Diagnostic targeted sequencing panel for hepatocellular carcinoma genomic screening

Viola Paradiso[1#], Andrea Garofoli[1#], Nadia Tosti[1], Manuela Lanzafame[1], Valeria Perrina[1], Luca Quagliata[1], Matthias S. Matter[1], Stefan Wieland[2], Markus H. Heim[2,3], Salvatore Piscuoglio[1], Charlotte K. Y. Ng[1,2*] and Luigi M. Terracciano[1*]

[1]Institute of Pathology, University Hospital Basel, Basel, Switzerland;

[2]Department of Biomedicine, University of Basel, Basel, Switzerland;

[3]Department of Gastroenterology and Hepatology, University Hospital Basel, Switzerland.

# equally contributed as first authors

* equally contributed as senior authors

Text pages: 32 (including references and legends); **Word count:** 5,705**; Number of Tables:** 0**;** **Number of Figures:** 5

Running title: HCC panel for genomic screening

FINANCIAL SUPPORT: This study was funded in part by the Swiss Cancer League (Oncosuisse) KLS-3639-02-2015 to L.M.T. and KFS-3995-08-2016 to S.P. and in part by Krebsliga beider Basel KLbB-4183-03-2017 to C.K.Y.N.; S.P. is supported by Swiss National Science Foundation (Ambizione grant number PZ00P3_168165). V.Pa. is supported by the Swiss Centre for Applied Human Toxicology (SCAHT). C.K.Y.N. and M.H.H. acknowledge support from the European Research Council (ERC Synergy Grant No. 609883). Funding bodies had no role in the design of the study, collection, analysis and interpretation of the data or the writing of the manuscript.

Disclosures: None declared

Correspondence: Prof. Luigi M. Terracciano: Institute of Pathology, University Hospital Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland. Tel: +41612652849; Fax: +41612653194. E-mail: Luigi.Terracciano@usb.ch

ABSTRACT (204 words)

Commercially available targeted panels miss genomic regions frequently altered in hepatocellular carcinoma (HCC). We sought to design and benchmark a sequencing assay for genomic screening in HCC. We designed an AmpliSeq custom panel targeting all exons of 33 protein-coding and 2 long non-coding RNA genes frequently mutated in HCC, *TERT* promoter, and 9 genes with frequent copy number alterations (CNA). Using this panel, the profiling of DNA from fresh-frozen (n=10, 1495x) and/or formalin-fixed paraffin-embedded (FFPE) tumors with low-input DNA (n=36, 530x) from 39 HCCs identified at least one somatic mutation in 90% of the cases. Median of 2.5 (0-74) and 3 (0-76) mutations were identified in fresh-frozen and FFPE tumors, respectively. Benchmarked against the mutations identified from Illumina whole-exome sequencing (WES) of the corresponding fresh-frozen tumors (105x), 98% (61/62) and 100% (104/104) of the mutations from WES were detected in the 10 fresh-frozen tumors and the 36 FFPE tumors, respectively, using the HCC panel. Additionally, we identified 18 and 70 somatic mutations in coding and non-coding genes, respectively, not found by WES using our HCC panel. CNAs between WES and our HCC panel showed an overall concordance of 86%. In conclusion, we established a cost-effective assay for the detection of genomic alterations in HCC.

Keywords: Hepatocellular carcinoma; somatic mutation; copy number alteration; targeted sequencing.

INTRODUCTION

Sequencing technologies have allowed the discovery of genetic alterations essential in the diagnosis and treatment of human cancer or approval of new targeted therapies.[1] Additionally, the presence of subclonal mutations has direct implications in the development of drug resistance.[2, 3] In the era of precision medicine, the development of rapid, accurate, high-throughput and cost-effective genomic assays to accommodate the increasingly genotype-based therapeutic approaches is required.[4, 5] Currently, the costs of whole-genome and whole-exome sequencing (WES) are still prohibitive in the clinical setting, especially for small institutions. Furthermore, while DNA from fresh-frozen tissue is ideal for genomic screening, it is not part of routine diagnostic practice at most hospitals and institutions. Instead, DNA from formalin-fixed paraffin-embedded (FFPE) material is frequently the only option. Moreover, DNA from small tumors, after reserving materials for histopathologic analyses, may be extremely limited. For research institutes, being able to exploit and re-visit archival materials associated with long-term follow-up but whose DNA may potentially be degraded is also highly desirable. Given these limitations, PCR-based sequencing panels may be more broadly applicable than capture-based solutions.

Existing commercial sequencing panels, such as the amplicon-based Ion Torrent Oncomine Comprehensive Assay® v3 (Thermo Fisher Scientific, MA, USA) and the capture-based Foundation Medicine FoundationOne assay, are broadly applicable to common cancer types. Compared to other common cancer types, however, hepatocellular carcinoma (HCC) has a distinct mutational profile. While HCC driver genes *TP53* and *CTNNB1* are also frequently mutated in cancers such as those of the lungs, the breasts and colon,[6] genes such as *APOB*, *ALB*, *HNF1A*, *HNF4A* are significantly mutated only in HCC.[7-17] The distinct mutational landscape of HCC is likely a result of the unique biology of hepatocyte differentiation and liver functions. Importantly, the frequently altered *APOB*, *ALB* and *HNF4A* are not targeted by most commercial assays. In the non-coding regions, recent commercially available panels include *TERT* promoter mutation hotspot (c.-124C>T). However, long non-coding RNA (lncRNA) genes frequently mutated in HCC, such *MALAT1* and *NEAT1,*[16] have yet to be included in commercial panels or in exome capture panels.

Recent whole-genome studies have also uncovered mutation clusters in promoter regions of genes such as *MED16*, *WDR74* and *TFPI2*[16, 18] that are not covered in commercial panels.

In this study, we designed a high-throughput and cost-effective amplicon-based sequencing panel specifically to screen for somatic mutations and copy number alterations (CNAs) in HCC. Our panel includes genes and regions frequently altered in HCC, including those not currently covered by commercial panels. We tested the sequencing panel using fresh-frozen and FFPE materials with low-input DNA to evaluate the feasibility of this panel in routine diagnostics.

MATERIALS AND METHODS

Targeted panel design and generation

A custom targeted sequencing panel focusing on the most frequently altered genes in HCC[7-18] was designed using Ion Ampliseq Designer (Thermo Fisher Scientific). The panel (hereafter the "HCC panel") covers all exons of 33 protein-coding genes, recurrently mutated lncRNA genes *MALAT1* and *NEAT1* and the recurrently mutated promoter regions of *TERT, WDR74, MED16* and *TFPI2* (Figure 1A and Supplementary Table S1).[7-18] Nine genes frequently altered by copy number alterations (CNAs) as well as mutation hotspots in seven cancer genes are also covered (Figure 1A and Supplementary Table S1).[7-18] The HCC panel was designed using the FFPE option for smaller amplicon size. The nine genes for CNA profiling were designed to be covered by at least 10 non-overlapping amplicons evenly distributed across the length of the genes. The designed panel was further inspected by the white glove service (Thermo Fisher Scientific) for primer specificity in a multiplex PCR reaction. The HCC panel consists of 2120 amplicons split into two primer pools and covers genomic regions of ~203kb.

Tissue samples

Human tissues were obtained from patients undergoing diagnostic liver biopsy at the University Hospital Basel, Basel, Switzerland. Written informed consent was obtained from all included patients. Ultrasound-guided needle biopsies were obtained from tumor lesion(s) and adjacent non-

tumoral liver tissue (Figure 1B). The study was approved by the Ethics committee of the north-western part of Switzerland (Protocol Number EKNZ 2014-099). For all patients except cases 2, 6, 7 and 9, a single tumor biopsy was included (Supplementary Table S2). For cases 6 and 7, two tumor biopsies were included, and for cases 2 and 9, three tumor biopsies were included. A portion of each biopsy was formalin-fixed paraffin-embedded for clinical purposes and the remaining portion of each biopsy was snap-frozen and stored at -80° for research purposes. For this study, 45 fresh-frozen tumor biopsies and 39 fresh-frozen non-tumor biopsies from 39 patients were included. FFPE tissue samples that remained after diagnostic routine (36 tumor biopsies and 31 non-tumor biopsies from 36 patients) were included. Pathologic assessment of tumor content was performed by two expert hepatopathologists (M.S.M. and L.M.T.) using diagnostic hematoxylin-and-eosin slides.

DNA extraction

DNA from fresh-frozen biopsies was extracted using the ZR-Duet DNA/RNA MiniPrep Plus kit (Zymo Research, CA, USA) following the manufacturer's instructions. Prior to extraction, tissue samples were crushed in liquid nitrogen to facilitate lysis. For DNA extraction from FFPE samples, one 5μm-thick slide was cut directly in the tube and DNA extracted with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions as previously described.[19, 20] DNA was quantified using the Qubit Fluorometer (Thermo Fisher Scientific).

Library preparation and deep sequencing using the HCC panel

Library preparation for the HCC panel was performed using the Ion AmpliSeq library kit 2.0 (Thermo Fisher Scientific) according to the manufacturer's guidelines. For cases 2, 6, 7, and 9, DNA extracted from multiple fresh-frozen tumor biopsies was pooled equimolar prior to library preparation (Supplementary Table S2). In total, 20 fresh-frozen samples (10 tumor samples and 10 non-tumoral counterparts) and 67 FFPE samples (36 tumor biopsies and 31 non-tumoral counterparts) were sequenced using the HCC panel.

The HCC panel consists of two pools of amplification primers. 10ng of DNA per sample was used

for library preparation for each pool. Amplification was performed according to the manufacturer's guidelines. The amplicons from the two pools were combined and treated to digest the primers and to phosphorylate the amplicons. The amplicons were then ligated to Ion Adapters (Thermo Fisher Scientific) using DNA ligase. Finally, cleaning and purification of the generated libraries were performed with Agencourt AMPure XP (Beckman Coulter, CA, USA) according to the manufacturer's guidelines. Quantification and quality control were performed with Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific). Samples were diluted to reach the concentration of 40pmol and then were pooled for sequencing. 25µl of the pooled libraries was loaded on Ion 530 Chip (Thermo Fisher Scientific) and processed in Ion Chef Instrument (Thermo Fisher Scientific). Sequencing was performed on Ion S5 XL system (Thermo Fisher Scientific).

Sequence data analysis for the HCC panel

Sequence reads were aligned to the human reference genome hg19 using TMAP within the Torrent Suite Software (v5.4) for the Ion S5XL system. Coverage analysis was performed using Picard's CollectTargetedPcrMetrics tool (http://broadinstitute.github.io/picard/, v2.4.1, Supplementary Table S3). Uniformity of sequencing was defined as the proportion of target bases covered at >20% of mean amplicon coverage for a given sample. Comparison of the coverage for the two primer pools was performed using paired Wilcoxon test.

Somatic mutations were identified using Torrent Variant Caller (v5.0.3, Thermo Fisher Scientific). For fresh-frozen samples, the corresponding fresh-frozen non-tumoral samples were used as the germline control. For FFPE samples, FFPE non-tumoral samples were used as the matched germline sample where available. Where FFPE non-tumoral samples were not available, the corresponding fresh-frozen non-tumoral samples were used as germline control. Mutations at hotspot residues were white-listed.[21, 22] We filtered out mutations supported by <8 reads, and/or those covered by <10 reads in the tumor or <10 reads in the matched non-tumoral counterpart. Only those for which the tumor variant allele fraction (VAF) was >10 times that of the matched non-tumoral VAF were retained to ensure the somatic nature of the variants. Due to the repetitive nature

and the high GC content of the *TERT* promoter region, *TERT* mutation hotspots (chr5:1295228 and chr5:1295250) were additionally screened. *TERT* promoter mutations were considered present if supported by at least 5 reads or variant allele fraction of at least 5%. All mutations were manually inspected using the Integrative Genomics Viewer (v2.3.69).[23]

CNAs were defined as follows. For each sample, end-to-end sequence reads were extracted separately for the two amplicon pools. A copy number reference for each pool was generated using all non-tumoral samples to estimate overall read depth, $\log_2$ ratio and variability using the 'reference' function from CNVkit (v0.9.0).[24] Amplicons with <100 read depth, absolute $\log_2$ ratio >1.5 or spread >1 were removed from copy number analysis. Protein-coding genes for which the complete coding region was included in the panel or for which amplicons were specifically designed for copy number analysis were included. Samples with excessive residual copy number $\log_2$ ratio (segment interquartile range >0.8) were excluded, as previously described.[25]

For each tumor/non-tumor pairs, $\log_2$ ratio was computed for each amplicon, separately for the two amplicon pools using Varscan2 (v2.4.3).[26] Log_2 ratios for the two pools were separately centered then merged for segmentation using circular binary segmentation.[27] CNAs were determined adopting a previously described approach.[20] In brief, standard deviation (SD) of the $\log_2$ ratios of the 40% of the central positions ordered by their $\log_2$ ratios was computed. Copy number gains and amplifications/ high gains were defined as +2SDs and +6SDs, respectively. Copy number losses and deep deletions were defined as -2.5SDs and -7SDs, respectively. All gene amplifications and deep deletions were visually inspected using $\log_2$ ratio plots.

To evaluate the impact of tumor purity on CNA analysis, we performed an *in silico* simulation on 12 cases (6 frozen and 6 FFPE, selected on the basis of the presence of gene amplification/ high gain or deep deletion), by replacing tumor reads with reads sampled from the normal samples to simulate tumor content 5%, 10%, 20% up to the actual tumor content for the samples. CNA analysis was performed as described above.

Whole exome sequencing (WES)

WES was performed for DNA extracted from the 45 tumor biopsies and 39 non-tumoral counterparts from the 39 patients (Supplementary Table S2). Whole exome capture was performed using the SureSelectXT Clinical Research Exome (Agilent, CA, USA) platform according to the manufacturer's guidelines. Sequencing (2x101bp) was performed at the Genomics Facility of ETH Zurich Department of Biosystems Science and Engineering (Basel, Switzerland) using Illumina HiSeq 2500 (Illumina, CA, USA) according to the manufacturer's guidelines. Sequence reads were aligned to the reference human genome GRCh37 using Burrows-Wheeler Aligner-MEM (v0.7.12).[28] Local realignment, duplicate removal and base quality adjustment were performed using the Genome Analysis Toolkit (v3.6)[29] and Picard (http://broadinstitute.github.io/picard/, v2.4.1).

For WES samples, sequence reads overlapping with the target regions of the HCC panel were extracted for further comparative analyses. Sequencing statistics were evaluated for the overlap of the target regions of the WES and the HCC panel. In addition, for cases 2, 6, 7, and 9, for which DNA from multiple fresh-frozen tumor biopsies was pooled prior to sequencing using the HCC panel, WES reads from the multiple biopsies were merged to facilitate downstream comparisons. For all four cases, the number of reads obtained from WES of individual biopsies was comparable (Supplementary Table S3).

Somatic single nucleotide variants and small insertions and deletions were detected using MuTect (v1.1.4)[30] and Strelka (v1.0.15)[31], respectively. We filtered out single nucleotide variants, and small insertions and deletions outside of the target regions, those with variant allelic fraction of <1% and/or those supported by <3 reads. We only retained variants for which the tumor VAF was >5 times that of the matched non-tumoral VAF. We further excluded variants identified in at least two of a panel of 123 non-tumoral liver tissue samples, including the 39 non-tumoral samples in the current study, captured and sequenced using the same protocols using the artifact detection mode of MuTect2 implemented in Genome Analysis Toolkit (v3.6).[29] All indels were manually inspected

using the Integrative Genomics Viewer [23]. Copy number analysis was performed using FACETS (v0.5.13),[32] and genes targeted by amplifications or deep deletions defined using the same thresholds as above.

Pairwise comparisons between mutations identified by whole exome sequencing, fresh frozen and formalin-fixed paraffin-embedded tissues

Pairwise comparisons of the somatic mutations identified by WES and by the HCC panel were performed, according to the originating biopsies (Supplementary Table S2). Discordant variants were re-evaluated and interrogated for their presence by supplying Torrent Variant Caller (v5.0.3) with their positions as the 'hotspot list' (for Ion Torrent sequencing) or by Genome Analysis Toolkit (v3.6) Unified Genotyper using the GENOTYPE_GIVEN_ALLELES mode.

Sanger sequencing

To validate the discordant variants, Sanger sequencing was performed on both DNA from the fresh-frozen and the corresponding FFPE tumor biopsies. PCR amplification of 5ng of genomic DNA was performed using the AmpliTaq 360 Master Mix Kit (Thermo Fisher Scientific) on a Veriti Thermal Cycler (Thermo Fisher Scientific) as previously described[20] (Supplementary Table S4). PCR fragments were purified with ExoSAP-IT (Thermo Fisher Scientific). Sequencing reactions were performed on a 3500 Series Genetic Analyzer instrument using the ABI BigDye Terminator chemistry (v3.1, Thermo Fisher Scientific) according to manufacturer's instructions. All analyses were performed in duplicate. Sequences of the forward and reverse strands were analyzed using MacVector software (MacVector, Inc, MA, USA).[20]

Analysis of The Cancer Genome Atlas (TCGA) data

To determine the frequencies of high-level copy number gains/ focal amplifications, and deep deletions/ focal homozygous deletions in HCC, we obtained the GISTIC 2.0 copy number calls for the TCGA HCC cohort from the cBioPortal.[33] High-level gains and deep deletions were defined as those with GISTIC copy number state 2 and -2, respectively. Focal amplifications and focal

homozygous deletions were defined as high-level gains and deep deletions that affected <25% of a given chromosome arm. For the 37 genes included in the copy number analysis, we computed the frequencies of high-level gains/deep deletions and of focal amplifications/focal homozygous deletions.

Statistical analysis

Correlation analyses were performed using Pearson's r and r^2. Statistical analyses were performed in R (v3.4.2).

RESULTS

HCC-specific custom targeted sequencing panel design and quality assessment.

We designed an HCC sequencing panel specifically targeting genes and genomic regions frequently altered in HCC[7-18] (Figure 1A and Supplementary Table S1). The HCC panel consists of complete coding regions of 33 genes involved in several pathways implicated in HCC pathogenesis, including the WNT pathway (*CTNNB1, AXIN1*), chromatin remodelling (*ARID1A*, *ARID2* and *BAP1*), cell cycle regulation (*CDKN1A*, *CDKN2A*, *CDKN2B*, *CCND1*, *RPS6KA3*, *RB1* and *TP53*), inflammatory response (*IL6R*, *IL6ST*) and hepatocyte differentiation (*ALB*, *APOB*, *HNF1A, HNF4A*). Additionally, the HCC panel also targets recurrently mutated lncRNA genes *MALAT1* and *NEAT1* and recurrently mutated promoter regions of *TERT*, *WDR74*, *MED16* and *TFPI2.* Genes frequently altered by copy number alterations (CNAs, e.g. *CCNE1*, *VEGFA*, *TERT*), and mutation hotspots in *BRAF*, *EEF1A1*, *HRAS*, *IL6ST*, *KRAS*, *NRAS* and *PIK3CA* are also targeted. To enable the efficient profiling of DNA samples derived from potentially degraded FFPE materials, the panel was designed using the FFPE option for smaller amplicon size, with a mean amplicon size of 118bp (range 63bp-252bp, Figure 2A). We tested the HCC panel on the DNA extracted from 20 fresh-frozen samples (10 from tumor biopsies and 10 from non-tumoral counterparts) and 67 FFPE samples (36 from tumor biopsies and 31 from non-tumoral counterparts) obtained from 39 patients (Figure 1B and Supplementary Table S2).

We first performed a coverage analysis of the HCC panel using the 10 fresh-frozen and 31 FFPE non-tumoral DNA samples. In the fresh-frozen and FFPE non-tumoral DNA samples, we achieved a mean coverage of 1478x (range 925x-2420x) and 580x (range 263x-1300x), respectively (Figure 2B and Supplementary Table S3). There was no difference between the depth of coverage of the two pools of amplicons (P=0.9879, paired Wilcoxon test, Supplementary Figure S1A). At least 96.8% and 91.1% of the amplicons were covered at >30x and at least 98.7% and 95.6% of the amplicons were covered at >10x in the fresh-frozen and FFPE non-tumor samples (Figure 2C and Supplementary Figure S1B). Median uniformity (defined as the proportion of target bases covered at >20% of the mean amplixcon coverage of a given sample) was 89.9% (range 86.8%-91.5%) in the fresh-frozen samples and 89.0% (range 73.3%-92.3%) in the FFPE samples (Figure 2D). As expected, depth of sequencing of the amplicons was associated with GC content, with reduced depth at extreme GC content (Figure 2E).

HCC panel captured somatic mutations concordant with WES and identified additional mutations

Next, we evaluated the somatic mutations identified using the 10 fresh-frozen tumor-non-tumoral pairs sequenced using the HCC panel. We achieved a median sequencing depth of 1495x (range 1026x-1855x) in the tumor samples (Figure 2B, Supplementary Table S3). A median of 2.5 (range 0-74) somatic mutations were identified, including a median of 2 (range 0-52) mutations in protein-coding genes (Figure 3A and Supplementary Table S4). No somatic mutations were identified for 2/10 cases (cases 3 and 12), although both cases had ≥50% tumor cell content (Supplementary Table S2). One case (case 9) exhibited a hypermutator phenotype with 74 somatic mutations identified.

To evaluate the somatic mutations defined using the HCC panel, we used the somatic mutations derived from whole-exome sequencing (WES) using the orthogonal Illumina technology of the same DNA aliquots from the fresh-frozen tumors and matched non-tumor samples as our benchmark (Figure 1B). Considering only the coding regions covered by our HCC panel, the median depths of

WES was 114x (range 92x-345x) and 51x (range 45x-84x) in the fresh-frozen tumors and matched non-tumor samples, respectively (Supplementary Table S3). WES analysis confirmed that no mutations were present within the targeted protein coding regions in cases 3 and 12 and that case 9 was hypermutated (Figure 3B). Of the 62 mutations in the coding region identified from WES analysis, 61 (98%) were also called by our HCC panel analysis (Figure 3B). One *NRAS* Q61K hotspot mutation (case 6) was missed using our HCC panel. Manual review of this position revealed that the mutation had variant allele fraction of 2.5% by WES and 2.0% by the HCC panel (Supplementary Figure S2 and Supplementary Table S4). It should, however, be noted that 2% is very close to the detection limit of the current sequencing technologies.

Compared to the WES analysis, our HCC panel analysis revealed an additional six mutations in the coding regions, including five in case 9 and one in case 11 (Figure 3B). Manual review of the WES data revealed that all six mutations were in fact supported by at least one read in WES, but those positions were covered at reduced depth, with 4/6 covered by ≤40 reads (including 3 in *LRP1B*) and 5/6 ≤80 reads (Supplementary Figure S2C and Supplementary Table S4). This suggests that the increased sensitivity in our HCC panel analysis is likely due to the increased depth achieved.

Additional to the mutations in the protein coding regions, our HCC panel also targeted the lncRNA genes *MALAT1* and *NEAT1*, as well as the promoter regions of *TERT*, *WDR74*, *MED16* and *TFPI2* (Figure 1A). Within these non-coding regions, we identified an additional 32 mutations across the 10 cases, representing a 48% gain of information compared to sequencing the protein coding genes alone (Figure 3B). *TERT* promoter mutations were found in 60% (6/10) of cases and 16 somatic mutations in the lncRNA gene *NEAT1* were identified in 40% (4/10) of cases (Figure 3B and Supplementary Table S4).

Taken together, for the protein coding genes frequently mutated in HCC, our HCC panel analysis produced highly reliable results compared to WES. Given the increased sequencing depth achieved using the HCC panel, we identified somatic mutations that were missed by WES. Importantly, our

HCC panel analysis enabled us to identify somatic mutations in promoter regions and frequently mutated lncRNA genes.

HCC panel analysis identified somatic mutations in FFPE diagnostic biopsies with low input DNA

Nucleic acids from diagnostic specimens are frequently derived from small FFPE samples. Therefore, we sought to determine whether our HCC panel could also be used for somatic mutational screening using low-input DNA (20ng) extracted from FFPE samples. We subjected the DNA extracted from 36 diagnostic FFPE tumor biopsies to HCC panel sequencing to a median depth of 530x (range 192x-1257x, Figures 1A and 2B-C, Supplementary Table S3). The median tumor content for these 36 cases was 90% (range 5%-100%, Supplementary Table S2), thus representative of the distribution of tumor content in diagnostic samples in clinical practice. We identified a median of 3 mutations (range 0-76) per sample, including a median of 2 (range 0-53) mutations in the coding regions (Figure 4, Supplementary Figure S3 and Supplementary Table S4). No somatic mutations were identified for 8% (3/36) of cases (cases 7, 12 and 37), indicating that at least one somatic mutation could be detected in 92% of HCC diagnostic samples. Of note, while we were unable to detect somatic mutations in the one biopsy with 5% tumor content, we were able to detect somatic alterations in samples with 30%-40% tumor content.

We compared the mutations identified in protein-coding genes from these 36 FFPE diagnostic biopsies to those identified by WES of the DNA from the corresponding fresh-frozen biopsies. All 104 mutations identified from WES analysis were also called based on our HCC panel analysis (Figure 4 and Supplementary Figure S3), with 21/36 (58%) of our cases harboring *CTNNB1* mutations, a higher proportion than the TCGA and other HCC cohorts that is likely due to the higher percentage of alcohol-associated HCC (Supplementary Tables S1 and S2).[15] Additionally, we identified 18 mutations in the coding regions that were not found in the WES analysis in 11 cases. Of these 18, 13 were evident in WES but were not identified as mutations in the WES analysis, predominantly due to low sequencing depth (Supplementary Figures S2D and S3). The remaining

five mutations were verified to be present in the corresponding FFPE samples but absent in the fresh-frozen samples by Sanger sequencing (Supplementary Figure S4 and Supplementary Table S4), indicating that they were genuine discordances between the fresh-frozen and FFPE DNA and not false positive calls from the HCC panel assay. Of note, 2/5 mutations validated to be absent from the fresh-frozen DNA affected mutation hotspots in *CTNNB1* (p.Asp32Asn and p.Ser45Ala, Figure 4 and Supplementary Figure S4). The increased number of detected mutations by our HCC panel analysis was likely due to a combination of intra-tumor heterogeneity and the higher sequencing depth achieved.

Considering the 36 FFPE diagnostic biopsies, our HCC panel identified 70 somatic mutations in lncRNA genes and promoter regions, including 22 *TERT* promoter mutations (Figure 4 and Supplementary Table S4). Somatic mutations in lncRNA genes and promoter regions accounted for 37% of the total number of somatic mutations identified in the FFPE samples.

Compared to the very high correlation of VAF between the sequencing platforms used in the fresh-frozen samples ($r=0.89$, $r^2=0.79$, Pearson correlation), the correlation between WES from fresh-frozen samples and HCC panel using FFPE samples was more modest ($r=0.67$, $r^2=0.45$, Pearson correlation, Supplementary Figure S2A-B). We observed that mutations with large deviations in VAFs between the sequencing platforms used in the fresh-frozen samples tended to be covered at reduced depths on either platform (Supplementary Figure S2C). Similar observations could be made between VAFs of exome (fresh-frozen) and HCC panel (FFPE, Supplementary Figure S2D). The deviations in the latter may be more noticeable by the overall lower depth achieved in the FFPE samples compared to the HCC panel sequencing of the fresh-frozen samples. Intra-tumor heterogeneity between the fresh-frozen and FFPE aliquots likely contributed to the reduced correlation.

Taken together these results suggest that our HCC panel analysis has high specificity and sensitivity in somatic mutation detection. Furthermore, somatic mutations in promoter regions (*TERT* promoter) and lncRNA genes (*MALAT1* and *NEAT1*) highly mutated in HCC can also be detected.

Copy number analysis of the HCC panel reveals high concordance with WES

We sought to determine whether our HCC panel could also be used to detect CNAs. Of the genes targeted on the panel, we evaluated our ability to detect CNAs in 42 genes (complete coding regions covered and genes with amplicons tiled across the length of the genes for CNA detection, Figure 1A and Supplementary Table S1). Using the 41 non-tumoral samples, we assessed the variability of the depth of coverage in the amplicons targeting the 42 genes (Methods). After removing amplicons with low depth of coverage or high variability, 1,483 amplicons were used for CNA profiling. To assess our ability to evaluate per-gene CNA detection, we further paired each non-tumoral sample with two others randomly selected, gender-matched non-tumoral samples. We observed that the copy number $\log_2$ ratio of five genes, namely *LRP1B, ALB, BRD7, ACVR2A* and *IRF2*, was variable (SD>0.3) and therefore these genes were excluded from further CNA analyses. 37 genes were included in the CNA analysis.

We compared the copy number profiles of matched fresh-frozen tumor-non-tumor pairs and those derived from WES. Of the 10 fresh-frozen pairs sequenced using the HCC panel, one was excluded for excessive residual copy number $\log_2$ ratio (segment interquartile range >0.8).[25] For the nine evaluable samples, we found a correlation of r=0.80 (r^2=0.64) between the copy number $\log_2$ ratio of the two platforms (Figure 5A). When we compared the copy number profiles of the 34 evaluable FFPE tumors with the matched profiles from WES, we observed a correlation of r=0.73 (r^2=0.54) between the copy number $\log_2$ ratios (Figure 5A). Overall, 86% of the evaluable genes had concordant copy number states (Figure 5B).

It has previously been reported that tumor purity had an impact on the ability to make CNA calls.[25, 34] We therefore evaluated the impact of tumor purity on CNA analysis using an *in silico* simulation on 12 cases (6 fresh-frozen and 6 FFPE, selected on the basis of the presence of gene amplification/ high gain or deep deletion), by replacing tumor reads with reads sampled from the normal samples to simulate tumor content 5%, 10%, 20% up to the actual tumor content for the samples. We observed that amplifications/ high gains were readily detected at 5% tumor content in many cases and at 20% in all cases (Supplementary Figure S5). In our cohort, deep deletions could not be detected at tumor content <40%.

Taken together, our results demonstrate that, despite profiling only a small number of genes, our HCC panel is able to detect CNAs in genes frequently gained or lost in HCC in both fresh-frozen and FFPE tumor samples with low input DNA.

DISCUSSION

HCC has a distinct mutational landscape compared to the major tumor entities. Numerous genes have been found to be mutated frequently in HCC but rarely in other tumors, such as those important for hepatocyte differentiation (*ALB*, *APOB*, *HNF1A*, *HNF4A*) and inflammatory response (*IL6R*, *IL6ST*). Given the relative rarity of HCC, these genes are currently not targeted or are only partially targeted in commercial panels (e.g. Oncomine Comprehensive Panel v3®) and in panels used by sequencing services (e.g. FoundationOne assay, Supplementary Table S1). Thus, the currently available commercial assays for genomic profiling have suboptimal utility for HCC and a targeted sequencing panel specifically designed for HCC is warranted.

In this study, we designed a custom Ion Torrent AmpliSeq sequencing panel, targeting all exons of 33 protein-coding genes, two lncRNA genes, promoter regions of four genes previously found to be recurrently mutated in HCC, nine genes frequently affected by copy number alterations (CNAs), and mutation hotspots in seven cancer genes.[7-17] Importantly, a number of the genes targeted using our HCC panel are not currently on these two commercial panels. Of the 39 cases profiled with the HCC

panel (including both fresh-frozen and FFPE samples), we detected at least one somatic mutation in 90% (35/39) of cases. Of the mutations in coding genes found using our panel, 22% (42/189) would have missed by both Oncomine Comprehensive Panel v3® and the FoundationOne assay. Additionally, recent whole-genome studies of HCC have revealed frequent mutations in lncRNA genes *NEAT1* and *MALAT1*, both of which are not currently targeted by commercial panels. In fact, we found that around 1/3 of the mutations on the HCC panel were within the promoter and lncRNA regions.

We benchmarked our mutation screening and copy number profiling results from the HCC panel against those obtained from whole-exome sequencing (WES) by the orthogonal Illumina sequencing technology. We demonstrated that all but one mutation identified from WES were detected using our HCC panel. We identified an additional 10-15% of mutations within the coding regions. The majority of these additional mutations were in fact supported by few reads by WES, thus our increased sensitivity was likely a direct result of the increased sequencing depth of both the tumor and the matched normal samples achieved. Crucially, however, we found evidence of intra-tumor genetic heterogeneity between the adjacent fresh-frozen and FFPE biopsies, including two *CTNNB1* mutations, suggesting that in these cases, the *CTNNB1* mutations were not trunk mutations.

While CNA detection using capture-based methods has been successful for targeted sequencing panel of several hundred genes,[35] CNA detection using amplicon-based targeted sequencing has proven more difficult. A recent study investigated the use of an amplicon-based sequencing strategy targeting all exons of 113 genes related to DNA repair.[25] The authors demonstrated that, with an appropriate analysis strategy and quality control, amplicon-based sequencing strategy is feasible and cost-effective for CNA profiling in FFPE samples.[25] In the current study, the strategy of computing and centering the $\log_2$ ratios for the primer two pools separately, prior to merging and segmentation proved to be an effective strategy in resolving issues associated with variable amplification efficiencies, with 86% of the genes showing concordant copy number states.

Considering the few studies investigating the use of small targeted sequencing panel for CNA profiling, further benchmarking studies comparing analysis strategies and including larger sample size will likely improve the accuracies.

In the clinical setting, the quality, type and amount of input materials for genomic profiling are crucial considerations, particularly in light of the smaller tumors being detected in screening programs. Here we demonstrated that the HCC panel could be used for genomic screening with high sensitivity and specificity with very low input DNA (20ng) derived from FFPE samples without compromising the results. Although based on an analysis of the TCGA HCC cases, 92% and 85% of the cases would have exhibited at least 1 non-synonymous mutation using the FoundationOne and the Oncomine assays, respectively, our HCC panel holds the advantage of much lower input requirement than that required for commercial panels (e.g. >40micron tissue samples for the FoundationOne assay) and for capture-based targeted sequencing strategies.[35] We further demonstrated that somatic genetic alterations (somatic mutations and amplifications) could be detected from tumor samples with as low as 30% tumor content. Considering that we could not detect mutations in the one sample with 5% tumor content, we contend that 30% may be the lower limit of successful genomic profiling. Although lower limits (~20%) have also been reported,[36] we did not have the samples to verify this. The samples included in this study are *de facto* samples obtained from routine diagnostic practice and we demonstrated that our low input DNA requirement facilitates genomic profiling from very small biopsies.

Driver genetic alterations have not yet become a tangible tool in clinical decision making for the treatment of HCC, thus the immediate clinical application of our panel may be limited. However, recent studies have described the association of *TERT* promoter and *CTNNB1* exon 3 mutations with increased risk of malignant transformation of hepatocellular adenomas,[37, 38] more frequent *HNF1A* and *IL6ST* mutations in hepatocellular adenomas than HCCs,[37] as well as *TP53* mutation as a poor prognostic indicator in HCC.[39-41] These associations suggest a potential utility of genomic profiling in prognostication for hepatocellular adenomas and HCCs, in tissues or in even in cell-free

DNA.[41, 42] In terms of potential targetable alterations, three somatic mutations identified in our cohort of HCC are molecular targets in other cancer types according to OncoKB.[43] These include *ATM* loss of function mutation using olaparib in prostate cancer (level 4; biological evidence), *NRAS* hotspot mutation with binimetinib or in combination with ribociclib in melanoma (level 3; clinical evidence) and *TSC2* mutation with everolimus in central nervous system cancer (level 2; standard of care).[43] Application of our panel in clinical decision may become feasible in the future.

This study has several limitations. Firstly, the targeted nature of our HCC panel means that copy number profiling is not genome-wide and is restricted to the genes included on the panel. Clinically, focal amplifications, compared to gains of chromosome arm, are more likely to be true driver genetic event and may be considered drug targets. The targeted nature of the HCC panel means it may be difficult to distinguish the two scenarios. However, a re-analysis of the TCGA data suggests that high-level gains of chr11q13.3 (encompassing *CCND1*, *FGF19*, *FGF3*, *FGF4*) are almost always focal amplifications (>93%), while 50-70% of high-level gains of *TERT* and *VEGFA* are focal amplifications (Supplementary Table S5). By contrast, high-level gains of chr1q (*SETDB1* and *IL6R*) and chr8q (*NCOA2*, *MYC* and *PTK2*) are frequently non-focal (<10%), consistent with the frequent high-level gain of entire arms of chr1q and chr8q.[44] For deletions, most deep deletions are focal deletions, including all deletions (100%) in *ARID2*, *AXIN1*, *CDKN2A/B*, *PTEN* and *TSC1/2*. These results suggest that for CNAs affecting some of the most promising drug targets on the HCC panel are frequently true focal CNAs. Secondly, given that we identified a median of 2-3 mutations per tumor, we would not be able to accurately define tumor mutational burden, a putative biomarker for response to immune therapy.[45] Thirdly, the HCC panel does not include unique molecular identifiers, which would be useful to assess library complexity, particularly for samples with low input DNA. We envisage that the addition of unique molecular identifiers would be particularly beneficial for the study of cell-free DNA from HCC patients.[41, 42] Fourthly, we designed the panel specific for HCC. Recent studies have revealed that mixed HCC/cholangiocarcinoma and cholangiocarcinoma have recurrent mutations in genes such as *IDH1/2,*[46] while *FRK* mutations decrease in frequency from hepatocellular adenoma to HCC.[37] These genes are not covered by the HCC panel. However,

as an amplicon-based sequencing panel, adding amplicons to include genes that may assist in the differential diagnosis of HCC is straightforward.

In conclusion, our study demonstrated that our HCC panel is a cost-effective strategy for mutation screening and copy number profiling for routine diagnostic HCC samples with low input DNA.

ACKNOWLEDGMENTS

None

AUTHOR'S CONTRIBUTION

S.P., C.K.Y.N., and L.M.T. conceived and supervised the study; L.Q, M.S.M., S.P, C.K.Y.N. and L.M.T. performed literature search and designed the sequencing panel; S.W and M.H.H. provided the samples and the whole exome sequencing data; V.Pa., N.T., V.Pe., M.L. and S.P. performed DNA extraction, library preparation and sequencing; A.G. and C.K.Y.N. developed the bioinformatic pipeline for mutation calling; V.Pa., A.G., S.P., C.K.Y.N. and L.M.T. analysed the results and wrote the manuscript.

REFERENCES

1. Chin L, Andersen JN, Futreal PA: Cancer genomics: from discovery science to personalized medicine. Nat Med 2011, 17:297-303.

2. Mok TS, Wu YL, Ahn MJ, Garassino MC, Kim HR, Ramalingam SS, Shepherd FA, He Y, Akamatsu H, Theelen WS, Lee CK, Sebastian M, Templeton A, Mann H, Marotti M, Ghiorghiu S, Papadimitrakopoulou VA, Investigators A: Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. N Engl J Med 2017, 376:629-640.

3. Toy W, Weir H, Razavi P, Lawson M, Goeppert AU, Mazzola AM, Smith A, Wilson J, Morrow C, Wong WL, De Stanchina E, Carlson KE, Martin TS, Uddin S, Li Z, Fanning S, Katzenellenbogen JA, Greene G, Baselga J, Chandarlapaty S: Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists. Cancer Discov 2017, 7:277-287.

4. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba, II, Varella-Garcia M, Franklin WA, Aronson SL, Su PF, Shyr Y, Camidge DR, Sequist LV, Glisson BS, Khuri FR, Garon EB, Pao W, Rudin C, Schiller J, Haura EB, Socinski M, Shirai K, Chen H, Giaccone G, Ladanyi M, Kugler K, Minna JD, Bunn PA: Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA 2014, 311:1998-2006.

5. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, Dogan S, Borsu L, Hameed M, Nafa K, Arcila ME, Ladanyi M, Berger MF: Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn 2015, 17:251-264.

568 6. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF,
569 Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ,
570 Wilson RK, Raphael BJ, Ding L: Mutational landscape and significance across 12 major
571 cancer types. Nature 2013, 502:333-339.

572 7. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, Bioulac-
573 Sage P, Letexier M, Degos F, Clement B, Balabaud C, Chevet E, Laurent A, Couchy G,
574 Letouze E, Calvo F, Zucman-Rossi J: Integrated analysis of somatic mutations and focal
575 copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat
576 Genet 2012, 44:694-698.

577 8. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo
578 M, Miya F, Arai Y, Takahashi H, Shirakihara T, Nagasaki M, Shibuya T, Nakano K,
579 Watanabe-Makino K, Tanaka H, Nakamura H, Kusuda J, Ojima H, Shimada K, Okusaka T,
580 Ueno M, Shigekawa Y, Kawakami Y, Arihiro K, Ohdan H, Gotoh K, Ishikawa O, Ariizumi S,
581 Yamamoto M, Yamada T, Chayama K, Kosuge T, et al.: Whole-genome sequencing of liver
582 cancers identifies etiological influences on mutation patterns and recurrent mutations in
583 chromatin regulators. Nat Genet 2012, 44:760-764.

584 9. Cleary SP, Jeck WR, Zhao X, Chen K, Selitsky SR, Savich GL, Tan TX, Wu MC, Getz G,
585 Lawrence MS, Parker JS, Li J, Powers S, Kim H, Fischer S, Guindi M, Ghanekar A, Chiang
586 DY: Identification of driver genes in hepatocellular carcinoma by exome sequencing.
587 Hepatology 2013, 58:1693-1702.

588 10. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, Hao K, Willard MD, Xu J,
589 Hauptschein R, Rejto PA, Fernandez J, Wang G, Zhang Q, Wang B, Chen R, Wang J, Lee
590 NP, Zhou W, Lin Z, Peng Z, Yi K, Chen S, Li L, Fan X, Yang J, Ye R, Ju J, Wang K, Estrella
591 H, Deng S, Wei P, Qiu M, Wulur IH, et al.: Whole-genome sequencing identifies recurrent
592 mutations in hepatocellular carcinoma. Genome Res 2013, 23:1422-1433.

11. Ahn SM, Jang SJ, Shim JH, Kim D, Hong SM, Sung CO, Baek D, Haq F, Ansari AA, Lee SY, Chun SM, Choi S, Choi HJ, Kim J, Kim S, Hwang S, Lee YJ, Lee JE, Jung WR, Jang HY, Yang E, Sung WK, Lee NP, Mao M, Lee C, Zucman-Rossi J, Yu E, Lee HC, Kong G: Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology 2014, 60:1972-1982.

12. Jhunjhunwala S, Jiang Z, Stawiski EW, Gnad F, Liu J, Mayba O, Du P, Diao J, Johnson S, Wong KF, Gao Z, Li Y, Wu TD, Kapadia SB, Modrusan Z, French DM, Luk JM, Seshagiri S, Zhang Z: Diverse modes of genomic alteration in hepatocellular carcinoma. Genome Biol 2014, 15:436.

13. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, Tsuji S, Donehower LA, Slagle BL, Nakamura H, Yamamoto S, Shinbrot E, Hama N, Lehmkuhl M, Hosoda F, Arai Y, Walker K, Dahdouli M, Gotoh K, Nagae G, Gingras MC, Muzny DM, Ojima H, Shimada K, Midorikawa Y, Goss JA, Cotton R, Hayashi A, Shibahara J, Ishikawa S, Guiteau J, Tanaka M, Urushidate T, Ohashi S, Okada N, et al.: Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet 2014, 46:1267-1273.

14. Shiraishi Y, Fujimoto A, Furuta M, Tanaka H, Chiba K, Boroevich KA, Abe T, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Shibuya T, Nakano K, Sasaki A, Maejima K, Kitada R, Hayami S, Shigekawa Y, Marubashi S, Yamada T, Kubo M, Ishikawa O, Aikata H, Arihiro K, Ohdan H, Yamamoto M, Yamaue H, Chayama K, Tsunoda T, Miyano S, Nakagawa H: Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. PLoS One 2014, 9:e114263.

15. Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F, Calatayud AL, Pinyol R, Pelletier L, Balabaud C, Laurent A, Blanc JF, Mazzaferro V, Calvo F, Villanueva A, Nault JC, Bioulac-Sage P, Stratton MR,

618 Llovet JM, Zucman-Rossi J: Exome sequencing of hepatocellular carcinomas identifies new
619 mutational signatures and potential therapeutic targets. Nat Genet 2015, 47:505-511.

620 16. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, Tanaka H, Taniguchi H,
621 Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Wardell CP, Hayami S, Nakamura T, Aikata H,
622 Arihiro K, Boroevich KA, Abe T, Nakano K, Maejima K, Sasaki-Oku A, Ohsawa A, Shibuya
623 T, Nakamura H, Hama N, Hosoda F, Arai Y, Ohashi S, Urushidate T, Nagae G, Yamamoto
624 S, Ueda H, Tatsuno K, Ojima H, et al.: Whole-genome mutational landscape and
625 characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016,
626 48:500-509.

627 17. Cancer Genome Atlas Research Network. Electronic address wbe, Cancer Genome Atlas
628 Research N: Comprehensive and Integrative Genomic Characterization of Hepatocellular
629 Carcinoma. Cell 2017, 169:1327-1341 e1323.

630 18. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W: Genome-wide analysis of noncoding
631 regulatory mutations in cancer. Nat Genet 2014, 46:1160-1165.

632 19. Ng CK, Piscuoglio S, Geyer FC, Burke KA, Pareja F, Eberle C, Lim R, Natrajan R, Riaz N,
633 Mariani O, Norton L, Vincent-Salomon A, Wen YH, Weigelt B, Reis-Filho JS: The Landscape
634 of Somatic Genetic Alterations in Metaplastic Breast Carcinomas. Clin Cancer Res 2017.

635 20. Piscuoglio S, Ng CK, Murray MP, Guerini-Rocco E, Martelotto LG, Geyer FC, Bidard FC,
636 Berman S, Fusco N, Sakr RA, Eberle CA, De Mattos-Arruda L, Macedo GS, Akram M,
637 Baslan T, Hicks JB, King TA, Brogi E, Norton L, Weigelt B, Hudis CA, Reis-Filho JS: The
638 Genomic Landscape of Male Breast Cancers. Clin Cancer Res 2016, 22:4045-4056.

639 21. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, Gao J, Socci ND, Solit
640 DB, Olshen AB, Schultz N, Taylor BS: Identifying recurrent mutations in cancer reveals
641 widespread lineage diversity and mutational specificity. Nat Biotechnol 2016, 34:155-163.

642 22. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, Zhang H, Solit DB,
643 Taylor BS, Schultz N, Sander C: 3D clusters of somatic mutations in cancer reveal
644 numerous rare mutations as functional targets. Genome Med 2017, 9:4.

645 23. Thorvaldsdottir H, Robinson JT, Mesirov JP: Integrative Genomics Viewer (IGV): high-
646 performance genomics data visualization and exploration. Brief Bioinform 2013, 14:178-192.

647 24. Talevich E, Shain AH, Botton T, Bastian BC: CNVkit: Genome-Wide Copy Number Detection
648 and Visualization from Targeted DNA Sequencing. PLoS Comput Biol 2016, 12:e1004873.

649 25. Seed G, Yuan W, Mateo J, Carreira S, Bertan C, Lambros M, Boysen G, Ferraldeschi R,
650 Miranda S, Figueiredo I, Riisnaes R, Crespo M, Rodrigues DN, Talevich E, Robinson DR,
651 Kunju LP, Wu YM, Lonigro R, Sandhu S, Chinnayan A, de Bono JS: Gene Copy Number
652 Estimation from Targeted Next-Generation Sequencing of Prostate Cancer Biopsies:
653 Analytic Validation and Clinical Qualification. Clin Cancer Res 2017, 23:6070-6077.

654 26. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding
655 L, Wilson RK: VarScan 2: somatic mutation and copy number alteration discovery in cancer
656 by exome sequencing. Genome Res 2012, 22:568-576.

657 27. Olshen AB, Venkatraman ES, Lucito R, Wigler M: Circular binary segmentation for the
658 analysis of array-based DNA copy number data. Biostatistics 2004, 5:557-572.

659 28. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform.
660 Bioinformatics 2009, 25:1754-1760.

29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010, 20:1297-1303.

30. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G: Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013, 31:213-219.

31. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK: Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012, 28:1811-1817.

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

33. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013, 6:pl1.

34. Grasso C, Butler T, Rhodes K, Quist M, Neff TL, Moore S, Tomlins SA, Reinig E, Beadling C, Andersen M, Corless CL: Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. J Mol Diagn 2015, 17:53-63.

35. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, Cheng DT, Chandramohan R, Mohanty AS, Ptashkin RN, Jayakumaran G, Prasad M, Syed MH, Rema AB, Liu ZY, Nafa K, Borsu L, Sadowska J, Casanova J, Bacares R, Kiecka IJ, et al.: Mutational landscape of

684 metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med
685 2017, 23:703-713.

686 36. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White
687 J, Sanford EM, An P, Sun J, Juhn F, Brennan K, Iwanik K, Maillet A, Buell J, White E, Zhao
688 M, Balasubramanian S, Terzic S, Richards T, Banning V, Garcia L, Mahoney K, Zwirko Z,
689 Donahue A, Beltran H, Mosquera JM, Rubin MA, Dogan S, Hedvat CV, Berger MF, Pusztai
690 L, Lechner M, Boshoff C, et al.: Development and validation of a clinical cancer genomic
691 profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013, 31:1023-
692 1031.

693 37. Pilati C, Letouze E, Nault JC, Imbeaud S, Boulai A, Calderaro J, Poussin K, Franconi A,
694 Couchy G, Morcrette G, Mallet M, Taouji S, Balabaud C, Terris B, Canal F, Paradis V,
695 Scoazec JY, de Muret A, Guettier C, Bioulac-Sage P, Chevet E, Calvo F, Zucman-Rossi J:
696 Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations
697 and the mechanisms of malignant transformation. Cancer Cell 2014, 25:428-441.

698 38. Nault JC, Couchy G, Balabaud C, Morcrette G, Caruso S, Blanc JF, Bacq Y, Calderaro J,
699 Paradis V, Ramos J, Scoazec JY, Gnemmi V, Sturm N, Guettier C, Fabre M, Savier E,
700 Chiche L, Labrune P, Selves J, Wendum D, Pilati C, Laurent A, De Muret A, Le Bail B,
701 Rebouissou S, Imbeaud S, Investigators G, Bioulac-Sage P, Letouze E, Zucman-Rossi J:
702 Molecular Classification of Hepatocellular Adenoma Associates With Risk Factors, Bleeding,
703 and Malignant Transformation. Gastroenterology 2017, 152:880-894 e886.

704 39. Goossens N, Sun X, Hoshida Y: Molecular classification of hepatocellular carcinoma:
705 potential therapeutic implications. Hepat Oncol 2015, 2:371-379.

706 40. Desert R, Rohart F, Canal F, Sicard M, Desille M, Renaud S, Turlin B, Bellaud P, Perret C,
707 Clement B, Le Cao KA, Musso O: Human hepatocellular carcinomas with a periportal

phenotype have the lowest potential for early recurrence after curative resection. Hepatology 2017, 66:1502-1518.

41. Kancherla V, Abdullazade S, Matter MS, Lanzafame M, Quagliata L, Roma G, Hoshida Y, Terracciano LM, Ng CKY, Piscuoglio S: Genomic Analysis Revealed New Oncogenic Signatures in TP53-Mutant Hepatocellular Carcinoma. Front Genet 2018, 9:2.

42. Ng CKY, Di Costanzo GG, Tosti N, Paradiso V, Coto-Llerena M, Roscigno G, Perrina V, Quintavalle C, Boldanova T, Wieland S, Marino-Marsilia G, Lanzafame M, Quagliata L, Condorelli G, Matter MS, Tortora R, Heim MH, Terracciano LM, Piscuoglio S: Genetic profiling using plasma-derived cell-free DNA in therapy-naive hepatocellular carcinoma patients: a pilot study. Ann Oncol 2018.

43. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, Rudolph JE, Yaeger R, Soumerai T, Nissan MH, Chang MT, Chandarlapaty S, Traina TA, Paik PK, Ho AL, Hantash FM, Grupe A, Baxi SS, Callahan MK, Snyder A, Chi P, Danila D, Gounder M, Harding JJ, Hellmann MD, Iyer G, Janjigian Y, Kaley T, Levine DA, Lowery M, Omuro A, Postow MA, Rathkopf D, Shoushtari AN, Shukla N, et al.: OncoKB: A Precision Oncology Knowledge Base. JCO Precis Oncol 2017, 2017.

44. Cancer Genome Atlas Research: Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell 2017, 169:1327-1341 e1323.

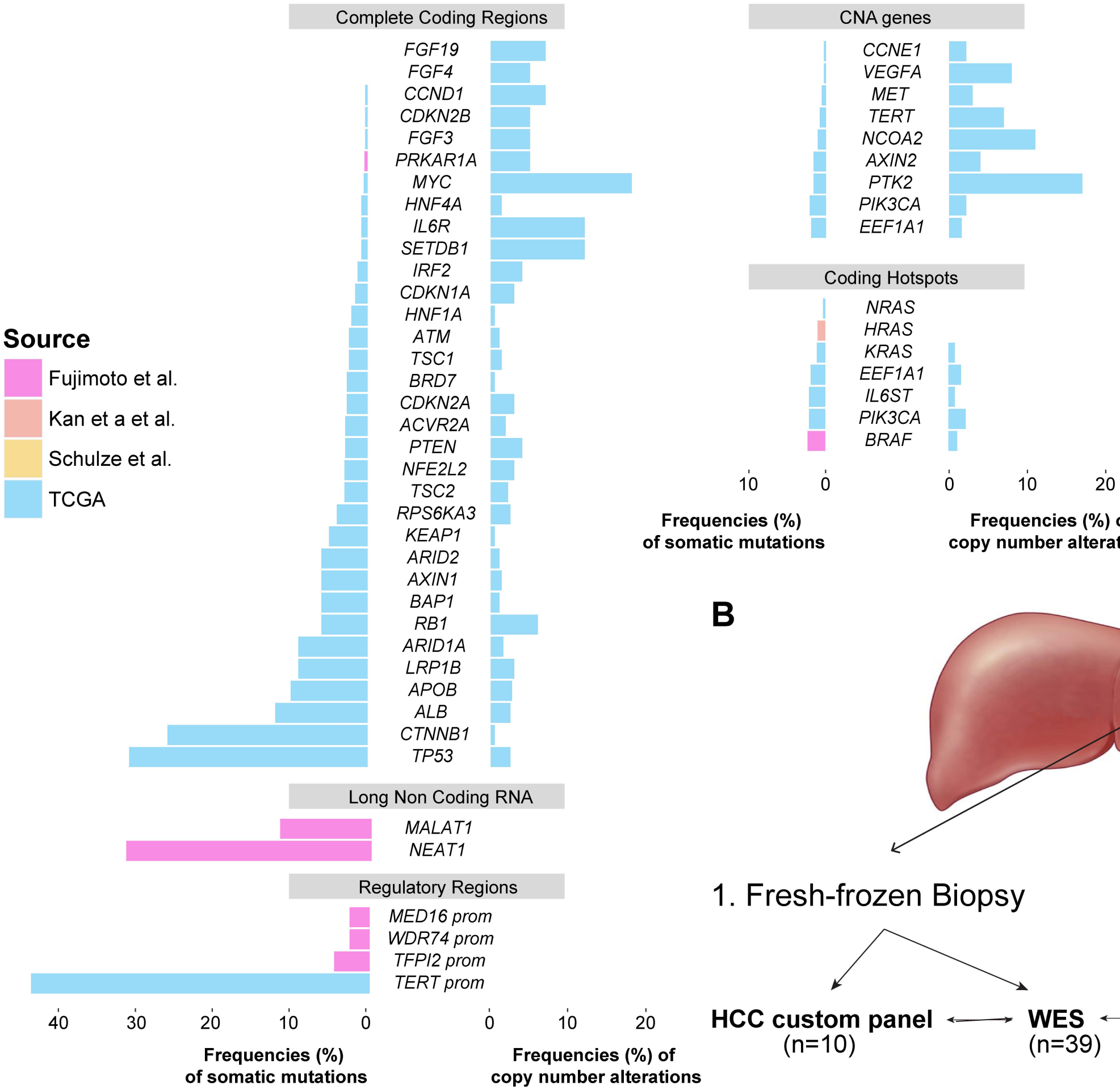
45. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R: Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther 2017, 16:2598-2608.

46. Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, Robertson AG, Hinoue T, Hoadley KA, Gibb EA, Roszik J, Covington KR, Wu CC, Shinbrot E, Stransky N, Hegde A, Yang JD,

731 Reznik E, Sadeghi S, Pedamallu CS, Ojesina AI, Hess JM, Auman JT, Rhie SK, Bowlby R,
732 Borad MJ, Cancer Genome Atlas N, Zhu AX, Stuart JM, Sander C, Akbani R, Cherniack AD,
733 Deshpande V, Mounajjed T, Foo WC, Torbenson MS, et al.: Integrative Genomic Analysis of
734 Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. Cell Rep 2017,
735 19:2878-2880.
736
737

Figure 1

A



B

1. Fresh-frozen Biopsy

2. FFPE Biopsy

HCC custom panel (n=10) ↔ **WES** (n=39) ↔ **HCC custom panel** (n=36)

Figure 2

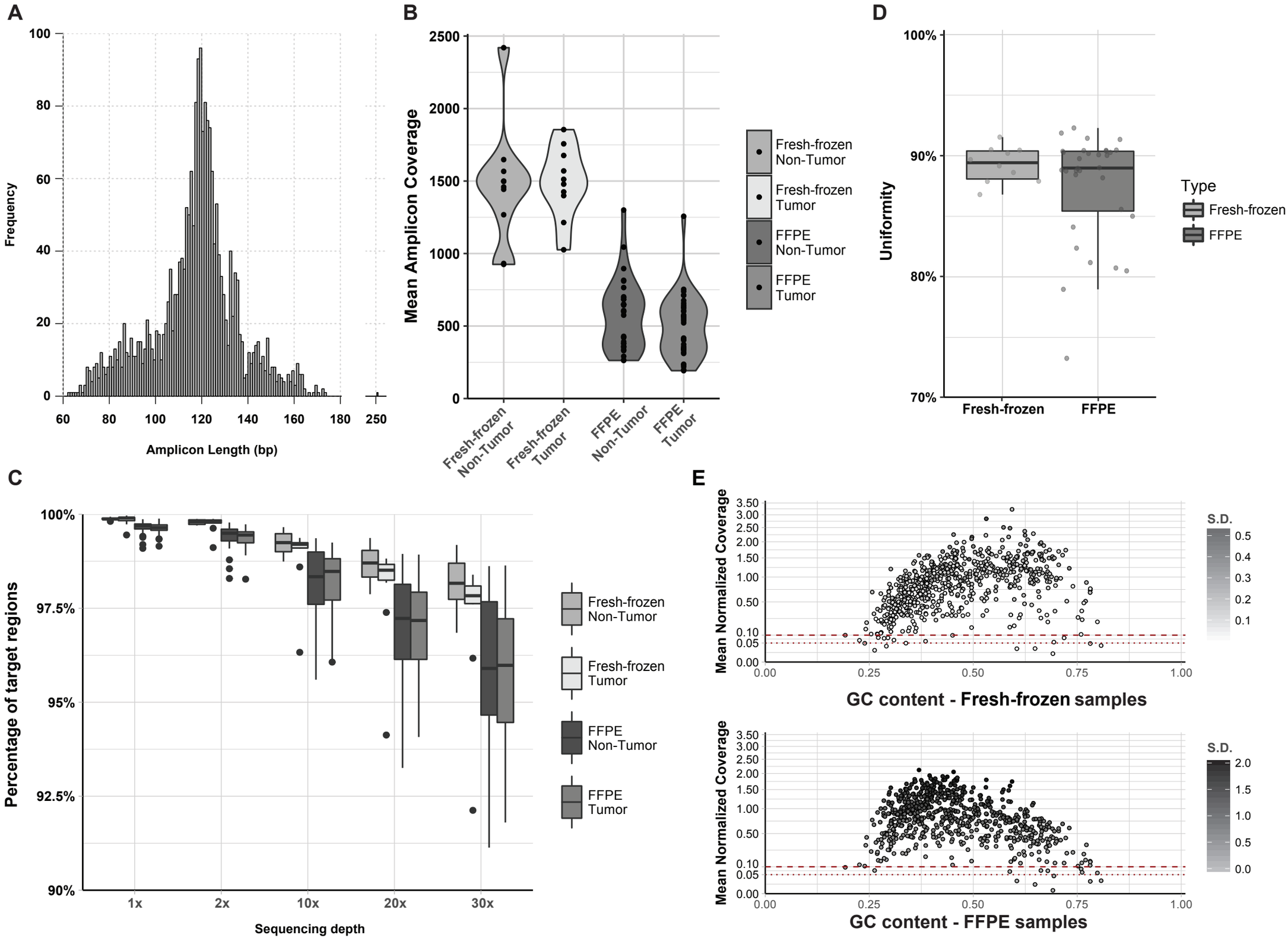

A
Frequency
Amplicon Length (bp)
B
Mean Amplicon Coverage
Fresh-frozen Non-Tumor
Fresh-frozen Tumor
FFPE Non-Tumor
FFPE Tumor
D
Uniformity
Type
Fresh-frozen
FFPE
C
Percentage of target regions
Sequencing depth
E
Mean Normalized Coverage
S.D.
GC content - Fresh-frozen samples
GC content - FFPE samples

Figure 3

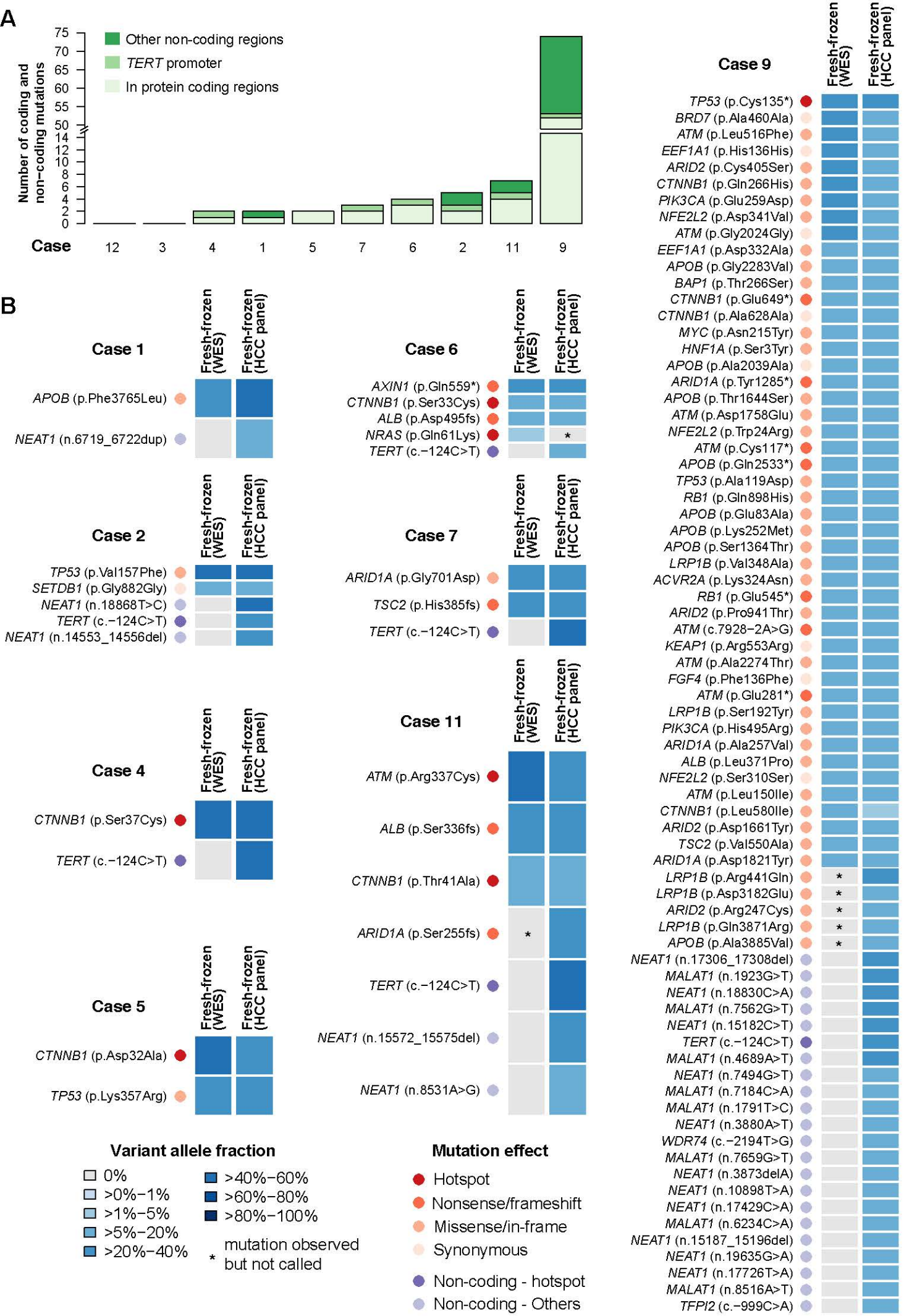

A
Other non-coding regions
TERT promoter
In protein coding regions
Number of coding and non-coding mutations
Case
12
3
4
1
5
7
6
2
11
9
B
Case 1
APOB (p.Phe3765Leu)
NEAT1 (n.6719_6722dup)
Case 2
TP53 (p.Val157Phe)
SETDB1 (p.Gly882Gly)
NEAT1 (n.18868T>C)
TERT (c.−124C>T)
NEAT1 (n.14553_14556del)
Case 4
CTNNB1 (p.Ser37Cys)
TERT (c.−124C>T)
Case 5
CTNNB1 (p.Asp32Ala)
TP53 (p.Lys357Arg)
Case 6
AXIN1 (p.Gln559*)
CTNNB1 (p.Ser33Cys)
ALB (p.Asp495fs)
NRAS (p.Gln61Lys)
TERT (c.−124C>T)
Case 7
ARID1A (p.Gly701Asp)
TSC2 (p.His385fs)
TERT (c.−124C>T)
Case 11
ATM (p.Arg337Cys)
ALB (p.Ser336fs)
CTNNB1 (p.Thr41Ala)
ARID1A (p.Ser255fs)
TERT (c.−124C>T)
NEAT1 (n.15572_15575del)
NEAT1 (n.8531A>G)
Fresh-frozen (WES)
Fresh-frozen (HCC panel)
Variant allele fraction
0%
>0%−1%
>1%−5%
>5%−20%
>20%−40%
>40%−60%
>60%−80%
>80%−100%
* mutation observed but not called
Mutation effect
Hotspot
Nonsense/frameshift
Missense/in-frame
Synonymous
Non-coding - hotspot
Non-coding - Others
Case 9
TP53 (p.Cys135*)
BRD7 (p.Ala460Ala)
ATM (p.Leu516Phe)
EEF1A1 (p.His136His)
ARID2 (p.Cys405Ser)
CTNNB1 (p.Gln266His)
PIK3CA (p.Glu259Asp)
NFE2L2 (p.Asp341Val)
ATM (p.Gly2024Gly)
EEF1A1 (p.Asp332Ala)
APOB (p.Gly2283Val)
BAP1 (p.Thr266Ser)
CTNNB1 (p.Glu649*)
CTNNB1 (p.Ala628Ala)
MYC (p.Asn215Tyr)
HNF1A (p.Ser3Tyr)
APOB (p.Ala2039Ala)
ARID1A (p.Tyr1285*)
APOB (p.Thr1644Ser)
ATM (p.Asp1758Glu)
NFE2L2 (p.Trp24Arg)
ATM (p.Cys117*)
APOB (p.Gln2533*)
TP53 (p.Ala119Asp)
RB1 (p.Gln898His)
APOB (p.Glu83Ala)
APOB (p.Lys252Met)
APOB (p.Ser1364Thr)
LRP1B (p.Val348Ala)
ACVR2A (p.Lys324Asn)
RB1 (p.Glu545*)
ARID2 (p.Pro941Thr)
ATM (c.7928−2A>G)
KEAP1 (p.Arg553Arg)
ATM (p.Ala2274Thr)
FGF4 (p.Phe136Phe)
ATM (p.Glu281*)
LRP1B (p.Ser192Tyr)
PIK3CA (p.His495Arg)
ARID1A (p.Ala257Val)
ALB (p.Leu371Pro)
NFE2L2 (p.Ser310Ser)
ATM (p.Leu150Ile)
CTNNB1 (p.Leu580Ile)
ARID2 (p.Asp1661Tyr)
TSC2 (p.Val550Ala)
ARID1A (p.Asp1821Tyr)
LRP1B (p.Arg441Gln)
LRP1B (p.Asp3182Glu)
ARID2 (p.Arg247Cys)
LRP1B (p.Gln3871Arg)
APOB (p.Ala3885Val)
NEAT1 (n.17306_17308del)
MALAT1 (n.1923G>T)
NEAT1 (n.18830C>A)
MALAT1 (n.7562G>T)
NEAT1 (n.15182C>T)
TERT (c.−124C>T)
MALAT1 (n.4689A>T)
NEAT1 (n.7494G>T)
MALAT1 (n.7184C>A)
MALAT1 (n.1791T>C)
NEAT1 (n.3880A>T)
WDR74 (c.−2194T>G)
MALAT1 (n.7659G>T)
NEAT1 (n.3873delA)
NEAT1 (n.10898T>A)
NEAT1 (n.17429C>A)
MALAT1 (n.6234C>A)
NEAT1 (n.15187_15196del)
NEAT1 (n.19635G>A)
NEAT1 (n.17726T>A)
MALAT1 (n.8516A>T)
TFPI2 (c.−999C>A)

Figure 4

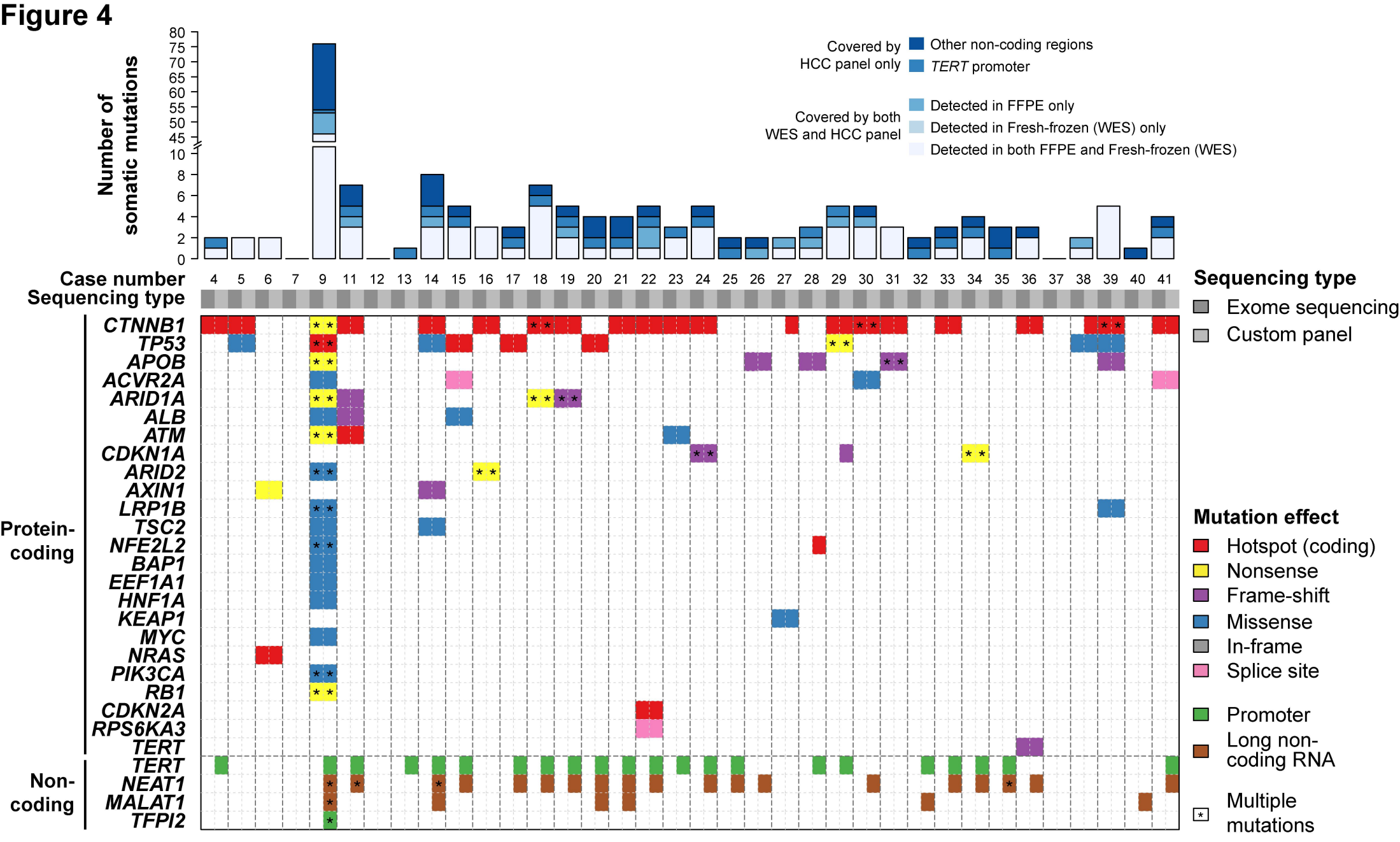

Number of somatic mutations
Covered by HCC panel only
Other non-coding regions
TERT promoter
Covered by both WES and HCC panel
Detected in FFPE only
Detected in Fresh-frozen (WES) only
Detected in both FFPE and Fresh-frozen (WES)
Case number
Sequencing type
Protein-coding
Non-coding
CTNNB1
TP53
APOB
ACVR2A
ARID1A
ALB
ATM
CDKN1A
ARID2
AXIN1
LRP1B
TSC2
NFE2L2
BAP1
EEF1A1
HNF1A
KEAP1
MYC
NRAS
PIK3CA
RB1
CDKN2A
RPS6KA3
TERT
TERT
NEAT1
MALAT1
TFPI2
Sequencing type
Exome sequencing
Custom panel
Mutation effect
Hotspot (coding)
Nonsense
Frame-shift
Missense
In-frame
Splice site
Promoter
Long non-coding RNA
Multiple mutations

Figure 5

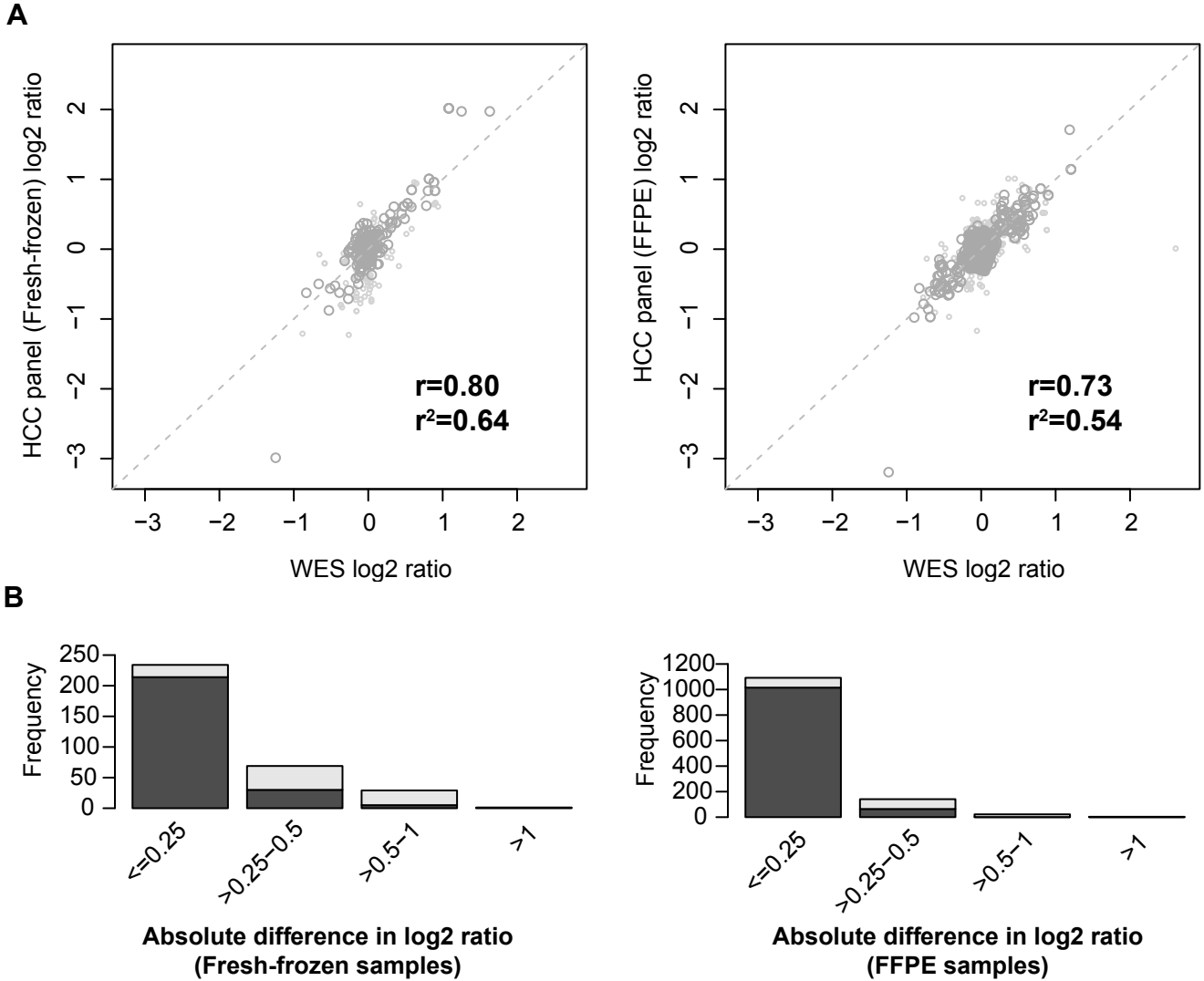

A
HCC panel (Fresh-frozen) log2 ratio
WES log2 ratio
r=0.80
r²=0.64
HCC panel (FFPE) log2 ratio
WES log2 ratio
r=0.73
r²=0.54
B
Frequency
<=0.25
>0.25–0.5
>0.5–1
>1
Absolute difference in log2 ratio
(Fresh-frozen samples)
Frequency
<=0.25
>0.25–0.5
>0.5–1
>1
Absolute difference in log2 ratio
(FFPE samples)

FIGURE LEGENDS

Figure 1: Design of the HCC sequencing panel and the study. (A) Frequencies of somatic mutations and copy number alterations in the genes included on the HCC panel according to previously published studies. **(B)** Outline of the study with the number of samples for each analysis performed. CNA: copy number alteration; FFPE: formalin-fixed paraffin-embedded; HCC: hepatocellular carcinoma; prom: promoter; WES: whole-exome sequencing.

Figure 2: Coverage analyses and statistics of the HCC panel. (**A)** Distribution of the amplicon sizes on the HCC panel. (**B)** Violin plots of the mean amplicon coverage across fresh-frozen non-tumor, fresh-frozen tumor, FFPE non-tumor and FFPE tumor samples. (**C)** Percentages of target regions covered at various depths (1x, 2x, 10x, 20x and 30x) across fresh-frozen non-tumor, fresh-frozen tumor, FFPE non-tumor and FFPE tumor samples. (**D)** Coverage uniformity, defined as the percentage of target bases covered at >20% of the mean coverage, in fresh-frozen and FFPE non-tumor samples. (**E)** Scatter plot of GC content and mean normalized coverage for all amplicons in fresh-frozen and FFPE samples. Color of the dots indicates the standard deviation of mean normalized coverage within each group. FFPE: formalin-fixed paraffin-embedded; SD: standard deviation.

Figure 3: Comparison of somatic mutations defined by whole-exome sequencing and HCC panel in fresh frozen tissues. **(A)** Number of coding and non-coding mutations per case identified in 10 fresh-frozen biopsies using the HCC panel. **(B)** Comparison of somatic coding and non-coding mutations found by WES and the HCC panel in the fresh-frozen samples. Heatmaps indicate the variant allele fractions of the somatic mutations (blue, see color key) or their absence (grey) in the 8 cases in which at least one somatic mutation was identified. Mutation types are indicated as colored dots according to the color key. Mutations that were not called by mutation caller but were supported by at least 1 sequencing read are indicated by an asterisk. HCC: hepatocellular carcinoma; WES: whole-exome sequencing.

765

766 **Figure 4: Comparison of somatic mutations defined by whole-exome sequencing and HCC**
767 **panel in FFPE tissue**. Barplot illustrates the number of somatic coding and non-coding mutations
768 found in 36 FFPE tumor biopsies using the HCC panel. In the main panel, each row represents a
769 gene on the HCC panel and each column represents a sample. The mutations identified by WES in
770 the fresh-frozen biopsies and those defined by sequencing the corresponding FFPE samples using
771 the HCC panel are placed next to each other. Mutation types are color coded according to the color
772 key. The presence of multiple mutations in the same gene is illustrated by an asterisk. Non-coding
773 regions below the dotted line were not covered by WES. FFPE: formalin-fixed paraffin-embedded;
774 WES: whole-exome sequencing.

775

776 **Figure 5: Copy number profiling using the HCC panel.** (**A**) Scatter plots illustrate the copy
777 number $\log_2$ ratio of WES and HCC panel sequencing of the fresh-frozen (left) and the FFPE (right)
778 tumor samples. (**B**) Barplots illustrate the number of genes with concordant (dark grey) or
779 discordant (light grey) copy number states, binned by the absolute difference in copy number $\log_2$
780 ratio between WES and HCC panel sequencing of the fresh-frozen (left) and FFPE (right) samples.
781 FFPE: formalin-fixed paraffin-embedded; HCC: hepatocellular carcinoma; WES: whole-exome
782 sequencing.