Characterisation of dysplastic liver nodules using low-pass DNA sequencing and detection of chromosome arm-level abnormalities in blood-derived cell-free DNA

Fateen, Waleed; Johnson, Philip; Wood, Henry; Zhang, Han; He, Shan; El-Meteini, Mahmoud ; Wyatt, Judy I. ; Aithal, Guruprasad P.; Quirke, Philip

DOI:
10.1002/path.5734

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Fateen, W, Johnson, P, Wood, H, Zhang, H, He, S, El-Meteini, M, Wyatt, JI, Aithal, GP & Quirke, P 2021, 'Characterisation of dysplastic liver nodules using low-pass DNA sequencing and detection of chromosome arm-level abnormalities in blood-derived cell-free DNA', *Journal of Pathology*, vol. 255, no. 1, pp. 30-40. https://doi.org/10.1002/path.5734

Journal of Pathology
J Pathol 2021; **00:** 000–000
Published online 7 July 2021 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.5734

ORIGINAL PAPER

Characterisation of dysplastic liver nodules using low-pass DNA sequencing and detection of chromosome arm-level abnormalities in blood-derived cell-free DNA

Waleed Fateen[1,2,3,4]*, Philip J Johnson[5], Henry M Wood[2], Han Zhang[6], Shan He[6], Mahmoud El-Meteini[3], Judy I Wyatt[7], Guruprasad P Aithal[1,4] and Philip Quirke[2]

[1] *NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and The University of Nottingham, Nottingham, UK*
[2] *Pathology and Data Analytics, Leeds Institute of Medical Research, St James's University Hospital, University of Leeds, Leeds, UK*
[3] *Ain Shams Centre for Organ Transplant, Ain Shams University, Cairo, Egypt*
[4] *Nottingham Digestive Diseases Centre, School of Medicine, University of Nottingham, Nottingham, UK*
[5] *Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK*
[6] *School of Computer Science, Centre for Computational Biology, The University of Birmingham, Birmingham, UK*
[7] *Histopathology Department, Leeds Teaching Hospitals NHS Trust, Leeds, UK*

Correspondence to: W Fateen, Nottingham Digestive Diseases Centre, E Floor, West Block, Queens Medical Centre, Derby Road, Nottingham NG7 2UH, UK. E-mail: waleed.fateen@nottingham.ac.uk

Abstract

High-grade dysplasia carries significant risk of transformation to hepatocellular carcinoma (HCC). Despite this, at the current standard of care, all non-malignant hepatic nodules including high-grade dysplastic nodules are managed similarly. This is partly related to difficulties in distinguishing high-risk pathology in the liver. We aimed to identify chromosome arm-level somatic copy number alterations (SCNAs) that characterise the transition of liver nodules along the cirrhosis–dysplasia–carcinoma axis. We validated our findings on an independent cohort using blood-derived cell-free DNA. A repository of non-cancer DNA sequences obtained from patients with HCC ($n = 389$) was analysed to generate cut-off thresholds aiming to minimise false-positive SCNAs. Tissue samples representing stages from the multistep process of hepatocarcinogenesis ($n = 184$) were subjected to low-pass whole genome sequencing. Chromosome arm-level SCNAs were identified in liver cirrhosis, dysplastic nodules, and HCC to assess their discriminative capacity. Samples positive for 1q+ or 8q+ arm-level duplications were likely to be either HCC or high-grade dysplastic nodules as opposed to low-grade dysplastic nodules or cirrhotic tissue with an odds ratio (OR) of 35.5 (95% CI 11.5–110) and 16 (95% CI 6.4–40.2), respectively ($p < 0.0001$). In an independent cohort of patients recruited from Nottingham, UK, at least two out of four alterations (1q+, 4q−, 8p−, and 8q+) were detectable in blood-derived cell-free DNA of patients with HCC ($n = 22$) but none of the control patients with liver cirrhosis ($n = 9$). Arm-level SCNAs on 1q+ or 8q+ are associated with high-risk liver pathology. These can be detected using low-pass sequencing of cell-free DNA isolated from blood, which may be a future early cancer screening tool for patients with liver cirrhosis.

Keywords: hepatocellular carcinoma; copy number; premalignant lesions; tumour pathology; cell-free DNA; early cancer

Received 30 November 2020; Revised 6 May 2021; Accepted 21 May 2021

Conflict of interest statement: PQ declares consulting with Amgen, Origin Sciences, Eisai, and Roche, and grant/research support from Roche, Ventana, Affymetrix, Leica, and Eisai. GPA is a member of the advisory boards for Aegerion, GSK, and Pfizer. WF and GPA have filed patents/patent applications based on the data generated from this work (reference: PCT/GB2021/051102). No other conflicts of interest were declared.

Introduction

Early events in hepatocarcinogenesis involve the progression of dysplastic nodules (DNs) to overt hepatocellular carcinoma (HCC). Such a contention has been supported by the frequent appearance of nodule-in-nodule lesions containing both dysplasia and cancer [1]. In comparison to regenerative nodules, DNs have an increased risk of malignant transformation [2]. Genomic features common to hepatic dysplasia and carcinoma demonstrate a step-wise increase in chromosome arm-level somatic copy number alterations (SCNAs) [3,4], chromosomal instability [3,5], and *TERT* promoter mutations [6,7]. Among dysplastic nodules, a histological distinction can be drawn between those that are 'low-grade' DNs and those that are 'high-grade'. The latter are assumed to be the immediate precursors of overt HCC. The rate of HCC development was found to be significantly higher in high-grade DNs than in low-grade DNs and regenerative nodules [8].

Accurate histological diagnosis of cancer or nodules that have a high likelihood of evolving into cancer is required. Such diagnosis is currently based on morphology, supplemented with immunohistochemistry where available, but the diagnosis is subjective and remains challenging in routine clinical practise [9]. The difficulty in histological diagnosis is focused on differentiating low-grade DNs from high-grade DNs and high-grade DNs from histologically well-differentiated HCC and there is a need for additional diagnostic tests which could be applied to fixed tissue or blood samples. Previous studies with focus on the prevalence of arm-level SCNAs identified 1q+ and 8q+ as common events in HCC [10], significantly less common in DNs [3,4], and rare events in cirrhosis [11]. However, their discriminatory role has not been investigated previously. We aimed, therefore, to identify the utility of arm-level SCNAs as an aid for stratifying high- and low-risk pathology along the cirrhosis–dysplasia–carcinoma axis using low-pass sequencing.

High-quality deep sequencing analysis of the HCC genome has been conducted in various contexts and has led to thorough characterisation of its main genomic features [10,12–15]. However, costly and time-consuming deep sequencing is not required for detection of arm-level SCNAs. For several decades, conventional karyotyping and chromosomal microarrays were standard [16]. More recently, low-pass next-generation sequencing (NGS) was validated with as little as 5 ng of DNA needed to generate a copy number karyogram [17–19] and reports of higher accuracy with an average coverage of 0.25x and mosaic levels as low as 20% [20].

Cell-free DNA refers to non-encapsulated DNA circulating freely in the blood stream. In cancer patients, a variable proportion of cell-free DNA originates from tumour cells. HCC is known to possess relatively high amplitude gains in copy number on 1q and 8q in comparison to other parts of the genome [10,11]. Therefore, both chromosomal arms offer natural signal amplification, boosting detectability within the plasma pool of cell-free DNA. Therefore, we tested the applicability of our study findings on a blood-derived cell-free DNA cohort.

Materials and methods

Study design

The Cancer Genome Atlas (TCGA; https://portal.gdc.cancer.gov/) is a public repository which includes high-quality genomic sequencing data from cancer tissue as well as matched non-cancer tissue of patients with cancer. We first downloaded all the 'cancer' and 'non-cancer' copy number segment files of patients with liver hepatocellular carcinoma (TCGA-LIHC) which were used as the derivation cohort (Figure 1) [10]. We used

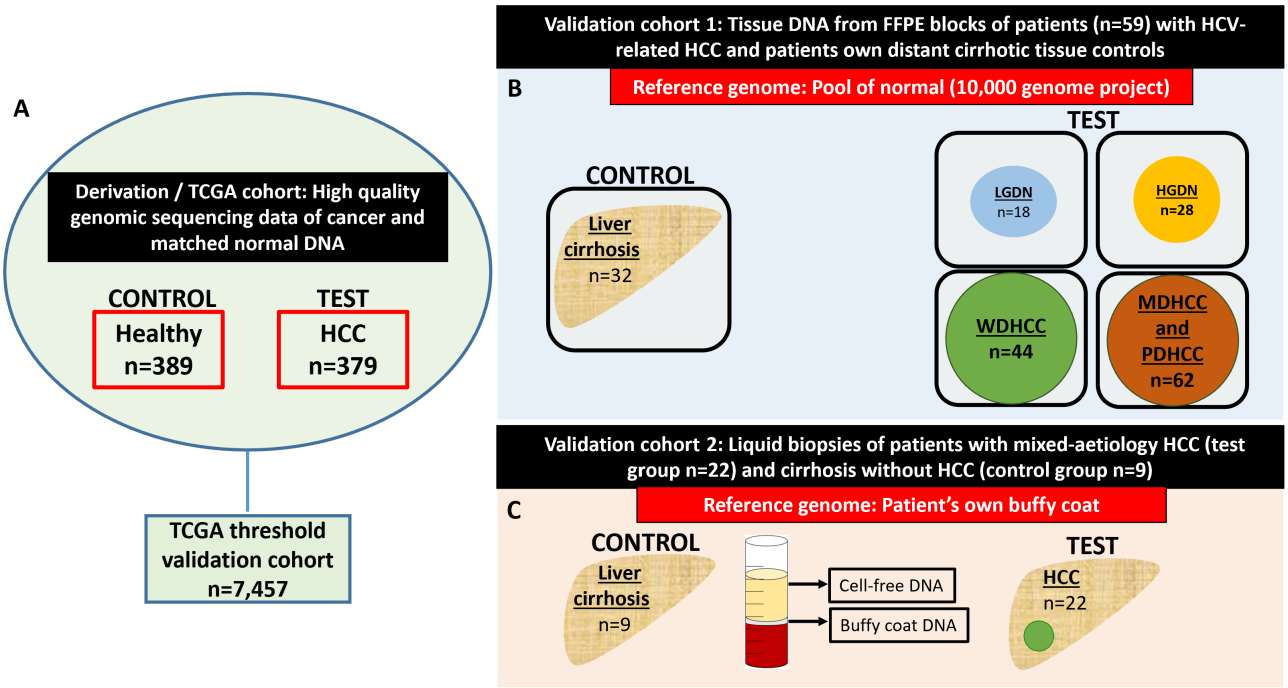

Figure 1. Description of study cohorts. (A) Derivation/TCGA cohort: TCGA-LIHC non-cancer sub-cohort ($n = 389$) was used to generate thresholds above or below which copy number gains or losses were called in test samples, respectively. TCGA-LIHC cancer sub-cohort ($n = 379$) was used to identify key arm-level SCNAs in HCC. The thresholds were validated for false positives against 7457 patients constituting the TCGA threshold validation sub-cohort, which is non-cancer tissue DNA of patients diagnosed with 26 different types of cancer but not HCC. (B) Validation cohort 1 (formalin-fixed, paraffin-embedded) was used to validate the ability of key arm-level SCNAs identified from the derivation cohort in classifying low- and high-risk liver pathology. (C) Validation cohort 2 (blood-derived cell-free DNA). The total number of samples per cohort and the main clinical features of samples in each cohort are highlighted. HGDN, high-grade dysplastic nodules; LGDN, low-grade dysplastic nodules; MDHCC, moderately differentiated HCC; PDHCC, poorly differentiated HCC; WDHCC, well-differentiated HCC.

the TCGA-LIHC non-cancer sub-cohort to generate thresholds above or below which copy number gains or losses were called in cancer samples, respectively. Next, we identified key arm-level SCNAs in the TCGA-LIHC cancer sub-cohort. Using the first validation cohort, we analysed the ability of key arm-level SCNAs to discriminate between low- and high-risk pathology. Lastly, we used the second validation cohort to test the principle that low-pass sequencing of cell-free DNA from blood samples offers a future early cancer screening tool for patients with liver cirrhosis using liquid biopsies.

Ethical approval and sample collection

The study was approved by the UK National Research Ethics Service (NRES, reference: 10/H1310/61) and Ain Shams University Hospitals local research ethics committee (LREC, reference: FMASU/485/2009). Anonymised samples were obtained from the Nottingham Digestive Diseases Biomedical Research Unit Tissue Bank (NRES, Ref 14/WA/1234). Tissue and blood samples were collected from three hospitals: St James's University Hospital, Leeds (UK); Nottingham University Hospitals NHS Trust, Nottingham (UK); and Ain Shams Specialised Hospital, Cairo (Egypt). Samples used in this study were divided into two cohorts:

- Validation cohort 1 – formalin-fixed, paraffin-embedded (FFPE) tissue from explanted liver specimens of 59 patients with HCC and/or dysplastic nodules complicating cirrhosis due to hepatitis C (supplementary material, Table S1 for clinico-pathological characteristics and Table S2 for histopathology). Patients were recruited between August 1999 and April 2013. FFPE tissue was processed as detailed in Supplementary materials and methods. All H&E slides were reviewed by a single experienced liver histopathologist (JIW), who identified clear examples to represent the range of hepatocellular lesions, outlined the lesions on the glass slides, and recorded their differentiation and morphological pattern according to the World Health Organisation (WHO) classification [21]; the annotated H&E slides were then scanned using an automated scanning system with a $\times 20$ objective to produce digital images and uploaded to a digital pathology server: https://bit.ly/31QugTa (supplementary material, Table S2) [22]. DNA was extracted from the nodules and from cirrhotic tissue geographically distant to the nodules. HCC nodules ($n = 106$, including WD-HCC $n = 44$) and DNs ($n = 46$, including high-grade DNs, $n = 28$, and low-grade DNs, $n = 18$) were retrieved from 59 patient livers.
- Validation cohort 2 – Blood-derived cell-free DNA was obtained from 31 patients (HCC, $n = 22$, and cirrhosis without HCC, $n = 9$) from Nottingham, UK with various aetiologies (Figure 1 and supplementary material, Table S1 for clinico-pathological characteristics). Patients were recruited between September 2016 and September 2017. HCC was diagnosed radiologically according to European Association for the Study of the Liver (EASL) criteria [23].

Sample processing

Detailed DNA extraction, library preparation, quality control, sequencing, and bioinformatics methodology can be found in Supplementary materials and methods and Tables S3 and S4.

In brief, for validation cohort 1, test DNA was extracted from FFPE dysplastic or malignant tissue. Control DNA was extracted from the patient's own cirrhotic tissue geographically distant to the nodule. Test and control DNA were characterised in comparison to a reference pool of normal DNA downloaded from the 1000 Genomes Project (see Supplementary materials and methods) [24].

For validation cohort 2, test cell-free DNA was extracted from the plasma of patients diagnosed with HCC. Control cell-free DNA was extracted from the plasma of patients diagnosed with liver cirrhosis and found to be free of cancer on imaging for 6 months after sample collection. Test and control samples were characterised in comparison to the patient's own buffy coat genome.

The average coverage per genome was calculated by multiplying the number of aligned reads by the read length in base pairs, or double the read length if paired-end, and dividing the result by 3 giga base pairs (supplementary material, Table S3). We identified the predicted proportion of tumour DNA within the eluted pool of extracted DNA using ABSOLUTE (https://software.broadinstitute.org/cancer/cga/absolute; supplementary material, Table S5) [25]. The proportion of tumour DNA within TCGA data was directly downloaded from Genomics Data Commons (https://gdc.cancer.gov/about-data/publications/pancanatlas) and can be found in supplementary material, Table S6.

Plasma cell-free DNA sequencing

An 8.5-ml blood sample was obtained from each patient and centrifuged within 2 h. The plasma and buffy coat portions were extracted into different cryovials. Both were re-centrifuged followed by re-extraction and storage at -80 °C until further analysis. DNA concentration was measured fluorometrically (PicoGreen®; Thermo Fisher Scientific, Waltham, MA, USA) and fragmentation assessed using an Agilent 2200 TapeStation (Agilent Technologies, Santa Clara, CA, USA). DNA libraries were prepared using tagged primers as detailed in Supplementary materials and methods and then labelled using unique 6-bp tags to enable multiplexing of libraries. Equal quantities of DNA libraries that passed quality control were pooled for cluster amplification and multiplexed on the same sequencing lane. Two DNA library pools were prepared: a cell-free DNA sample pool and a reference sample pool. The cell-free DNA sample pool included cell-free DNA samples from the HCC (test) group and the liver cirrhosis (control) group.

Ten nanograms of DNA library per sample was included in the cell-free DNA pool and sequenced (Illumina HiSeq3000; Illumina, San Diego, CA, USA) using paired-end sequencing with a read length of 151 and average coverage of 3–4x (IQR 2.8x–3.8x) per sample. A description of the reference sample pool can be found in Supplementary materials and methods.

FastQ files were output by the sequencer, two files per sample, a file for each read. File integrity was verified using MD5checksum and quality controlled using FastQC (Babraham Institute, UK, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor sequences were trimmed using Cutadapt [26]. Nucleotide sequences were aligned against the human genome assembly 19 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/) using Burrows–Wheeler aligner [27]. Sequences with poor mapping qualities less than 37 were not used. The number of aligned reads can be found in supplementary material, Table S3. The aligned reference read lengths were trimmed to match aligned test read lengths. Each genome was divided into 100-kbp non-overlapping windows. The ratio of test to reference number of reads per window was normalised according to the most abundant ratio using CNAnorm [28] (supplementary material, Figure S1), where the most abundant ratio was considered as ratio 1 and the other normalised ratios were calculated accordingly. GC correction was performed using CNAnorm and breakpoints were called using DNAcopy [29].

Detection and definition of arm-level somatic copy number alterations

For the derivation cohort, copy number segment files were downloaded from TCGA, https://gdc.cancer.gov/, including the TCGA-LIHC cancer sub-cohort ($n = 379$) and the TCGA-LIHC non-cancer sub-cohort ($n = 305$ blood-derived and 84 solid tissue).

For the validation cohorts, copy number segment files were generated using CNAnorm [28]. Detailed bioinformatics may be found in Supplementary materials and methods. Centromeric locations were obtained from the University of California Santa Cruz (UCSC) genome browser and the mean value of normalised test to reference ratio for each autosomal arm was calculated (supplementary material, Table S7).

Generation of thresholds using the TCGA-LIHC non-cancer sub-cohort ($n = 389$)

Two thresholds were generated for each autosomal arm to mark gains and losses. The thresholds were set at the fifth and 95th centile values for losses and gains, respectively (supplementary material, Table S8). Therefore, less than 5% of the TCGA-LIHC non-cancer sub-cohort exceeded the threshold for a gain or had values inferior to the threshold for a loss.

Validation of thresholds using the TCGA threshold validation sub-cohort ($n = 7457$)

The TCGA threshold validation sub-cohort comprised non-cancer DNA sequences obtained from all TCGA patients who had 26 different types of cancer (supplementary material, Table S9). To investigate the potential for false-positive results, the thresholds were tested against copy number segment files of the TCGA threshold validation sub-cohort ($n = 7457$).

Proportion of TCGA-HCC samples passing the threshold

The TCGA-LIHC cancer sub-cohort was used to identify the proportion of tumour samples passing the threshold for each autosomal arm. Autosomal arms where the proportion of tumour samples passing thresholds was higher than 75% were identified. Figure 2 shows two examples of autosomal arms where the proportions of TCGA-HCC samples passing the threshold were higher than 75%.

Statistical analyses

Cumulative frequency incorporates the frequency and amplitude of autosomal arm alterations within a group of lesions. This was used to display whole genome karyotype for groups within both validation cohorts (supplementary material, Figure S2).

To assess the capacity of key SCNAs identified using the TCGA/derivation cohort in discriminating low- and high-risk pathology, we analysed their prevalence within each group and calculated the odds ratio (OR) with 95% confidence interval (CI) according to Altman [30].

Results

Identification of key arm-level SCNAs

Key arm-level SCNAs were defined as those prevalent in more than 75% of the TCGA-LIHC cancer sub-cohort. This included 1q+ (90%), 4q− (77%), 8p− (76%), 8q+ (81%), and 17p− (79%). Table 1 shows the prevalence of 1q+, 4q−, 8p−, 8q+, and 17p− within the study groups explained in Figure 1.

Comparison of arm-level SCNAs in low- and high-grade dysplastic nodules

To identify the potential for arm-level SCNAs in aiding the discrimination of high-grade DNs from low-grade DNs, the prevalence of key arm-level SCNAs within both groups was examined. The likelihood of a dysplastic nodule being high grade ($n = 28$) as opposed to low grade ($n = 18$) in the presence or absence of 1q+, 4q−, 8p−, 8q+, and 17p− was measured using the OR. Development of high-grade dysplasia was associated with 1q+ OR = 8 (95% CI 1.5–41.5) and 8q+ OR = 5.8 (95% CI 1.4–24.5), as shown in Figure 3.

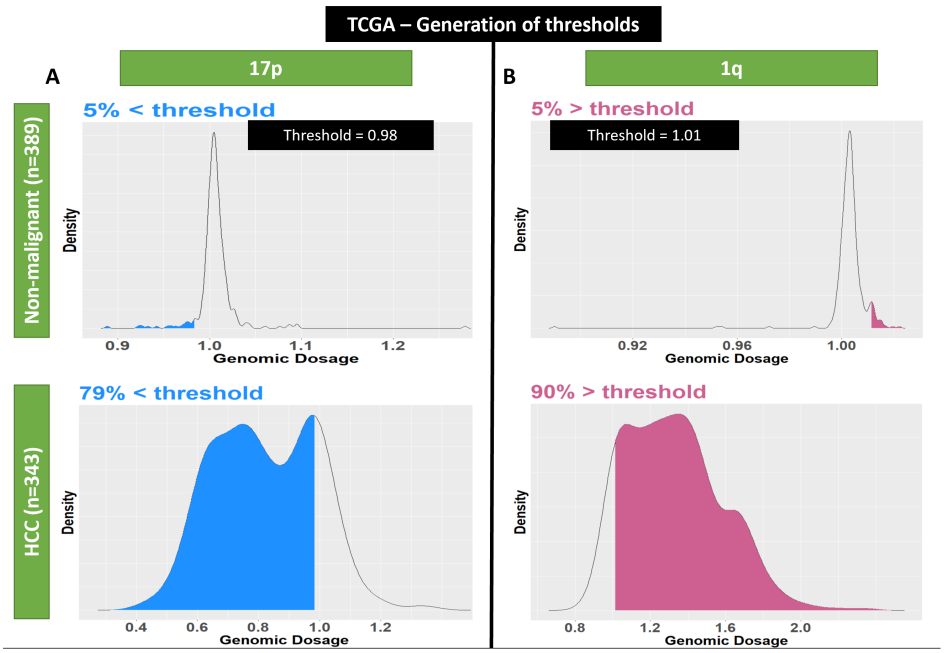

Figure 2. Generation of thresholds. The TCGA-LIHC non-cancer sub-cohort ($n = 305$ blood-derived and 84 solid tissue) was used to generate the thresholds which were set as the fifth and 95th centile values for each autosomal arm to mark losses and gains, respectively, in test samples. The TCGA-LIHC cancer sub-cohort ($n = 379$) included genomic sequencing of HCC. A shows the threshold for 17p set as 0.98 (fifth centile) of the TCGA-LIHC non-cancer sub-cohort and 79% of the TCGA-LIHC cancer sub-cohort had lower mean values for 17p. B shows the threshold for 1q set as 1.01 (95th centile) of the TCGA-LIHC non-cancer sub-cohort and 90% of the TCGA-LIHC cancer sub-cohort had higher mean values for 1q. Thresholds are marked using vertical black lines. Blue-shaded areas identify copy number losses and red-shaded areas identify copy number gains.

Comparison of arm-level SCNAs in high-grade dysplastic nodules and well-differentiated HCC

To identify the potential for arm-level SCNAs in aiding the discrimination of well-differentiated HCC from high-grade DNs, we examined the prevalence of key arm-level SCNAs within both groups. The likelihood of nodules being well-differentiated HCC ($n = 44$) as opposed to high-grade DNs ($n = 28$) in the presence or absence of 1q+, 4q−, 8p−, 8q+, and 17p− was measured using the OR. Development of early malignant features was associated with 1q+ OR = 13.7 (95% CI 3.4–54.7), 4q− OR = 3.8 (95% CI 1.4–10.6), 8p− OR = 5 (95% CI 1.6–15.7), and 8q+ OR = 4.9 (95% CI 1.5–13.7), as shown in Figure 3.

Comparison of 1q+ and 8q+ in low- and high-risk pathology

We defined liver cirrhosis and low-grade dysplasia as 'low-risk pathology', while high-grade dysplasia and cancer were defined as 'high-risk pathology'. As 1q+ and 8q+ were detected in high-grade DNs significantly more than in low-grade DNs, we aimed to explore their utility for discriminating between 'low-risk' versus

Table 1. Prevalence of 1q+, 4q−, 8p−, 8q+, and 17p−.

Group	1q+	4q−	8p−	8q+	17p−
Derivation cohort (*TCGA*)					
TCGA-LIHC cancer sub-cohort ($n = 379$)	340/379 (90%)	290/377 (77%)	285/377 (76%)	303/376 (81%)	271/343 (79%)
TCGA threshold validation sub-cohort ($n = 7457$)	445/7457 (6%)	396/7457 (5.3%)	226/7457 (3%)	404/7457 (5.4%)	NA
Validation cohort 1 (*FFPE*)					
HCC ($n = 106$)	95/106 (90%)	64/106 (60%)	59/106 (56%)	87/106 (82%)	53/106 (50%)
Well-differentiated HCC ($n = 44$)	41/44 (93%)	27/44 (61%)	23/44 (52%)	37/44 (84%)	22/44 (50%)
Moderately and poorly differentiated HCC ($n = 62$)	54/62 (87%)	37/62 (60%)	36/62 (58%)	50/62 (81%)	31/62 (50%)
DNs ($n = 46$)	16/46 (35%)	13/46 (28%)	7/46 (15%)	18/46 (39%)	14/46 (30%)
Low-grade DNs ($n = 18$)	2/18 (11%)	2/18 (11%)	2/18 (11%)	3/18 (17%)	5/18 (28%)
High-grade DNs ($n = 28$)	14/28 (50%)	11/28 (39%)	5/28 (18%)	15/28 (54%)	9/28 (32%)
Liver cirrhosis ($n = 32$)	2/32 (6%)	8/32 (25%)	0/32 (0%)	4/32 (13%)	1/32 (3%)
Validation cohort 2 (*blood-derived cell-free DNA*)					
HCC ($n = 22$)	19/22 (86%)	21/22 (96%)	15/22 (68%)	12/22 (55%)	7/22 (32%)
Cirrhosis ($n = 9$)	0/9 (0%)	4/9 (44%)	0/9 (0%)	0/9 (0%)	3/9 (33%)

DN, dysplastic nodule; FFPE, formalin-fixed, paraffin-embedded; HCC, hepatocellular carcinoma; TCGA-LIHC, The Cancer Genome Atlas – liver HCC.

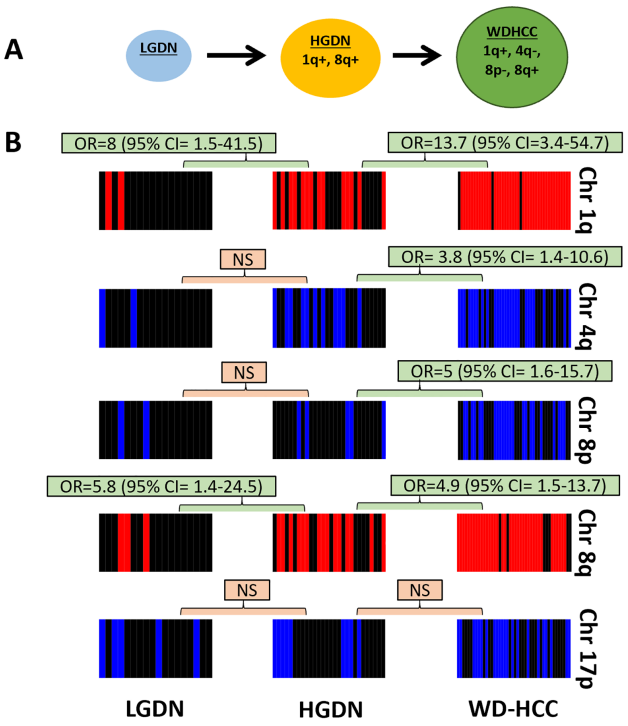

Figure 3. 1q+, 4q−, 8q−, and 8q+ can classify low-grade DNs, high-grade DNs, and well-differentiated HCC. (A) Genomic structural features along the dysplasia–carcinoma axis in HCC. (B) Heatmap showing gains (red) in 1q and 8q as well as losses (blue) in 4q, 8q, and 17p in low-grade DNs, high-grade DNs, and well-differentiated HCC. HGDN, high-grade dysplastic nodules; LGDN, low-grade dysplastic nodules; NS, not significant; WDHCC, well-differentiated HCC.

'high-risk' pathology. We therefore examined the prevalence of 1q+ and 8q+ within both groups. The odds of pathology being 'high risk' ($n = 134$) as opposed to 'low risk' ($n = 50$) in the presence or absence of 1q+ or 8q+ was measured using the OR. Development of 'high-risk' pathology was associated with 1q+ OR = 35.5 (95% CI 11.5–110) and 8q+ OR = 16 (95% CI 6.4–40.2).

Validation of thresholds using the TCGA threshold validation sub-cohort ($n = 7457$)

To investigate the potential for false-positive results, the thresholds were tested against non-cancer DNA sequencing obtained from all TCGA patients who had 26 different types of cancer ($n = 7457$) but not liver cancer. In this cohort, the thresholds were crossed in 445 (6%) of 1q, 404 (5.4%) of 8q, 396 (5.3%) of 4q, and 226 (3%) of 8p. None of the cases concurrently crossed more than one of the four thresholds.

Arm-level SCNAs associated with early features of malignancy are detectable using blood-derived cell-free DNA

Cell-free DNA from the plasma and reference DNA from the buffy coat of patients with HCC (test group, $n = 22$) was compared with that from patients with liver cirrhosis (control group, $n = 9$). The test group included patients within Milan criteria ($n = 7$) and patients with AFP < 20 ng/ml ($n = 9$) (Figure 4). None of the patients recruited to the control group developed HCC after a median follow-up of 22.4 months.

Supplementary material, Table S10 shows the sensitivity and specificity for each of the key arm-level SCNAs. At least two or three of 1q+, 4q−, 8p−, and 8q+ were present in 22/22 and 16/22 of the patients with HCC, respectively (Figure 4). All patients in the control group were negative for 1q+, 8p−, and 8q+, while four patients were positive for 4q−. None of the control group patients had more than one of the four arm-level SCNAs. Such patterns were in close resemblance to those observed in the TCGA threshold validation subcohort and validation cohort 1 (Table 1 and Figure 5).

Discussion

The European Association for the Study of the Liver (EASL) recommends the development of tools to stratify patients at high, intermediate, and low risk for HCC [23]. In this study, we found that among five key arm-level SCNAs in HCC, 1q+ and 8q+ were significantly associated with either early cancer or high-grade dysplasia, whereas such events were rare in low-grade dysplasia or cirrhosis. We demonstrated that the alterations can be detected using low-pass sequencing in blood-derived cell-free DNA of patients with HCC but not cirrhotic patients with no known HCC.

The detection on blood-derived cell-free DNA is likely related to two factors. Firstly, the high prevalence of arm-level SCNAs in HCC, TCGA data has revealed that arm-level SCNAs are more prevalent than SCNAs at gene level and more prevalent than common mutations such as in *TERT* and *CTNBB1* [10,31]. Secondly, the natural signal amplification at duplication hotspots such as 1q and 8q within a pool of cell-free DNA makes the signal more detectable despite the low proportion of circulating-tumour DNA. Cell-free DNA was directly compared against buffy coat DNA from the same patient as a reference, thus focusing on the cancer-related alterations. Previous studies found variable concordance between matching pairs of tumour and liquid biopsies, as recently reviewed [32,33]. This was not the focus of our study, which was on testing the utility of detecting specific chromosomal arm-level SCNAs using low-pass sequencing of cell-free DNA from patients with HCC. The prevalence of such SCNAs in cell-free DNA may be different to the tumour or the background liver tissue. Similarly, area under receiver operating curve analysis was not performed; the study was designed as a test of feasibility and not designed for in-depth analysis, which is the focus of currently ongoing work designed using PRoBE (prospective specimen collection, retrospective blind evaluation) [34].

There is variability in the literature regarding the optimal method of identifying a true copy number event [4,11,35]. The conventional z-score analysis is known

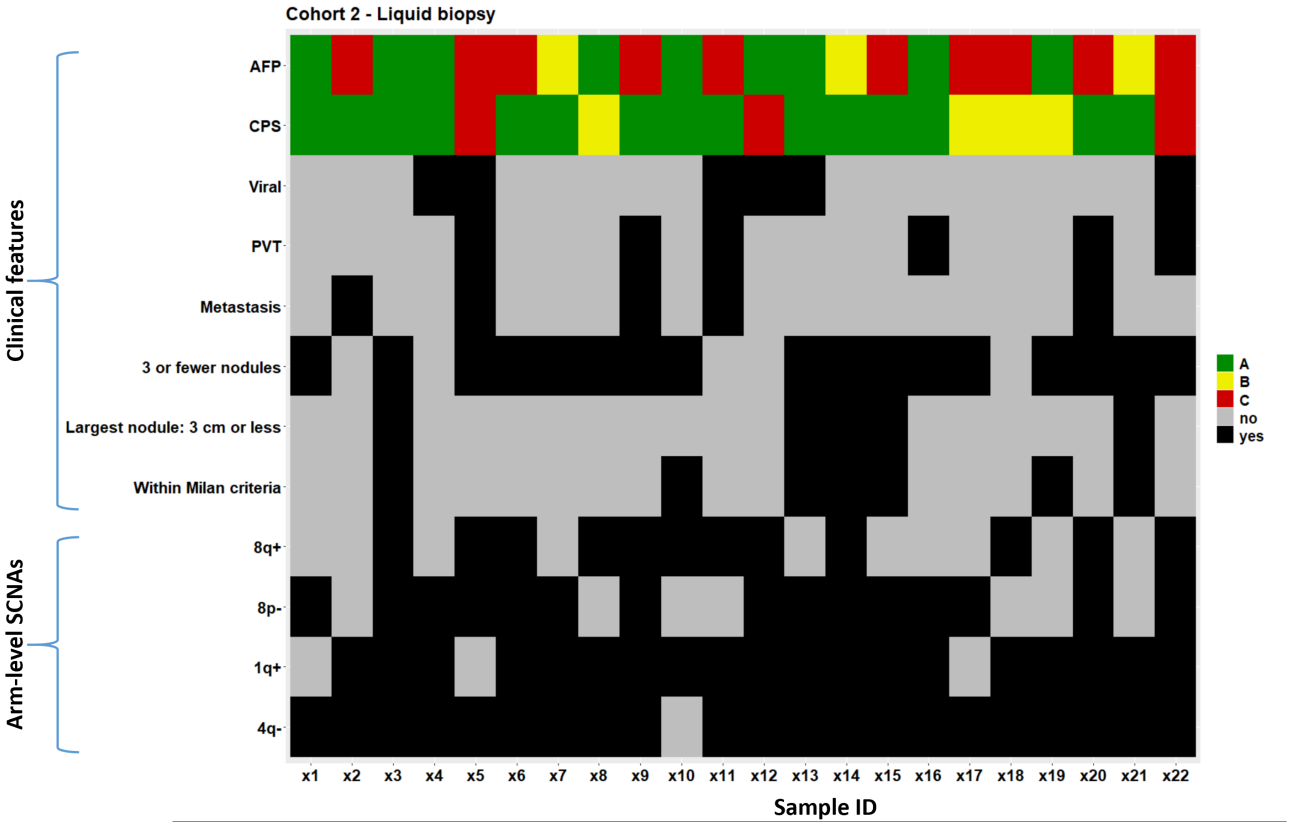

Figure 4. Validation cohort 2: clinical characteristics of 22 patients with HCC and prevalence of 1q+, 4q−, 8p−, and 8q+ using blood-derived cell-free DNA. (A) AFP < 20 ng/ml; (B) AFP 20–200 ng/ml. (C) AFP > 200 ng/ml. AFP, alpha-fetoprotein; CPS, Child–Pugh score; PVT, portal venous thrombosis; SCNAs, somatic copy number alterations.

to be affected by the depth of sequencing, due to its reliance on the standard deviation of the sequenced read density from the reference group [11,36–38]. We have generated fixed thresholds using TCGA sequencing data of matched non-cancer tissue obtained from patients diagnosed with HCC. The thresholds were validated on three independent datasets (TCGA dataset, an FFPE cohort, and a cell-free DNA cohort). More than 94% of 1q or 8q ratios obtained from non-cancer tissue ($n = 7457$) were within a tight range between 0.997 and 1.013, and with similar patterns observed across two independent datasets of patients with liver cirrhosis and no cancer sequenced in-house (Table 1 and Figure 5). This indicates a false-positive rate of less than 6% but much lower if both thresholds for 1q and 8q are crossed as this was not observed in any of the DNA

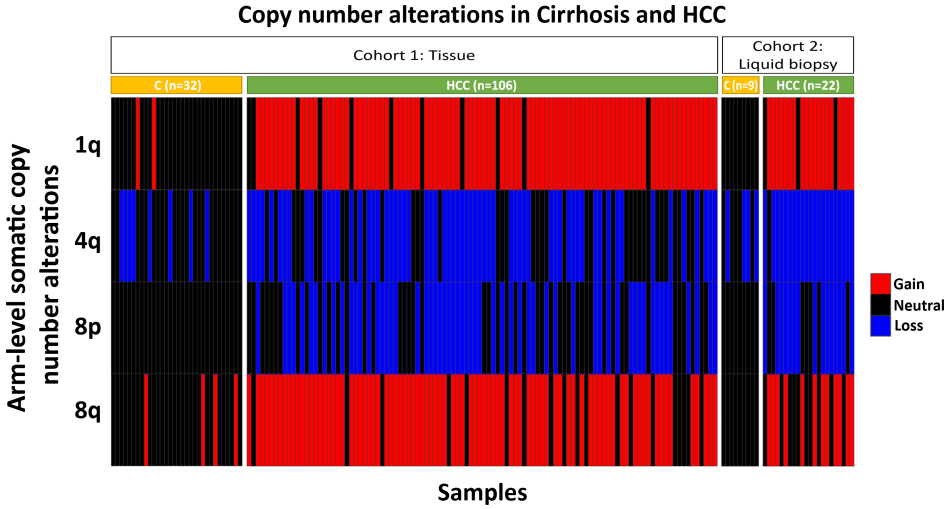

Figure 5. 1q+, 4q−, 8p−, and 8q+ are common in HCC and uncommon in cirrhosis. Similar patterns are observed in tissue DNA and blood-derived cell-free DNA. C, cirrhosis; HCC, hepatocellular carcinoma.

tested across three independent control cohorts. On the other hand, either 1q or 8q threshold was crossed in 104/106 HCCs in validation cohort 1 and in 20/22 HCCs in validation cohort 2, while both 1q and 8q thresholds were crossed in 78/106 HCCs in validation cohort 1 and in 11/22 HCCs in validation cohort 2, which indicates a potentially satisfactory negative predictive value (supplementary material, Tables S7 and S8). The thresholds are easily reproducible, unaffected by the depth of sequencing, and external validation highlights the potential for promising accuracy (supplementary material, Table S9).

A landmark study from Hong Kong investigating patients with hepatitis B virus found clear evidence of SCNAs in plasma DNA of 84.4% of patients who had HCC and 22.2% of patients with cirrhosis [11]. Our study showed similar results (Table 1) for patients with diverse background aetiologies and using low-pass sequencing. Our study agrees with deep sequencing studies and SNP arrays reporting the common copy number features in HCC as shown in supplementary material, Figure S2 [10–12,14,39]. Two earlier studies investigating gene transcriptional profiles have proposed molecular markers targeted at discriminating stages of hepatocarcinogenesis. This included dysplastic nodules and early HCC. However, in both studies, most of the dysplastic nodules included were low-grade DNs ($n = 10/16$). Therefore, it is not clear if the same signatures would discriminate between high-grade DNs and early HCC [40,41]. The rate of HCC development is significantly higher in high-grade DNs than in low-grade DNs and regenerative nodules [8]. Moreover, an international consensus panel did not find difficulty in differentiating between low-grade DNs and early HCC. The diagnostic discrepancy arose in the discrimination of low-grade DNs from high-grade DNs and of high-grade DNs from well-differentiated HCC [1]. Our study included a considerable number of high-grade DNs ($n = 28$) aiming to address this issue. Previous studies identified the increase in chromosomal instability from DNs to HCC [3,5]. More recently, Torrecilla *et al* [4] reported copy number data on low-grade DNs ($n = 14$), high-grade DNs ($n = 15$), and small HCC ($n = 17$) and identified 1q+, 8q+, and 8q− as potential 'gate-keeper' events, due to their prevalence in dysplastic nodules. Our work agreed, characterising 1q+, 8p−, and 8q+, as well as 4q− and 17p−, as early events in hepatocarcinogenesis.

EASL clinical practice guidelines currently recommend biopsy for liver nodules larger than 1 cm that do not show typical HCC features on at least one out of two imaging modalities [23]. The American Association for the Study of Liver Diseases (AASLD) endorses the Liver Imaging Reporting And Data System (LI-RADS) and recommends biopsy for lesions classed as 'probably HCC (LR-4)' or 'malignant but not HCC (LR-M)' [42]. Histopathologists face inherent difficulties in discriminating between high-grade DNs and well-differentiated HCC [43,44]. EASL recommends the use of a panel of three immunohistochemistry (IHC) antibodies to aid the histological discrimination of high-grade DNs from well-differentiated HCC [23]. However, a recent collaboration of Eastern and Western expert pathologists validating the EASL recommended panel found a sensitivity of 52% and recommended two further IHC antibodies to raise the sensitivity to 93%, suggesting that this may be difficult to apply in routine clinical practice [9]. Our study showed that 1q+ and 8q+ were significantly more prevalent in high-grade DNs than in low-grade DNs. 1q+, 4q−, 8p−, and 8q+ were significantly more prevalent in well-differentiated HCC than in high-grade DNs. Our study suggests that identification of arm-level SCNAs (1q+, 4q−, 8p−, and 8q+) has the potential to improve the current inter-observer agreement on tissue histopathological distinction between high-grade DNs and well-differentiated HCC [1,9].

Our study made an attempt to characterise low-grade DNs, high-grade DNs, and well-differentiated HCC according to their broad chromosomal structure. The detection of such chromosomal features is technically simple and applicable in day-to-day clinical practise. For instance, cytogenetic analysis of *BCR/ABL1* translocation, *HER2* amplification, and *ALK* rearrangement is the current standard of care for guiding the management of chronic myeloid leukaemia [45], breast cancer [46], and lung adenocarcinoma [47], respectively. Unlike pre-malignant lesions in other cancers such as Barrett's oesophagus [48] and colonic polyps [49], high-grade dysplasia in HCC currently is treated similarly to any other 'low-risk' lesion, such as regenerative nodules. This could be related partially to the difficulty in identifying and discriminating different stages of pre-malignant hepatic nodules. With rising indications for targeted liver biopsy, and known histopathological challenges even amongst experts, incorporating cytogenetic examinations for 1q+ and 8q+ may provide more objective discrimination even for non-expert pathologists. Moreover, liquid biopsy is a future non-invasive tool as such features can be detectable using low-pass sequencing (3–4x coverage). This is significantly less costly and less time-consuming in comparison to deep sequencing, which is not required for the purpose of detection of key arm-level SCNAs. Low-pass sequencing is still likely to be superior to alternative techniques as it enables delicate discrimination of ratios between 0.99 and 1.01 using small quantities of starting DNA.

Our study had some limitations; most dysplastic nodules ($n = 42/46$) were extracted from livers that harboured HCC as well, and the results may be different in livers harbouring DNs without HCC. Conversely, such livers may or may not ever develop cancer and further longitudinal studies on the natural history of DNs are required to identify the most appropriate DNs for study of genomic predictors of malignant transformation. Torrecilla *et al* recently found significantly lower prevalence of arm-level SCNAs in DNs that were retrieved from cirrhotic livers that did not have cancer. The lower prevalence reported by Torrecilla *et al* may be related to the biology of such DNs or higher thresholds for calling SCNAs [4]. Moreover, a recent study

based on phylogenetic analysis of single nucleotide variants and copy number profiles found evidence of independent growth of DNs and HCC within the same patient liver [50]. The majority, but not all ($n = 34/44$), of well-differentiated HCCs in validation cohort 1 were small (i.e. ≤2 cm; IQR = 11–21 mm). Our study was designed to discriminate histopathological features rather than nodule size. Our patient group in validation cohort 2 included generally more advanced cancers (only seven out of 22 were within Milan criteria, two of whom had liver transplantation and five were considered for transplants). Further work is required to determine whether early HCC which has less vascular invasion, or indeed dysplastic nodules, also releases cell-free DNA with detectable arm-level SCNAs into the peripheral circulation. Lastly, validation cohort 1 was all related to HCV, with a question about the generalisability, but the findings had no obvious correlation with aetiology in the cell-free DNA dataset.

In conclusion, 1q+ or 8q+ is associated with high-risk liver pathology, e.g. cancer or high-grade dysplasia, but not low-risk pathology, e.g. low-grade dysplasia and liver cirrhosis. Detection of 1q+, 4q−, 8p−, and 8q+ in the tissue may aid in distinguishing types of liver nodules and in subsequent decision making. Arm-level SCNAs can be detected in blood-derived cell-free DNA using low-pass sequencing, which may be useful as a tool for the surveillance, diagnosis, and monitoring of HCC in patients with cirrhosis.

Acknowledgements

We thank Dr Charles Millson for his contribution to the concept and design of the study, Dr Stefano Berri for his contribution to the bioinformatics, Professor John Armour for critical review of the manuscript, Dr Mathew Hoare for critical review of the manuscript, and Dr Augusto Villanueva for his valuable comments. We thank patients and staff at Leeds Teaching Hospitals NHS Trust, Liver Unit, specifically the clinical investigators including Dr Mervyn Davis, Dr Rebecca Jones, Dr Mark Aldersley, and Dr Darren Treanor. We thank staff at Leeds Institute of Cancer and Pathology, specifically participating investigators including Ms Catherine Daly, Ms Morag Raynor, Ms Gemma Hemmings, Dr Christopher Watson, Dr Sally Harrison, Dr Katie Southward, and Dr Ian Carr. We thank patients and staff at Nottingham Digestive Diseases Centre, University of Nottingham, specifically participating investigators Ms Melanie Lingaya, Mrs Yirga Falcone, Dr Jane Grove, and clinical investigators Dr Martin James, Professor Neil Guha, and Professor Steve Ryder. We thank patients and staff at Ain Shams University Centre for Organ Transplant.

The work was supported by the following awards to WF: (1) Leeds Teaching Hospitals Charitable Foundation; (2) Flexifund award from the MRC/EPSRC Nottingham Molecular Pathology Node (MR/N005953/1). PQ is supported by a National Institute of Health Research Senior Investigators Award. WF and GPA are supported by the National Institute of Health Research Nottingham Biomedical Research Centre (Reference No: BRC-1215-20003). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Author contributions statement

WF, JIW, MEM and PQ conceived and designed the study. WF, JIW, MEM and PQ acquired data. WF, PJ, HMW, JW, GPA and PQ analysed and interpreted data. WF, HMW, HZ, SH and PJ were responsible for bioinformatics. All the authors critically revised the manuscript and read and approved the final version.

Data availability statement

Aligned sequence reads have been deposited in the European Nucleotide Archive, accession number PRJEB20592. Histopathology slides are available for review using virtual microscopy on our digital pathology server accessible via the link https://bit.ly/3bGIVXH and instructions on using the viewer via https://bit.ly/3eZBWuQ.

References

1. International Consensus Group for Hepatocellular Neoplasia, Kojiro M, Wanless IR, *et al*. Pathologic diagnosis of early hepatocellular carcinoma: a report of the International Consensus Group for Hepatocellular Neoplasia. *Hepatology* 2009; **49:** 658–664.
2. Borzio M, Fargion S, Borzio F, *et al*. Impact of large regenerative, low grade and high grade dysplastic nodules in hepatocellular carcinoma development. *J Hepatol* 2003; **39:** 208–214.
3. Moinzadeh P, Breuhahn K, Stützer H, *et al*. Chromosome alterations in human hepatocellular carcinomas correlate with aetiology and histological grade – results of an explorative CGH meta-analysis. *Br J Cancer* 2005; **92:** 935–941.
4. Torrecilla S, Sia D, Harrington AN, *et al*. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. *J Hepatol* 2017; **67:** 1222–1231.
5. Maggioni M, Coggi G, Cassani B, *et al*. Molecular changes in hepatocellular dysplastic nodules on microdissected liver biopsies. *Hepatology* 2000; **32:** 942–946.
6. Nault JC, Calderaro J, Di Tommaso L, *et al*. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* 2014; **60:** 1983–1992.
7. Nault JC, Zucman-Rossi J. *TERT* promoter mutations in primary liver tumors. *Clin Res Hepatol Gastroenterol* 2016; **40:** 9–14.
8. Kobayashi M, Ikeda K, Hosaka T, *et al*. Dysplastic nodules frequently develop into hepatocellular carcinoma in patients with chronic viral hepatitis and cirrhosis. *Cancer* 2006; **106:** 636–647.
9. Sciarra A, Di Tommaso L, Nakano M, *et al*. Morphophenotypic changes in human multistep hepatocarcinogenesis with translational implications. *J Hepatol* 2016; **64:** 87–93.

10. The Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 2017; **169:** 1327–1341.e23.
11. Jiang P, Chan CW, Chan KC, *et al*. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* 2015; **112:** E1317–E1325.
12. Schulze K, Imbeaud S, Letouzé E, *et al*. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015; **47:** 505–511.
13. Fujimoto A, Furuta M, Totoki Y, *et al*. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* 2016; **48:** 500–509.
14. Guichard C, Amaddeo G, Imbeaud S, *et al*. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* 2012; **44:** 694–698.
15. Cleary SP, Jeck WR, Zhao X, *et al*. Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology* 2013; **58:** 1693–1702.
16. Wapner RJ, Martin CL, Levy B, *et al*. Chromosomal microarray versus karyotyping for prenatal diagnosis. *N Engl J Med* 2012; **367:** 2175–2184.
17. Dong Z, Zhang J, Hu P, *et al*. Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. *Genet Med* 2016; **18:** 940–948.
18. Wood HM, Belvedere O, Conway C, *et al*. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. *Nucleic Acids Res* 2010; **38:** e151.
19. Wood HM, Daly C, Chalkley R, *et al*. The genomic road to invasion – examining the similarities and differences in the genomes of associated oral pre-cancer and cancer samples. *Genome Med* 2017; **9:** 53.
20. Chau MHK, Wang H, Lai Y, *et al*. Low-pass genome sequencing: a validated method in clinical cytogenetics. *Hum Genet* 2020; **139:** 1403–1415.
21. Bosman FT, World Health Organization, International Agency for Research on Cancer. *WHO Classification of Tumours of the Digestive System*. International Agency for Research on Cancer: Lyon, 2010.
22. Randell R, Ruddle RA, Mello-Thoms C, *et al*. Virtual reality microscope versus conventional microscope regarding time to diagnosis: an experimental study. *Histopathology* 2013; **62:** 351–358.
23. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. *J Hepatol* 2018; **69:** 182–236.
24. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, *et al*. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467:** 1061–1073.
25. Carter SL, Cibulskis K, Helman E, *et al*. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012; **30:** 413–421.
26. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011; **17:** 10.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25:** 1754–1760.
28. Gusnanto A, Wood HM, Pawitan Y, *et al*. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics* 2012; **28:** 40–47.
29. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 2007; **23:** 657–663.
30. Altman DG. *Practical Statistics for Medical Research*. Chapman and Hall: London, 1991.
31. Ciriello G, Miller ML, Aksoy BA, *et al*. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet* 2013; **45:** 1127–1133.
32. Corcoran RB, Chabner BA. Application of cell-free DNA analysis to cancer treatment. *N Engl J Med* 2018; **379:** 1754–1765.
33. Wan JCM, Massie C, Garcia-Corbacho J, *et al*. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 2017; **17:** 223–238.
34. Pepe MS, Feng Z, Janes H, *et al*. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst* 2008; **100:** 1432–1438.
35. Zhang Z, Hao K. SAAS-CNV: a joint segmentation approach on aggregated and allele specific signals for the identification of somatic copy number alterations with next-generation sequencing data. *PLoS Comput Biol* 2015; **11:** e1004618.
36. Chan KC, Jiang P, Chan CW, *et al*. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci U S A* 2013; **110:** 18761–18768.
37. Chan KC, Jiang P, Zheng YW, *et al*. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* 2013; **59:** 211–224.
38. Leary RJ, Sausen M, Kinde I, *et al*. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* 2012; **4:** 162ra154.
39. Wang K, Lim HY, Shi S, *et al*. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* 2013; **58:** 706–717.
40. Llovet JM, Chen Y, Wurmbach E, *et al*. A molecular signature to discriminate dysplastic nodules from early hepatocellular carcinoma in HCV cirrhosis. *Gastroenterology* 2006; **131:** 1758–1767.
41. Wurmbach E, Chen YB, Khitrov G, *et al*. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* 2007; **45:** 938–947.
42. Marrero JA, Kulik LM, Sirlin CB, *et al*. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* 2018; **68:** 723–750.
43. Forner A, Vilana R, Ayuso C, *et al*. Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. *Hepatology* 2008; **47:** 97–104.
44. Kojiro M. Focus on dysplastic nodules and early hepatocellular carcinoma: an Eastern point of view. *Liver Transpl* 2004; **10:** S3–S8.
45. Hochhaus A, Saussele S, Rosti G, *et al*. Chronic myeloid leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017; **28:** iv41–iv51.
46. Senkus E, Kyriakides S, Ohno S, *et al*. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015; **26**(Suppl 5): v8–v30.
47. Wu YL, Planchard D, Lu S, *et al*. Pan-Asian adapted Clinical Practice Guidelines for the management of patients with metastatic non-small-cell lung cancer: a CSCO–ESMO initiative endorsed by JSMO, KSMO, MOS, SSO and TOS. *Ann Oncol* 2019; **30:** 171–210.
48. Fitzgerald RC, di Pietro M, Ragunath K, *et al*. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014; **63:** 7–42.
49. Lieberman DA, Rex DK, Winawer SJ, *et al*. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 2012; **143:** 844–857.
50. Joung JG, Ha SY, Bae JS, *et al*. Nonlinear tumor evolution from dysplastic nodules to hepatocellular carcinoma. *Oncotarget* 2017; **8:** 2076–2082.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Karyogram with associated density plot for sample representing normalised, smoothed, and segmented ratios

Figure S2. Cumulative frequency of genome-wide arm-level SCNAs in HCC tissue and cell-free DNA as well as dysplastic nodules

Figure S3. Tagged adaptors (referred to in Supplementary materials and methods)

Figure S4. Forward and reverse adaptors ligated to DNA inserts (referred to in Supplementary materials and methods)

Figure S5. Universal primer annealing with one of the two attachment sites to the flow cell (P5) (referred to in Supplementary materials and methods)

Figure S6. Index primer annealing to the enrichment PCR reaction mix; it adds the tag as well as the second flow-cell attachment site (P7) (referred to in Supplementary materials and methods)

Figure S7. Full DNA library design (referred to in Supplementary materials and methods)

Figure S8. The distribution of number of reads in test and reference genomes (referred to in Supplementary materials and methods)

Figure S9. Marked reduction in noise levels after trimming off poorly mapping windows (referred to in Supplementary materials and methods)

Table S1. Patient demographics

Table S2. Scanned slide identifiers, histopathology description, and source of tissue

Table S3. Sequencing. DNA quantity, DNA library quantity, quantity of library added to the pool, size selection, PCR enrichment, type of tag, paired sequencing, read length, sequencer model, total reads, and aligned reads

Table S4. Primer tags. Full primer sequences (mentioned only in Supplementary materials and methods)

Table S5. Tumour content of study cohorts. Predicted proportion of tumour DNA within the eluted pool of extracted DNA using ABSOLUTE

Table S6. Tumour content of the TCGA cohort. The proportion of tumour DNA within the eluted pool of DNA. Downloaded from Genomics Data Commons (https://gdc.cancer.gov/about-data/publications/pancanatlas)

Table S7. Test to reference ratio. The mean value of normalised and GC-corrected test to reference ratio for each autosomal arm

Table S8. Thresholds. A table outlining the 5th and 95th percentile thresholds for each chromosomal arm, below and above which losses and gains were called, respectively

Table S9. TCGA threshold validation sub-cohort

Table S10. Validation of frequent arm-level SCNAs on the blood-derived cell-free DNA cohort

Table S11. Data frame outlining normalised, smoothed, and segmented data within each window (referred to in Supplementary materials and methods)