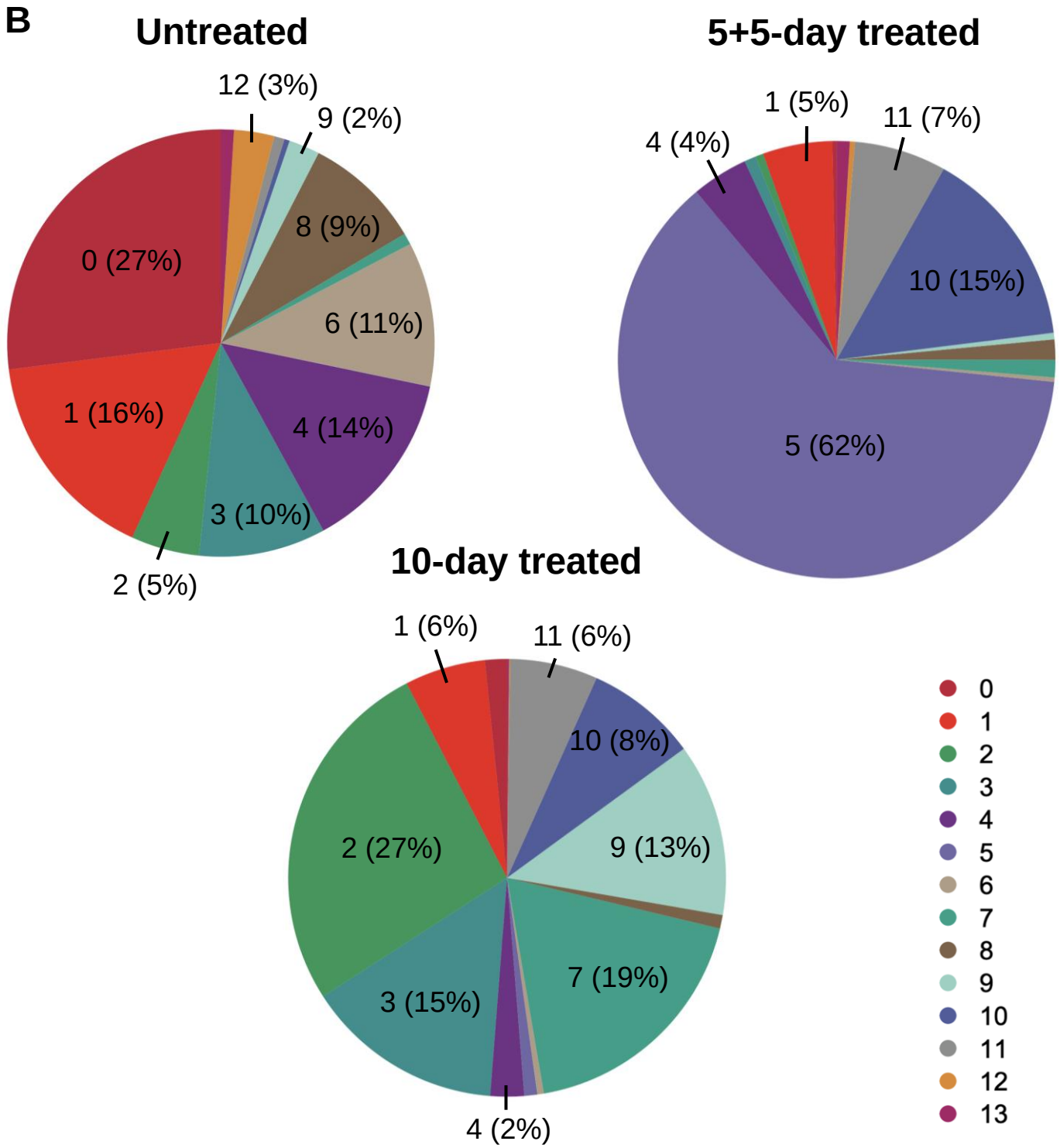


Figure S5



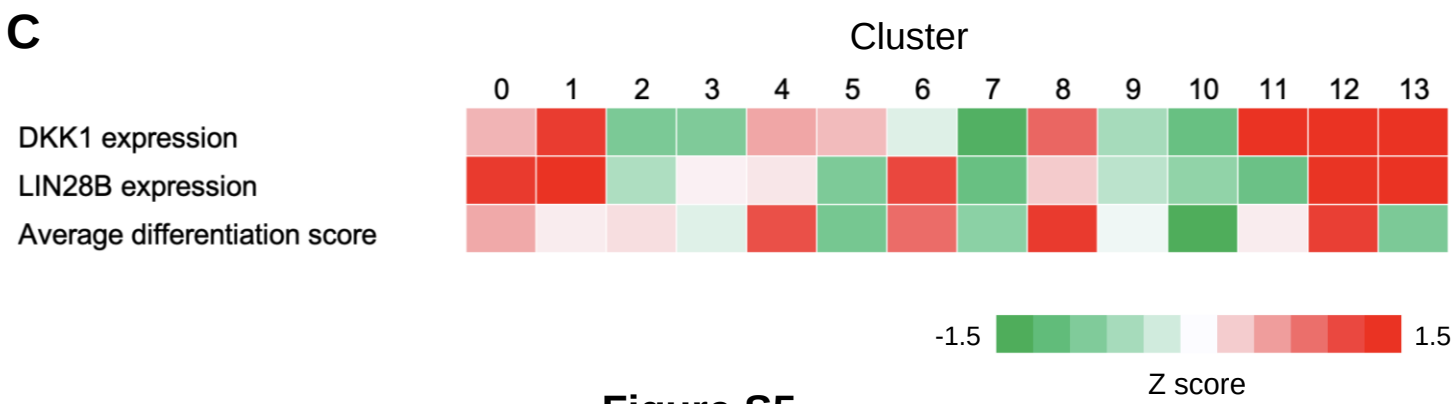
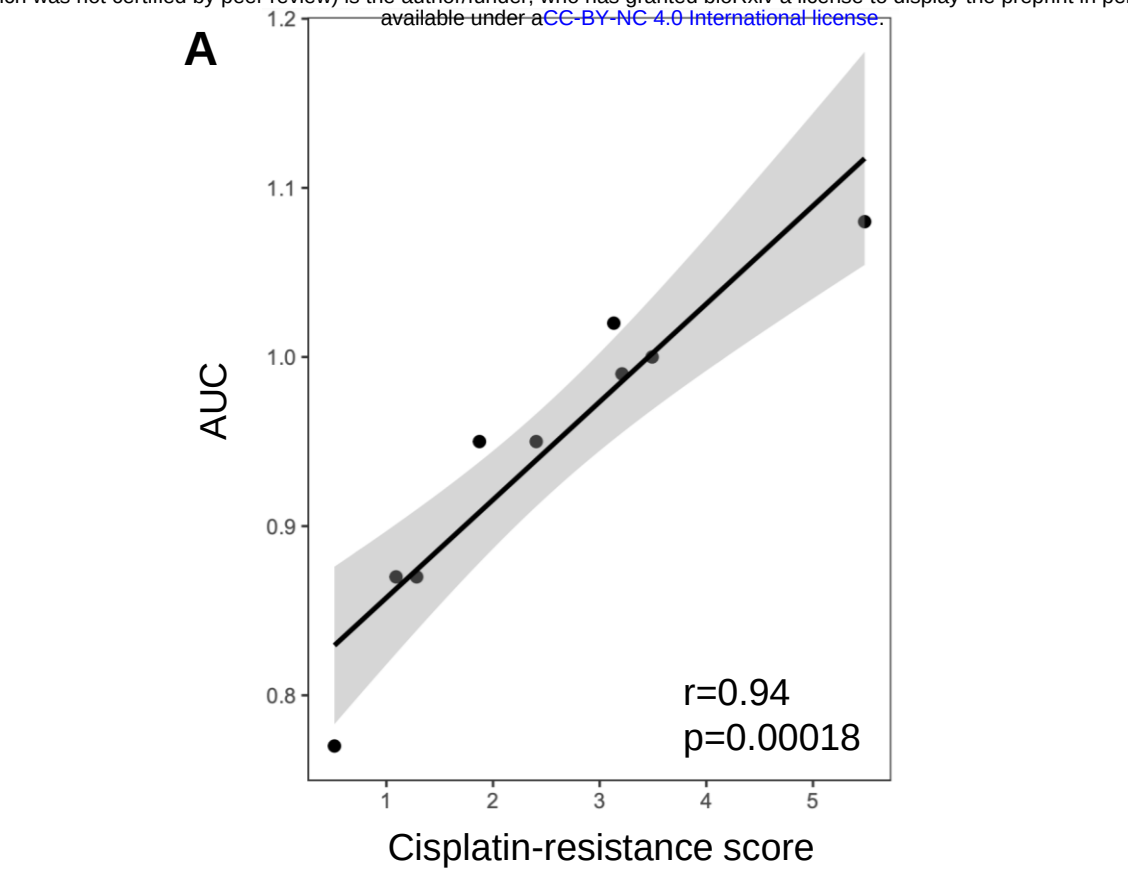


Figure S5



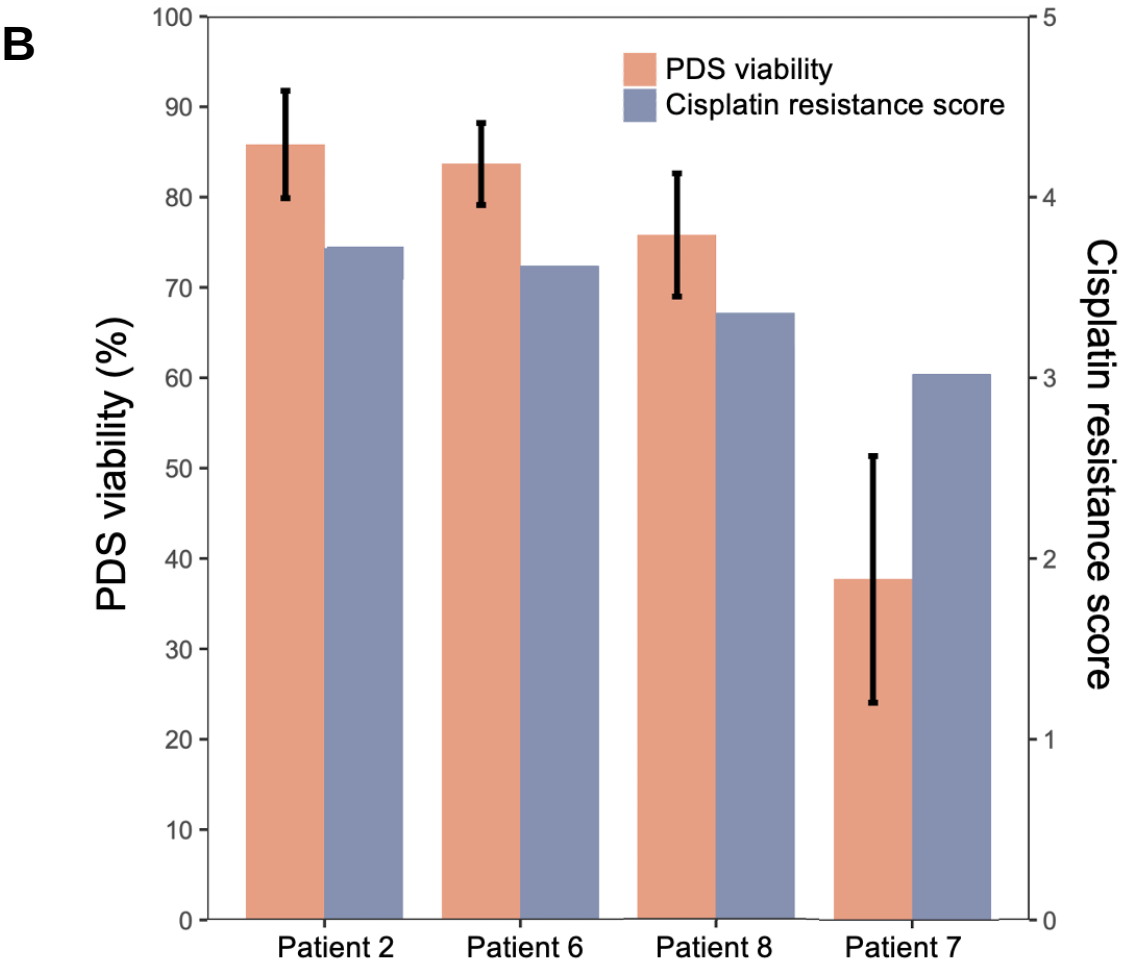


Figure S6

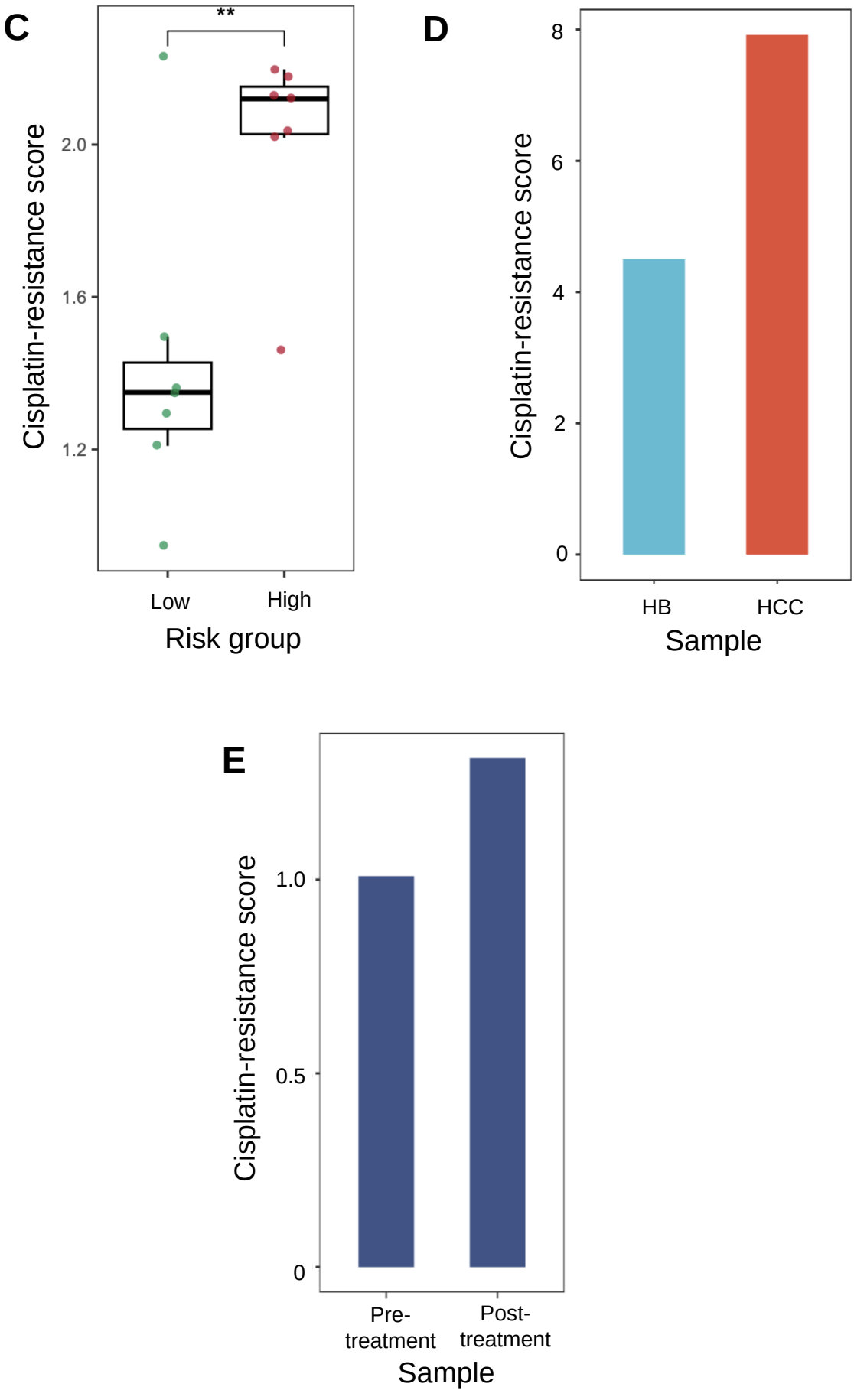
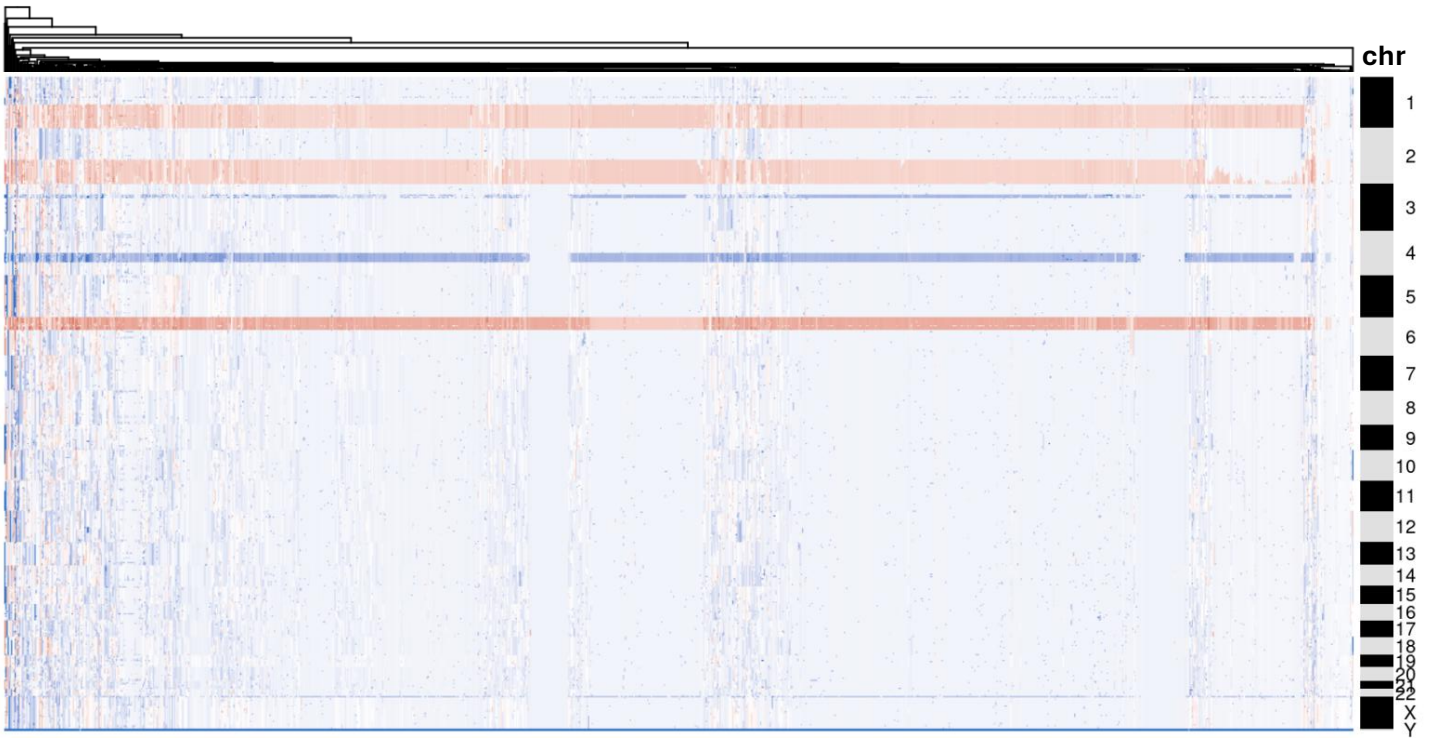
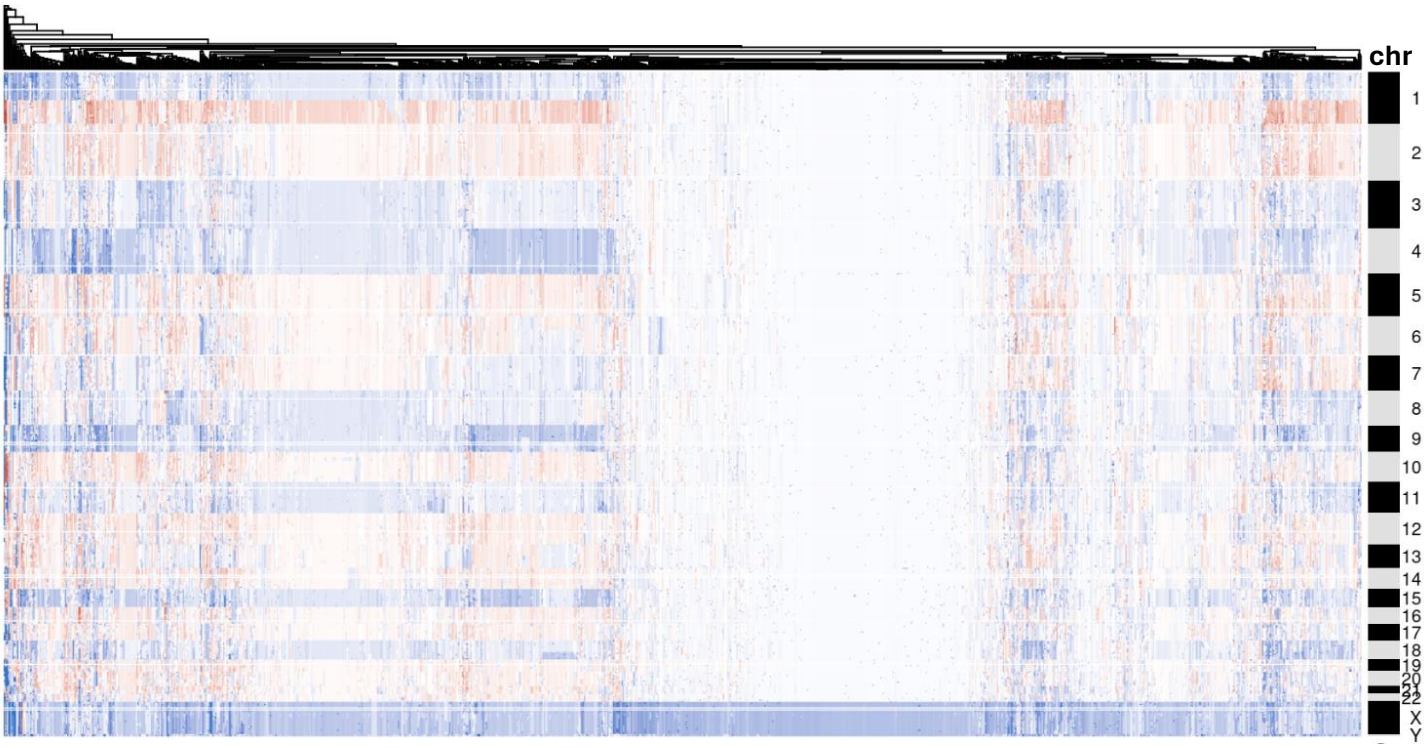


Figure S6

A



B



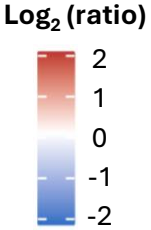
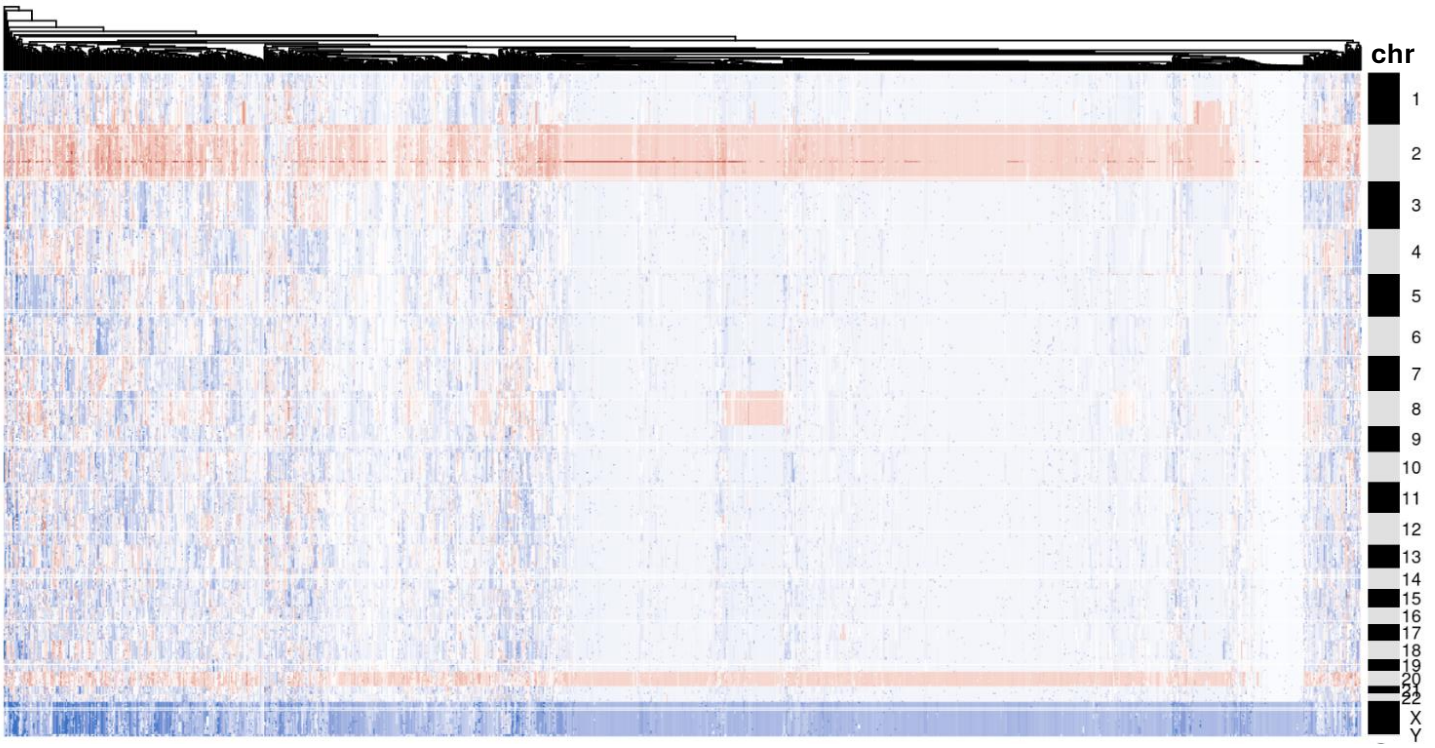
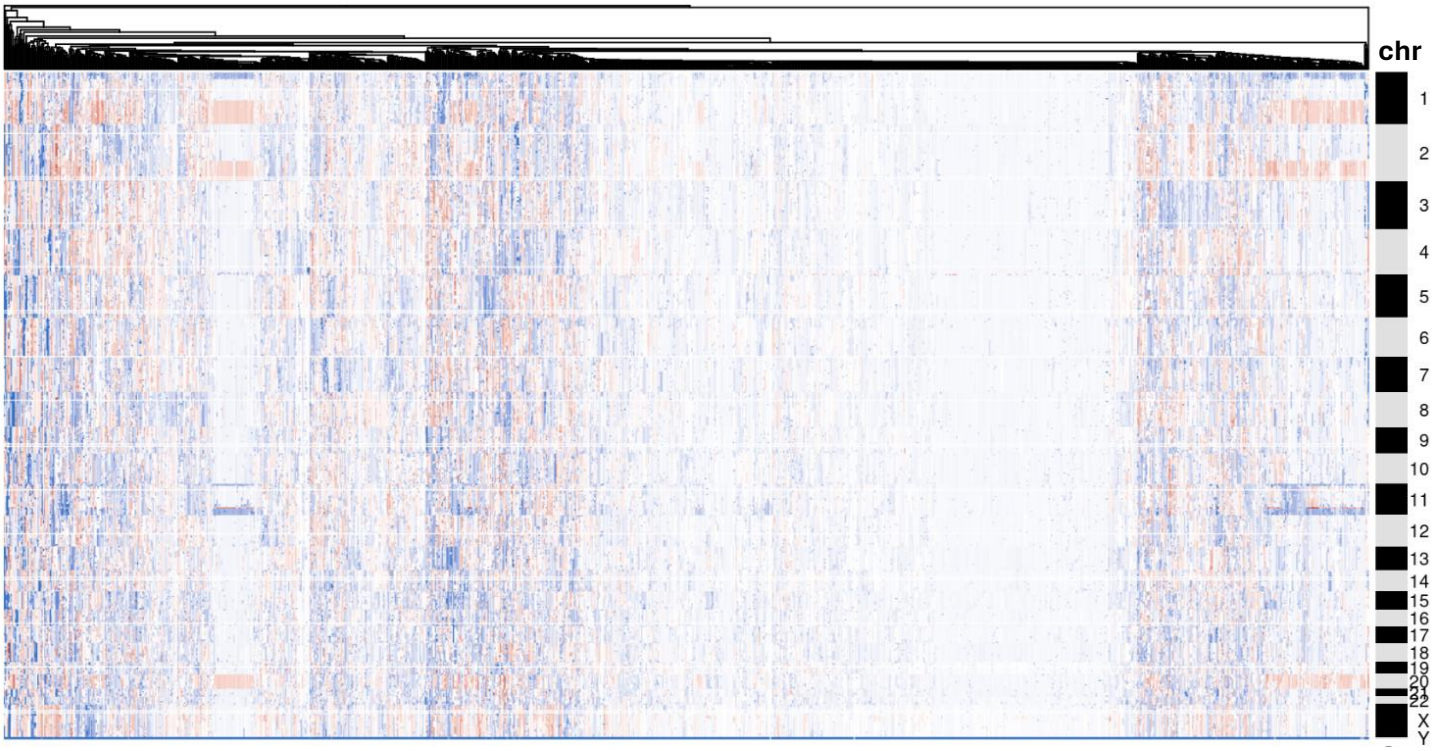


Figure S7

C



D



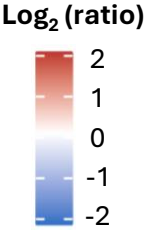


Figure S7

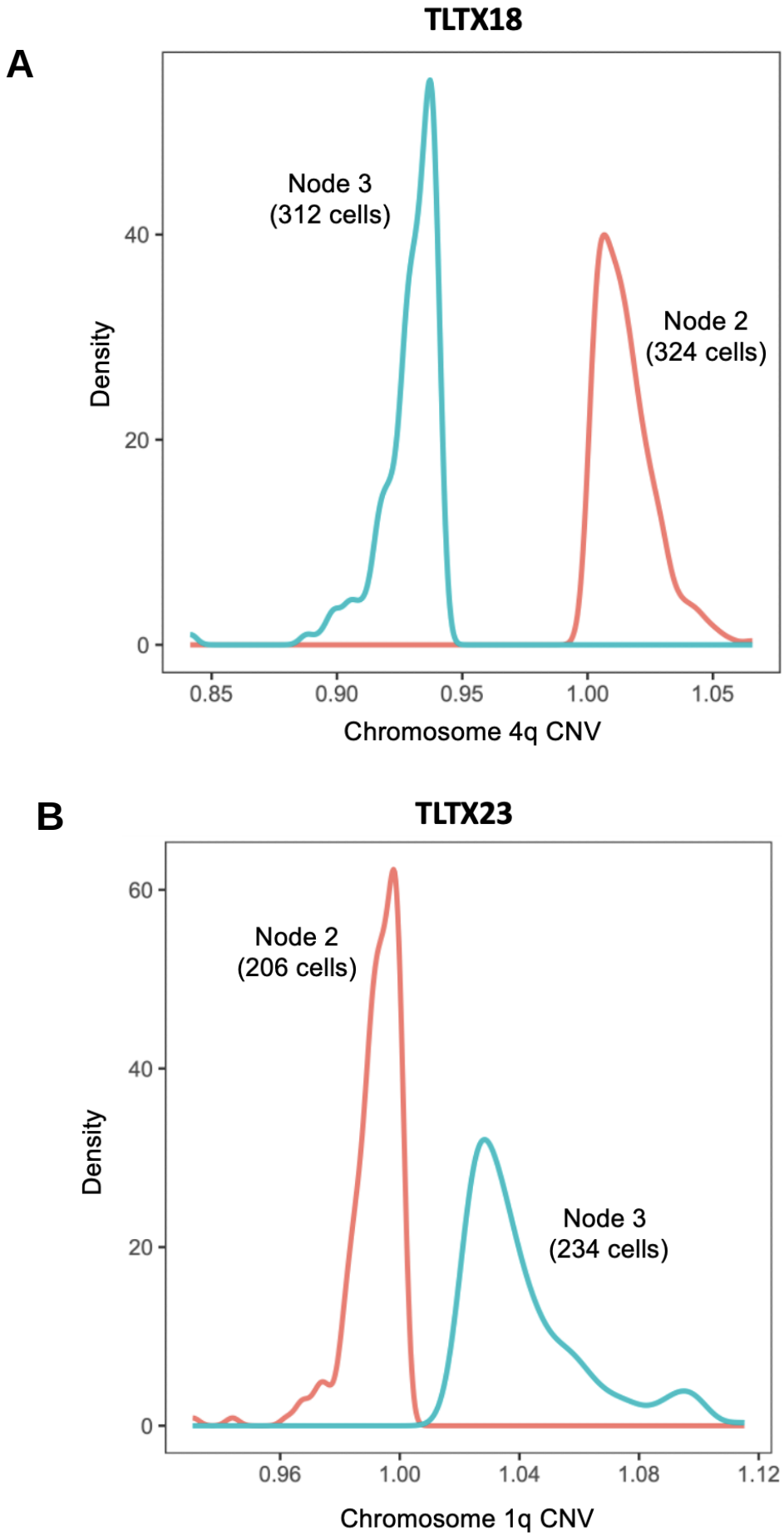
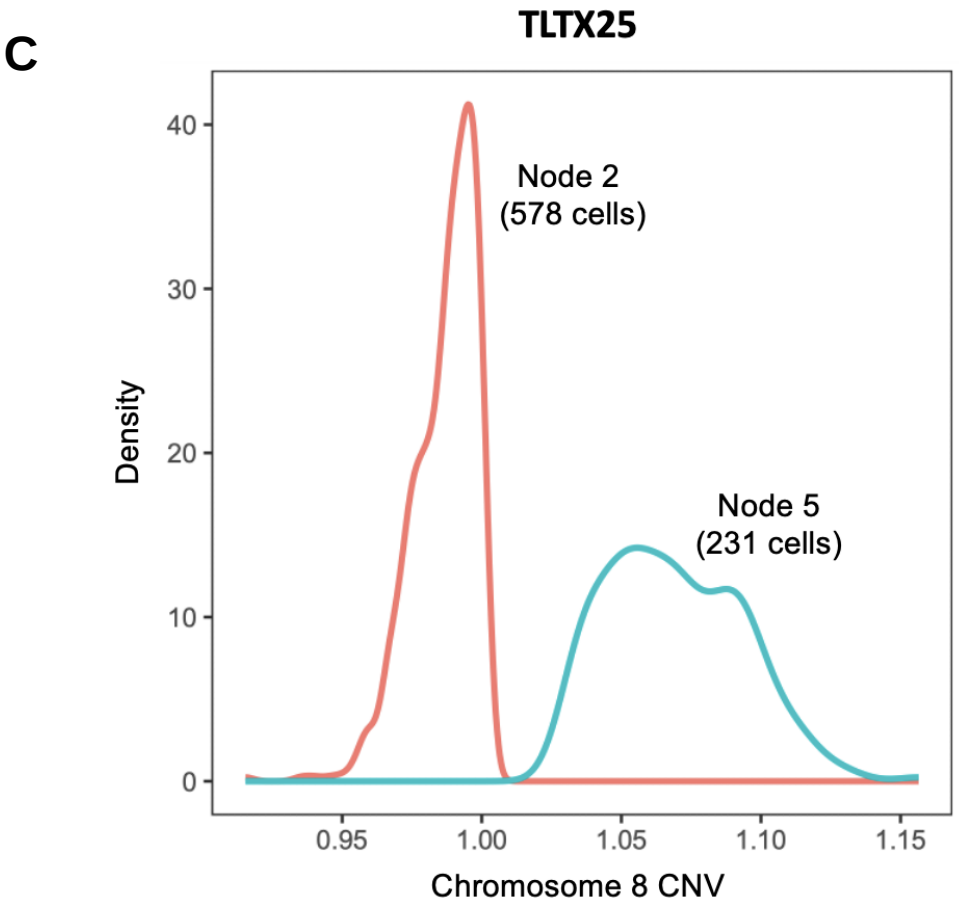


Figure S8



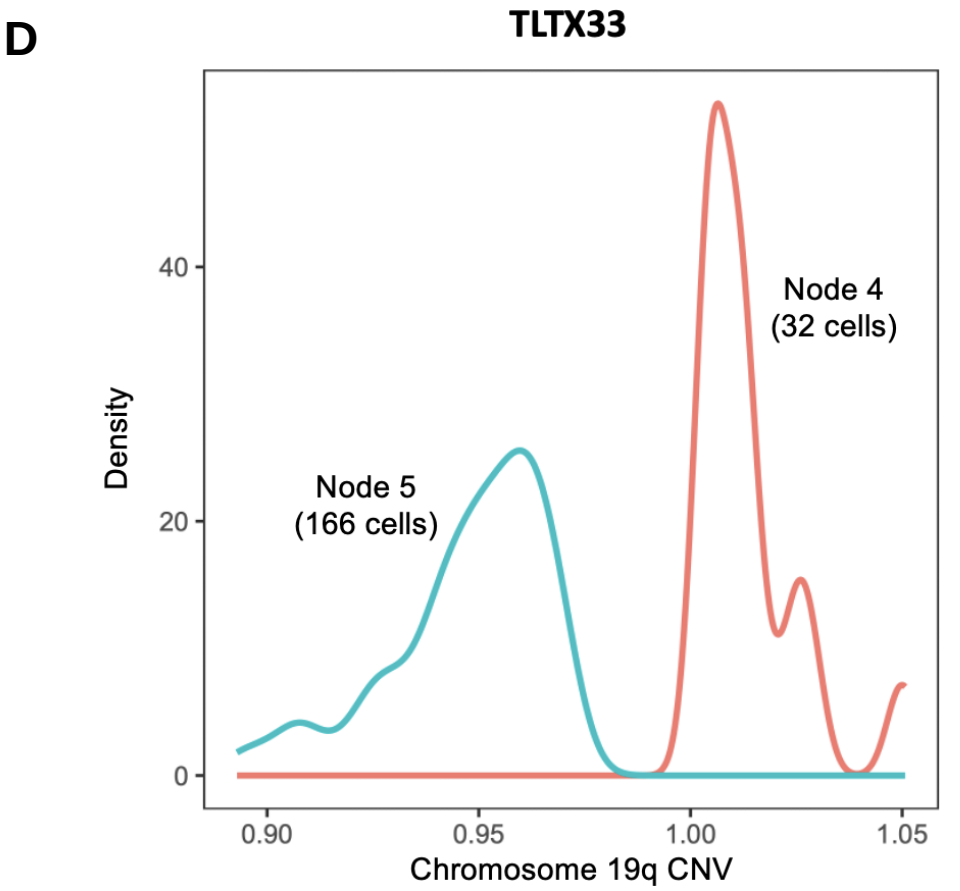


Figure S8

E

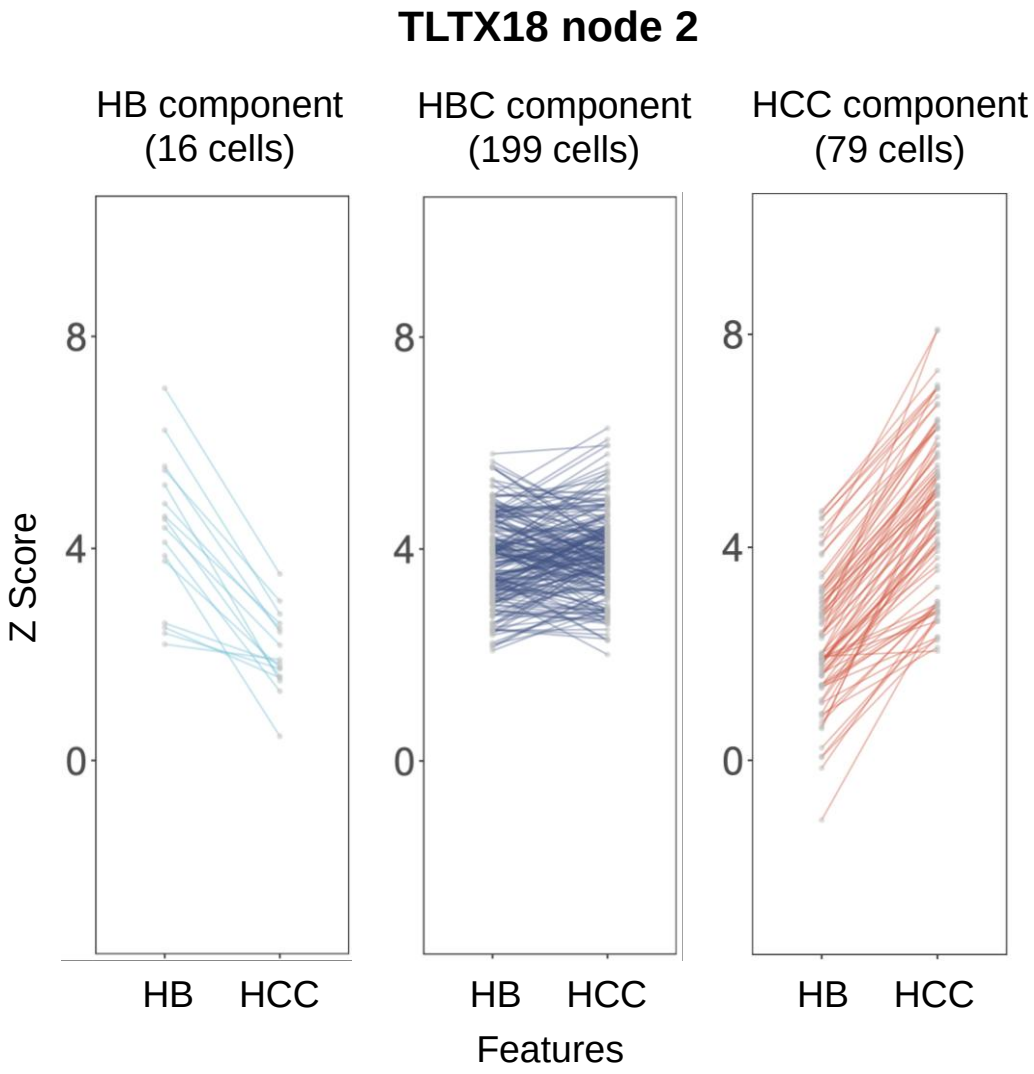


Figure S8

F

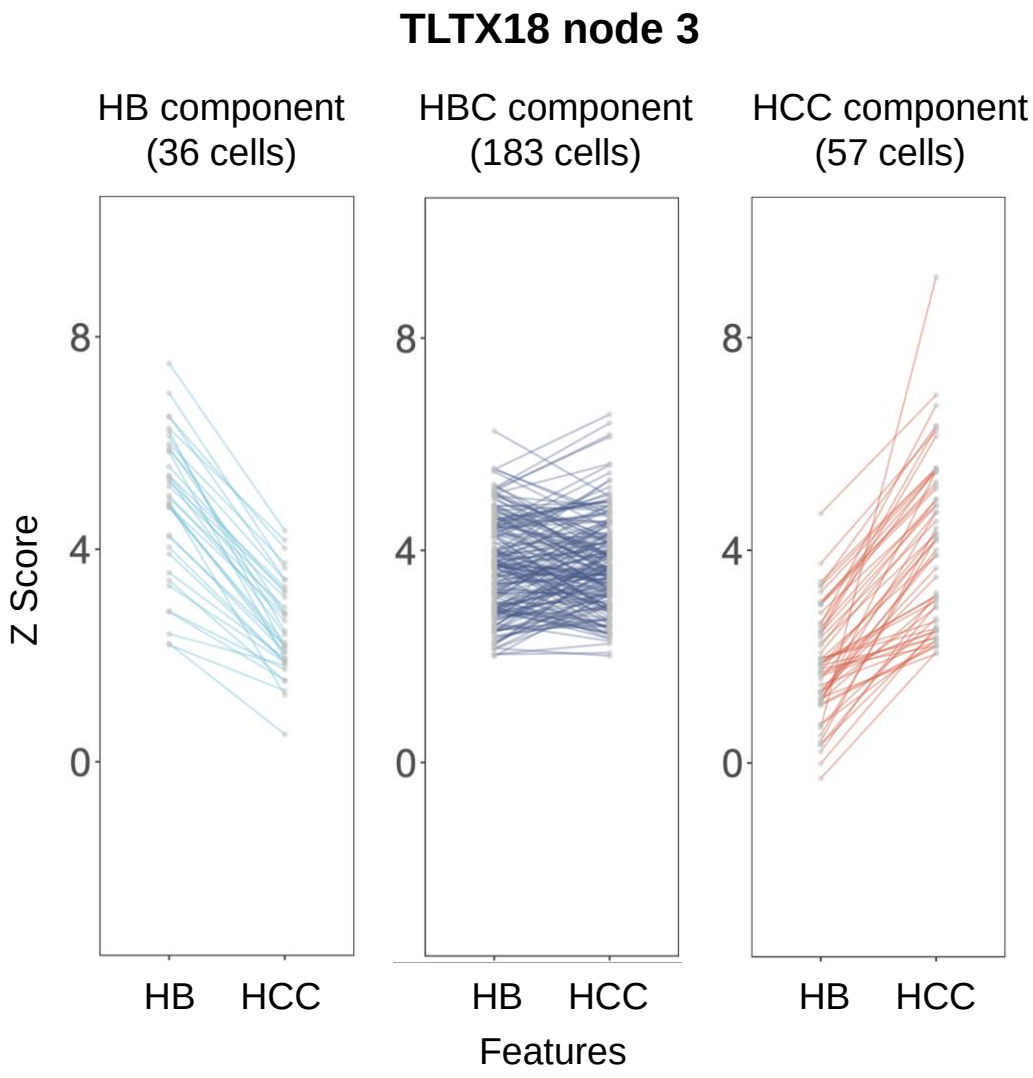


Figure S8

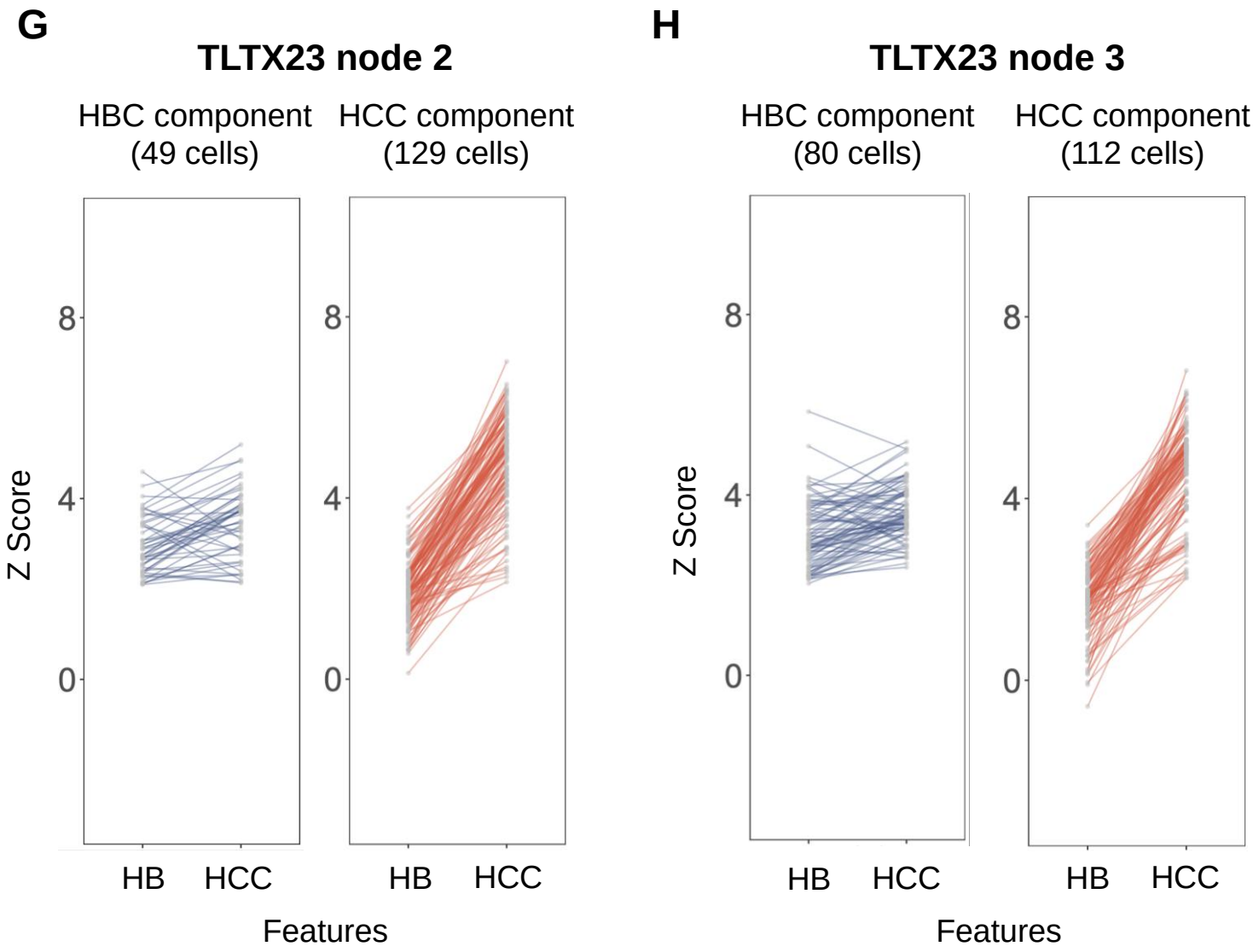


Figure S8

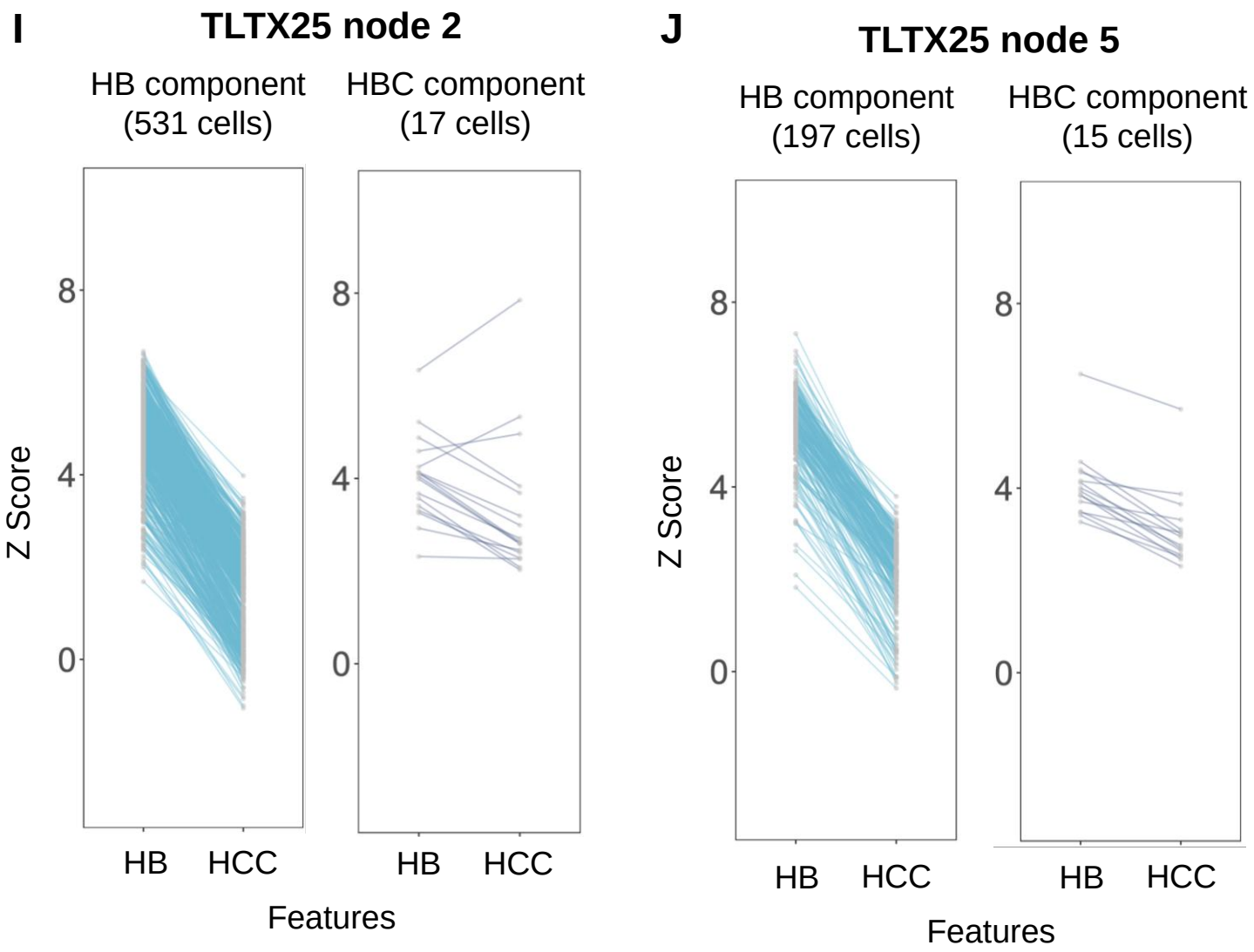


Figure S8

K

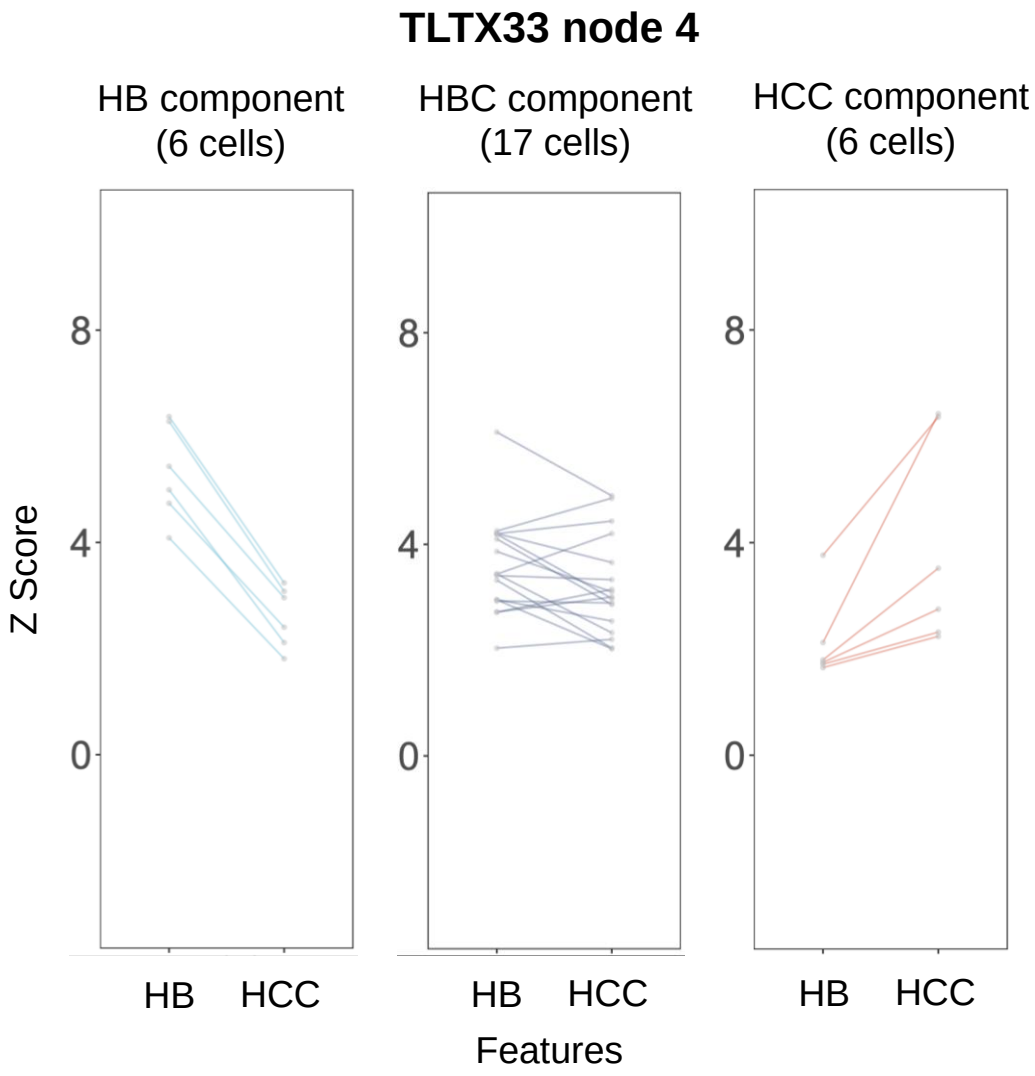


Figure S8

L

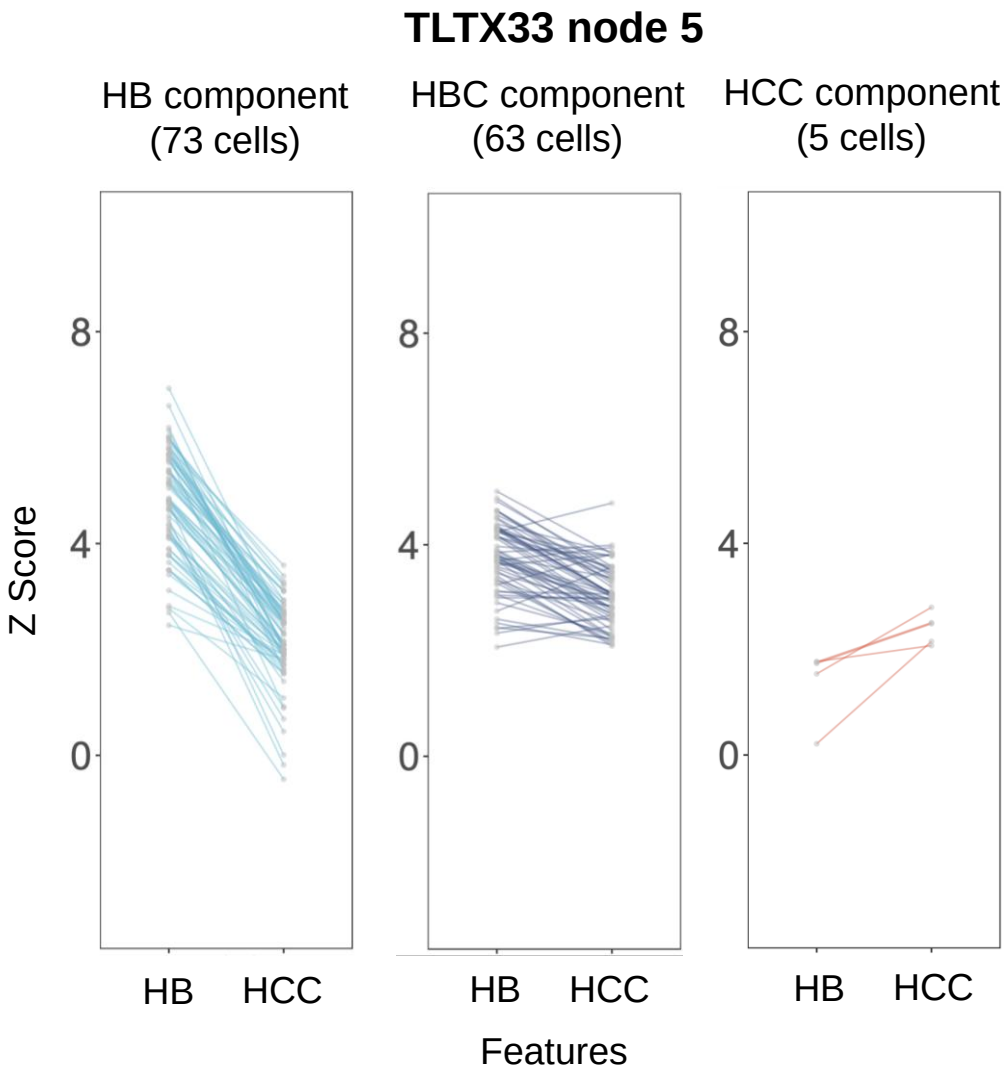
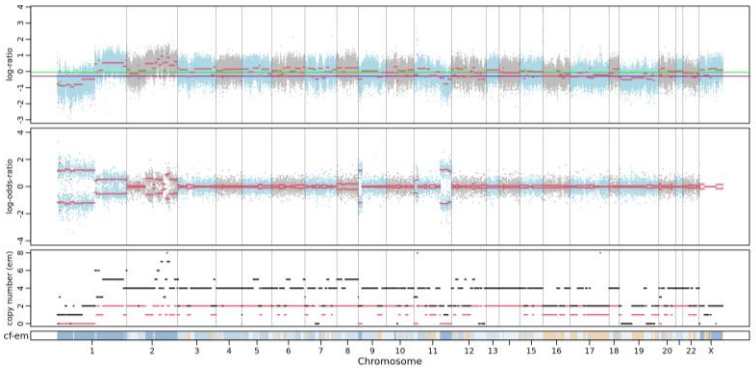
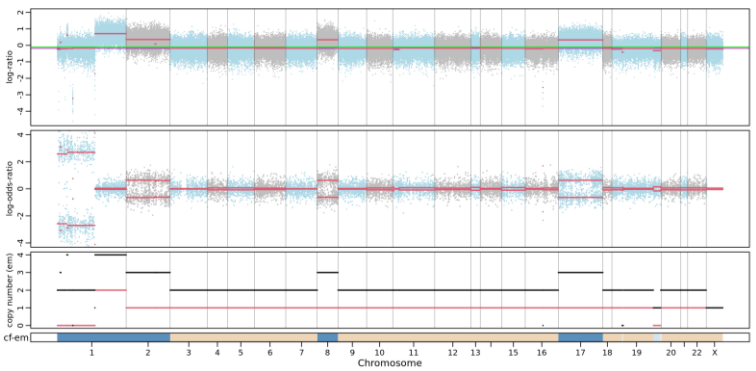


Figure S8

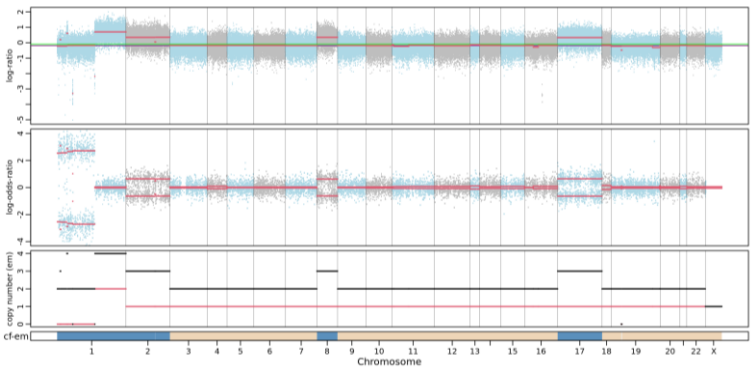
TLTX36.A1



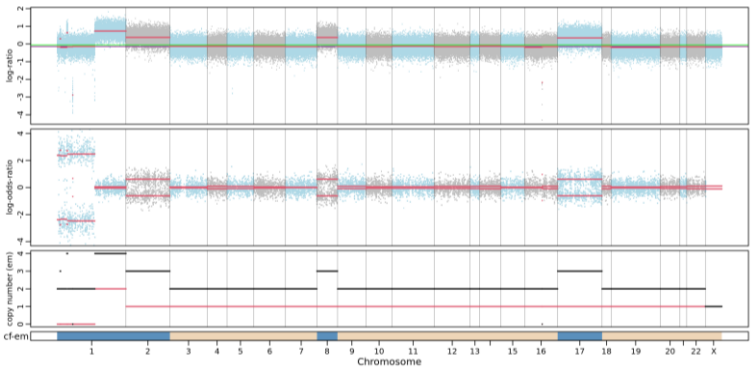
TLTX36.A14



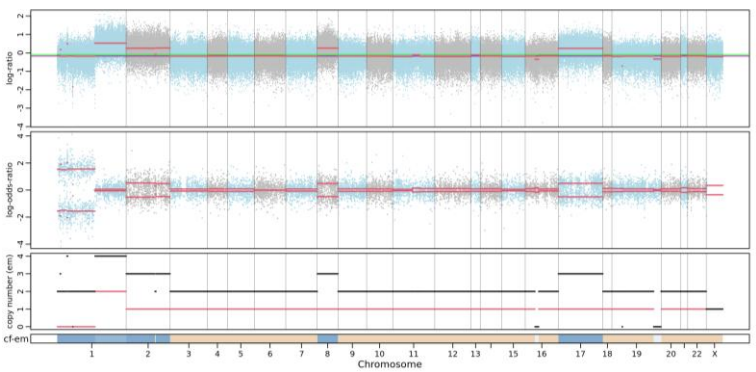
TLTX36.A17



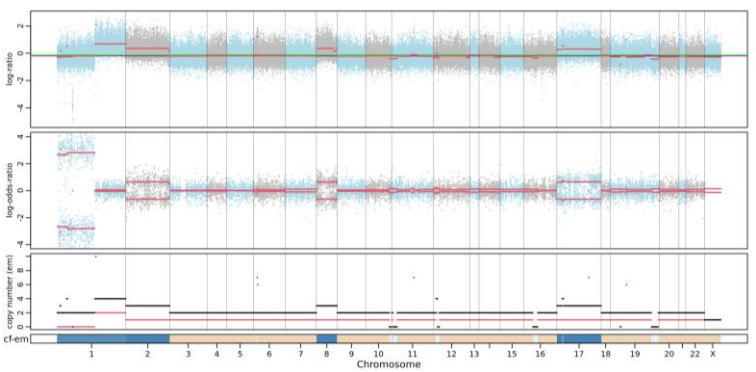
TLTX36.A18.B



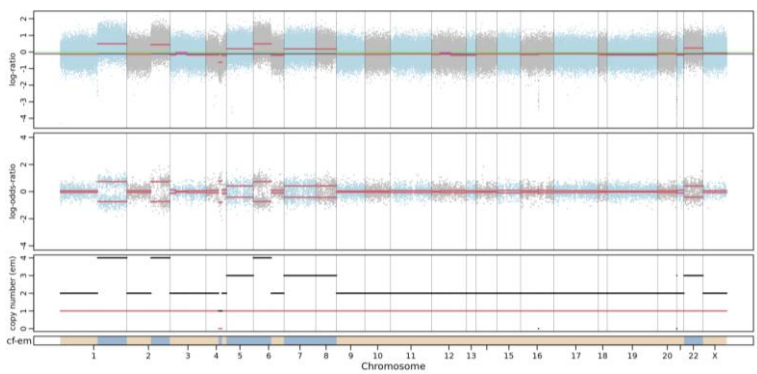
TLTX36.A11.A



TLTX36.A18.A



TLTX18.A9



TLTX33.A2

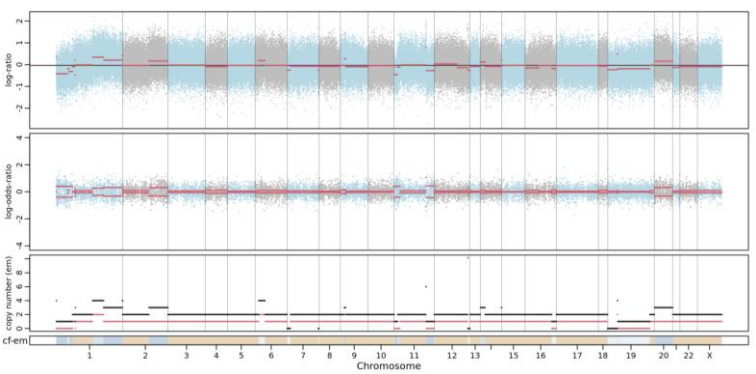
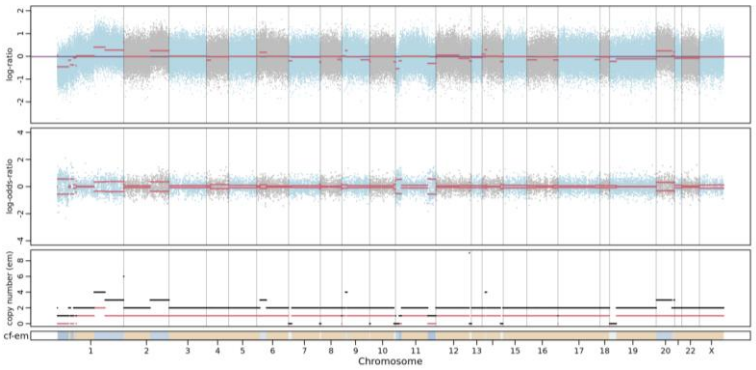
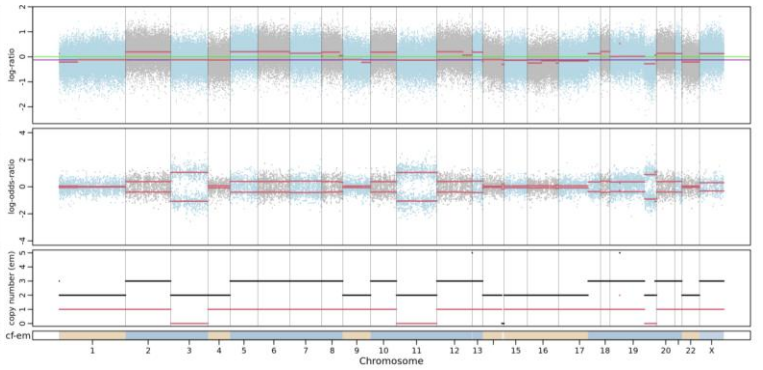


Figure S9

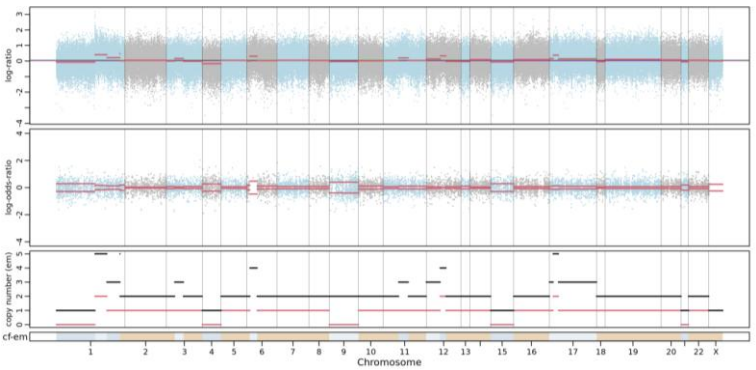
TLTX33.A12



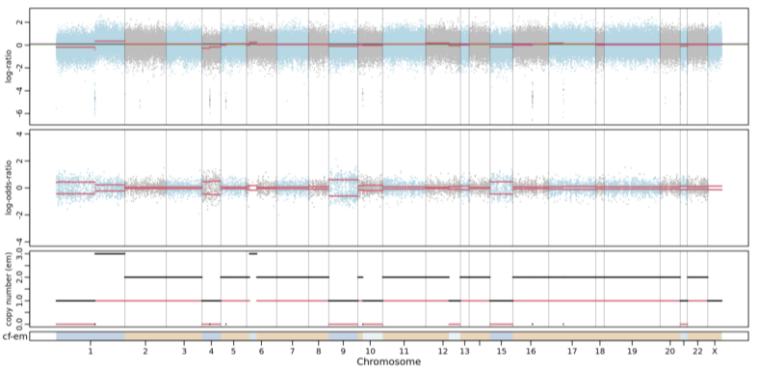
TLTX33.A15



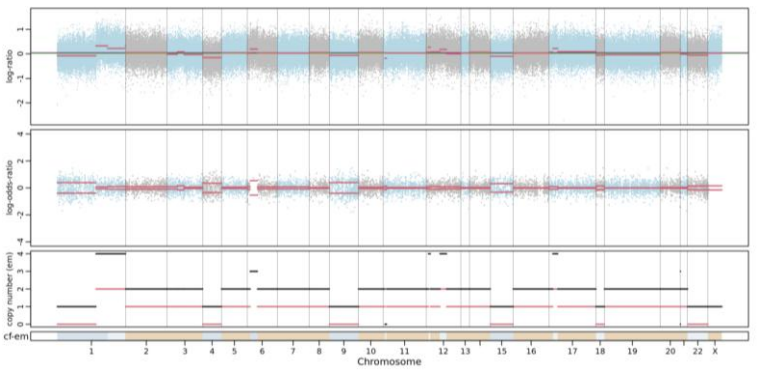
TLTX23.A1



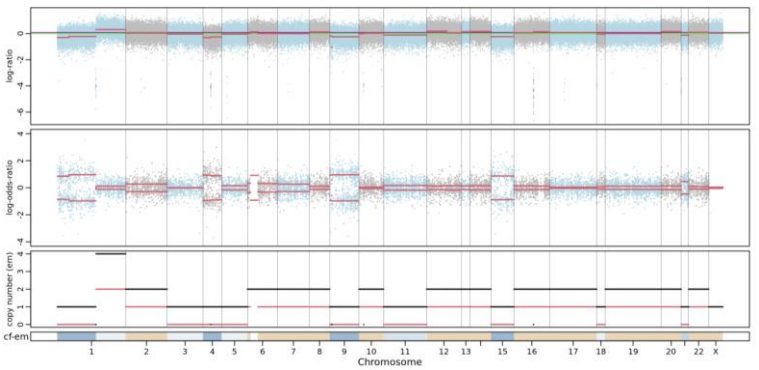
TLTX23.A6



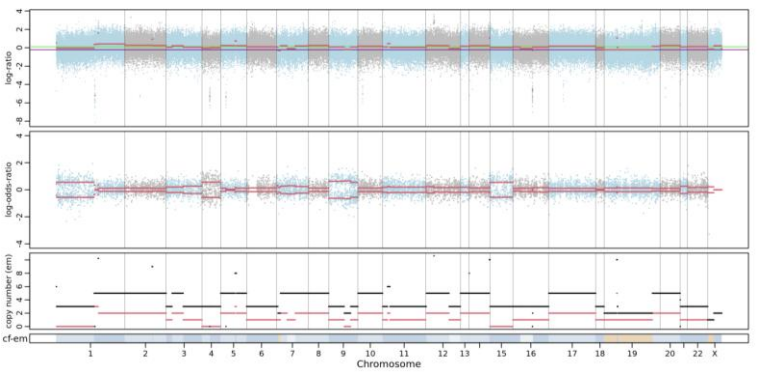
TLTX23.A7



TLTX23.A11



TLTX23.FT



TLTX23.biopsy

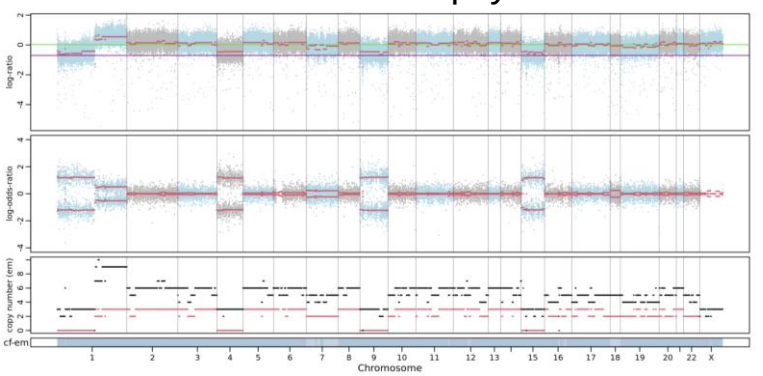
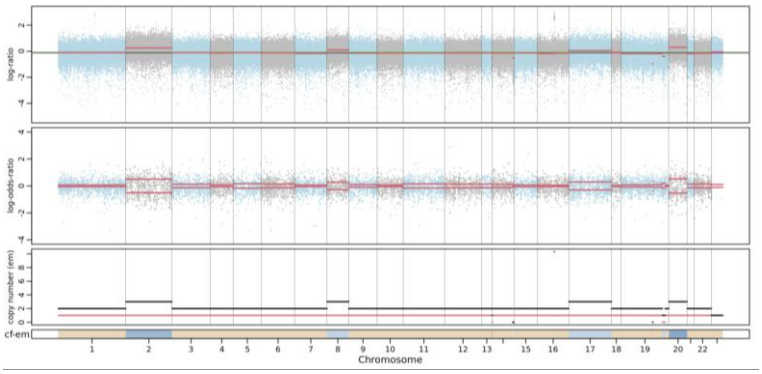
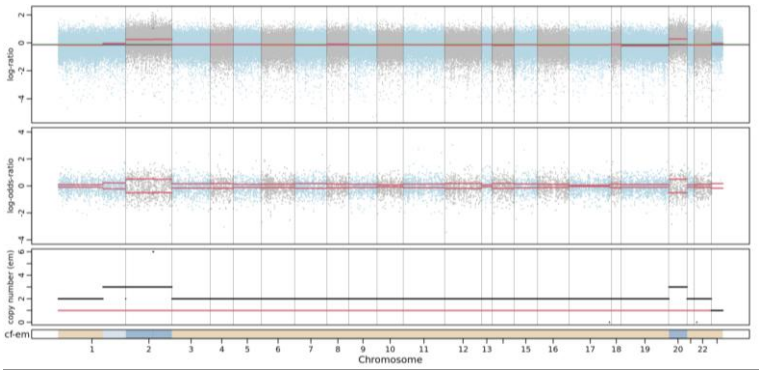


Figure S9

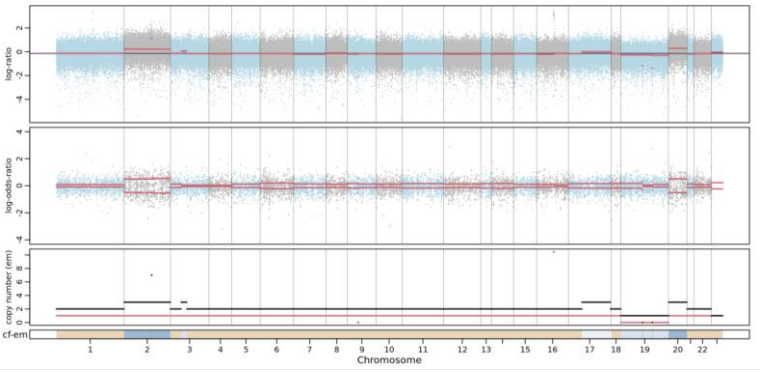
TLTX25.C3.A



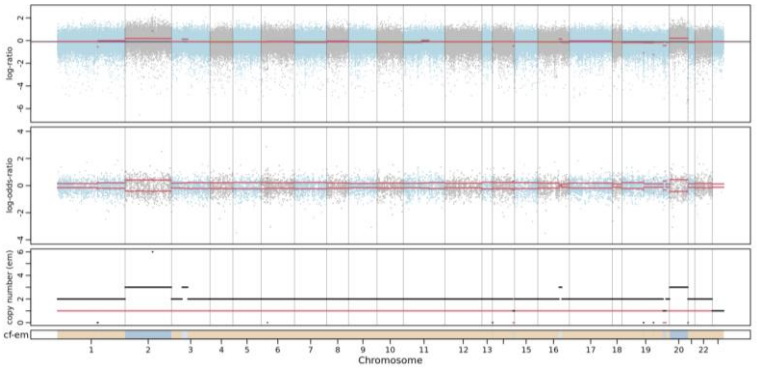
TLTX25.C3.B



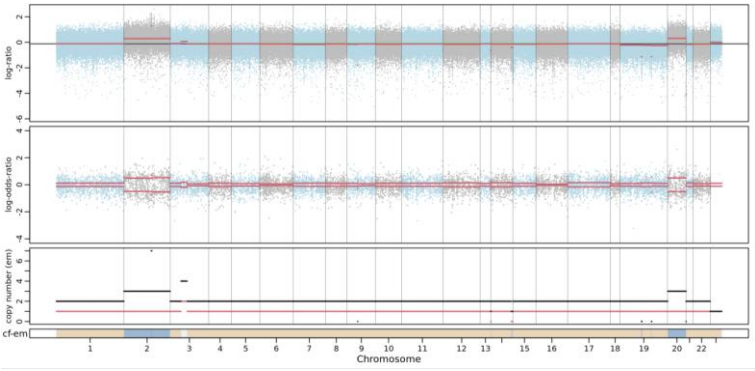
TLTX25.C2.A



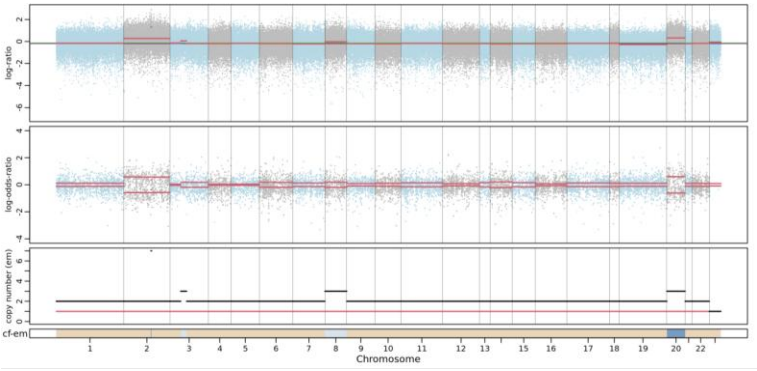
TLTX25.C2.B



TLTX25.C8



TLTX25.C10



TLTX25.FT

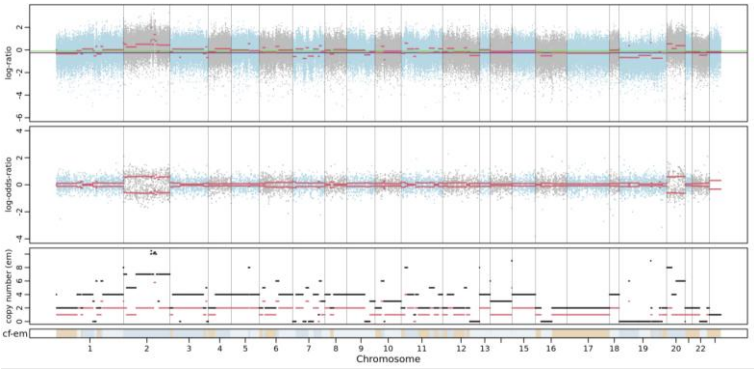


Figure S9