Bridge Capture Permits Cost-Efficient, Rapid and Sensitive Molecular Precision Diagnostics

Simona Adamusová[1*], Anttoni Korkiakoski[1*], Tatu Hirvonen[1], Anna Musku[1], Tuula Rantasalo[1], Nea Laine[1], Jukka Laine[1,2,3], Juuso Blomster[1,2], Juha-Pekka Pursiheimo[1, **], Manu Tamminen[1,3, **, ***]

[1] Genomill Health Inc, Turku, Finland

[2] University of Turku, Finland

[3] Turku University Hospital, Finland

*, ** These authors contributed equally to the work

*** Corresponding author: manu@genomill.com

Author Contributions

Conceptualization and experimental design (SA, AK, TH, NL, JPP, MT), supervision and funding acquisition (JPP, MT), experiments (SA, TH, AM, TR), bioinformatics (AK, NL), visualization (AK, NL, SA), writing - original draft (SA, AK, JPP, MT), writing - review and editing (JB, JL). All authors commented and agreed on the manuscript.

Abstract

Liquid biopsies are gaining popularity as a less invasive alternative to tissue biopsies that have been the mainstay of cancer diagnostics to date. Recently, the quantification of mutations in circulating tumor DNA (ctDNA) by next-generation sequencing (NGS) has been gaining popularity. Targeted NGS approaches are preferable in ctDNA analysis as they provide greater sequencing depth and affordability compared to whole genome NGS. Targeted NGS can be achieved through various library preparation methods, each with distinct advantages and limitations. Here we introduce Bridge Capture, a novel technology that combines the advantages of market-leading liquid biopsy technologies while eliminating the need to compromise between scalability, cost-efficiency, sensitivity, or panel size. We compared Bridge Capture to leading commercial technologies currently available in cancer diagnostics; Archer™ LIQUIDPlex™ and AmpliSeq™ Cancer HotSpot Panel v2 for Illumina®. We found high mutant allele frequency (MAF) concordance as well as the lowest MAF among the three technologies on matched contrived colorectal biospecimens mimicking ctDNA. We showed the reproducibility of Bridge Capture by observing a high correlation between results from two independent laboratories. Additionally, we demonstrate the capability of Bridge Capture to affordably utilize bench-top sequencers for low MAF patient samples. Therefore, we believe that Bridge Capture will considerably enhance cancer diagnostics as a cost efficient, simple, rapid and sensitive precision diagnostic tool.

Keywords: liquid biopsy, ctDNA, cfDNA, mutant allele frequency, cancer diagnostics, NGS

Abbreviations

ANOVA	Analysis of variance
cfDNA	Cell-free DNA
CRC	Colorectal cancer
ctDNA	Circulating tumour DNA
dsDNA	Double-stranded DNA
gDNA	Genomic DNA
MAF	Mutant allele frequency
MIP	Molecular inversion probe
NGS	Next-generation sequencing
O/N	Overnight
PCR	Polymerase chain reaction
RCA	Rolling circle amplification
SNV	Single-nucleotide variant
ssDNA	Single-stranded DNA
UMI	Unique molecular identifier
WGS	Whole genome sequencing

Introduction

Until recently, tissue biopsies have been the primary method for diagnosing and monitoring cancer. However, a new family of techniques, collectively known as liquid biopsies, is gaining popularity as a less invasive alternative. Liquid biopsies, typically based on a simple blood draw, permit more feasible longitudinal disease monitoring which is crucial for tracking the disease progression and the effectiveness of treatments over time[1]. Moreover, liquid biopsies are unaffected by tumor heterogeneity – the spatial variation within a tumor – which is a downside for traditional biopsies[2].

A key component of liquid biopsies is circulating tumor DNA (ctDNA), which is a part of cell-free DNA (cfDNA) found in blood. ctDNA originates from tumor cells shedding their DNA either through apoptosis, necrosis, or active release and reflects the genetic makeup of the tumor in its entirety[2,3]. A key aspect of ctDNA analysis is quantifying mutant allele frequencies (MAFs). A given MAF indicates the proportion of a mutation to the non-mutated DNA in cfDNA. A higher MAF generally suggests a larger tumor burden, providing insight into the extent of the cancer in the body[3–6].

Next-generation sequencing (NGS) is a crucial tool for precision cancer diagnostics and is roughly divided into whole genome sequencing (WGS) and targeted sequencing. WGS plays a pivotal role for reference building and mutation discovery. When applied to cancer diagnostics, WGS is impractical due to high sequencing costs (typically 99.9 % of the reads obtained by WGS are irrelevant for cancer detection and prognosis) and complexity of data interpretation[7]. Targeted sequencing, on the other hand, focuses only on the regions of interest relevant to the disease, therefore permitting more affordable sequencing costs through improved sequencing depth utilization, and decreased data storage and analysis requirements. Importantly, targeting of the sequencing effort permits detecting mutations present at very low MAFs[8].

Targeted sequencing can be achieved through various approaches such as hybridization-based, amplicon-based, and molecular inversion probe-based NGS library preparation methods, each with distinct advantages and limitations. Amplicon-based methods target specific regions of interest via specifically designed primer pairs, typically followed by PCR[9,10], while hybridization methods use chemically modified oligonucleotide probes to enrich pre-specified parts of NGS libraries[11–13]. Amplicon-based techniques are generally limited in scalability due to primer cross-reactivity and are prone to false variant reads owing to increased error-rates coming from PCR amplification and NGS[7,14]. Hybridization-based techniques are generally more expensive and time-consuming, but permit extensive panels, and can more efficiently provide information from difficult genomic regions such as repeat sequences[8]. Molecular Inversion Probes (MIP) are single-stranded oligonucleotides that hybridize to target genomic region with terminal ends, and the resulting gap between the probe ends is filled and the molecule is circularized[15]. MIPs offer cost efficiency compared to hybridization-based methods, and improved panel size and scalability compared to amplicon-based methods. Their downsides include non-uniform coverage, high probe synthesis cost and increased noise compared to amplicon-based methods[16]. Given the shortcomings of the described methodologies, we recognize the need for a technology that combines the advantages of aforementioned methods, eliminating the need to compromise between scalability, cost-efficiency, sensitivity, or extensive panel size.

In this publication, we demonstrate analytical validation of Bridge Capture, a patented NGS library preparation workflow[17] that matches the performance of commercially available technologies while providing a workflow that combines the speed and simplicity of the amplicon- and MIP-based methods with the scalability of the hybrid capture methods.

Materials and Methods

Sample Preparation

All biospecimens were obtained from Indivumed GmbH (Hamburg, Germany). Total DNA from the tissue samples was isolated using QIAamp DNA Mini Kit (50) (Qiagen, Hilden, Germany). The extracted DNA was fragmented using Bioruptor® Pico (Daigenode, Liege, Belgium). 50 µL of sample was loaded in the 0.1 mL Bioruptor® Microtube (Diagenode) and sample was fragmented using 60 cycles with cycle conditions of 30”/30” to mimic size distribution of cfDNA. Processed samples were frozen at -80 °C and double-stranded DNA (dsDNA) concentration of the sample was measured the following day using Qubit 4 Fluorometer and Qubit dsDNA HS (high sensitivity) Assay Kit (ThermoFisher Scientific, Waltham, MA).

Panel Preparation

Bridge oligomer, left and right probes were synthesized by Integrated DNA Technologies (IDT, Coralville, IA). Corresponding left and right probes were mixed with the bridge oligomer in 96-well plate using automatized pipetting robot Opentrons OT-2 (Opentrons, Brooklyn, NY) and afterwards annealed in C1000 Touch Thermal Cycler (Bio-Rad, Hercules, CA). Each annealed probe pair and bridge were pooled by Opentrons OT-2 to prepare 25 nM panel stock solution for 282 probe panel and 10 nM panel stock solution for 887 probe panel. The list of 84 genes included in the 282 probe panel and number of probes covering the mutation of interest in the gene are provided in **Supplementary Table 1**. Similarly, the list of genes presented in 887 probe panel is available in **Supplementary Table 2**.

Bridge Capture and Library Preparation

320 ng of DNA (~100 000 copies of gDNA) was used as input for the analysis. Samples underwent overnight (O/N) target capture using Genomill's proprietary 282 probe panel (**Supplementary Table 1**). Gaps between the probes were filled using DNA polymerase and ligase. Non-circularized probe complexes were digested by thermolabile exonucleases. Fully circularized probe complexes were amplified by thermolabile Rolling Circle Amplification (RCA) enzyme. Long DNA concatemers were digested by thermolabile restriction enzyme. Sequencing libraries were made from digested molecules using a hot start DNA polymerase and indexed using customized Illumina primers and bead purified with Agencourt Ampure XP beads (Beckman Coulter, Brea, CA). Sequencing libraries were quantified by Qubit 4 Fluorometer and Qubit dsDNA HS (high sensitivity) Assay Kit (ThermoFisher Scientific), normalized and processed by Illumina sequencing platform (Illumina, United States).

Technology Comparison and Inter-lab Reproducibility

Two patient tissue CRC specimens were selected and underwent sample preparation. To confirm the processed samples contained the expected mutations, the samples were analyzed by Bridge Capture and were sequenced with iSeq 100 sequencer using iSeq 100 i1 Reagent v2 (300-cycle) 2 x 150 bp in paired-end run (Illumina). Presence of *APC* p.Q1406*, *KRAS* p.G13D, *PIK3CA* p.E545K and *APC* p.E1309Dfs*4, *KRAS* p.G12D, *TP53* p.Y126D was confirmed in sample 1 and sample 2, respectively. Patient samples were used as is or they were diluted with gDNA in various ratios to mimic different levels of MAFs in tenfold, hundredfold, and thousandfold dilution (**Supplementary Table 3**). Each sample was tested in triplicates for each technology using Bridge Capture, Archer™ LIQUIDPlex™ (IDT), AmpliSeq™ Cancer HotSpot Panel v2 for Illumina® (Illumina). For the Bridge Capture test, 282 probe panel was

used. The starting material load for Bridge Capture and Archer™ LIQUIDPlex™ was 320 ng. To stay within the recommended load range, 100 ng of starting material was used for AmpliSeq™ Cancer HotSpot Panel v2 for Illumina®. All the samples were processed by each technology at an independent diagnostic service provider in the UK. For Archer™ LIQUIDPlex™ (IDT) and AmpliSeq™ Cancer HotSpot Panel v2 for Illumina® sequencing was performed using Illumina NextSeq 500 and for Bridge Capture NextSeq 550 was used. NextSeq 500/550 High Output Kit v2.5 (300-cycle), 2 x 151 bp paired-end run (Illumina) was used for all three sequencing runs.

Aliquots from the same samples were kept at Genomill for later processing by Bridge Capture to evaluate inter-lab reproducibility. Sequencing libraries prepared at Genomill were sequenced using the MiSeq system, with MiSeq Reagent Kit v3 using a 2 x 150 bp paired-end run (Illumina). Both laboratories processed the samples using a ready-made Bridge Capture kit, containing the probe mixture, reagents, and protocol.

Panel Characteristics

To characterize the panel and evaluate the panel size scalability, 282 probe panel and 887 probe panel were compared. 320 ng of gDNA was used as starting material. In the experiment, 10 technical replicates of gDNA were analyzed with both panels by Bridge Capture. Obtained libraries were normalized and sequenced using the MiSeq system, with MiSeq Reagent Kit v3 using a 2 x 150 bp paired-end run (Illumina).

Hybridization Duration Comparison

CRC specimen covering single-nucleotide variants (SNVs): *KRAS* p.G12V, *APC* R876* and *APC* E1408* was diluted tenfold with gDNA, to simulate MAF ten times lower than the original. To determine the impact of hybridization duration on detection limit of the presented

method, 320 ng of diluted specimen was analyzed by Bridge Capture using 282 probe panel under various incubation durations - O/N, 4 hours, 2 hours, 1 hour and 0.5 hours. Each timepoint was carried out in five replicates. Produced libraries were sequenced using the MiSeq system, with MiSeq Reagent Kit v3 using a 2 x 150 bp paired-end run (Illumina).

Data Analysis

Bridge Capture reads were merged using VSEARCH[18] (v2.15.2_linux_x86_64) with the following parameters: --fastq_minovlen 10 --fastq_maxdiffs 15 --fastq_maxee 1 --fastq_allowmergestagger. A proprietary pipeline utilising Unique Molecular Identifier (UMI) based error-correction was used to process the Bridge Capture data. Both Archer™ LIQUIDPlex™ and AmpliSeq™ Cancer HotSpot Panel v2 for Illumina® data were processed by the independent diagnostic service provider in the UK by using ArcherDX's Archer Analysis Unlimited (v6.0.3.1) and Illumina's DNA Amplicon (v2.1.1) data processing pipelines with default parameters (SNV cutoff LIQUIDPlex 0.5 %, Ampliseq 1 %). For the figures and statistics, Python version 3.11.5 was used. The figures were drawn using matplotlib (v3.7.2) library, R^2 scores were calculated using sklearn (v1.2.2) library, and scipy (v1.11.1) library was used for linear regressions and Analysis of variance (ANOVA).

In silico sub-sampling was done without replacement to include 10 %, 1 % and 0.1 % of the original forward and reverse reads and then processed as mentioned before.

Results

Bridge Capture

Bridge Capture is a targeted NGS library preparation method that uses oligonucleotide probe constructs to target specific genomic regions of interest[17]. The probe construct consists of a bridge oligomer annealed to left and right probes (**Fig. 1a**). The bridge includes sequences complementary to the probes, a restriction enzyme site and various modifications protecting the bridge from exonuclease activity or unwanted elongation by polymerase. Left and right probes contain bridge binding sites, binding sites for sequencing platform specific adapters, UMI sequences, and the target specific binding sites.

In the first step of the Bridge Capture, probe constructs are introduced to a sample containing the targets of interest (cfDNA/DNA isolated from blood and other sources; step A in **Fig. 1b**). Successful targeting by Bridge Capture construct results in a gap between the left and right target specific probes with the mutation of interest located within the gap region of targeted DNA strand (step B in **Fig. 1b**). Next, the gap is filled by DNA polymerase, starting from the 3' end of left probe and continuing towards the 5' end of right probe. The newly transcribed sequence is ligated to the 5' end of the right probe by a DNA ligase. The gap between 5' end of the left and 3' end of the right probe, held together by bridge oligo, is also filled by a DNA polymerase and ligated by a DNA ligase. This circularizes the probe construct and captures the mutation(s) of interest. Afterwards, all the non-circularized constructs are digested by exonucleases (step C in **Fig. 1b**). Rolling circle amplification (RCA) is initiated from 3' end of bridge oligomer of circularized construct (step D in **Fig. 1b**). RCA generates multiple copies of the circular construct by creating long concatemeric single-stranded DNA (ssDNA). After RCA, ssDNA concatemer is digested into monomers using a restriction enzyme (the restriction recognition site being provided in the bridge oligomer; step E in **Fig. 1b**). Sequencing platform specific adapters are introduced to the monomers through

limited-cycle indexing PCR (step F in **Fig. 1b**). The resulting libraries are purified, pooled, and sequenced on the NGS platform of choice (step G in **Fig. 1b**).

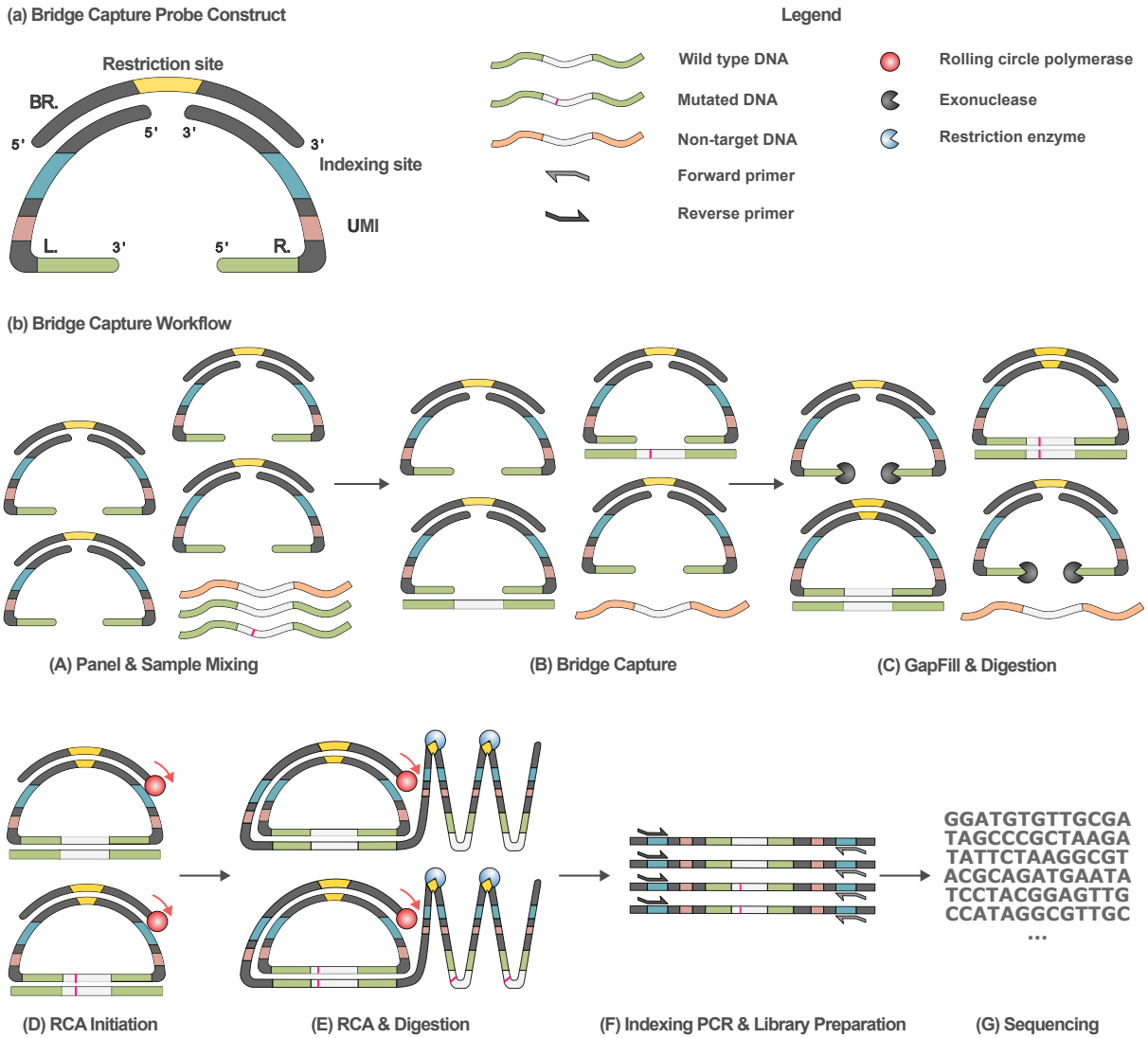


Figure 1. Bridge Capture in summary. (**a**) Bridge Capture probe construct consists of a bridge oligomer (BR.) with binding sites to left (L.) and right probe (R.). The bridge contains the restriction recognition site for restriction enzyme (yellow). Left and right probe contain sequences for binding of dual index primers (blue), UMIs (pink) and target specific regions (green). (**b**) Bridge Capture workflow. The probe constructs are mixed with a sample containing sequences of interest (step A – Panel & Sample Mixing), to which the Bridge Capture probe constructs hybridize (step B – Bridge Capture). After Bridge Capture is complete, DNA polymerase and DNA ligase are used to fill all the gaps in the probe constructs to create circularized molecules. The constructs that failed to hybridize to target DNA are digested by addition of exonucleases (step C – GapFill & Exonuclease Digestion). RCA is initiated from the bridge oligomer of the circularized construct (step D – RCA Initiation). A long single stranded concatemer is synthetized by RCA and subsequently digested by a restriction enzyme creating multiple copies of a monomer (step E - RCA & Digestion). Newly produced monomers are indexed with dual indexing primers to create a library viable for sequencing (step F – Indexing PCR & Library Preparation). As the final part of the workflow, the library is sequenced (step G – Sequencing).

Technology Concordance

The performance of Bridge Capture was assessed by comparing it to commercially available cancer diagnostics technologies; Archer™ LIQUIDPlex™ (from now on referred to as LIQUIDPlex) and AmpliSeq™ Cancer HotSpot Panel v2 for Illumina® (from now on referred to as AmpliSeq), to establish the concordance between MAFs (**Fig. 2**) and sensitivities (**Table 1**). The comparison was performed using dilutions of CRC biospecimens into gDNA (respective to genomic copies) to mimic different levels of MAFs (**Supplementary Table 3**), resulting in 24 samples in total.

MAFs determined by the commercially available technologies were strongly correlated to the MAFs determined by Bridge Capture, with R^2 = 0.9951 between LIQUIDPlex and Bridge Capture, and R^2 = 0.9879 between AmpliSeq and Bridge Capture (**Fig. 2**). LIQUIDPlex did not target any *APC* mutations and AmpliSeq did not target *APC* p.Q1406*, and therefore these mutations were not included in the comparison.

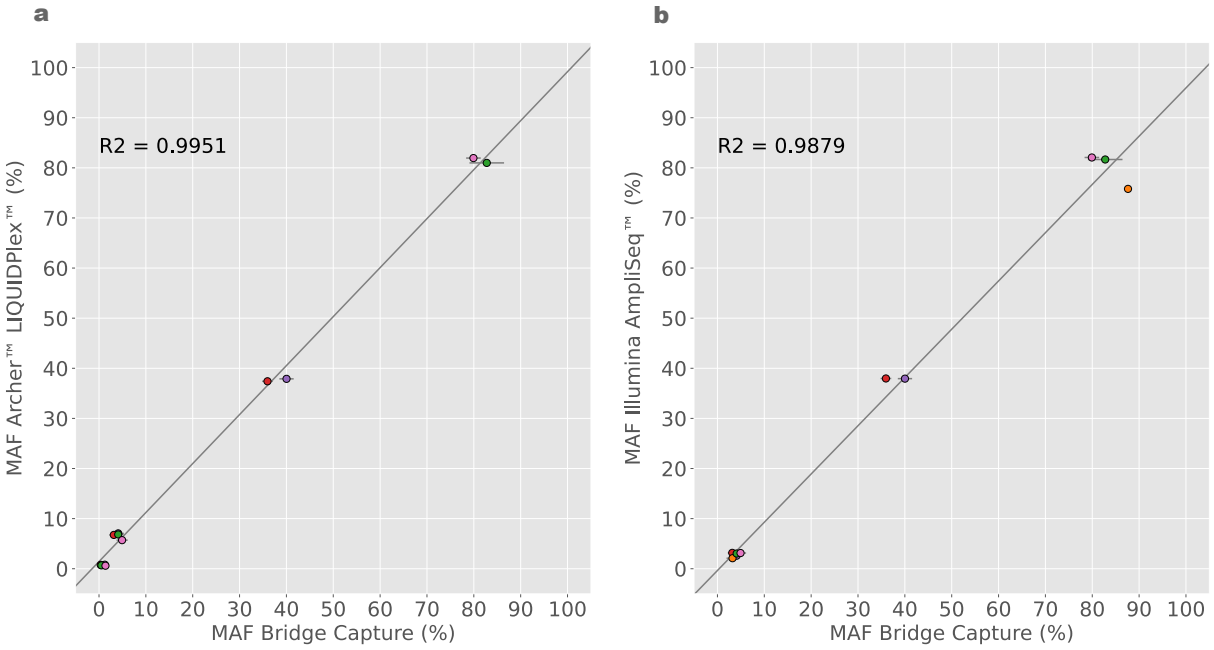


Figure 2. Concordance between MAFs detected by Bridge Capture and commercially available technologies. (**a**) Concordance between MAFs detected by Archer™ LIQUIDPlex™ assay plotted on y-axis and corresponding MAFs detected by Bridge Capture reported on x-axis. (**b**) Concordance between MAFs detected by AmpliSeq™ Cancer HotSpot Panel v2 for Illumina® reported on y-axis and corresponding MAFs detected by Bridge Capture on x-axis.

All technologies consistently detected single nucleotide variants (SNVs) > 2 % MAF. The lowest MAF identified by Bridge Capture, LIQUIDPlex and AmpliSeq were 0.08 %, 0.58 % and 2.0 %, respectively. The only deletion (*APC* p.E1309Dfs*4) was detected by both Bridge Capture and AmpliSeq at MAF above 2 % (**Table 1**). All technologies exhibited high linear correlation between the sample dilution factor and MAF, implying consistent linear performance across wide range of MAFs. Average R^2 of the detected variants was 0.9610, 0.9996, and 0.9950 for Bridge Capture, Ampliseq, and LIQUIDPlex, respectively.

Table 1. Table of detected MAFs of Bridge Capture, Archer™ LIQUIDPlex™ and AmpliSeq™ Cancer HotSpot Panel v2 for Illumina®. MAF is indicated for each dilution (No dilution, 10x, 100x, 1000x) of each specified gene and its corresponding variant, along with the coefficient of determination R^2 value.

Gene	Variant	Detected MAF (%)														
		Bridge Capture					Archer™ LIQUIDPlex™					Illumina AmpliSeq™				
		No dilution	10x	100x	1000x	R^2	No dilution	10x	100x	1000x	R^2	No dilution	10x	100x	1000x	R^2
APC	p.Q1406*	73.9 ± 2.2	6.0 ± 0.7	0.5 ± 0.0	-	0.999	-	-	-	-	-	-	-	-	-	-
KRAS	p.G13D	40.0 ± 1.5	4.1 ± 1.2	1.2 ± 0.2	0.2	0.974	37.9 ± 0.5	7.1 ± 0.2	0.8 ± 0.1	-	0.992	37.9 ± 0.3	2.6 ± 0.0	-	-	1.0
PIK3CA	p.E545K	36.0 ± 1.1	3.2 ± 1.0	0.4 ± 0.2	0.1	0.980	37.4 ± 0.3	6.7 ± 0.2	0.8 ± 0.1	-	0.993	38.0 ± 0.1	3.2 ± 0.1	-	-	1.0
APC	p.E1309Dfs*4	87.6 ± 0.8	3.2 ± 1.3	-	-	0.981	-	-	-	-	-	75.8 ± 0.3	2.1 ± 0.1	-	-	1.0
KRAS	p.G12D	82.8 ± 3.7	4.1 ± 1.1	0.5 ± 0.3	0.5	0.875	81.0 ± 0.5	6.8 ± 0.1	0.7 ± 0.1	-	0.998	81.7 ± 0.7	3.0 ± 0.4	-	-	0.998
TP53	p.Y126D	79.9 ± 1.6	4.9 ± 1.2	1.1 ± 0.3	-	0.957	81.9 ± 0.4	5.7 ± 0.1	0.6	-	0.997	82.1 ± 0.4	3.1 ± 0.1	-	-	1.0

To determine the sequencing depth required by Bridge Capture for detecting low MAFs, the raw sequencing data was sub-sampled (**Table 2**). At 10 % subsampling, corresponding to approximately 750 000 reads, Bridge Capture could detect MAFs below 0.1 % for *KRAS* p.G13D, *PIK3CA* p.E545K and *KRAS* p.G12D and results were comparable with the non-subsampled data. At 1 % subsampling, corresponding to approximately 75 000 reads, Bridge Capture could detect MAFs down to 1 % for all mutations except *APC* p.E1309Dfs*4.

Table 2. Table representing required sequencing depth by Bridge Capture for detecting low MAFs. Raw reads obtained by sequencing (~7.5 million reads) were subsampled by 10 % (~750 000 reads), 1 % (~75 000 reads) and 0.1 % (~7 500 reads).

Gene	Variant	Detected MAF (%) by Bridge Capture														
		10% subsampling (~750 000 reads)					1% subsampling (~75 000 reads)					0.1 % subsampling (~7500 reads)				
		No dilution	10x	100x	1000x	R^2	No dilution	10x	100x	1000x	R^2	No dilution	10x	100x	1000x	R^2
APC	p.Q1406*	73.7 ± 2.6	5.9 ± 0.8	0.5 ± 0.1	-	0.997	75.2 ± 3.0	6.4 ± 0.8	0.5 ± 0.3	-	0.984	74.1 ± 8.9	10.5 ± 5.4	3.8	-	0.925
KRAS	p.G13D	39.0 ± 0.7	4.1 ± 1.3	1.2 ± 0.1	0.1	0.97	39.5 ± 3.1	3.9 ± 2.4	1.0	-	0.922	32.3 ± 7.5	6.7	11.1	-	0.599
PIK3CA	p.E545K	36.1 ± 0.7	2.9 ± 0.9	0.4 ± 0.1	0.1	0.988	34.3 ± 3.7	2.7 ± 1.3	0.4 ± 0.2	-	0.948	30.4 ± 2.8	5.7	2.6	-	0.971
APC	p.E1309Dfs*4	86.9 ± 0.3	3.2 ± 0.7	-	-	0.994	82.4 ± 2.5	2.2 ± 0.6	-	-	0.994	78.6 ± 10.4	-	-	-	-
KRAS	p.G12D	82.8 ± 3.2	3.8 ± 1.2	0.5 ± 0.3	0.6	0.852	86.2 ± 3.0	2.9 ± 2.1	-	0.8	0.723	91.7 ± 14.4	-	-	-	-
TP53	p.Y126D	79.7 ± 1.7	4.6 ± 0.9	1.2 ± 0.4	-	0.949	77.4 ± 1.3	4.3 ± 0.4	1.8 ± 0.5	-	0.908	81.4 ± 10.0	4.3 ± 1.6	5.6	-	0.709

Reproducibility and Hybridization Duration

The reproducibility of Bridge Capture was assessed by comparing the results from an assay performed by an independent diagnostic service provider in the UK (site A) to the results from Genomill laboratory (site B) (**Fig. 3a**). The detected MAFs were strongly correlated between the laboratories ($R^2 = 0.9792$).

The effect of hybridization time on Bridge Capture performance was tested by detecting MAFs using six different incubation times (overnight (O/N), 4 h, 2 h, 1.5 h, 1 h, 0.5 h) in five replicates. Mutations of interest were successfully identified in all five replicates, excluding one replica in the 0.5 h incubation time point, where the *APC* E1408* was not detected (**Fig. 3b**). The standard deviation of the MAFs detected between the five replicates is lowest for replicates incubated O/N and 4 h and the standard deviation between replicates is increasing with shorter incubation time and its peak can be observed at 0.5 h time point. Only the detection of *KRAS* p.G12V was significantly affected by the incubation time (one-way ANOVA test, $p < 0.05$).

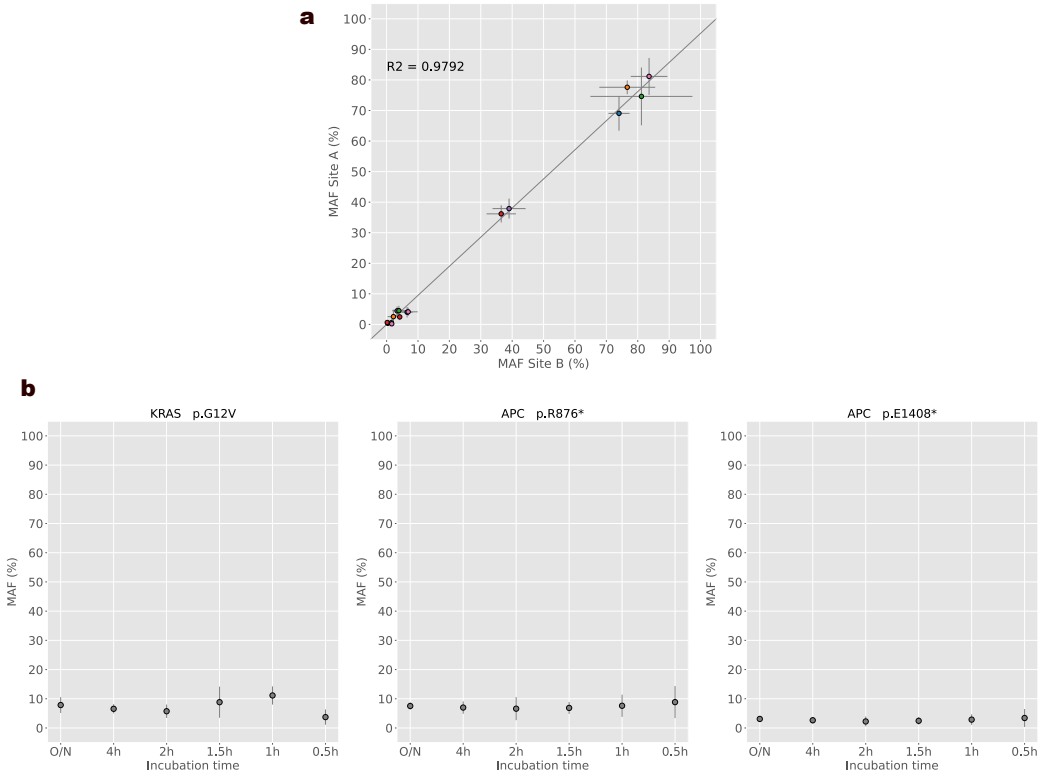


Figure 3. (**a**) The inter-lab reproducibility of Bridge Capture was confirmed by detecting a strong correlation of MAFs from an independent diagnostic service provider laboratory site in UK (site A) and Genomill laboratory, Turku, Finland (site B) ($R^2 = 0.9792$). **(b)** Effect of hybridization time on Bridge Capture performance was tested by detecting MAFs, performing incubation for overnight (O/N), 4 h, 2 h, 1.5 h, 1 h and 0.5 h. Average MAF of a given incubation time point was calculated based on 5 replicates and standard deviation between replicates is displayed as error bars. The incubation time had a significant effect on *KRAS* p.G12V (one-way ANOVA, $p < 0.05$) and did not have a significant effect on the other mutations.

Panel Characteristics

Signal intensity was established for each probe from ten replicates of healthy gDNA (**Fig. 4a** and **Fig. 4b**) for panels of 282 and 887 probes, by dividing the read count per probe by the total read count of each replicate. Probe signal intensity was variable between the probes, with median intensity 0.324, compared to theoretical intensity of 0.355 assuming a uniform performance for the 282-probe panel. In comparison, for the 887-probe panel, the average intensity was 0.113 compared to the theoretical intensity of 0.087. Signal intensity per probe

between replicates was notably consistent, with an average standard deviation of 0.0267 per probe between ten replicates for the 282-probe panel and 0.0091 for the 887-probe panel.

A relationship between probe signal intensity and the respective GC content of the target region (defined as 300 nt around the probe binding site) was observed (**Fig. 4c** and **Fig. 4d**). Most of the probes with average probe signal intensity had a target region GC content around 30–54 %. Probe signal intensity started decreasing around target region GC content of 55–59 % and 25–29 %. Significant decline was observed for target region GC content > 60 % and was severely impaired in the range of 70–79 %.

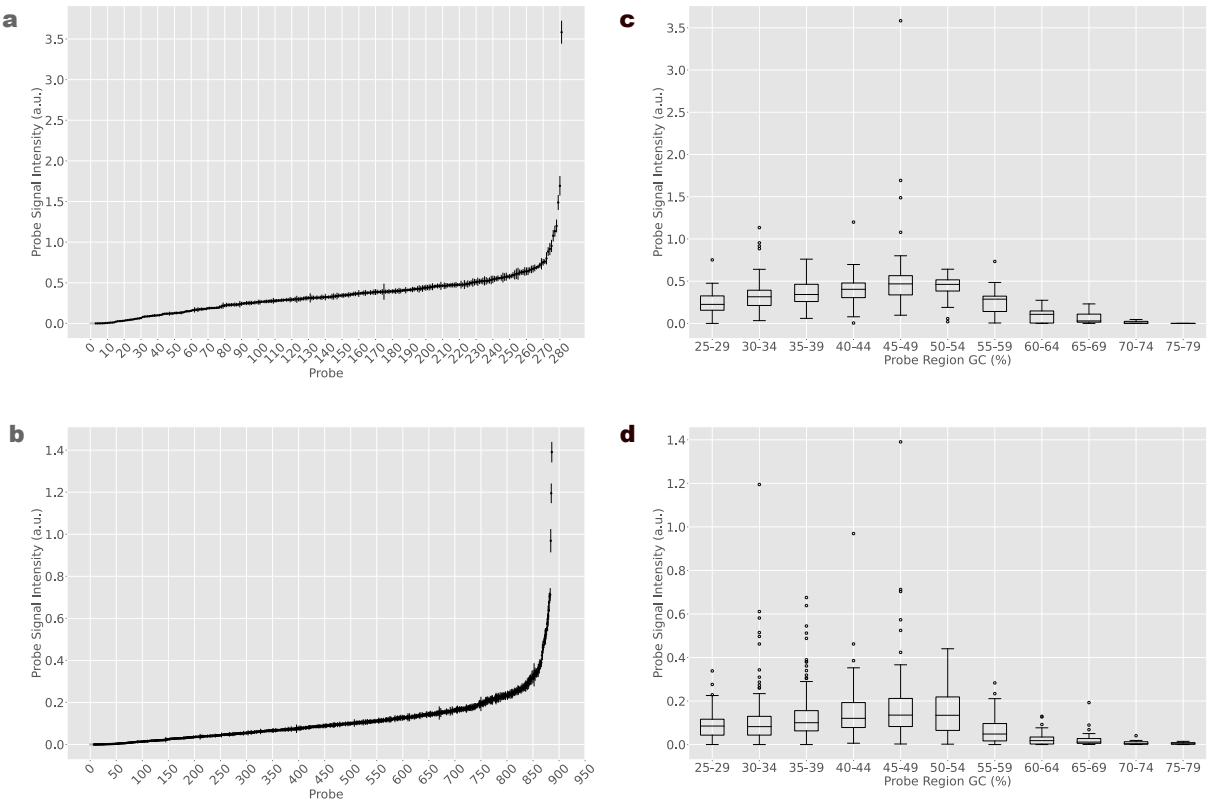


Figure 4. Probe signal intensity across ten technical replicates for (**a**) 282 probe panel and (**b**) 887 probe panel. Probe signal intensity (y-axis) was calculated by dividing the read count per probe by the total read count of the replicate and mean from all the replicates was determined. Standard deviation between the replicates was drawn as error bar. The relationship between probe signal intensity and GC content of the probe target region (300 nt around the probe binding site) of (**c**) 282 probe panel and (**d**) 887 probe panel.

The 282-probe panel covers 64 cancer types and associated driver genes obtained from intOgen-framework[21] and their respective mutations of 1 and 2 significance which are based on COSMIC database, CMC v.99. A majority of cancer type mutations are covered at least 50 % by the 282-probe panel (**Supplementary Table 1**). Expanded 878-probe panel covers almost fully 71 driver genes (**Supplementary Table 2**) with a majority of the cancer types (65/71) covered by at least 90 % including most common cancer types such as colorectal adenocarcinoma, breast carcinoma, non-small cell lung cancer, prostate adenocarcinoma, and ovarian cancer.

Discussion

Various methods are available for NGS-based liquid biopsies, each with their advantages and limitations. For instance, amplicon-based technologies such as AmpliSeq and TruSeq Amplicon are simple and rapid in terms of their workflow but provide limited scalability to a large number of targets[22–24]. Hybrid capture methods such as AVENIO and FoundationACT, provide scalability to a very large number of target genes but are slow, cumbersome and expensive[12,13,25,26]. MIP-based workflows are rapid and simple but suffer from high probe synthesis costs as well as poor uniformity[16,27].

Bridge Capture addresses the challenges of hybridization-, amplicon-and MIP-based technologies and offers a simple, cost-efficient, single-day, automatable workflow with high performance. We demonstrated that Bridge Capture matches or surpasses the performance of commercially available leading technologies for cancer diagnostics such as LIQUIDPlex and AmpliSeq, for example Bridge Capture identified certain mutations that were below cutoff limit of LIQUIDPlex (> 0.5 %) and AmpliSeq (> 1 %). We also demonstrated that Bridge Capture is easily adoptable in a new laboratory setting, shown by the highly correlated MAFs detected by two independent laboratories. High correlation was achieved despite use of

different sequencing platforms, over 9-fold difference in sequencing depth, and operators having different levels of experience.

The sequencing depth required by Bridge Capture is low, with approximately 750 000 reads sufficient for detection of below 0.1 % MAF mutations with a panel of 282 probes. The low sequencing depth requirement permits pooling tens of samples even on benchtop devices such as Illumina MiniSeq, and up to thousands of samples on production-scale sequencers such as NovaSeq X. This is of high interest to small and medium scale hospitals and laboratories where the weekly or monthly sample volume would not be sufficient for filling up runs on production-scale sequencers.

We demonstrate panel expansion from 282 to 887 probes. Despite the panel coverage increasing by more than 300 %, the panel uniformity remains unchanged, implying that the Bridge Capture panels can be scaled without any theoretical or practical upper limit. Performance between probes is variable, with certain probes over- or underperforming, and could be mitigated by probe design. We observed that the probe performance is related to the high or low GC content of the probe target region. This performance effect could result from the indexing PCR[28–31], since PCR has been shown to deplete loci with GC content > 65 %[30,31]. Another potential source for the probe performance imbalance could be the sequencing by synthesis on Illumina platforms, which is also negatively impacted by high-GC content[29,30,32]. Outliers above the average performance could partly be explained by the binding site overrepresentation by pseudogenes, for instance the probe targeting NF1 gene, which is a known homologous pseudogene located through the human genome[19,20].

We foresee certain research directions to further improve and optimize the Bridge Capture technology. For instance, we expect to further increase the sensitivity of Bridge Capture by refining our probe design and optimizing the reaction process to exhaustively target all DNA molecules in a sample.

Acknowledgements

The authors would like to express their gratitude to Voima Ventures, Avohoidon Tutkimussäätiö and Almaral for support and funding.

References

1. Heitzer, E., Haque, I. S., Roberts, C. E. S. & Speicher, M. R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88 (2019).
2. Cheng, F., Su, L. & Qian, C. Circulating tumor DNA: a promising biomarker in the liquid biopsy of cancer. *Oncotarget* **7**, 48832–48841 (2016).
3. McEvoy, A. C. *et al.* Correlation between circulating tumour DNA and metabolic tumour burden in metastatic melanoma patients. *BMC Cancer* **18**, 726 (2018).
4. Bittla, P. *et al.* Exploring Circulating Tumor DNA (CtDNA) and Its Role in Early Detection of Cancer: A Systematic Review. *Cureus* **15**, e45784.
5. Strijker, M. *et al.* Circulating tumor DNA quantity is related to tumor volume and both predict survival in metastatic pancreatic ductal adenocarcinoma. *Int. J. Cancer* **146**, 1445–1456 (2020).
6. Valpione, S. *et al.* Plasma total cell-free DNA (cfDNA) is a surrogate biomarker for tumour burden and a prognostic biomarker for survival in metastatic melanoma patients. *Eur. J. Cancer* **88**, 1–9 (2018).
7. Song, P. *et al.* Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics. *Nat. Biomed. Eng.* **6**, 232–245 (2022).
8. Bewicke-Copley, F., Arjun Kumar, E., Palladino, G., Korfi, K. & Wang, J. Applications and analysis of targeted genomic sequencing in cancer studies. *Comput. Struct. Biotechnol. J.* **17**, 1348–1359 (2019).

9. Vendrell, J. A. *et al.* Detection of known and novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches. *Sci. Rep.* **7**, 12510 (2017).

10. Zheng, Z. *et al.* Anchored multiplex PCR for targeted next-generation sequencing. *Nat. Med.* **20**, 1479–1484 (2014).

11. Lu, W., Zhu, M., Chen, Y. & Bai, Y. A novel approach to improving hybrid capture sequencing targeting efficiency. *Mol. Cell. Probes* **46**, 101424 (2019).

12. Cottrell, C. E., Bredemeyer, A. J. & Al-Kateb, H. Chapter 19 - Targeted Hybrid-Capture for Somatic Mutation Detection in the Clinic. in *Clinical Genomics* (eds. Kulkarni, S. & Pfeifer, J.) 321–341 (Academic Press, 2015). doi:10.1016/B978-0-12-404748-8.00019-8.

13. Clark, T. A. *et al.* Analytical Validation of a Hybrid Capture–Based Next-Generation Sequencing Clinical Assay for Genomic Profiling of Cell-Free Circulating Tumor DNA. *J. Mol. Diagn.* **20**, 686–702 (2018).

14. Poh, J. *et al.* Analytical and clinical validation of an amplicon-based next generation sequencing assay for ultrasensitive detection of circulating tumor DNA. *PLoS ONE* **17**, e0267389 (2022).

15. Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M. & Babik, W. Molecular Inversion Probes for targeted resequencing in non-model organisms. *Sci. Rep.* **6**, 24051 (2016).

16. Biezuner, T. *et al.* An improved molecular inversion probe based targeted sequencing approach for low variant allele frequency. *NAR Genomics Bioinforma.* **4**, lqab125 (2022).

17. Pursiheimo, J.-P., Hirvonen, T., Tamminen, M. & Korkiakoski, A. Highly sensitive methods for accurate parallel quantification of nucleic acids. (2022).

18. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

19. Rossi, S. *et al.* Neurofibromin C terminus-specific antibody (clone NFC) is a valuable tool for the identification of NF1-inactivated GISTs. *Mod. Pathol.* **31**, 160–168 (2018).

20. Cunha, K. S. *et al.* Hybridization Capture-Based Next-Generation Sequencing to Evaluate Coding Sequence and Deep Intronic Mutations in the NF1 Gene. *Genes* **7**, 133 (2016).

21. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).

22. Wei, J. *et al.* Clinical Value of EGFR Copy Number Gain Determined by Amplicon-Based Targeted Next Generation Sequencing in Patients with EGFR-Mutated NSCLC. *Target. Oncol.* **16**, 215–226 (2021).

23. Chang, F. & Li, M. M. Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genet.* **206**, 413–419 (2013).

24. Wing, M. R. *et al.* Analytic validation and real-time clinical application of an amplicon-based targeted gene panel for advanced cancer. *Oncotarget* **8**, 75822–75833 (2017).

25. Verma, S. *et al.* Analytical performance evaluation of a commercial next generation sequencing liquid biopsy platform using plasma ctDNA, reference standards, and synthetic serial dilution samples derived from normal plasma. *BMC Cancer* **20**, 945 (2020).

26. Zulato, E. *et al.* Implementation of Next Generation Sequencing-Based Liquid Biopsy for Clinical Molecular Diagnostics in Non-Small Cell Lung Cancer (NSCLC) Patients. *Diagnostics* **11**, 1468 (2021).

27. Turner, E. H., Lee, C., Ng, S. B., Nickerson, D. A. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 individuals. *Nat. Methods* **6**, 315–316 (2009).

28. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques* **52**, 87–94 (2012).

29. Kozarewa, I. *et al.* Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* **6**, 291–295 (2009).

30. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).

31. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72 (2012).

32. Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y. & Hwang, C.-C. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *PLOS ONE* **8**, e62856 (2013).

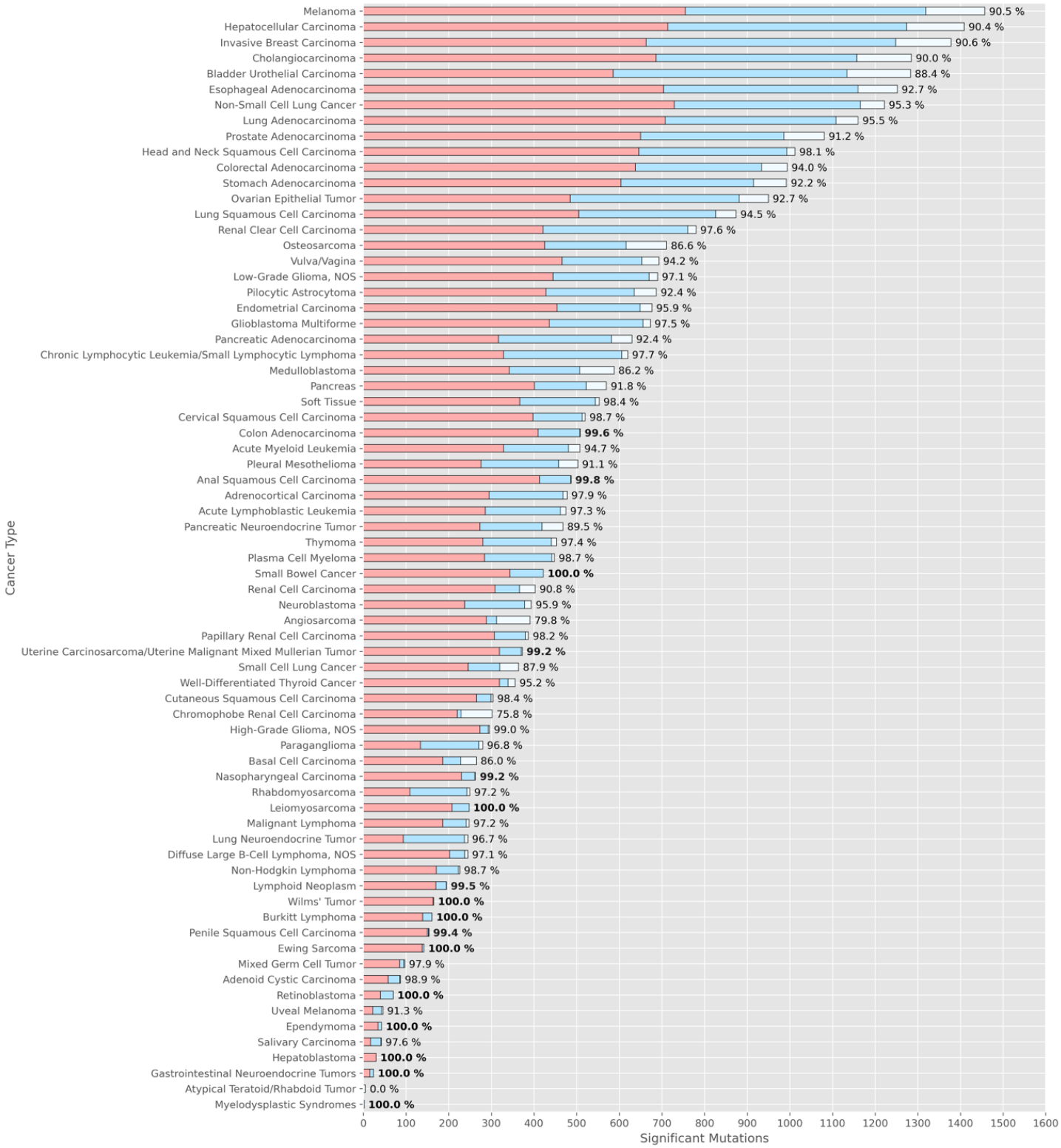
Supplementary Table 1. Composition of 282 probe panel. Table represent the list of genes covered in the panel, alongside the corresponding number of probes.

84 gene panel consisting of 282 probes.							
Gene	Probes	Gene	Probes	Gene	Probes	Gene	Probes
ABL1	1	CTNNB1	2	KMT2D	5	PTPN11	1
ALK	4	CUL3	2	KRAS	3	RAC1	1
APC	39	DICER1	1	MAP2K1	1	RAF1	1
AR	1	DNMT3A	7	MET	1	RB1	12
ASXL1	6	EGFR	4	MLH1	6	RBM10	1
ATM	11	EP300	1	MSH2	10	RUNX1	1
ATR	1	ERBB2	1	MSH6	2	SETD2	1
B2M	1	ERBB3	2	MTOR	1	SF3B1	3
BAP1	1	EZH2	1	MUTYH	1	SMAD2	1
BCL2L2	1	FANCA	1	NBN	1	SMAD4	9
BRAF	2	FBXW7	5	NF1	18	SPEN	1
BRCA1	6	FGFR2	2	NF2	2	SPOP	2
BRCA2	5	FGFR3	1	NFE2L2	2	STAG2	1
CBL	1	FLT3	2	NFKBIA	1	STK11	2
CDC73	1	GNAQ	1	NOTCH1	4	TERT	1
CDH1	3	GNAS	2	NRAS	2	TP53	16
CDK12	1	HRAS	1	PALB2	3	TSC1	2
CDKN1B	1	IDH1	1	PDGFRA	1	U2AF1	2
CDKN2A	3	JAK2	1	PIK3CA	4	VHL	4
CHEK2	5	KEAP1	1	PTCH1	3	WT1	1
CREBBP	2	KIT	2	PTEN	10	XPO1	1

Supplementary Table 2. Composition of 887 probe panel. Table represent the list of genes covered in the panel, alongside the corresponding number of probes.

123 gene panel consisting of 887 probes.							
Gene	Probes	Gene	Probes	Gene	Probes	Gene	Probes
ABL1	2	CTNNB1	2	KEAP1	1	PPP2R1A	1
AKT1	1	CUL3	2	KIT	2	PTCH1	3
ALK	4	CYLD	4	KLF4	1	PTEN	19
APC	81	DICER1	12	KMT2D	6	PTPN11	3
AR	3	DNMT3A	9	KRAS	3	RAC1	1
ARID1A	1	EGFR	9	LZTR1	1	RAF1	1
ARID1B	1	EIF3E	1	MAP2K1	2	RB1	36
ASXL1	6	EP300	9	MAP3K1	1	RBM10	1
ATM	83	ERBB2	6	MAPK1	1	RET	1
ATR	1	ERBB3	4	MAX	4	RHOA	1
B2M	1	ESR1	1	MEN1	8	RNF43	2
BAP1	17	EZH2	1	MET	1	RUNX1	8
BCL2L2	1	FANCA	1	MLH1	8	SETD2	3
BCL9	1	FBXW7	5	MSH2	11	SF3B1	4
BCOR	1	FGFR1	2	MSH6	3	SMAD2	1
BMPR1A	7	FGFR2	4	MTOR	4	SMAD4	18
BRAF	3	FGFR3	3	MUTYH	1	SMARCA4	16
BRCA1	47	FH	1	MYC	1	SPEN	1
BRCA2	70	FLCN	5	NBN	1	SPOP	2
BTK	1	FLT3	2	NF1	100	STAG2	1
CASP8	3	GATA1	1	NF2	14	STK11	14
CBL	1	GATA3	1	NFE2L2	2	TERT	1
CDC73	1	GNA11	1	NFKBIA	1	TGFBR2	1
CDH1	19	GNAQ	2	NOTCH1	5	TP53	19
CDK12	1	GNAS	2	NOTCH2	1	TSC1	2
CDK4	1	HRAS	2	NRAS	2	TSC2	1
CDKN1B	2	IDH1	1	PALB2	3	U2AF1	2
CDKN2A	8	IDH2	2	PDGFRA	1	VHL	9
CHEK2	5	IRS4	1	PIK3CA	13	WT1	1
CREBBP	23	JAK2	1	PIK3R1	4	XPO1	1
CTCF	1	KDM6A	1	PMS2	12		

Supplementary Table 3. List of the contrived CRC samples analysed by Bridge Capture, Archer™ LIQUIDPlex™ and AmpliSeq™ Cancer HotSpot Panel v2 for Illumina® to evaluate the concordance between technologies and establish the detection limits of the technologies. Table describes composition of each contrived sample and displays original patient MAFs detected in the tissue sample and how the representation of the original MAFs would change with the dilution with Healthy genomic DNA.

Experimental design				
Sample number	**Patient**	**Patient copies**	**Healthy genomic DNA copies**	**Representation of % of original patient MAF detected in tissue**
1 2 3	Patient 1	100 000	-	85.9% APC p.Q1406*, 43.2% KRAS p.G13D, 35% PIK3CA p.E545K
4 5 6	Patient 1	10 000	90 000	8.59% APC p.Q1406*, 4.32% KRAS p.G13D, 3.5% PIK3CA p.E545K
7 8 9	Patient 1	1 000	100 000	0.859% APC p.Q1406*, 0.432% KRAS p.G13D, 0.35% PIK3CA p.E545K
10 11 12	Patient 1	100	100 000	0.0859% APC p.Q1406*, 0.0432% KRAS p.G13D, 0.035% PIK3CA p.E545K
13 14 15	Patient 2	100 000	-	80% APC p.E1309Dfs*4, 80% KRAS p.G12D, 80% TP53 p. Y126D
16 17 18	Patient 2	10 000	90 000	8.0% APC p.E1309Dfs*4, 8.0% KRAS p.G12D, 8.0% TP53 p. Y126D
19 20 21	Patient 2	1 000	100 000	0.80% APC p.E1309Dfs*4, 0.80% KRAS p.G12D, 0.80% TP53 p. Y126D
22 23 24	Patient 2	100	100 000	0.080% APC p.E1309Dfs*4, 0.080% KRAS p.G12D, 0.080% TP53 p. Y126D

Supplementary Figure 1. Number of significant mutations associated with driver genes in 282 probe panel (red bars) and 887 probe panel (blue bars) that was expanded from the 282 probe panel. White bars represent non-targeted significant mutations of the driver genes not included in either of the panels. The percentage is a ratio of number of significant mutations targeted by 887 probe panel and total amount of significant mutations for that specific cancer type. Cancer types and their associated driver genes are obtained from intOgen-framework (https://www.intogen.org) v.2023.05.31. Significant mutations of the driver genes are mutations of significance 1 or 2 in COSMIC database, CMC v.99 (https://cancer.sanger.ac.uk/cosmic).