

Supplemental Discussion

The following comparison of enrichment methods for targeted sequencing was added at the request of the reviewers to address concerns that the reader be able to compare the performance of the microdroplet PCR enrichment process to other already established methods. We wish to note that each of the published technical studies has selected different targets (with different GC contents and homologs in the genome), have sequenced the captured DNA to different levels of coverage, and have performed different analyses to measure performance. Furthermore, the definition of “target”, “background” and “fraction of reads mapping on target” are used in different contexts in the various technical studies describing enrichment methods. Thus, we recognize that there are limitations to being able to fairly compare performance measures of targeting efficiencies between the different technologies.

A. Comparison of enrichment methods for targeted sequencing

Here we clearly define “target”, “background” and “fraction of reads mapping on target” used for the microdroplet PCR enrichment method as well as for a recently published hybridization method and a recent publication on molecular inversion probes. In addition to defining these terms for all three technologies we provide published performance measures (Supplemental Table 8).

Microdroplet PCR

For the scale-up phase after trimming primer sequences from the amplicons there is a total of 1.35 Mb of targeted bases. The fraction of the 10,603,854 total reads generated that map to the targeted bases is 65.1%. The fraction of targeted bases that are covered by at least 1 base is 99.8%, and the fraction that passes our threshold for calling variants (covered by 5 or more reads and a consensus score for 50 or greater) is 94.5%. The average coverage was 185 and variant concordance based on comparison with 2226 HapMap genotypes was 98.8%. The uniformity of coverage: 96.6% of bases within 25-fold

Hybridization-based methods

We use the data from Gnirke et al.¹ for the basis of comparing the Solution Hybridization enrichment method. The study used 22,000 bait sequences of 170 bases in length for a total of 3.74 Mb of targeted bases. The authors report that 321 Mb of the total sequence (851 Mb) and of uniquely mapping sequence (492 Mb) map directly on the bait itself. Due to the fact that Gnirke et al. uses all sequence reads while we use filtered sequence reads for performance measures, direct comparisons is difficult and may result in an underrepresentation for solution

hybridization performance. Therefore to perform this comparison we acquired the hybridization sequence raw data (851 Mb)² and mapped back to HG18 using the same analysis methods we used to assess micro-droplet PCR performance. We obtained similar mapping efficiency for the unfiltered reads (523 Mb uniquely mapping to HG18 and 331 Mb uniquely mapping on target) as described in Gnirke et al. with the slight variation due to differing mapping strategies. We filtered the reads assuming the GERALD filter flag is still properly represented in the archive obtaining 524 Mb. Of the filtered reads 439 Mb mapped uniquely to HG18 of which 278 Mb fell on the baited region. Thus, 53.1% (278/524) of the filtered reads mapped on to the baited regions. For all other coverage statistics we used values given in the original paper. The fraction of targeted bases that are covered by at least 1 read is 88.0%, and the fraction that passes the threshold for calling variants (best-to-next-best ratio of 10^5 or greater) is 64% of the exonic sequence in the GA-I run and 89% in the GA-II run. The average coverage was 86x and variant concordance based on 7712 HapMap genotypes was 99.6% for the GA-I run, concordance was not reported for the GA-II run. The uniformity of coverage: 80% of bases within 25-fold.

Molecular Inversion probes

We use the data from Turner et al.³ for the basis of comparing the Molecular Inversion Probe (MIP) enrichment method. Our comparisons are based on the sequence data generated using the End Sequencing approach (13,000 MIPs), which had greater targeting efficiency than the Shotgun approach (55,000 MIPs). The study used 13,000 MIPs ranging in length from 100-191 bp, which total ~1.7 Mb. However, only 1.4 Mb were considered the targeted bases because 0.3 Mb were not accessible by end sequencing. The MIPs method showed high specificity with 90% of reads mapping on target. However, the study generated 76 bp reads with the first 20 bp of each read corresponding to the MIP primer, thus 56 of the bases are informative. $90 \times (56/76) = 66.3\%$ = the proportion of reads mapping to targeted bases. The fraction of targeted bases that are covered by at least 1 read is 98.0%, and the fraction that passes the threshold for calling variants (covered by 8 or more reads) is 75%. In Turner et al. there is no direct mention of average sequence depth per sample. We calculate an estimated average coverage of ~224X coverage per sample by multiplying the average number of reads mapping on target (8.4 million reads x 90% mapping on target) by the ratio of the read that was on target (56 bases of 76 bp read) divided by the total sequence captured (1.4 Mb). The variant concordance based on 9489 HapMap genotypes was 99.7%. The uniformity of coverage: 58% of bases with 10-fold and 88% of bases within 100-fold.

B. Reduction in DNA requirements for microdroplet PCR

In the RDT 1000: Merge step described above we started with 4.5 ug of genomic DNA in 20 ul of buffer and divided it evenly into 1.5 million droplets of 14 pl. Each microdroplet thus contains ~3 pg of genomic DNA, corresponding roughly to 1 haploid genome equivalent. Hence, if the amount of starting genomic DNA were to be halved, the number of productive microdroplets would also be halved. However, in some cases, researchers have a need to be able to work with less starting DNA and therefore we have examined the effects of whole genome amplification on data quality and are developing the ability to perform microdroplet PCR with multiple primer pairs in each droplet.

Whole-genome amplification

The effects of whole-genome amplification (“WGA”) on data quality were examined. Two WGA kits were used for this experiment: Rapisome pWGA Kit (Biohelix, H0300S) which utilizes a primase and helicase for WGA; and REPLI-g Mini kit (Qiagen, 150023) which utilizes random hexamers and phi29 for WGA. 100 ng of input DNA (Coriell HapMap sample NA18858) was amplified according to the manufacturers’ instructions for both kits. Quality of WGA template was assessed by agarose gel electrophoresis, followed by WGA template purification using QIAmp Mini DNA Kit (Qiagen) according to manufacturer’s instructions. Purified WGA DNA samples were then sheared to 2 to 4 kb by nebulization (Invitrogen) using an air pump at 10 psi. Following nebulization, a small aliquot of the sheared samples was assessed by agarose gel electrophoresis to determine correct sizing of the sheared DNA fragments. Sheared WGA samples were then precipitated by isopropyl alcohol precipitation, and the air-dried DNA pellets were resuspended by vigorous pipetting in 1 mM Tris-HCl pH 8.0. Samples were resuspended in 12 µL and DNA concentrations were determined by picogreen fluorescence using 1 µL of the DNA sample.

The purified, sheared WGA DNA templates (4 ug) were amplified using the same 3976 primer pair library and sequence enrichment protocol as used in the scale-up phase alongside unamplified genomic DNA from HapMap sample NA18858. The SNP calls made from the sequencing data were compared with the data in the HapMap database and the two WGA samples had similar call rates (97%) and the same concordance rate (98.7%) as the unamplified genomic sample. These data demonstrate that equivalently accurate SNP calling can be performed using the microdroplet PCR sequence enrichment procedure whether the DNA is unamplified or whole genome amplified.

Expanded content microdroplet PCR

Another possibility exists to increase the number of active PCR droplets without increasing the starting amount of template DNA. If one starts with 1/4 of a haploid genome per droplet and includes 4 sets of primer pairs in each of the primer droplets, essentially all of the droplets will generate a single PCR product and 2.0 million active reactions can be achieved from 1.5 ug of input DNA. To demonstrate this, we prepared 21 uL of template buffer solution with 375 ng of template DNA and a 384-member library where each library droplet contains 4 sets of primer pairs; 96 different types of droplets. Gel electrophoresis curves indicate that the yield is consistent with what is expected if all droplets are producing a single amplification product.

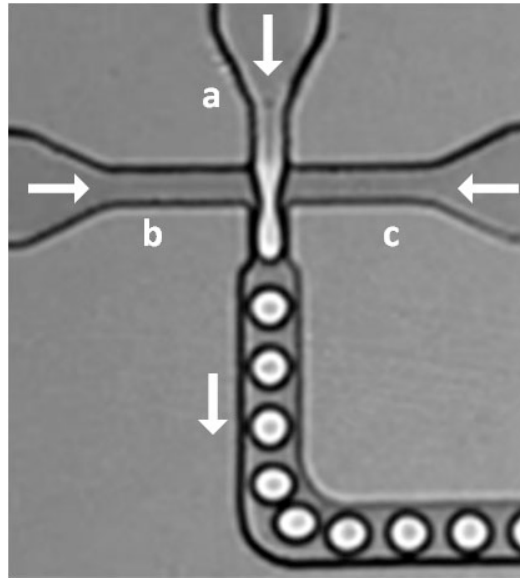
Additionally, we generated a 1536 primer pair droplet library such that each of 384 different types of droplets contained four primer pairs. The procedure for generating the library was identical to the procedure used for the library generation described above except that four primer pairs instead of one were mixed in the solutions used to generate the library droplets. The 1536 primer pair library was merged with Coriell sample NA18858 using the sequence enrichment procedure under the same conditions as described in the corresponding section above. The sample was sequenced on an Illumina GAI instrument and the sequencing data was used to make SNP calls. The SNP calls were compared to those in the HapMap database. The concordant rate for the droplets containing 4 sets of primers was similar (98.8%) as for the individually encapsulated primers (98.7%). This data demonstrates that using up to 4 primer pairs in the droplet library does not affect the performance of primer droplet libraries while allowing a significant decrease in required template. When the expanded content format is scaled to 5 to 10 primer pairs in each droplet we anticipate that the current libraries with 4000 types of primer droplets will be sufficient to capture 20,000 to 40,000 amplicons representing 1/10th to 1/5th of the exome.

References

1. Gnirke, A. et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* **27**, 182-189 (2009).
2. http://www.broadinstitute.org/annotation/hybrid_selection/hybrid_selection.html.
3. Turner, E.H., Lee, C., Ng, S.B., Nickerson, D.A. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*, 1-2 (2009).

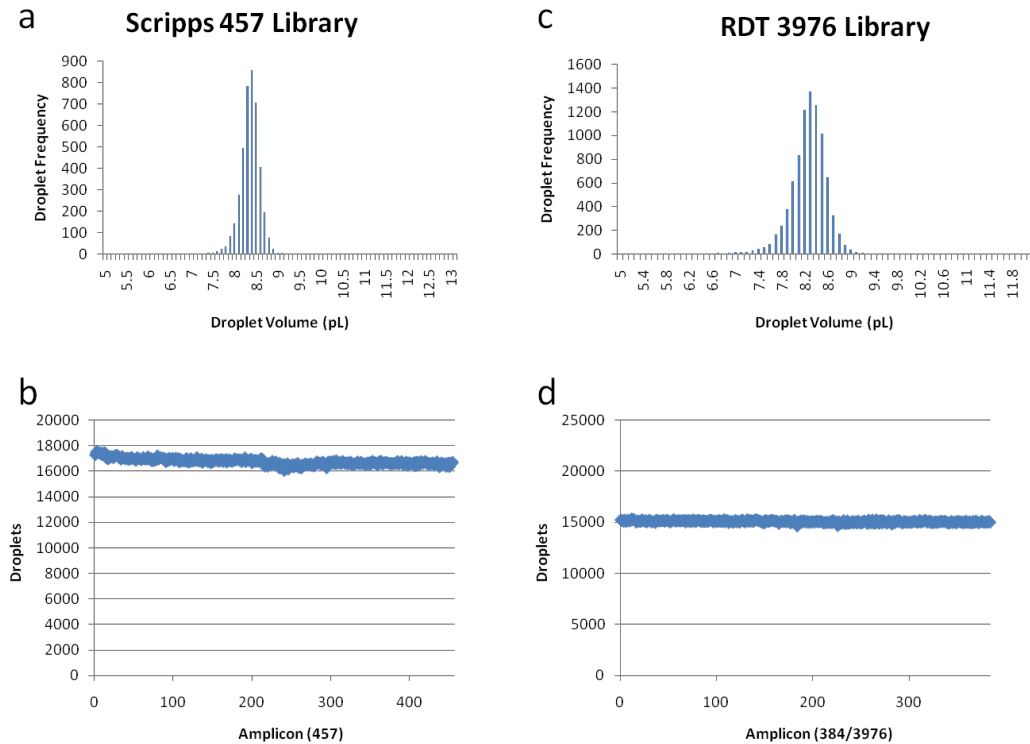
Supplemental Figure 1.

Microfluidic nozzle for generation of primer pair droplets. An aqueous solution of the forward and reverse primers for a single amplicon is infused through channel (a) into the nozzle where opposing streams of a fluorocarbon oil (b) and (c) focus the flow through the orifice to generate uniform primer droplets.



Supplemental Figure 2.

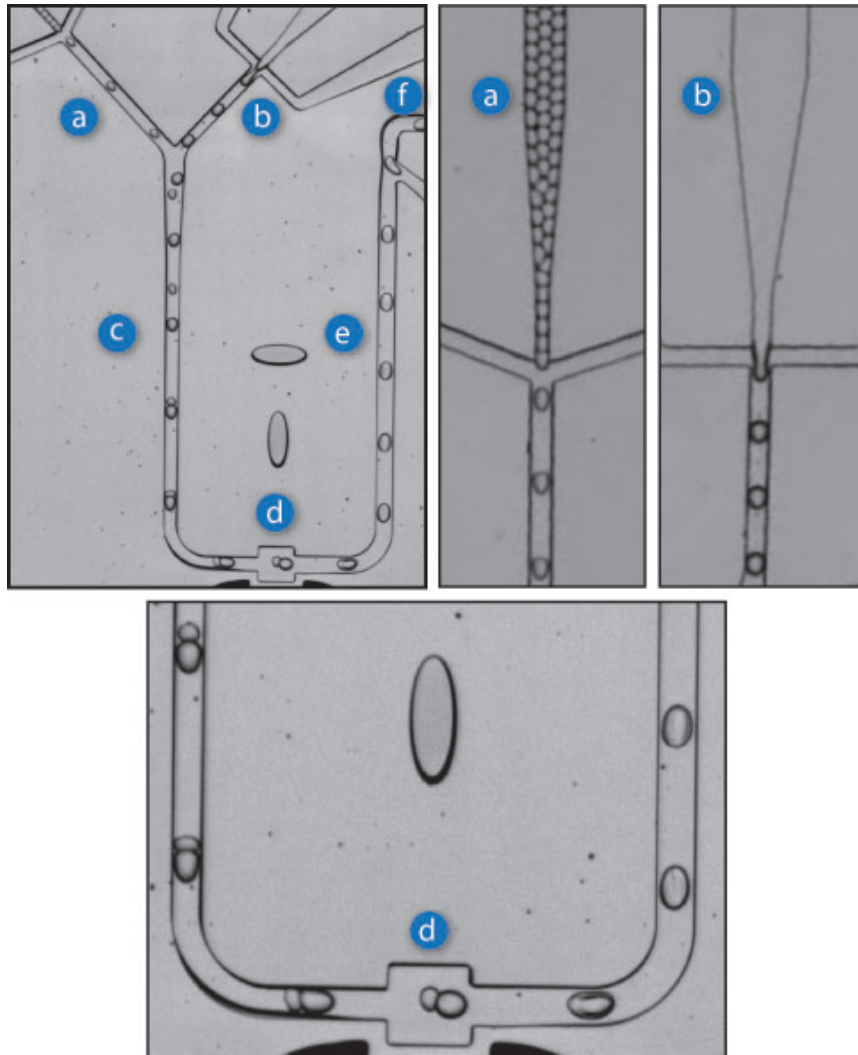
Primer Library Quality Control for the validation phase 457 amplicon and the scale-up phase 3976 amplicon libraries. Primer libraries are tested for primer pair droplet size and uniformity as well as library element representation within the primer library. **(a)** droplet size and uniformity data for the validation phase 457 amplicon library: Avg. droplet volume = 8.3 pL, CV = 3.4%. **(b)** library element representation data within the validation phase 457 amplicon library: Min = 16069; Max = 17574; Avg. = 16773; Std. Dev. = 239.523; CV = 8.9% **(c)** droplet size and uniformity data for the scale-up phase 3976 amplicon library: Avg. droplet volume = 8.2 pL, CV = 4.4%. **(d)** library element representation data within the scale-up phase 3976 amplicon library: Min = 14653; Max = 15323 Avg. = 15084; Std. Dev. = 93.84; CV = 3.6%.



Supplemental Figure 3

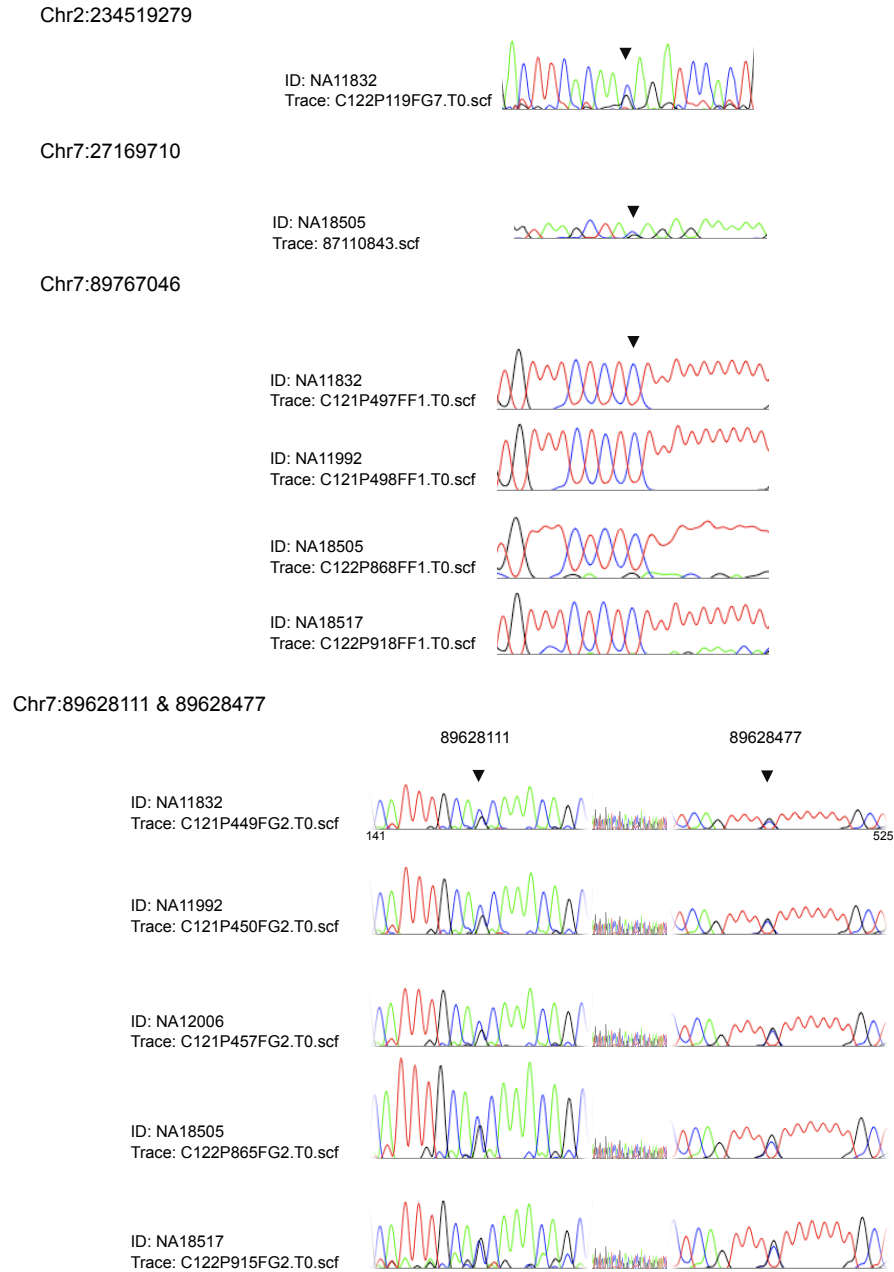
Processing droplets in a microfluidic chip. The optical image shows all key microfluidic elements for processing droplets. These include: **(a)** Spacing of the Library droplets with carrier-oil. **(b)** Generation of Template droplets of uniform size. **(c)** One-to-one pairing of Template and Library droplets. Pairing is achieved by actively matching the rate of library droplet introduction to the rate of template droplet generation and by using the property that small droplets move faster than large droplets and will catch-up to the preceding large droplet. **(d)** Merging of Template and Library droplets by field-induced coalescence to make individual PCR droplets. **(e)** PCR droplets are transported to a collect line. **(f)** PCR droplets are collected off chip into a standard 0.2 ml PCR tube. Enlarged regions of the image show: **(a)** library droplets before they are spaced, **(b)** template droplets being generated and **(d)** the region where droplets merge in an electric field.

The PCR droplet generation can be viewed in the attached movie (Supplemental Movie 1).



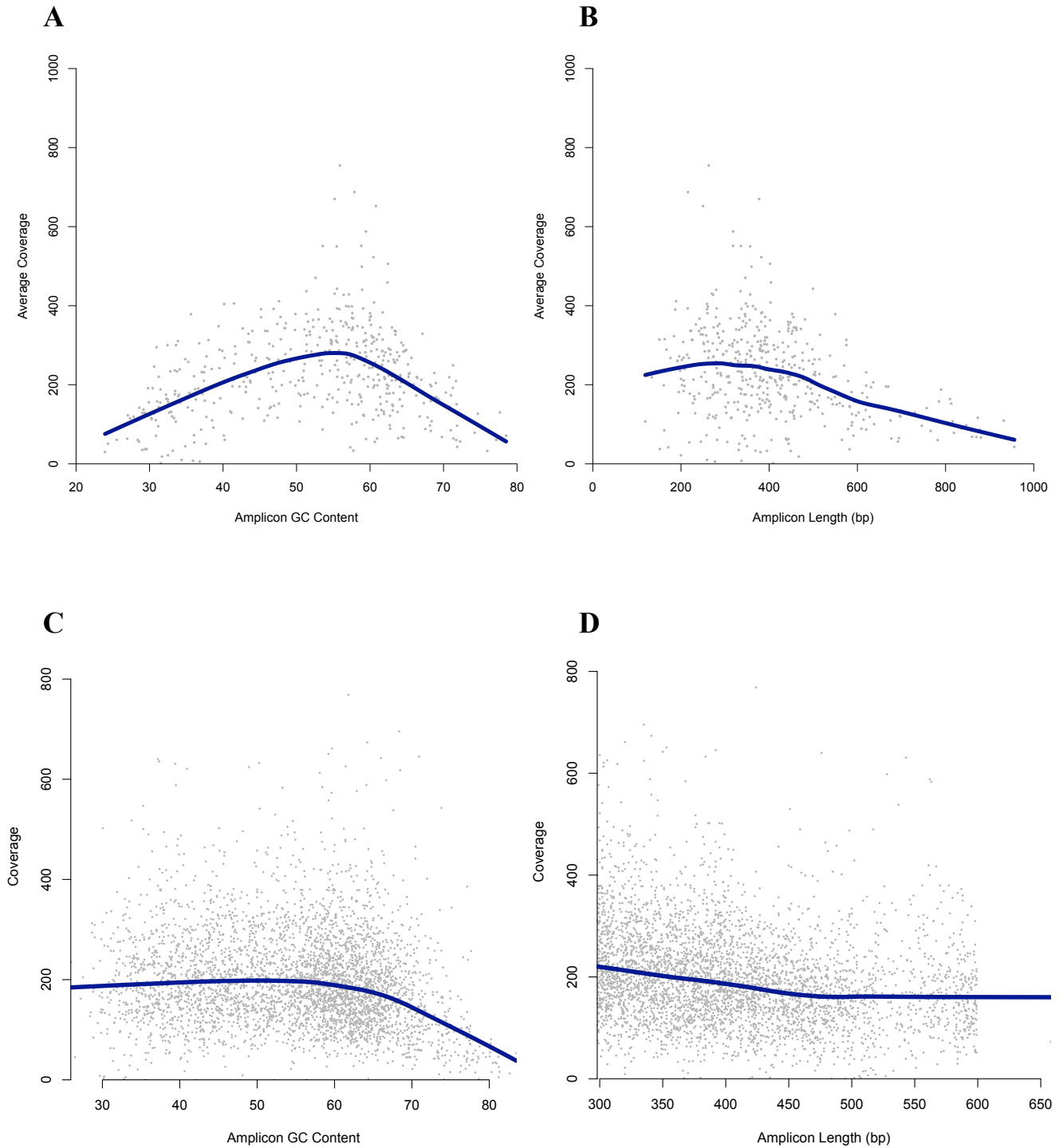
Supplemental Figure 4

HapMap Sequence traces for discordant alleles in the ENCODE regions.



Supplemental Figure 5

Effect of GC% and amplicon length on average sequence coverage. Average sequence coverage of the primer trimmed amplicons for the 456 amplicons in the validation phase (A,B) and 3976 amplicons in the scale-up phase (C,D) plotted against the full amplicons GC% (A,C) and length (B,D).



Supplemental Table 1. Summary of targeted genes in Validation Phase. In total, 435 exons from 47 genes were targeted. The 435 exons were split into 457 amplicons, each under 1 kb in total length (Supplemental Methods).

Chr.	Gene	No. Exon	No. Amplicon	Median Size (bp)	Size Range (bp)	Category
Chr2	HJURP	9	11	358	135-814	ENCODE
Chr2	UGT1A10	5	6	444	362-604	ENCODE
Chr7	C7orf63	22	22	344.5	193-583	ENCODE
Chr7	CLDN12	1	2	627.5	623-632	ENCODE
Chr7	EVX1	3	3	767	589-783	ENCODE
Chr7	GTPBP10	8	8	335	153-479	ENCODE
Chr7	HOXA1	2	3	584	580-589	ENCODE
Chr7	HOXA2	2	3	621	583-697	ENCODE
Chr7	HOXA3	2	3	700	699-861	ENCODE
Chr7	HOXA4	2	3	621	522-747	ENCODE
Chr7	HOXA5	2	2	658	621-695	ENCODE
Chr7	HOXA6	2	2	576	554-598	ENCODE
Chr7	HOXA7	2	2	527.5	445-610	ENCODE
Chr7	HOXA9	2	2	690.5	500-881	ENCODE
Chr7	HOXA10	3	4	575.5	375-693	ENCODE
Chr7	HOXA11	2	3	434	360-588	ENCODE
Chr7	HOXA13	2	3	566	499-957	ENCODE
Chr7	STEAP1	4	4	450	246-879	ENCODE
Chr7	STEAP2	6	6	528	220-676	ENCODE
Chr9	C9orf106	1	2	510.5	445-576	ENCODE
Chr9	CRAT	14	14	349	152-444	ENCODE
Chr9	NUP188	44	44	336	177-516	ENCODE
Chr9	DOLPP1	8	8	317	235-461	ENCODE
Chr9	FAM73B	15	15	380	242-539	ENCODE
Chr9	IER5L	1	3	731	515-820	ENCODE
Chr9	PHYHD1	11	11	284	120-391	ENCODE
Chr9	PPP2R4	11	11	338	188-588	ENCODE
Chr9	SH3GLB2	11	11	383	209-592	ENCODE
Chr11	APOA1	3	4	488.5	221-568	Haemostatis/Thrombosis
Chr16	CETP	16	16	362.5	162-457	Haemostatis/Thrombosis
Chr5	F2R	2	3	874	401-933	Haemostatis/Thrombosis
Chr17	GP1BA	1	4	711.5	576-791	Haemostatis/Thrombosis
Chr1	LRP8	19	19	396	164-678	Haemostatis/Thrombosis
Chr20	PLTP	15	15	297	216-418	Haemostatis/Thrombosis
Chr2	PROC	8	8	393	163-848	Haemostatis/Thrombosis
Chr20	PROCR	4	4	402.5	347-561	Haemostatis/Thrombosis
Chr20	THBD	1	3	836	817-871	Haemostatis/Thrombosis
Chr9	UGCG	9	9	392	308-473	Haemostatis/Thrombosis
Chr8	TRPA1	27	27	329	197-504	TRP Channels
Chr4	TRPC3	11	12	451	330-666	TRP Channels
Chr11	TRPM5	24	24	400	262-542	TRP Channels
Chr2	TRPM8	24	24	347	209-671	TRP Channels & ENCODE
Chr17	TRPV1	15	15	439	281-639	TRP Channels
Chr17	TRPV2	14	14	400	172-612	TRP Channels
Chr17	TRPV3	17	17	349	144-515	TRP Channels
Chr12	TRPV4	15	15	396	193-620	TRP Channels
Chr12	PAH	13	13	330	168-462	Phenylketonuria
Total	47	435	457	434	120-957	

Supplemental Table 2. Primer Design Strategy for Validation Phase. The primer sets were designed using five stages. The parameters used in the first stage were the most stringent and were loosened in the following stages to allow for primer design of all 457 amplicons (Supplemental Methods).

Stage		Masking		Padding Sequence (bp)	Primer3 Parameter* Tm (°C)	Number of Primer Sets																																																
		Repeat	SNP			Designed	Failed																																															
1	5'-primer	+	+	200	59.5-60.5	424	33																																															
	3'-primer	+	+					2	5'-primer	+	+	200	59.5-60.5	10	23	3'-primer	-	+	3	5'-primer	-	+	200	59.5-60.5	17	6	3'-primer	+	+	4	5'-primer	+	+	300	57.5-62.5	5	1	3'-primer	-	+	5	5'-primer	-	+	300	57.5-62.5	1	0	3'-primer	+	+	Total		
2	5'-primer	+	+	200	59.5-60.5	10	23																																															
	3'-primer	-	+					3	5'-primer	-	+	200	59.5-60.5	17	6	3'-primer	+	+	4	5'-primer	+	+	300	57.5-62.5	5	1	3'-primer	-	+	5	5'-primer	-	+	300	57.5-62.5	1	0	3'-primer	+	+	Total						457							
3	5'-primer	-	+	200	59.5-60.5	17	6																																															
	3'-primer	+	+					4	5'-primer	+	+	300	57.5-62.5	5	1	3'-primer	-	+	5	5'-primer	-	+	300	57.5-62.5	1	0	3'-primer	+	+	Total						457																		
4	5'-primer	+	+	300	57.5-62.5	5	1																																															
	3'-primer	-	+					5	5'-primer	-	+	300	57.5-62.5	1	0	3'-primer	+	+	Total						457																													
5	5'-primer	-	+	300	57.5-62.5	1	0																																															
	3'-primer	+	+					Total						457																																								
Total						457																																																

*Primer length was kept constant at 18-27 bp.

Supplemental Table 4. Coverage statistics per sample for validation phase.

Percentage of all bases that had a sequence depth of 1/5 the mean coverage or greater, less than 5 times the mean, and fell between 1/5 and 5 times the mean.

Sample	PCR Method	Mean Coverage	Proportion of Mapped Bases		
			>1/5 Mean	<5x Mean	> 1/5 & <5x
NA11832	Traditional	294	93.5%	100%	93.5%
NA11832	Microdroplet	274	89.93%	99.91%	89.83%
NA11992	Traditional	235	92.78%	99.96%	92.73%
NA11992	Microdroplet	188	89.95%	99.98%	89.93%
NA12006	Traditional	247	93.46%	100%	93.46%
NA12006	Microdroplet	186	90.08%	99.98%	90.06%
NA18505	Traditional	242	91.16%	99.97%	91.13%
NA18505	Microdroplet	193	90.03%	99.96%	89.99%
NA18517 ^a	Traditional	165	91.06%	99.91%	90.97%
NA18517 ^a	Microdroplet	82	89.14%	99.98%	89.11%
NA18489	Traditional	271	94.88%	99.98%	94.86%
NA18489	Microdroplet	342	91.77%	99.9%	91.67%

a - NA18517 had fewer reads than the other 5 samples for both traditional and microdroplet PCR. This

Supplemental Table 5. Discordant variants between sequence data and HapMap genotypes in validation phase.

Method	Sample	Chr	Position	Reference Allele	HapMap Alleles		Sequencing Alleles		Variant Quality	Neighbor Quality	Base Frequency ^a	Amplicon	Amplicon ^d
					A	B	A	B					
Microdroplet	NA11832	2	234519279	G	G	G	C	G	213	62	46.31	TRPM8_6_1	
Traditional	NA11832	2	234519279	G	G	G	C	G	209	62	52.12	TRPM8_6_1	
Microdroplet	NA18505	7	27169710 ^c	C	C	G	C	C	49	40	65.91	HOXA9_1_1	
Microdroplet	NA11832	7	89628111	C	G	G	C	C	209	62	90.43	STEAP1_2_1	
Traditional	NA11832	7	89628111	C	G	G	C	C	255	62	80.33	STEAP1_2_1	
Microdroplet	NA11992	7	89628111	C	G	G	C	C	165	62	80.34	STEAP1_2_1	
Traditional	NA11992	7	89628111	C	G	G	C	C	255	62	84	STEAP1_2_1	
Microdroplet	NA12006	7	89628111	C	G	G	C	C	255	62	90.79	STEAP1_2_1	
Traditional	NA12006	7	89628111	C	G	G	C	C	255	62	86.08	STEAP1_2_1	
Microdroplet	NA11832	7	89628477	C	G	G	C	C	255	62	100	STEAP1_2_1	
Traditional	NA11832	7	89628477	C	G	G	C	C	255	62	94.37	STEAP1_2_1	
Microdroplet	NA11992	7	89628477	C	G	G	C	C	255	62	100	STEAP1_2_1	
Traditional	NA11992	7	89628477	C	G	G	C	C	255	62	98.19	STEAP1_2_1	
Microdroplet	NA12006	7	89628477	C	G	G	C	C	255	62	99.21	STEAP1_2_1	
Traditional	NA12006	7	89628477	C	G	G	C	C	255	62	97.08	STEAP1_2_1	
Microdroplet	NA18505	7	89628477	C	G	G	C	C	199	62	100	STEAP1_2_1	
Traditional	NA18505	7	89628477	C	G	G	C	C	255	62	97.25	STEAP1_2_1	
Microdroplet	NA18517	7	89628477	C	G	G	C	C	202	62	100	STEAP1_2_1	
Traditional	NA18517	7	89628477	C	G	G	C	C	255	62	96	STEAP1_2_1	
Microdroplet	NA11832	7	89767046	C	C	T	C	C	255	62	93.57	FLJ21062_17_1	
Traditional	NA11832	7	89767046	C	C	T	C	C	255	62	89.93	FLJ21062_17_1	
Microdroplet	NA11992	7	89767046	C	C	T	C	C	140	62	88.52	FLJ21062_17_1	
Traditional	NA11992	7	89767046	C	C	T	C	C	127	62	92.73	FLJ21062_17_1	
Microdroplet	NA18505	7	89767046	C	C	T	C	C	54	56	81.48	FLJ21062_17_1	
Traditional	NA18505	7	89767046	C	C	T	C	C	84	62	94.12	FLJ21062_17_1	
Microdroplet	NA18517	7	89767046	C	C	T	C	C	86	62	96.43	FLJ21062_17_1	
Traditional	NA18517	7	89767046	C	C	T	C	C	96	62	85.07	FLJ21062_17_1	
Microdroplet	NA11832	8	73104764	G	G	T	T	T	177	62	18.71	TRPA1_4_1	
Traditional	NA11832	8	73104764	G	G	T	T	T	112	62	9.52	TRPA1_4_1	
Microdroplet	NA12006	8	73104764	G	G	T	T	T	132	62	14.43	TRPA1_4_1	
Traditional	NA12006	8	73104764	G	G	T	T	T	102	62	8.16	TRPA1_4_1	
Traditional	NA18517	8	73134059 ^c	T	G	T	T	T	65	22	79.22	TRPA1_20_1	
Microdroplet	NA18505	9	130979840	G	A	A	G	G	205	62	98.59	IER5L_1_2	IER5L_1_3
Traditional	NA18505	9	130979840	G	A	A	G	G	57	50	100	IER5L_1_2	IER5L_1_3
Microdroplet	NA18517	11	2396484	G	G	G	A	G	167	62	41.82	TRPM5_20_1	
Traditional	NA18517	11	2396484	G	G	G	A	G	165	62	52.5	TRPM5_20_1	
Microdroplet	NA11832	12	101764760	C	C	C	A	C	187	62	50.28	PAH_5_1	
Traditional	NA11832	12	101764760	C	C	C	A	C	138	62	52.45	PAH_5_1	
Microdroplet	NA11992	12	101764760	C	C	C	A	C	143	62	50.4	PAH_5_1	
Traditional	NA11992	12	101764760	C	C	C	A	C	121	62	49.15	PAH_5_1	
Microdroplet	NA18505	12	101764760	C	C	C	A	C	176	62	53.28	PAH_5_1	
Traditional	NA18505	12	101764760	C	C	C	A	C	149	62	44.35	PAH_5_1	
Microdroplet	NA18489	16	55562390	G	A	G	G	G	255	62	98.27	CETP_6_1	
Traditional	NA18489	16	55562390	G	A	G	G	G	255	62	97.88	CETP_6_1	

a - Frequency of reads mapped that contains the reference allele. **b** - class 1: homozygote reference by PCR, homozygote alternate in HapMap, class 2: heterozygote by PCR, homozygote reference in HapMap, class3: homozygote reference by PCR, heterozygote in HapMap. **c** - For variants at both 27169710 and 73134059 the PCR method not shown produced a no call at the position. **d** - Variants amplified by two PCR primer pairs

Supplemental Table 6. False positives in validation phase.

Chr	Position	Position	Method	Sample	Reference Allele		HapMap Alleles		Sequencing Alleles	
					Allele	A	B	A	B	
										A
2	234210504	234210504	Microdroplet	NA18517	T	T	T	C	T	
2	234210504		Traditional	NA18517	T	T	T	C	T	
2	234341152	234341152	Microdroplet	NA18505	T	T	T	C	T	
2	234341152		Traditional	NA18505	T	T	T	C	T	
2	234345991	234345991	Traditional	NA11992	T	T	T	C	T	
2	234414008	234414008	Microdroplet	NA18517	G	G	G	A	G	
2	234414008		Traditional	NA18517	G	G	G	A	G	
2	234427637	234427637	Microdroplet	NA18505	C	C	C	C	T	
2	234427637		Traditional	NA18505	C	C	C	C	T	
2	234427663	234427663	Microdroplet	NA11832	C	C	C	G	G	
2	234427663		Microdroplet	NA11992	C	C	C	C	G	
2	234427663		Microdroplet	NA12006	C	C	C	C	G	
2	234427663		Traditional	NA11832	C	C	C	G	G	
2	234427663		Traditional	NA11992	C	C	C	C	G	
2	234427663		Traditional	NA12006	C	C	C	C	G	
2	234427685	234427685	Microdroplet	NA11832	C	C	C	C	G	
2	234427685		Traditional	NA11832	C	C	C	C	G	
2	234511046	234511046	Traditional	NA11832	C	C	C	A	C	
2	234537961	234537961	Microdroplet	NA11992	G	G	G	A	G	
2	234537961		Microdroplet	NA12006	G	G	G	A	G	
2	234537961		Microdroplet	NA18517	G	G	G	A	G	
2	234537961		Traditional	NA11992	G	G	G	A	G	
2	234537961		Traditional	NA18517	G	G	G	A	G	
2	234542947	234542947	Traditional	NA12006	A	A	A	A	G	
2	234542947		Traditional	NA18505	A	A	A	A	G	
2	234542947		Traditional	NA18517	A	A	A	A	G	
7	27101313	27101313	Microdroplet	NA18505	A	A	A	T	T	
7	27101313		Traditional	NA18505	A	A	A	T	T	
7	27101883	27101883	Traditional	NA18517	C	C	C	C	G	
7	27106771	27106771	Microdroplet	NA18517	C	C	C	A	C	
7	27113998	27113998	Microdroplet	NA12006	C	C	C	A	C	
7	27114800	27114800	Microdroplet	NA12006	G	G	G	C	G	
7	27114800		Traditional	NA12006	G	G	G	C	G	
7	27136142	27136142	Microdroplet	NA11832	T	T	T	C	C	
7	27136142		Microdroplet	NA11992	T	T	T	C	C	
7	27136142		Microdroplet	NA12006	T	T	T	C	C	
7	27136142		Microdroplet	NA18505	T	T	T	C	C	
7	27136142		Microdroplet	NA18517	T	T	T	C	C	
7	27136142		Traditional	NA11832	T	T	T	C	C	
7	27136142		Traditional	NA11992	T	T	T	C	C	
7	27136142		Traditional	NA12006	T	T	T	C	C	
7	27136142		Traditional	NA18505	T	T	T	C	C	
7	27136142		Traditional	NA18517	T	T	T	C	C	
7	27136459	27136459	Microdroplet	NA11832	A	A	A	G	G	
7	27136459		Microdroplet	NA11992	A	A	A	G	G	
7	27136459		Microdroplet	NA12006	A	A	A	G	G	
7	27136459		Microdroplet	NA18505	A	A	A	G	G	
7	27136459		Microdroplet	NA18517	A	A	A	G	G	
7	27136459		Traditional	NA11832	A	A	A	G	G	
7	27136459		Traditional	NA11992	A	A	A	G	G	
7	27136459		Traditional	NA12006	A	A	A	G	G	
7	27136459		Traditional	NA18505	A	A	A	G	G	
7	27136459		Traditional	NA18517	A	A	A	G	G	
7	27136670	27136670	Microdroplet	NA11832	T	T	T	G	G	
7	27136670		Microdroplet	NA11992	T	T	T	G	G	
7	27136670		Microdroplet	NA12006	T	T	T	G	G	
7	27136670		Microdroplet	NA18505	T	T	T	G	G	
7	27136670		Microdroplet	NA18517	T	T	T	G	G	
7	27136670		Traditional	NA11832	T	T	T	G	G	
7	27136670		Traditional	NA11992	T	T	T	G	G	
7	27136670		Traditional	NA12006	T	T	T	G	G	
7	27136670		Traditional	NA18505	T	T	T	G	G	
7	27136670		Traditional	NA18517	T	T	T	G	G	
7	27149158	27149158	Microdroplet	NA18505	G	G	G	A	G	
7	27149158		Microdroplet	NA18517	G	G	G	A	G	
7	27149158		Traditional	NA18505	G	G	G	A	G	
7	27149158		Traditional	NA18517	G	G	G	A	G	
7	27152045	27152045	Microdroplet	NA18505	G	G	G	A	G	
7	27152045		Microdroplet	NA18517	G	G	G	A	G	
7	27152045		Traditional	NA18505	G	G	G	A	G	
7	27152045		Traditional	NA18517	G	G	G	A	G	
7	27162519	27162519	Microdroplet	NA12006	G	G	G	A	G	
7	27162519		Traditional	NA12006	G	G	G	A	G	
7	27171257	27171257	Microdroplet	NA11992	C	C	C	A	C	
7	27171257		Microdroplet	NA12006	C	C	C	A	C	
7	27171257		Traditional	NA11992	C	C	C	A	C	
7	27171257		Traditional	NA12006	C	C	C	A	C	
7	27171684	27171684	Traditional	NA18517	C	C	C	C	G	
7	27205216	27205216	Microdroplet	NA11832	C	C	C	G	G	
7	27205216		Microdroplet	NA11992	C	C	C	G	G	
7	27205216		Microdroplet	NA12006	C	C	C	G	G	
7	27205216		Microdroplet	NA18505	C	C	C	G	G	
7	27205216		Microdroplet	NA18517	C	C	C	G	G	
7	27205216		Traditional	NA11832	C	C	C	G	G	
7	27205216		Traditional	NA11992	C	C	C	G	G	
7	27205216		Traditional	NA12006	C	C	C	G	G	
7	27205216		Traditional	NA18505	C	C	C	G	G	
7	27205216		Traditional	NA18517	C	C	C	G	G	
7	27205718	27205718	Microdroplet	NA18517	C	C	C	T	T	
7	27205718		Traditional	NA18517	C	C	C	C	T	
7	27205726	27205726	Traditional	NA18517	G	G	G	G	T	

7	27206075	27206075	Microdroplet	NA18517	C	C	C	C	T
7	27206075		Traditional	NA18517	C	C	C	C	T
7	27249029	27249029	Traditional	NA11992	G	G	G	A	G
7	27249654	27249654	Microdroplet	NA11992	C	C	C	C	T
7	27249654		Traditional	NA11992	C	C	C	C	T
7	27252215	27252215	Microdroplet	NA18505	A	A	A	A	C
7	27252215		Traditional	NA18505	A	A	A	A	C
7	89631723	89631723	Traditional	NA11832	G	G	G	G	T
7	89699558	89699558	Traditional	NA11992	G	G	G	G	T
7	89729133	89729133	Microdroplet	NA12006	C	C	C	A	C
7	89744582	89744582	Microdroplet	NA18505	C	C	C	A	C
7	89753393	89753393	Microdroplet	NA11992	C	C	C	A	C
7	89753393		Traditional	NA18517	C	C	C	A	C
7	89774976	89774976	Microdroplet	NA18505	G	G	G	A	G
7	89839432	89839432	Microdroplet	NA12006	G	G	G	C	G
7	89839432		Traditional	NA12006	G	G	G	C	G
7	89852431	89852431	Microdroplet	NA18517	C	C	C	A	C
9	130738637	130738637	Microdroplet	NA18505	G	G	G	A	G
9	130738637		Traditional	NA18505	G	G	G	A	G
9	130749838	130749838	Microdroplet	NA18505	G	G	G	A	G
9	130749838		Traditional	NA18505	G	G	G	A	G
9	130749946	130749946	Microdroplet	NA18505	A	A	A	A	G
9	130749946		Traditional	NA18505	A	A	A	A	G
9	130749988	130749988	Microdroplet	NA18517	G	G	G	A	G
9	130749988		Traditional	NA18517	G	G	G	A	G
9	130782645	130782645	Microdroplet	NA18517	T	T	T	C	T
9	130782645		Traditional	NA18517	T	T	T	C	T
9	130788934	130788934	Microdroplet	NA18517	T	T	T	C	T
9	130788934		Traditional	NA18517	T	T	T	C	T
9	130800813	130800813	Microdroplet	NA11992	A	A	A	A	T
9	130800813		Traditional	NA11992	A	A	A	A	T
9	130803835	130803835	Microdroplet	NA18517	A	A	A	A	C
9	130803835		Traditional	NA18517	A	A	A	A	C
9	130808083	130808083	Microdroplet	NA18517	G	G	G	A	G
9	130808083		Traditional	NA18517	G	G	G	A	G
9	130811326	130811326	Microdroplet	NA11832	C	C	C	C	T
9	130811326		Microdroplet	NA11992	C	C	C	T	T
9	130811326		Microdroplet	NA12006	C	C	C	T	T
9	130811326		Traditional	NA11832	C	C	C	C	T
9	130811326		Traditional	NA11992	C	C	C	T	T
9	130811326		Traditional	NA12006	C	C	C	T	T
9	130830351	130830351	Microdroplet	NA18517	G	G	G	C	G
9	130830352	130830352	Microdroplet	NA11832	C	C	C	C	T
9	130830352		Microdroplet	NA11992	C	C	C	T	T
9	130830352		Microdroplet	NA12006	C	C	C	T	T
9	130830352		Microdroplet	NA18517	C	C	C	C	T
9	130830352		Traditional	NA11992	C	C	C	T	T
9	130830454	130830454	Microdroplet	NA11832	T	T	T	G	G
9	130830454		Microdroplet	NA11992	T	T	T	G	G
9	130830454		Microdroplet	NA12006	T	T	T	G	G
9	130830454		Microdroplet	NA18505	T	T	T	G	G
9	130830454		Microdroplet	NA18517	T	T	T	G	G
9	130830454		Traditional	NA11992	T	T	T	G	G
9	130851869	130851869	Microdroplet	NA18517	C	C	C	C	T
9	130851869		Traditional	NA18517	C	C	C	C	T
9	130887224	130887224	Microdroplet	NA18517	G	G	G	A	G
9	130887224		Traditional	NA18517	G	G	G	A	G
9	130888884	130888884	Microdroplet	NA18517	C	C	C	C	T
9	130888884		Traditional	NA18517	C	C	C	C	T
9	130897438	130897438	Microdroplet	NA11832	G	G	G	A	G
9	130897438		Traditional	NA11832	G	G	G	A	G
9	130900088	130900088	Microdroplet	NA18505	C	C	C	C	T
9	130900088		Traditional	NA18505	C	C	C	C	T
9	130902602	130902602	Microdroplet	NA11832	G	G	G	A	A
9	130902602		Microdroplet	NA11992	G	G	G	A	A
9	130902602		Microdroplet	NA12006	G	G	G	A	A
9	130902602		Microdroplet	NA18505	G	G	G	A	A
9	130902602		Traditional	NA11832	G	G	G	A	A
9	130902602		Traditional	NA11992	G	G	G	A	A
9	130902602		Traditional	NA12006	G	G	G	A	A
9	130902602		Traditional	NA18505	G	G	G	A	A
9	130902815	130902815	Microdroplet	NA18505	C	C	C	C	T
9	130902815		Traditional	NA18505	C	C	C	C	T
9	130922582	130922582	Microdroplet	NA11832	T	T	T	C	C
9	130922582		Microdroplet	NA11992	T	T	T	C	C
9	130922582		Microdroplet	NA12006	T	T	T	C	C
9	130922582		Microdroplet	NA18505	T	T	T	C	C
9	130922582		Microdroplet	NA18517	T	T	T	C	C
9	130922582		Traditional	NA11832	T	T	T	C	C
9	130922582		Traditional	NA11992	T	T	T	C	C
9	130922582		Traditional	NA12006	T	T	T	C	C
9	130922582		Traditional	NA18505	T	T	T	C	C
9	130922582		Traditional	NA18517	T	T	T	C	C
9	130925100	130925100	Traditional	NA18517	G	G	G	A	G
9	130938701	130938701	Microdroplet	NA18505	A	A	A	G	G
9	130938701		Microdroplet	NA18517	A	A	A	G	G
9	130938701		Traditional	NA18505	A	A	A	G	G
9	130938701		Traditional	NA18517	A	A	A	G	G
9	130939667	130939667	Microdroplet	NA11832	T	T	T	C	C
9	130939667		Microdroplet	NA11992	T	T	T	C	C
9	130939667		Microdroplet	NA12006	T	T	T	C	C
9	130939667		Traditional	NA11832	T	T	T	C	C
9	130939667		Traditional	NA11992	T	T	T	C	C
9	130939667		Traditional	NA12006	T	T	T	C	C

a - Frequency of reads mapped which the reference allele b - Variant calls between traditional and microdroplet PCR differed. For a

Supplemental Table 7. Scale up phase discordant variants between sequence data and HapMap genotypes.

Chr	Position	Reference Allele	HapMap Alleles		Sequencing Alleles				Illumina GAI					Roche 454			
			A	B	GA II A	GA II B	454 A	454 B	Variant Quality	Neighbor Quality	Base Frequency ^a	Base Called ^b	Base Discrepant	Variant Probability ^a	Base Frequency	Base Called ^b	Base Discrepant
1	35431339	G	C	C	G	G	G	G	255	62	99.05	yes	yes	-	100	yes	yes
1	159911794	G	G	G	G	G	C	G	255	62	98.76	yes	no	2.04E-27	68.57	yes	yes
1	226499840	A	A	G	A	A	A	A	230	62	99.1	yes	yes	-	100	yes	yes
1	226510920	T	G	G	T	T	T	T	253	62	100	yes	yes	-	100	yes	yes
3	12398113	G	G	G	C	G	C	G	165	62	48.1	yes	yes	6.61E-23	47.3	yes	yes
3	37510988	C	G	G	C	G	C	G	189	62	52.05	yes	yes	2.86E-10	43.37	yes	yes
3	136127326	G	T	T	T	T	G	T	255	62	0	yes	no	3.04E-09	33.33	yes	yes
3	185781762	G	A	G	A	G	A	A	53	62	22.94	yes	no	1.03E-14	17.14	yes	yes
5	176828423	G	G	G	A	G	A	G	165	62	57.14	yes	yes	3.36E-31	47.86	yes	yes
6	30967283	C	C	G	C	C	C	C	255	62	98.69	yes	yes	-	100	yes	yes
6	33051323	C	C	T	C	C	C	C	241	62	100	yes	yes	-	100	yes	yes
6	44326098	A	A	G	A	A	A	G	63	62	72	yes	yes	1.73E-08	34.62	yes	no
6	109091842	T	C	T	T	T	C	T	255	62	100	yes	yes	3.63E-25	72.19	yes	no
6	160401527	T	C	C	C	T	C	T	182	62	52.52	yes	yes	8.73E-19	60.94	yes	yes
7	5993514	G	G	G	A	G	A	G	62	62	57.36	yes	yes	8.45E-09	75	yes	yes
7	100248593	A	A	A	A	G	-	-	171	62	50.85	yes	yes	1.41E-05	52.38	no	-
7	150795122	A	G	G	A	A	A	A	255	62	100	yes	yes	-	100	yes	yes
9	133054166	T	C	C	C	C	C	T	255	62	0	yes	no	1.07E-32	28	yes	yes
10	76458866	C	C	T	-	-	C	C	1	62	81.93	no	-	3.27E-01	94.12	yes	yes
11	85369990	A	C	C	A	C	A	C	255	62	48.13	yes	yes	2.20E-27	39.34	yes	yes
11	116582149	C	C	C	C	C	C	T	255	62	99.3	yes	no	1.92E-06	63.16	yes	yes
11	116582159	G	A	A	G	G	-	-	255	62	97.76	yes	yes	2.47E-04	63.16	no	-
11	116582219	A	C	C	A	A	A	C	255	62	100	yes	yes	1.84E-09	61.54	yes	yes
12	4253419	T	C	C	C	C	C	T	255	62	1.09	yes	no	1.01E-06	25	yes	yes
15	82279640	A	G	G	G	G	A	G	255	62	0.46	yes	no	8.30E-08	45.1	yes	yes
17	21142143	T	T	T	C	T	C	T	162	62	60.34	yes	yes	4.15E-18	47.37	yes	yes
17	21142312	T	T	T	C	T	C	T	175	62	44.96	yes	yes	1.53E-20	55	yes	yes
17	21144883	G	G	G	G	G	C	G	97	62	83.97	yes	no	7.49E-13	77.06	yes	yes
17	21157990	A	A	A	A	G	A	G	120	62	68.25	yes	yes	7.88E-25	42.68	yes	yes
17	21158106	G	G	G	A	G	A	G	192	62	51.59	yes	yes	2.17E-65	42.42	yes	yes
18	7024464	T	A	T	A	A	A	A	255	62	0.67	yes	yes	6.99E-12	0	yes	yes
19	5953701	C	C	C	C	T	C	T	168	62	45.28	yes	yes	5.61E-09	30	yes	yes
19	50563876	C	T	T	C	T	T	T	159	62	61.73	yes	yes	6.72E-69	0.96	yes	no
20	35448117	G	G	G	A	G	A	G	181	62	53.96	yes	yes	6.04E-16	46.94	yes	yes
20	35464469	C	T	T	C	T	C	T	176	62	45.86	yes	yes	6.52E-11	51.16	yes	yes
X	107863596	G	C	G	C	G	G	G	175	62	54.29	yes	no	5.31E-11	80.56	yes	yes

a - Frequency of reads mapped which the reference allele. b - Filter criteria for variant calling explained are explained in the supplemental methods

Supplemental Table 8. Comparison of enrichment methods for targeted sequencing. Please see Supplemental Discussion starting on page 9 for supporting text.

Method	Targeted Bases	Fraction of Reads On Target		Coverage of targeted bases		Concordance	Uniformity of coverage	
		All Reads	Uniquely Mapping	≥ 1 read	Passed for Variant Detection			
Microdroplet PCR	1.35 Mb	64.2%	79.2%	99.8%	94.5%	98.8%	96.6% within 1/5 the mean (25-fold)	
MIP ¹	Shotgun	6.5 Mb	58.4%	99.0%	85.0%	58.8%	99.7 / 96.4 ²	N/A
	End Sequence	1.4 Mb ³	66.3% ⁴	73.1%	98.0%	75.0%	99.7%	58% within tenfold, 88% within 100-fold
Solution Hybridization ⁵	3.7 Mb	53.1% ⁶	65.2%	88.0%	64.0% / 89% ^{7,8}	99.6% / NA ⁸	80% within 1/5 the mean (25-fold)	

1 - MIP = Molecular inversion probe. Turner et al. N. Meth 2009. 2 - Concordance reported separately for homozygotes (99.7%) and heterozygotes (96.4%). 3 - 1.4 Mb represents a proportion of 1.7 Mb of the exonic sequence that could be acquired by 76 bp end sequencing 4 - Calculated by multiplying the proportion of the 76 base reads that did not consist of primer sequence (56/76 * 90.0). 5 - Gnirke et al. N. Biotech 2009 6 - Gnirke et al. mapped 37.7% of all reads. For a fairer assessment we report mapping of filtered reads using methods described in the supplemental discussion. 7 - Gnirke et al. only reported the call rate for exonic sequence (2.5 Mb) and not the entire bait. 8 - The concordance for the GA-II run was not reported, because of this exclusion we have provided call rates for the GA-I (first value) and GA-II runs (second value).