Enrichment of Resequencing Targets using Droplet-based PCR

James Brayer, Patricia David, Steve Kotsopoulos, Brian Libby, Darren Link, Jeff Olson, Lin Pham, Jason Warner
RainDance Technologies, Inc., Lexington, MA

Beyond Sequencing 2010

Introduction

The introduction of second generation sequencers has greatly increased the pace at which sequencing data can be obtained. One of the most common applications for high throughput sequencing is sequence variant or mutation detection in genes or genomic regions of interests, and sequence enrichment methods are employed to target the specific regions of the genome. Recent advances in droplet-based microfluidics technology have enabled the use of PCR to amplify as many as 20,000 different target loci simultaneously. This is accomplished by performing millions of single-plex PCR reactions in parallel within picoliter volume droplets inside a single PCR tube. These highly uniform droplets result in the generation of PCR amplicons in which the PCR success rate is high and the yield for all amplicons is very consistent. This results in data in which the majority of sequencing reads map to the regions of interest (specificity), the number of sequencing reads obtained for all regions is uniform (low bias) and the percentage of target regions in which sequencing reads were obtained is very high (completeness). These attributes result in the maximization of the sequencing efficiency of the second generation sequencers allowing researchers to obtain more sequencing data from their samples at a greater depth of coverage.

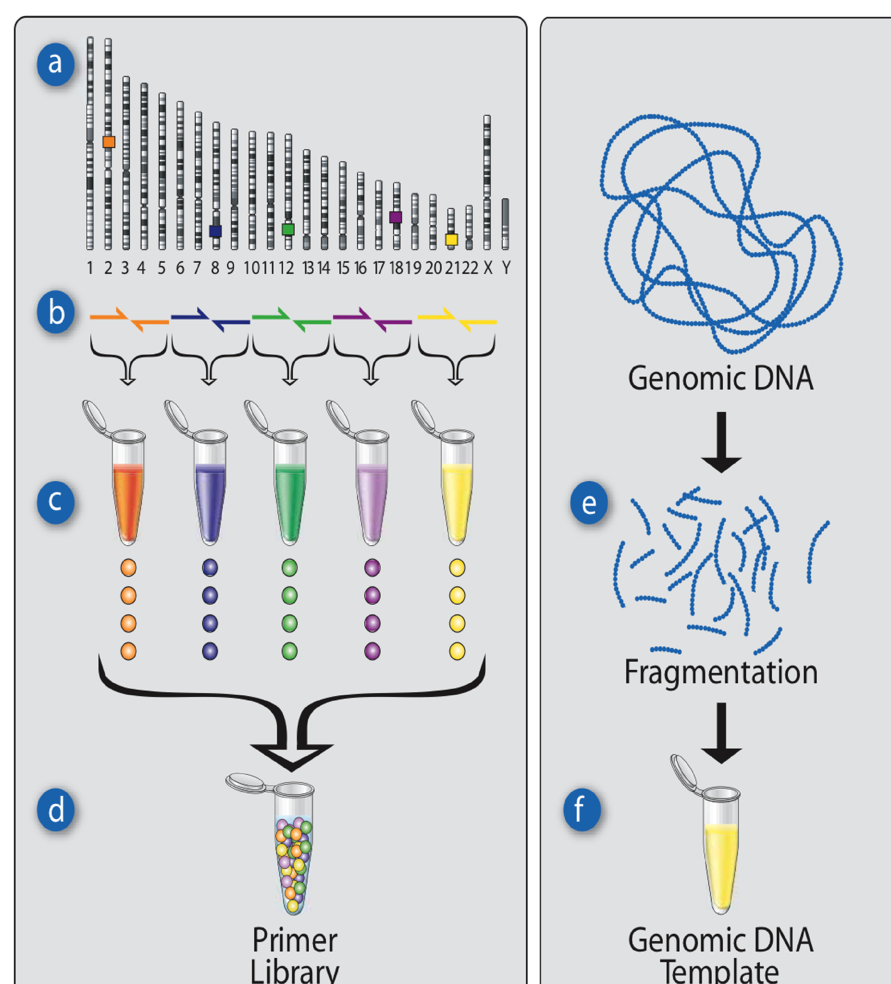
RainStorm™ Droplet-Based PCR

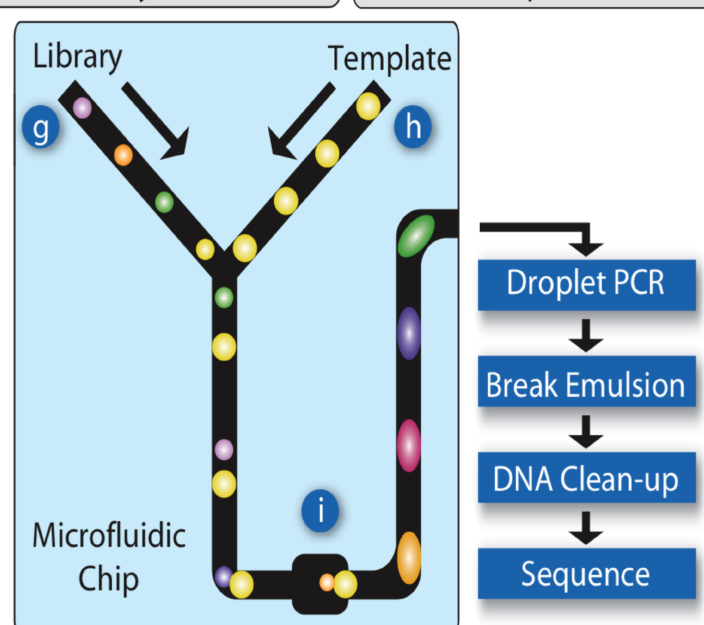
Primer Library Generation
(a) Identify targeted sequences of interest in the genome.
(b) Design and synthesize forward and reverse primer pairs for each targeted sequence.
(c) Generation of primer pair droplets. A microfluidic chip is used to encapsulate the aqueous PCR primers in inert fluorinated carrier oil with a block-copolymer surfactant to generate the equivalent of a picoliter scale test tube compatible with standard molecular biology.
(d) Primer pair droplets are mixed together so that each library element has an equal representation.

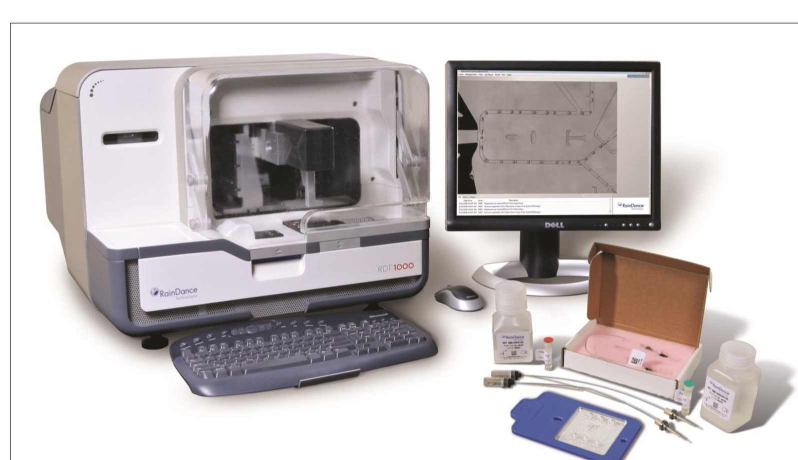
Genomic DNA Template Mix Preparation
(e) Genomic DNA is fragmented into 2 to 4 kb fragments and purified.
(f) Purified genomic DNA is mixed together with all of the components of the PCR reaction except the PCR primers.

Primer-Template Merge and PCR
(g) Primer Library droplets (~8pL) are dispensed to the microfluidic chip
(h) Genomic DNA Template is delivered as an aqueous solution and template droplets (~18pL) are formed within the microfluidic chip. The primer pair droplets and template droplets are then paired together in a 1:1 ratio.
(i) Paired droplets flow through the channel of the microfluidic chip to pass through a merge area where an electric field induces the two discrete droplets to coalesce into a single PCR droplet (~26 pL). Up to 2 million PCR droplets are collected into a single 0.2 ml PCR tube.

The collection of PCR droplets (PCR Library) is processed in a standard thermal cycler for targeted amplification, followed by breaking the emulsion of PCR droplets to release the PCR amplicons into solution for purification and next-generation sequencing.







The RDT 1000 instrument and associated reagents generate 2 million PCR reactions in less than one hour , targeting as many as 20,000 distinct regions in the genome for resequencing using next-generation sequencing technologies.

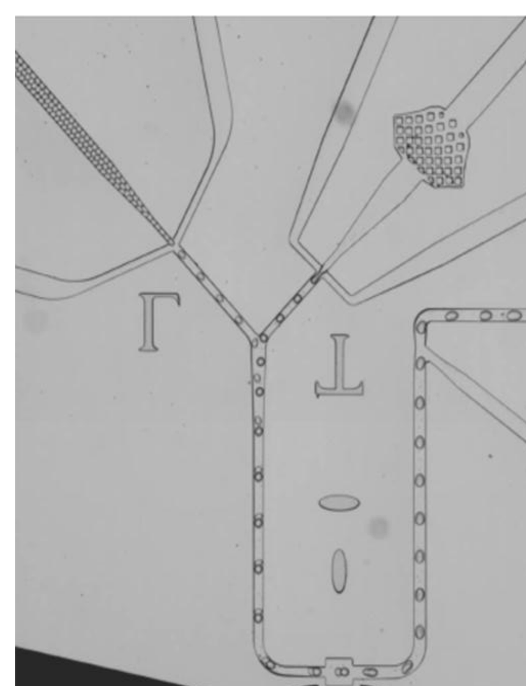
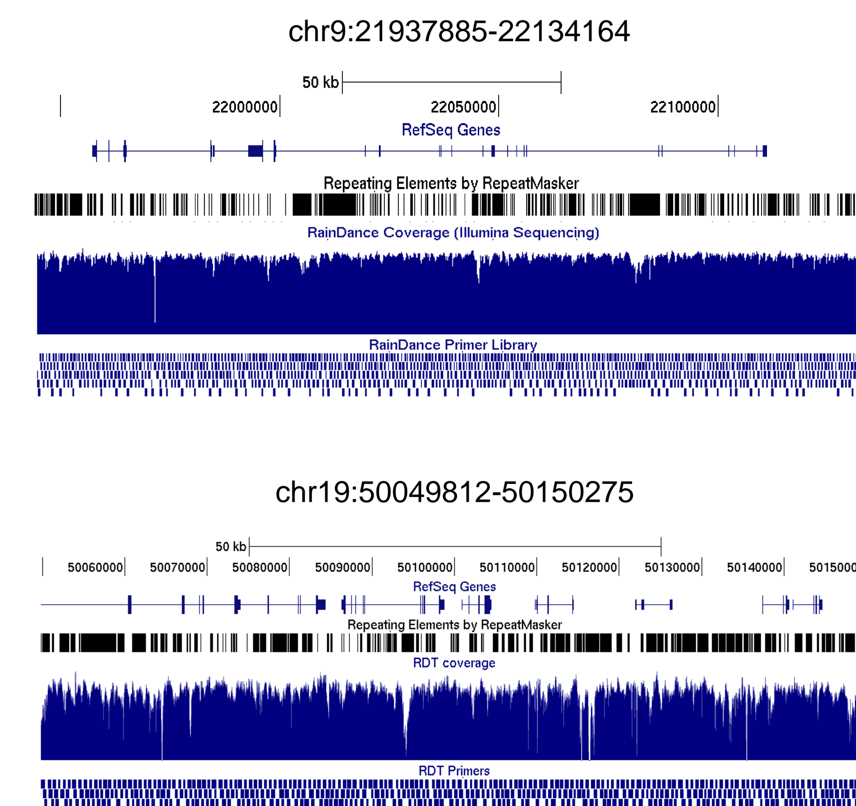


Image of RDT 1000 Sequence Enrichment Chip showing merge of Primer Library and DNA Template droplets.

GWAS Candidate Loci

PCR Primer Libraries were designed to enrich three contiguous regions from 8q24, 9p21, and 19q13. The samples were sequenced using 50 bp single end reads using the Illumina GAIIx system. The coverage (uniquely mapped reads) using the RainDance Sequence Enrichment assay ranged between 97-99%, compared with previously published results of 37-48% when using hybridization capture.

chr9:21937885-22134164



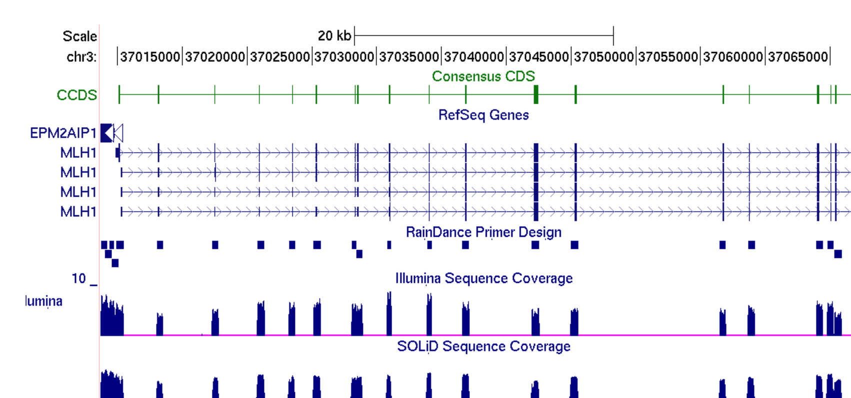
chr8:128459722-128585164

chr19:50049812-50150275

Gene Networks

PCR primers were designed to amplify coding exons, splice sites, 3'UTR and the promoter regions of 142 genes related to oncology. The resulting Primer Library comprised approximately 3,979 primer pairs totaling approximately 2 Mb of total sequence. Samples were processed on the RDT 1000 and sequenced using both Illumina GAIIx and Applied Biosystems SOLiD instruments. Concordance of variants calls was determined by comparison to genotype data from Illumina Infinium array.

ABL1	BUB1B	EP300	FLT3	MAP2K4	PTCH	NOTCH1	SYK
ABL2	CARD11	ERBB2	GATA1	MDM2	PTEN	NPM1	TAF15
AKT1	CBL	ERCC2	GATA2	MEN1	PTPN11	NRAS	TCF1
AKT2	CDH1	ERCC3	GPC3	MET	RAD51L1	NTRK1	TGFBR2
ALK	CDK4	ERCC4	GRAF	MLH1	RARA	NTRK3	TLX1
APC	CDK6	ERCC5	HIP1	MLH3	RB1	PALB2	TLX3
ARHH	CDKN2A	EWSR1	HRAS	MPL	RECQL4	PDGFRA	TOP1
ATM	CDX2	EXT1	IDH1	MSH2	REL	PDGFRB	TP53
AXIN2	CEBPA	EXT2	IL21R	MSH6	RET	PHOX2B	TSC1
BCL10	CHEK2	FANCA	IL6ST	MUTYH	ROS1	PIK3CA	TSC2
BCL2	COPEB	FANCC	ITK	MYC	RUNX1	PIK3R1	VHL
BCL9	CREB1	FANCE	JAK2	MYCN	RYR1	PIM1	WAS
BLM	CREBBP	FANCF	JAK3	MYH11	SH3GL1	PLAG1	WHSC1
BMPR1A	CTNNB1	FBXW7	KIT	MYH9	SMAD4	PML	WRN
BRAF	CYLD	FGFR1	KRAS	NBS1	SMARCB1	PMS1	WT1
BRCA1	EGFR	FGFR2	LCK	NF1	STK11	PRDM16	WTX
BRCA2	ENG	FGFR3	MAF	NF2	STK11IP	PRKAR1A	XPA
BRIP1		FLI1	MAFB	NFKB2	SUFU		XPC

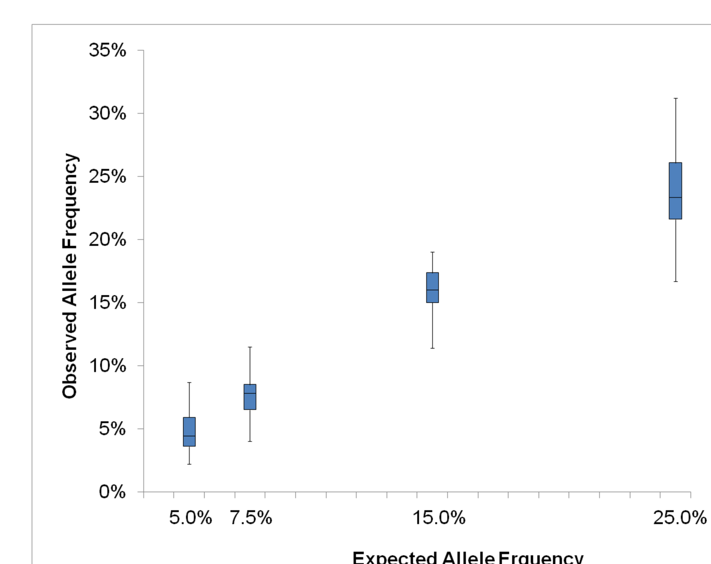


Representative sequence coverage (*MLH1* gene).

Performance Metric	Result
Coverage	**98%**
SNP Concordance (Homozygous)	**99.71%**
SNP Concordance (Heterozygous)	**99.29%**

Heterogeneous Samples

Two HapMap DNA samples were mixed together at varying ratios to represent sample pools of increasing complexity as a model of heterogeneous samples such as pooled DNA from multiple individuals or tumor samples with low tumor cellularity. The pooled samples were processed on the RDT 1000 using a PCR Primer Library of 1536 primer pairs representing 700kb of target sequence. The resulting PCR products were sequenced to an average depth of 1500X on the Illumina GAIIx. The figure on the right shows good correlation between the expected and observed allele frequency in the samples.



Summary

The combination of single-molecule PCR and large number of replicate independent PCR reactions enabled by the RDT 1000 Sequence Enrichment assay provides superior performance for targeted resequencing-based validation of genomic variation associated with complex diseases. The technique is ideally suited as a method for follow-up to GWAS studies, resequencing of candidate genes and deep resequencing of tumor samples.

References

Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, et al. (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol 27(11):1025-1031.

Tewhey R, Nakano M, Wang X, Pabón-Peña C, Novak B, et al. (2009) Enrichment of sequencing targets from the human genome by solution hybridization. Genome Biol. 2009;10(10): R116, Epub 2009 Oct 16.

For Research Use Only. Not for use in diagnostic procedures.