

1 **Rapid and highly-specific generation of targeted DNA**
2 **sequencing libraries enabled by linking capture probes**
3 **with universal primers**

4

5 Joel Pel¹, Amy Leung¹, Wendy W. Y. Choi¹, Milenko Despotovic¹, W. Lloyd Ung¹,
6 Gosuke Shibahara¹, Laura Gelinas¹, and Andre Marziali^{1,2,*}

7

8 1 Boreal Genomics Inc, Vancouver, British Columbia, Canada

9 2 University of British Columbia, Department of Physics and Astronomy, Vancouver, British Columbia,
10 Canada

11

12 *Corresponding Author

13 Email: andre@phas.ubc.ca (AM)

14 **Abstract**

15 Targeted Next Generation Sequencing (NGS) is being adopted increasingly broadly in
16 many research, commercial and clinical settings. Currently used target capture methods,
17 however, typically require complex and lengthy (sometimes multi-day) workflows that
18 complicates their use in certain applications. In addition, small panels for high
19 sequencing depth applications such as liquid biopsy typically have low on-target rates,
20 resulting in unnecessarily high sequencing cost.

21 We have developed a novel targeted sequencing library preparation method, named
22 Linked Target Capture (LTC), which replaces typical multi-day target capture workflows
23 with a single-day, combined ‘target-capture-PCR’ workflow. This approach uses
24 physically linked capture probes and PCR primers and is expected to work with panel
25 sizes from 100 bp to >10 Mbp. It reduces the time and complexity of the capture
26 workflow, eliminates long hybridization and wash steps and enables rapid library
27 construction and target capture. High on-target read fractions are achievable due to
28 repeated sequence selection in the target-capture-PCR step, thus lowering sequencing
29 cost.

30 We have demonstrated this technology on sample types including cell-free DNA
31 (cfDNA) and formalin-fixed, paraffin-embedded (FFPE) derived DNA, capturing a 35-
32 gene pan-cancer panel, and therein detecting single nucleotide variants, copy number
33 variants, insertions, deletions and gene fusions. With the integration of unique molecular
34 identifiers (UMIs), variants as low as 0.25% abundance were detected, limited by input
35 mass and sequencing depth. Additionally, sequencing libraries were prepared in less than
36 eight hours from extracted DNA to loaded sequencer, demonstrating that LTC holds
37 promise as a broadly applicable tool for rapid, cost-effective and high performance
38 targeted sequencing.

39 **Keywords**

40 Next generation sequencing

41 Targeted sequencing

42 Target Capture

43 Hybrid Capture

44 Copy Number Variant (CNV)

45 Single Nucleotide Variant (SNV)

46 Gene fusion

47 Structural variants

48 **Introduction**

49 Targeted Next Generation Sequencing (NGS) is common practice in many research,
50 commercial and clinical applications, as a faster and cheaper alternative to equivalent
51 depth whole-genome or whole-exome sequencing. As sequencing technologies continue
52 to become more accessible, the adoption of targeted NGS into more labs and markets is
53 likely to follow.

54 Existing targeted sequencing approaches generally fall into three categories: (i)
55 Multiplexed PCR; (ii) Hybridization and extension; and (iii) Hybridization and capture
56 (1), and are summarized briefly here. PCR is a well-known technique which can be very
57 effective for targeting small to mid-sized genomic regions. However, multiplex PCR is
58 generally challenging to design and does not scale easily to very large targets. Sample
59 splitting is generally required to tile large contiguous regions or reduce primer dimers,
60 subsequently reducing sensitivity to rare variants (2). Techniques aimed at mitigating
61 multiplexing challenges include using droplets to reduce primer dimer formation (3),
62 integrating special primer adapters to enable tiling without sample splitting (4), or linking
63 primers to increase specificity and reduce primer dimers (5, 6). While providing
64 improvements, these methods are generally more complex to design and use, and are still
65 limited in their multiplexing capabilities. Additionally, for many applications, including
66 diagnostics, PCR methods generally lose information compared to ligation-based
67 methods. For example, in multiplex PCR methods, the start and stop positions of genomic
68 fragments are lost, and integration of unique molecular identifiers (UMIs) for somatic
69 mutation detection can be challenging (7).

70 Hybridization and extension methods improve on PCR multiplexing limitations by using
71 a single ‘primer’ for each target that extends across a region of interest and reduces
72 primer dimers (8-12). The resulting products are then ligated and amplified by universal
73 primers to create sufficient material for sequencing. Despite the improvements in
74 multiplexing compared to PCR due to fewer primers, these methods have not achieved
75 the same widespread use compared to hybridization and capture methods. Potential
76 reasons may include high DNA input mass requirements, high cost and complexity, low
77 uniformity, or loss of sequence information under long priming regions (1, 4).

78 Perhaps the most common approach, hybridization and capture (13, 14), uses single-
79 stranded DNA or RNA probes that are designed to bind specifically to sequences of
80 interest. Probes containing biotin are annealed to targets during a lengthy incubation step,
81 after which avidin-biotin binding is used to extract the biotin-labeled probes, thus
82 enriching for the targets of interest. Hybridization and capture methods have many
83 advantages, including scalability to large panels, the ability to easily distinguish
84 duplicates on the sequencer through use of UMIs, and to retain insert start-stop positions
85 due to up-front ligation. Some of the main disadvantages, however, include low
86 sequencer on-target fraction, high cost, and complex and lengthy workflows (4).

87 Commercial hybridization and capture methods vary in speed, complexity and
88 performance. These methods typically start with a library preparation step (either by
89 ligation or transposase), followed by a universal pre-amplification PCR step and then one
90 or more hybridization capture steps, ranging from four to 72 hours. Following capture,
91 the targeted DNA is recovered via a series of pull-down and wash steps. Targeted DNA is
92 then amplified again and quantified prior to sequencing (15). In general, faster capture
93 times can only be achieved at the expense of lower on-target fractions. Also, as panel size
94 decreases from ~30 Mbp for whole exome captures to the 10 kbp -100 kbp range
95 commonly used for diagnostic applications, on-target fraction generally decreases as well
96 (16). Lower on-target results in lower depth of coverage and lower variant sensitivity
97 unless sequencing throughput (and cost) is increased (15, 17).

98 To the best of our knowledge, the IDT xGen workflow (Integrated DNA Technologies) is
99 the fastest available commercial assay, with a reported workflow time of nine hours.
100 However, this does not include library preparation or pre-amplification which generally
101 adds at least several more hours (depending on method), requiring the workflow to be
102 performed over multiple work days. Other common protocols can span two or more days,
103 such as Roche SeqCap (Roche Sequencing). The length and complexity of these
104 workflows limit their use, especially in clinical settings, where fast turn-around time and
105 ease of use are important.

106 We have developed Linked Target Capture (LTC), a novel target capture method with
107 broad application, designed to reduce hybridization workflows to less than eight hours

108 while retaining high performance over all panel sizes. LTC replaces existing
109 hybridization methods with a combined ‘target-capture-PCR’ workflow using linked
110 capture probes and universal amplification primers. Here we describe the LTC method,
111 and demonstrate its ability to rapidly deliver enriched sequencing libraries from multiple
112 sample types, including formalin-fixed, paraffin-embedded (FFPE) derived DNA,
113 plasma-derived cell-free DNA (cfDNA) and cell line DNA. Additionally, with the
114 integration of UMIs, we demonstrate LTC’s ability to detect low-level single nucleotide
115 variants, copy number variants, insertions/deletions and gene fusions.

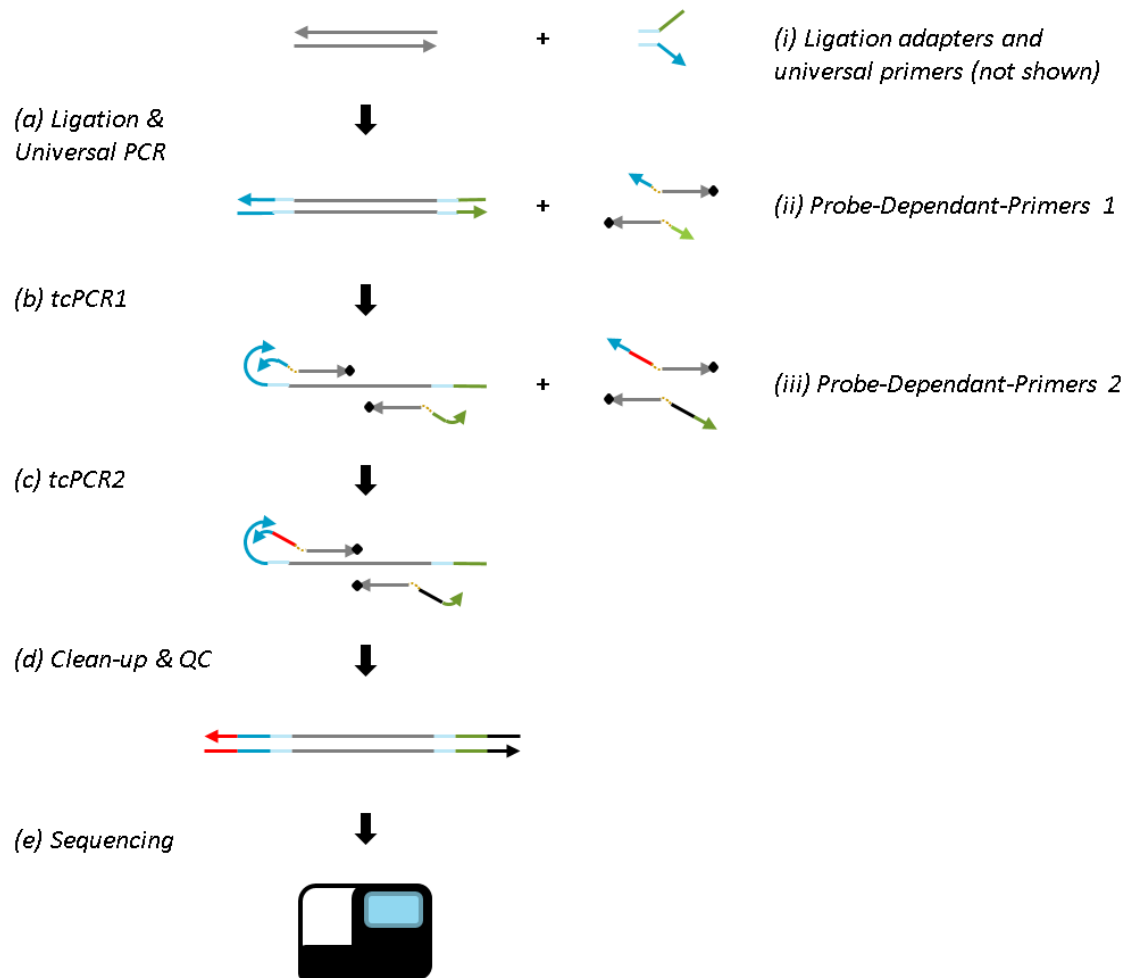
116

117 **Results**

118 **Linked Target Capture Concept**

119 The LTC method is illustrated in Figure 1 for Illumina sequencers, though it is expected
120 to be compatible with most sequencing platforms. The workflow begins with ligation of
121 short Y-adapters that contain truncated portions of the Illumina P5 and P7 flow cell
122 binding sequences, such that ligated molecules will not bind to the flow cell and be
123 sequenced without further processing. Following ligation and pre-amplification using
124 universal primers, two sequential target-capture-PCR (tcPCR) steps are performed with
125 Probe-Dependent-Primers (PDPs). PDPs consist of non-extendable DNA capture probes
126 linked 5’ to 5’ with a low melting-temperature universal primer complementary to a
127 portion of the ligated adapter (Figure 1 (ii) and (iii)). When bound to their targets, the
128 probes bring the universal primer into close proximity with the universal priming site on
129 the template, increasing the reaction rate of primer binding and initiating polymerase
130 extension. The polymerase displaces or digests the probe portion of the PDP to make a
131 copy of the entire target template, and the reaction proceeds to the next tcPCR cycle. To
132 create sequencer-compatible libraries, the second tcPCR integrates the full Illumina P5
133 and P7 sequences into the universal primer portion of the PDPs. Both tcPCR reactions are
134 performed above the melting temperature of the universal primers so that amplification is
135 heavily biased towards target-bound PDPs.

136 As described in the Materials and Methods, PDPs are made by reacting separately
137 synthesized probes and primers. PDP panels are made by linking probe sets to the
138 universal primers, making panel generation, expansion, and combination straightforward.
139



140
141 **Figure 1:** Linked Target Capture Workflow. (a) Custom adapters (i) are ligated to
142 template DNA and the resulting product is amplified with universal primers. (b) Target
143 regions are selectively amplified using custom probe-dependent-primers (PDPs) (ii)
144 which contain a recognition sequence (dark grey) with a 3' blocker (black diamond) and
145 are linked to an oligo containing a universal priming sequence for the first target capture
146 PCR reaction (tcPCR1). (c) A second set of PDPs (iii), which contain Illumina adapters
147 (red and black) between the probe and linked universal primer, are then added and a

148 second target capture PCR reaction (tcPCR2) is completed prior to (d), clean up and QC
149 and (e) loading on a sequencer.

150

151 **Assay Validation**

152 To validate the LTC workflow, PDPs were designed to cover relevant portions of 35
153 cancer-related genes, as described in Materials and Methods and listed in S1 Table. PDPs
154 were chosen to capture four major variant types, including single nucleotide variants
155 (SNVs), insertions/deletions (Indels), copy number variants (CNVs) and structural
156 rearrangements (ex: gene fusions). Libraries were created and sequenced in duplicate
157 from four sample types, as outlined in Table 1: mechanically sheared cell line DNA,
158 enzymatically sheared cell line DNA, cfDNA, and FFPE-derived DNA. Additionally, to
159 test lower input mass, duplicate libraries were created and sequenced from 5 ng of
160 mechanically sheared cell line DNA. The total time from extracted DNA to loaded
161 sequencer was eight hours, with about three hours of hands-on time.

162 All libraries were analyzed through the same pipeline (see Materials and Methods) and
163 down-sampled to a fixed number of sequencing clusters (or read pairs) for a given input
164 mass (2 M read pairs for 40 or 50 ng, 0.2 M for 5 ng). Fixing the number of read pairs is
165 important when comparing results, as the same sequencing data analyzed with different
166 numbers of read pairs produces different results (especially in coverage). This is
167 attributed to several factors, including insufficient reads for a given input mass (or a
168 given number of input genomes), and Poisson variation. Fixed-read results are shown in
169 Table 1. On-target fraction, mean target coverage and uniformity were calculated using
170 Picard CollectHSMetrics (broadinstitute.github.io/picard/), as described in Materials and
171 Methods.

172 These data demonstrate consistently high on-target fraction (86%-97%) and uniformity
173 (90%-96%) across a range of sample types and input mass relevant to clinical
174 applications of targeted sequencing. As a reference, commercially available Roche,
175 Illumina and Agilent methods have been compared using a 110 gene panel, and ranged in
176 performance from 75% to ~87% on-target (17). While not a direct comparison, this
177 reference provides a good indicator of relative performance, as it is typically easier to
178 achieve high on-target fraction on large panels (16) (a direct in-house comparison was

179 not undertaken due to the significant cost of capture panels). To demonstrate the
 180 scalability of LTC, we measured enrichment on four of the 35 gene targets (BRAF,
 181 EGFR, ERBB2 and TP53), using 50 ng mechanically-sheared cell line DNA. The
 182 measured on-target fraction was >97% in both replicates, higher than the same
 183 measurement for our 35-gene panel. Similarly-sized small panels using conventional
 184 single-round target capture reported ~5% on-target reads in (16) and (18).
 185 A comparison in uniformity can be made against the use of a SureSelect XT panel
 186 (Agilent Technologies) covering 231 SNV targets in 26 genes (19). For FFPE samples
 187 with similar coverage (>1000x) in (19), the authors report uniformity ranging from ~50%
 188 to 93%, whereas both FFPE replicates using LTC had a uniformity of 96%.

189

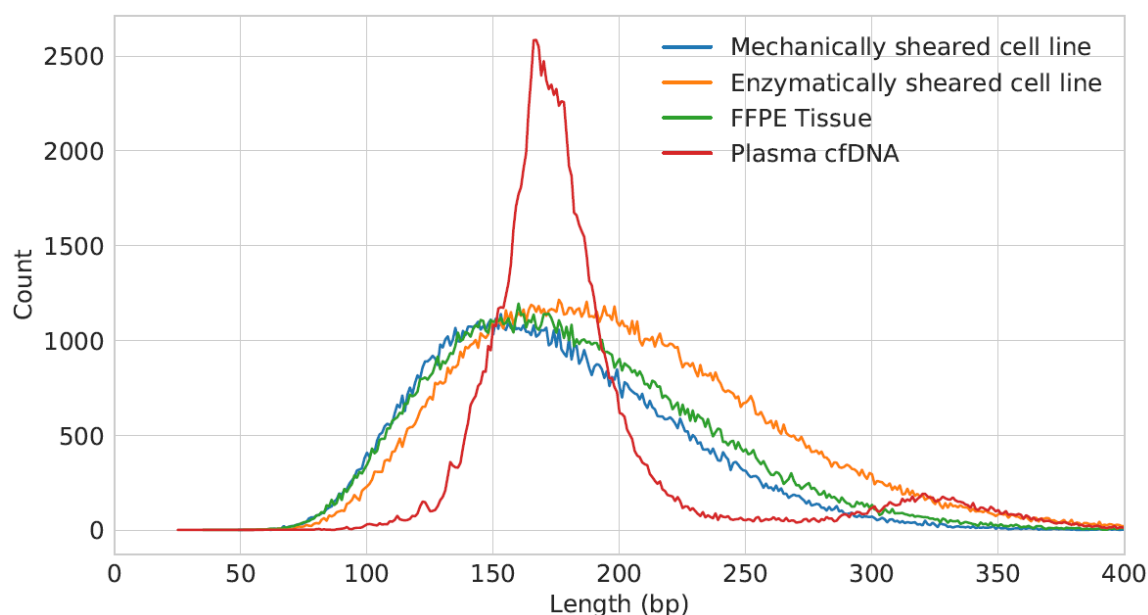
190 **Table 1:** LTC 35-gene sequencing performance data for two replicates of sample type
 191 and DNA mass. On-target fraction was defined as the fraction of total bases that aligned
 192 to the target regions. Mean target coverage was defined as the mean de-duplicated
 193 coverage over all target regions, and uniformity was defined as the fraction of on-target
 194 bases that were covered within two fold of the mean target coverage (i.e. between 0.5x
 195 and 2x of the mean). Median insert length was measured over all de-duplicated on-target
 196 inserts.

DNA Source	DNA Mass	Replicate	On-target Fraction	Mean Target Coverage	Uniformity	Median Insert Length
Mechanically sheared cell line	5	1	93%	157	95%	162
		2	93%	156	95%	162
	50	1	93%	1202	96%	168
		2	89%	1281	96%	169
Enzymatically sheared cell line	50	1	91%	1685	94%	193
		2	86%	1764	93%	190
FFPE tissue	40	1	91%	1358	96%	175
		2	86%	1444	96%	171
Plasma cfDNA	50	1	95%	1154	90%	174
		2	97%	1148	90%	175

197

198 Insert length distributions for each sample type were calculated using Picard
 199 CollectInsertSizeMetrics and are shown in Figure 2. Mechanically sheared cell lines were

200 created by the manufacturer to produce a majority of inserts in the range of ~100 bp to
201 250 bp (see Materials and Methods). The recovered insert lengths for these samples
202 represent a good match to the expected size distribution with 89% of targets between
203 100 bp and 250 bp. Enzymatically-sheared DNA samples produced slightly longer
204 inserts, likely a function of the shearing protocol. Additionally, the median insert size for
205 the cfDNA samples was 175 bp in a reasonably narrow distribution, in good concordance
206 with literature (20). A peak was also visible around ~325 bp, suggesting these long
207 fragments may have been wrapped twice around the histone. Finally, FFPE-derived DNA
208 samples also produced a short insert length distribution, as expected from the degradation
209 associated with FFPE samples combined with enzymatic shearing and repair.
210



211
212 **Figure 2:** Representative insert size distributions for each sample type used in this study.

213

214 **Variant Detection**

215 To enable the detection of low level variants with LTC, UMIs consisting of four random
216 bases in series were integrated into Illumina's 'Index 1' read position of the ligation
217 adapter. The UMIs were used in conjunction with the start and stop positions of the
218 inserts to uniquely identify individual starting template molecules and to create consensus
219 sequences (see Materials and Methods). A commercially available reference standard cell
220 line (HD786, Horizon Discovery) was used to assess the ability of LTC to detect variants

221 as it contains SNVs, CNVs, indels and fusions at levels characterized by the
222 manufacturer. The variants covered by the 35-gene panel are listed in Table 2, along with
223 the expected allele percentage as specified by the manufacturer for each of the different
224 samples used in this study.

225

226 **Table 2:** Reference standard variants. The expected allele percentage was measured and
227 specified by the manufacturer using digital PCR or next generation sequencing. Expected
228 allele percentages are given for stock samples (Sample 1), samples diluted to 1/10 of the
229 stock concentration (Sample 2), and wild-type samples (Sample 3).

Gene	Variant	Type	Expected Allele %		
			Sample 1 (1:1)	Sample 2 (1:10)	Sample 3 (WT)
PIK3CA	E545K	SNV	5.6%	0.56%	0.00%
KRAS	G13D	SNV	5.6%	0.56%	0.00%
EGFR	V769_D770- insASV	Insertion	5.6%	0.56%	0.00%
EGFR	Δ E746 - A750	Deletion	5.3%	0.53%	0.00%
MET	V237fs	Deletion	2.5%	0.25%	0.00%
RET	CCDC6/RET fusion	Fusion	5.0%	0.50%	0.00%
MET	amplification	CNV	4.5 copies/genome	1.35 copies/genome	1.00 copies/genome

230

231 To test variant detection, 50 ng of DNA was used from each cell line. DNA from the
232 reference standard (Sample 1) was analyzed in duplicate, along with duplicate analysis of
233 cell line from the same manufacturer known to be wild type for the variants of interest
234 (Sample 3). A ten-fold titration (Sample 2) of the reference standard was made with the
235 wild type cell line, and also tested in duplicate. Sequencing analysis and variant calling
236 was performed as outlined in Materials and Methods.

237 The measured variant fractions for detected SNVs, indels and fusions are plotted against
238 the expected fractions in Figure 3. All the variants that were detected were measured at
239 allele frequencies within ~3x of expected values. Expected variants as low as 0.25% were

240 detected (the lowest fraction tested in this study), which corresponds to ~38 mutant
241 fragments present in the initial 50 ng sample (assuming 3.3 pg of DNA per human
242 haploid genome). Since ligation yield in general is much lower than 100% (21), the actual
243 number of mutants entering capture could be considerably less than 38, and perhaps near
244 sampling error for some loci. Discrepancy between measured and expected values may
245 be attributed to a number of factors including the differences in variant calling methods,
246 titration of the reference standard, and the relative sequencing coverage of each variant,
247 all of which could potentially lead to sampling error.

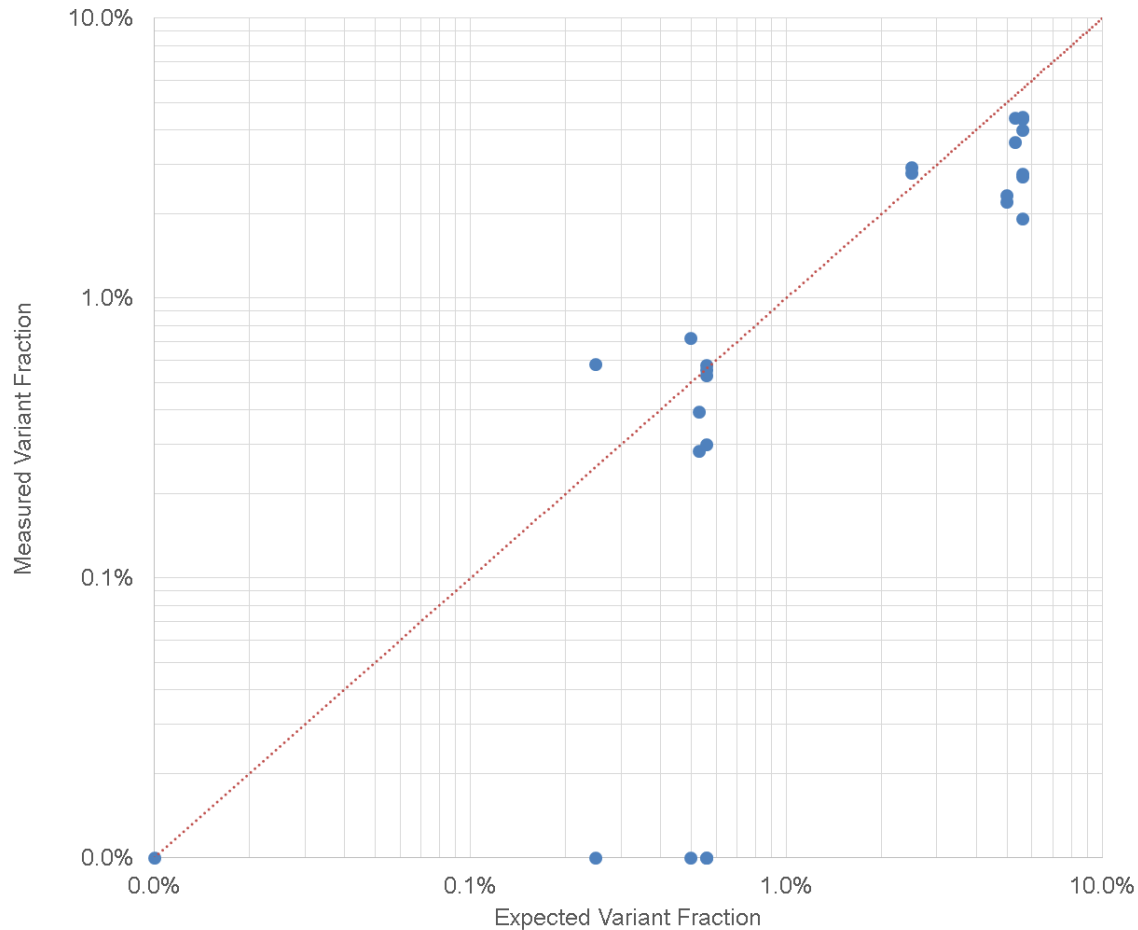
248 All twelve (100%) of the SNV, indel and fusion variants were correctly identified at
249 0.00% variant fraction in the wild-type cell line (Sample 3 replicates). Eight of the twelve
250 (67%) SNV, indel and fusion variants expected between 0.25% and 0.56% were detected
251 in the diluted reference standard (Sample 2 replicates), while all twelve (100%) of the
252 same variants expected between 2.5% and 5.6% were detected in the reference standard
253 (Sample 1 replicates). With further improvements to LTC (see Discussion), we expect
254 even higher sensitivity and lower detection limits to be possible.

255

256

257

258



259

260 **Figure 3:** Expected vs. measured SNV, indel and fusion fractions. The dotted line
261 represents a 1:1 ratio of expected vs. measured variants. Undetected and zero variant
262 fraction samples were reported at 0.0% for display purposes.

263

264 Copy number variation was assessed for the MET gene in all six samples (replicates of
265 Samples 1, 2 and 3) by our analysis pipeline, which was designed to identify samples as
266 “amplified”, “deleted” or “copy-number neutral” (see Materials and Methods). The MET
267 gene was identified as “amplified” in both replicates of Sample 1, and “copy-number
268 neutral” for both replicates of Samples 2 and 3. These results were consistent with
269 expectations, as the 4.5 copies of the MET gene present in each Sample 1 replicate
270 should be easily detectable above background, even when compared against only two
271 wild-type samples (Sample 3 replicates). On the other hand, the 1.35 copies of the MET
272 gene in each of the Sample 2 replicates would likely require many more measurements to

273 confidently detect a copy number variation above the wild-type background.
274 In general, all four variant types were detected as expected, demonstrating the capability
275 of LTC as a target capture tool for many different applications.

276 **Discussion**

277 In general, target capture performance and workflow improvements have the potential to
278 increase NGS and target capture usage in existing applications, and also to enable new
279 opportunities if workflow time, complexity or cost are reduced.
280 Arguably the most significant improvement of LTC over existing methods is the dramatic
281 decrease in workflow time. To the best of our knowledge, the IDT xGen workflow is the
282 fastest commercial hybridization capture method, at nine hours. This does not include
283 library preparation, which generally adds several additional hours, and requires the assay
284 to be run over two work days. In contrast, the LTC workflow was completed in eight
285 hours, including library preparation and loading of the sequencer.
286 Typical capture workflows are limited by the length and performance of the hybridization
287 step, which on its own can extend to 72 hours. Shorter hybridization steps typically
288 compromise performance resulting in either lower de-duplicated target coverage or
289 higher off-target fraction. LTC avoids this tradeoff and shortens this rate limiting step by
290 employing a combined target-capture-PCR (tcPCR) step. High de-duplicated target
291 coverage is achieved by specifically capturing both senses (see Materials and Methods)
292 and by operating at a relatively low temperature compared to the probe binding
293 temperature, while the on-target fraction is increased through many effective capture
294 cycles performed in each tcPCR reaction. An additional benefit of tcPCR is workflow
295 simplicity. In conventional target capture workflows, biotinylated probes typically
296 require binding to streptavidin coated beads to enrich for the target DNA. The subsequent
297 bead capture and wash steps are generally complicated, labor intensive and can be
298 difficult to automate (22), potentially limiting deployment of target capture workflows in
299 some cases. On the other hand, the LTC tcPCR setup and operation are analogous to a
300 standard PCR reaction, and thus are more familiar to a larger number of technicians, and
301 also easier to automate. Additionally, it should be noted that LTC could be paired with

302 any library preparation method that introduces the correct adapter sequences, such as
303 single stranded library prep (23) or transposition (24).
304 A related advantage of the combined PCR-capture step is the ability to produce consistent
305 sequencing performance from low input mass samples. Sequencing parameters, including
306 coverage, scaled as expected from the 50 ng samples down to 5 ng, suggesting that LTC
307 is able to recover molecules efficiently across a wide mass range. This is especially
308 important in applications where sample is limiting, and could be tested to even lower
309 limits in a future study.

310 It should be noted that the workflow time and complexity of LTC is comparable to
311 multiplexed PCR (such as AmpliSeq by ThermoFisher) and hybridization extension
312 methods (such as (11)). LTC holds a number of significant advantages over these
313 methods, however. First, since the LTC primers are universal, it does not require sample
314 splitting before amplification to prevent unwanted amplicon formation. This avoids loss
315 of sensitivity and the requirement for large DNA input mass. Second, since LTC probes
316 are displaced by the extended universal primer, sequence information at probe binding
317 sites is retained on the amplified molecules to be sequenced, thus capturing all of the
318 sequence information available from a single fragment. This is in contrast to PCR and
319 hybridization extension methods where any variants contained within a PCR primer
320 binding site are lost after the primer has bound and extended. LTC also retains fragment
321 start and stop positions, which are lost in PCR and hybridization extension methods, and
322 have been shown to provide useful biological information (25). Additionally, it is
323 generally much easier to integrate low variant detection in hybridization capture methods
324 like LTC compared to PCR methods. When UMIs are integrated in ligation as they are in
325 LTC, it is easier to avoid labelling a single molecule with multiple UMIs, which can
326 occur in PCR methods. Also, to our knowledge, it is not possible to integrate duplex
327 sequencing in a PCR-based UMI method, but this has been demonstrated with LTC.

328 Finally, because the challenges associated with multiplex PCR are reduced through the
329 use of universal primers, the LTC workflow can be used for a wide range of panel sizes,
330 including large panels for which multiplex PCR methods would not work. Small panels
331 have been demonstrated in this study, and initial work towards larger panels indicates that

332 exome-scale LTC panels may be possible. This is advantageous, as a single workflow
333 could be implemented for multiple assays or applications.

334 LTC has several other unique properties. Primers and probes can be oriented to capture a
335 specific strand of the target duplex DNA (ex: the sense strand, see Materials and
336 Methods), providing an advantage in rare variant detection, or in applications where it is
337 desirable to sequence only one strand of the starting template such as transcriptome
338 sequencing (26). In addition, LTC has been demonstrated in droplets, providing
339 multiplexing capabilities to droplet-based assays not achievable with standard capture
340 methods.

341 The sequencing statistics achieved using Linked Target Capture were excellent, with
342 greater than 91% average on-target and 94% average uniformity, providing cost-effective
343 sequencer usage and leaving little room to improve these metrics. Measuring how these
344 factors scale to much larger panels would be an important part of a future study. Mean
345 target coverage was lower than initially expected, by about two to three fold compared to
346 hybridization capture with similar analysis (27); we suspect this to be due to the lack of
347 LTC probe tiling. The 35-gene panel used in this work consisted of fairly sparse probe
348 placement to reduce panel cost, such that the probes covered less than 100% of bases in
349 the targets. Initial data from tiling two targets in the 35-gene panel to nearly 200%
350 demonstrated a more than 3-fold increase in mean target coverage, which agrees with
351 previously reported tiling improvements of at least two-fold (28). It is expected that tiling
352 will significantly improve mean target coverage as well as variant detection when applied
353 across the whole 35-gene panel.

354 Variant detection may be further improved through the use of lower error UMI designs.
355 Like hybridization and capture methods, the error rate of LTC is expected to be linked to
356 the UMI design used in a given assay. For example, integrating duplex UMIs into the
357 LTC ligation adapters is expected to further reduce the detection limit, similarly to the
358 reduction observed for duplex UMIs applied to hybridization and capture methods (27).
359 Increasing the input mass and sequencing depth are also expected to lower the detection
360 limit of LTC.

361

362 In summary, we have developed a novel target capture method with a rapid workflow and
363 efficient sequencer usage. With continued improvements in tiling and panel expansion,
364 we expect LTC to be a high performance target capture method applicable in many
365 settings.

366 **Materials and Methods**

367 **PDP Design and Conjugation**

368 In order to enable panel design flexibility, PDPs were made by conjugating target-
369 specific probes and universal primers. The universal primers (forward and reverse) were
370 manufactured by Integrated DNA Technologies (IDT) and contained a 5'
371 Dibenzocyclooctyl (DBCO) modification. The forward and reverse untailed primers for
372 the first target capture step were CACCGAGATCT and TACGAGATCGG respectively.
373 The forward and reverse tailed primers for the second target capture step were
374 AATGATACGGCGACCACCGAGATCT and
375 CAAGCAGAAGACGGCATAACGAGATCGG respectively.

376 Capture probes were designed to cover portions of 35 cancer-related genes, shown in S1
377 Table. Total sequence coverage was 11,473 bp. Probes were designed with adjacent
378 forward and reverse probes covering the desired regions, with zero gap between forward
379 and reverse probes, a minimum length of 30 bp, maximum length of 70 bp, and a melting
380 temperature of ~85 °C calculated using uMELT (29) with default conditions. Probes were
381 synthesized by IDT with a 5' azide modification to conjugate with the DBCO on the
382 primer and a 3' inverted dT base, to inhibit polymerase extension.

383 Pools of forward and reverse probes were conjugated with both forward and reverse
384 primers separately by mixing 22.5 µM primer with 10 µM total probe concentration, in
385 0.6x PBS. Each mixture was incubated at 60 °C for 16 hours. After incubation, the
386 conjugates were purified using a modified Agencourt AMPure XP Kit (Beckman
387 Coulter) and eluted in 20 µL 0.1x IDTE (IDT). A 2:1 bead to sample ratio was used
388 according to the manufacturer's instructions, except that prior to use, the bead buffer was
389 extracted and replaced with an equal volume of a custom formulated buffer. The custom
390 buffer consisted of 30% w/v PEG-8000, 1 M NaCl, 0.05% v/v Tween 20, 10 mM Tris-
391 HCl, and 1 mM EDTA (all reagents from Sigma-Aldrich). Following conjugation, PDPs
392 were quantified using the Qubit ssDNA Assay (ThermoFisher Scientific). Conjugates
393 were made and then stored at -20 °C. PDPs consisting of forward probes with forward
394 primers were labelled as FF, reverse probes with forward primers RF, and so on for all
395 four combinations.

396

397 **Sample Sources**

398 Four sample types were used in this study: mechanically sheared cell line DNA,
399 enzymatically sheared cell line DNA, plasma-derived cell-free DNA (cfDNA), and FFPE
400 -derived DNA. Mechanically sheared DNA was obtained from Horizon Discovery in
401 mutant (HD786) and wild-type (HD776) standards (Samples 1 and 3, respectively, from
402 Table 2). Mechanical shearing was performed by the manufacturer such that around 60%
403 of the templates were within 100 bp to 250 bp, with fragments up to 400 bp. Mutation
404 levels were measured by the manufacturer using droplet digital PCR. Enzymatically
405 sheared cell line was generated from genomic DNA (HD753, Horizon Discovery),
406 following the protocol described below. cfDNA was isolated from single donor human
407 plasma samples (IPLAS – K2 EDTA, Innovative Research), as described below. FFPE-
408 derived-DNA was obtained from Horizon Discovery, part number HD799.

409

410 **Cell-free DNA Extraction**

411 First, 5 mL of plasma was centrifuged for 10 min at 2,000g. cfDNA was isolated from
412 each sample using the QIAamp Circulating Nucleic Acid Kit (Qiagen) according to the
413 manufacturer's instructions. DNA was eluted from the column in 0.1x IDTE in a two-
414 step process to maximize elution yield: 50 µL of 0.1x IDTE was incubated in the column
415 for 10 min, followed by a 20,000g spin for 3 min; the column was then re-eluted after a 3
416 min incubation with another 50 µL 0.1x IDTE for a total elution volume of 100 µL. The
417 DNA sample was further purified to remove any potential inhibitors using the Agencourt
418 AmPure XP Kit (Beckman Coulter). A 1.4:1 bead to sample volumetric ratio was used as
419 per manufacturer's instructions, with the sample eluted in 0.1x IDTE. Extracted and
420 purified DNA was then used directly for library preparation, or in cases where library
421 preparation did not proceed within 24 hours, was frozen at -20 °C.

422 Following DNA extraction, sample concentration was measured using the Qubit dsDNA
423 HS kit (ThermoFisher Scientific) as per the manufacturer's instructions and used to
424 calculate the number of human genome equivalent copies in each sample.

425

426

427 **FFPE DNA Pre-treatment**

428 FFPE-derived DNA was pre-treated to reduce the impact of potential DNA damage,
429 before target capture. 100 ng of DNA, as quantified by the Qubit dsDNA HS kit
430 (ThermoFisher Scientific), was digested with 1 unit of UDG enzyme (New England
431 Biolabs (NEB)) in a 50 μ L reaction in 1X of the supplied reaction buffer (NEB). The
432 mixture was incubated at 37 °C for 10 minutes, cooled to 4 °C, and immediately purified
433 with the Agencourt AmPure XP Kit at a 3:1 bead to sample volumetric ratio as per
434 manufacturer's instructions. Samples were eluted in 20 μ L of 10 mM Tris-HCl, pH 8.
435 Total amplifiable DNA was quantified using KAPA hgDNA Quantification and QC Kit
436 (KAPA Biosystems) as per the manufacturer's instructions.

437

438 **Enzymatic DNA Shearing**

439 Prior to enzymatic shearing, a buffer exchange was performed with cell line and FFPE-
440 derived DNA samples with the Agencourt AmPure XP Kit at a 3:1 bead to sample
441 volumetric ratio according to the manufacturer's instructions. Samples were eluted in
442 40 μ L of 10 mM Tris-HCl, pH 8. Cell line and FFPE-derived DNA samples were then
443 enzymatically sheared immediately before ligation using the KAPA HyperPlus kit
444 (KAPA Biosystems) according to the manufacturer's instructions. 50 ng of cell line DNA
445 or 40 ng of FFPE-derived DNA in 35 μ L volume was added to 10 μ L KAPA fragmentase
446 (KAPA Biosystems) and topped up to a final volume of 50 μ L in 1x supplied reaction
447 buffer. Samples were incubated at 37 °C for 30 minutes, afterwards proceeding
448 immediately to the A-tailing step of adapter ligation (described below). Shearing
449 conditions were chosen as per manufacturer's instructions to achieve a mode fragment
450 length of 150 bp.

451

452

453 **Adapter Ligation**

454 The KAPA Hyper Prep kit (KAPA Biosystems) was used as per the manufacturer's
455 instructions, with a 15 minute ligation incubation and a 200:1 adapter to insert ratio.
456 Custom ligation adapter sequences were ordered for the LTC workflow (IDT), consisting
457 of
458 AGCACGCACCGAGATCTACACBBBBACACTCTTTCCCTACACGACGCTCTTCC

459 GATCTT annealed to AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
460 **BBBBNNNN** ACCGATCTCGTAACTCAGCGG, where BBBB indicates a four base
461 sample-specific barcode for multiplexing samples on the sequencer, and NNNN indicates
462 a four base UMI. The UMI-containing adapter was phosphorylated on its 5' and 3' ends.
463 The non-UMI adapter contained a phosphorothioate bond between the last two bases on
464 the 3' end of the adapter. After ligation, the ligation mixture was purified using the
465 Agencourt AMPure XP Kit (Beckman Coulter) as per manufacturer's specification, with
466 a 0.4:1 bead to sample volumetric ratio, and eluted in 40 μ L of 0.1x IDTE. After elution,
467 the sample was topped up to 100 μ L with 0.1x IDTE. An additional cleanup with the
468 Zymo Select-a-Size DNA Clean & Concentrator column (Zymo Research) was
469 performed, as per the manufacturer's instructions. A 5:1 binding buffer to ethanol ratio
470 was used to select the desired product size. The final product was eluted in 25 μ L of 0.1x
471 IDTE. After cleanup, the entire volume of ligated DNA was amplified with custom
472 primers TTTTtagcACGCACCGAGATCTACAC and
473 TTTTCCGCTGAGTTACGAGATCGGT. Amplification proceeded for eight cycles
474 with 0.3 μ M of each primer, in 1x KAPA HiFi HotStart ReadyMix (KAPA Biosystems).
475 Annealing was performed at 60 °C for 30 s, extension at 72 °C for 20 s, and denaturing at
476 98 °C for 20 s. The amplified products were cleaned up using the Agencourt AMPure XP
477 Kit as per manufacturer's specification, with a 1.2:1 bead to sample volumetric ratio, and
478 eluted in 20 μ L of 0.1x IDTE. The cleaned up template DNA was then quantified using
479 the Qubit dsDNA HS kit (ThermoFisher Scientific) as per the manufacturer's
480 instructions.

481

482 **Target Capture**

483 Target-capture-PCR (tcPCR) for the 35-gene panel was performed in two subsequent
484 steps, each consisting of two reactions per sample. In the first step, the PDPs with
485 untailed primers were used, split into two 50 μ L reactions such that in the first reaction
486 FF and RR PDPs were used to capture one sense of the target, and in the second reaction
487 FR and RF PDPs were used for the other sense. Each reaction consisted of 5 nM of each
488 individual PDP, 15 ng template DNA, 5 units of Platinum Taq polymerase
489 (ThermoFisher Scientific), 4 mM MgCl₂, 0.2 mM dNTP (Invitrogen) in 1x Platinum Taq

490 Buffer (ThermoFisher Scientific). 15 tcPCR cycles were performed with a 30 s
491 denaturing step at 95 °C followed by a combined annealing and extension step at 66 °C
492 for 105 s. The ramp rate was 4 °C/s between 95 °C and 85 °C, and then 0.2 °C/s from 85
493 °C to 66 °C. The second tcPCR was performed using 12.5 uL of the amplified material
494 from the first tcPCR, and was otherwise identical to the first step, with the following
495 exceptions: PDPs with the tailed primers were used, ramp rate was 4 °C/s throughout
496 cycling, 12 cycles were performed, and the combined annealing and extension step was
497 done at 68 °C. Following amplification, libraries were purified using two back-to-back
498 bead cleanups, using the Agencourt AMPure XP Kit as per manufacturer's specification,
499 with a 0.8:1 bead to sample volumetric ratio. Final libraries were eluted in 20 µL of 0.1x
500 IDTE. tcPCR for the 4-gene panel was performed using a similar but earlier version of
501 the protocol, that was the same with the exception of the following differences: reaction
502 volume was 25 µL for both tcPCRs, 20 cycles were used in the first tcPCR, and 18 in the
503 second, 6.25 µL of the first tcPCR was carried over into the second, the second tcPCR
504 was eluted in 15 µL after cleanup.

505

506 **Sequencing and Data Analysis**

507 Targeted libraries were sequenced on an Illumina MiSeq or MiniSeq with paired-end 2 x
508 150 bp reads as per manufacturer's instructions. Prior to sequencing, samples were
509 quantified using the KAPA Library Quant Kit (KAPA Biosystems) as per manufacturer's
510 instructions. Resulting FASTQ files were demultiplexed by sample barcode using
511 Fulcrum Genomic's FGBIO open source bioinformatics tool suite
512 (<https://github.com/fulcrumgenomics/fgbio>) and then adapter-trimmed using
513 Trimmomatic V0.36 (30). Trimmed read pairs were combined and aligned to the
514 GRCh38/hg38 reference sequence using BWA-MEM (<https://github.com/lh3/bwa>) and
515 output in BAM format. SAMtools (31) was then used for sorting and indexing BAM
516 files. The resulting BAM files were grouped into UMI consensus reads by FGBIO for
517 low level SNV detection. Picard Tools 2.9 (<https://github.com/broadinstitute/picard>) was
518 then used to collect hybrid selection metrics, including on-target fraction, mean coverage
519 and insert length distributions. SNV, CNV and indel mutation calling was achieved using
520 GATK4 (<https://software.broadinstitute.org/gatk/gatk4>). CNV detection was not

521 quantified, but CNVs were identified as “amplified”, “deleted” or “copy-number neutral”
522 by the GATK4 CallCopyRatioSegments caller. Fusion detection was measured by
523 comparing Picard de-duplicated reads containing alignments to both the CCDC6 and
524 RET genes. Analysis outputs for assay validation and variant detection can be found in
525 Supplementary Material S2 and S3, respectively.

526 **Acknowledgments**

527 The authors wish to thank Integrated DNA Technologies Inc. (IDT) for support in probe
528 synthesis. Additionally, the authors wish to acknowledge David Broemeling for his help
529 in creating the figures, and Matthew Wiggin for his helpful review and discussions of the
530 manuscript.

531

532 **Author Contributions**

533 Conceived and designed the experiments: JP, AL, LG, MD, WC, LU, AM

534 Performed the experiments: AL, LG, MD, WC, LU

535 Analyzed the data: JP, AL, LG, MD, WC, LU, GS

536 Wrote the paper: JP

537 Edited the paper: AM

538

539 **Competing Interests**

540 The authors have the following interests: JP, WC, GS, MD, AL, LU, LG and AM are
541 employed by Boreal Genomics, the funder of this study. Additionally, all Boreal
542 employees hold stock options in Boreal Genomics. However, this does not alter the
543 author's adherence to policies on data sharing.

544

545 **References**

- 546 1. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al.
547 Target-enrichment strategies for next-generation sequencing. *Nat Methods*.
548 2010;7(2):111-8.
- 549 2. Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, et al.
550 Combining highly multiplexed PCR with semiconductor-based sequencing for rapid
551 cancer genotyping. *J Mol Diagn*. 2013;15(2):171-6.
- 552 3. Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, et al.
553 Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol*.
554 2009;27(11):1025-31.
- 555 4. Schenk D, Song G, Ke Y, Wang Z. Amplification of overlapping DNA amplicons
556 in a single-tube multiplex PCR for targeted next-generation sequencing of BRCA1 and
557 BRCA2. *PLoS One*. 2017;12(7):e0181062.
- 558 5. Satterfield BC. Cooperative primers: 2.5 million-fold improvement in the
559 reduction of nonspecific amplification. *J Mol Diagn*.2014. p. 163-73.
- 560 6. Gentalen E, Chee M. A novel method for determining linkage between DNA
561 sequences: hybridization to paired probe arrays. *Nucleic Acids Res*. 1999;27(6):1485-91.
- 562 7. Ståhlberg A, Krzyzanowski PM, Jackson JB, Egyud M, Stein L, Godfrey TE.
563 Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection
564 in liquid biopsies using sequencing. *Nucleic Acids Res*. 2016;44(11):e105.
- 565 8. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al.
566 Anchored multiplex PCR for targeted next-generation sequencing. *Nat Med*.
567 2014;20(12):1479-84.
- 568 9. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M. Multiplex
569 amplification enabled by selective circularization of large sets of genomic DNA
570 fragments. *Nucleic Acids Res*. 2005;33(8):e71.
- 571 10. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, et al. Multiplex
572 amplification of large sets of human exons. *Nat Methods*. 2007;4(11):931-6.
- 573 11. Hopmans ES, Natsoulis G, Bell JM, Grimes SM, Sieh W, Ji HP. A programmable
574 method for massively parallel targeted sequencing. *Nucleic Acids Res*. 2014;42(10):e88.

- 575 12. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide
576 scalable SNP genotyping assay using microarray technology. *Nat Genet.* 2005;37(5):549-
577 54.
- 578 13. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al.
579 Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted
580 sequencing. *Nat Biotechnol.* 2009;27(2):182-9.
- 581 14. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, et al. Direct
582 selection of human genomic loci by microarray hybridization. *Nat Methods.*
583 2007;4(11):903-5.
- 584 15. Bodi K, Perera AG, Adams PS, Bintzler D, Dewar K, Grove DS, et al.
585 Comparison of commercially available target enrichment methods for next-generation
586 sequencing. *J Biomol Tech.* 2013;24(2):73-86.
- 587 16. Schmitt MW, Fox EJ, Prindle MJ, Reid-Bayliss KS, True LD, Radich JP, et al.
588 Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat*
589 *Methods.* 2015;12(5):423-5.
- 590 17. García-García G, Baux D, Faugère V, Moclyn M, Koenig M, Claustres M, et al.
591 Assessment of the latest NGS enrichment capture methods in clinical context. *Sci Rep.*
592 2016;6:20948.
- 593 18. Alcaide M, Yu S, Davidson J, Albuquerque M, Bushell K, Fornika D, et al.
594 Targeted error-suppressed quantification of circulating tumor DNA using semi-
595 degenerate barcoded adapters and biotinylated baits. *Sci Rep.* 2017;7(1):10574.
- 596 19. Lee C, Bae JS, Ryu GH, Kim NKD, Park D, Chung J, et al. A Method to Evaluate
597 the Quality of Clinical Gene-Panel Sequencing Data for Single-Nucleotide Variant
598 Detection. *J Mol Diagn.* 2017;19(5):651-8.
- 599 20. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al.
600 Fragment Length of Circulating Tumor DNA. *PLoS Genet.* 2016;12(7):e1006162.
- 601 21. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An
602 ultrasensitive method for quantitating circulating tumor DNA with broad patient
603 coverage. *Nat Med.* 2014;20(5):548-54.

- 604 22. Ware JS, John S, Roberts AM, Buchan R, Gong S, Peters NS, et al. Next
605 generation diagnostics in inherited arrhythmia syndromes : a comparison of two
606 approaches. *J Cardiovasc Transl Res.* 2013;6(1):94-103.
- 607 23. Wang Q, Wang X, Tang PS, O'leary GM, Zhang M. Targeted sequencing of both
608 DNA strands barcoded and captured individually by RNA probes to identify genome-
609 wide ultra-rare mutations. *Sci Rep.* 2017;7(1):3356.
- 610 24. Caruccio N. Preparation of next-generation sequencing libraries using Nextera™
611 technology: simultaneous DNA fragmentation and adaptor tagging by in vitro
612 transposition. *Methods Mol Biol.* 2011;733:241-55.
- 613 25. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA
614 Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell.*
615 2016;164(1-2):57-68.
- 616 26. Mamanova L, Turner DJ. Low-bias, strand-specific transcriptome Illumina
617 sequencing by on-flowcell reverse transcription (FRT-seq). *Nat Protoc.* 2011;6(11):1736-
618 47.
- 619 27. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al.
620 Integrated digital error suppression for improved detection of circulating tumor DNA.
621 *Nat Biotechnol.* 2016;34(5):547-55.
- 622 28. Tewhey R, Nakano M, Wang X, Pabón-Peña C, Novak B, Giuffre A, et al.
623 Enrichment of sequencing targets from the human genome by solution hybridization.
624 *Genome Biol.* 2009;10(10):R116.
- 625 29. Dwight Z, Palais R, Wittwer CT. uMELT: prediction of high-resolution melting
626 curves and dynamic melting profiles of PCR products in a rich web application.
627 *Bioinformatics.* 2011;27(7):1019-20.
- 628 30. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
629 sequence data. *Bioinformatics.* 2014;30(15):2114-20.
- 630 31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
631 Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
- 632