

Introduction

RainDance Technologies has developed an enrichment platform for targeted resequencing that leverages the sensitivity and the specificity of PCR to target genomic regions from either thousands of exons or large contiguous loci. The Sequence Enrichment application utilizes a novel microdroplet-based format to rapidly and reproducibly generate over 1 million independent PCR reactions. The process control achieved with this approach enables statistical modeling to tune the performance of the assay. Researchers are able to design experiments using pooled genomic DNA samples and have the confidence that they have the statistical power to detect variants at the frequency necessary for their biological models.



RDT 1000 Targeted Resequencing Workflow

Primer Library Generation

- Identify targeted sequences of interest in the genome.
- Design and synthesize forward and reverse primer pairs for each targeted sequence.
- 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.
- Primer pair droplets are mixed together so that each library element has an equal representation.

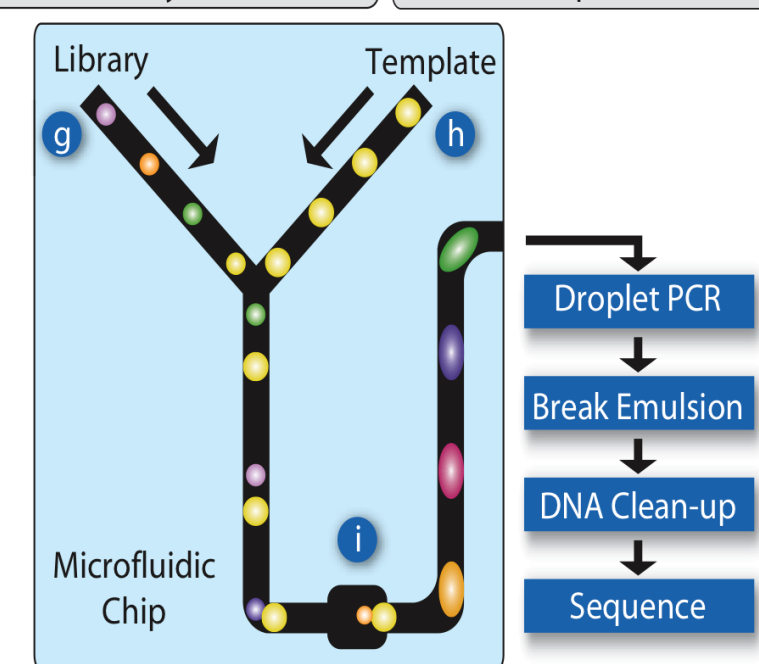
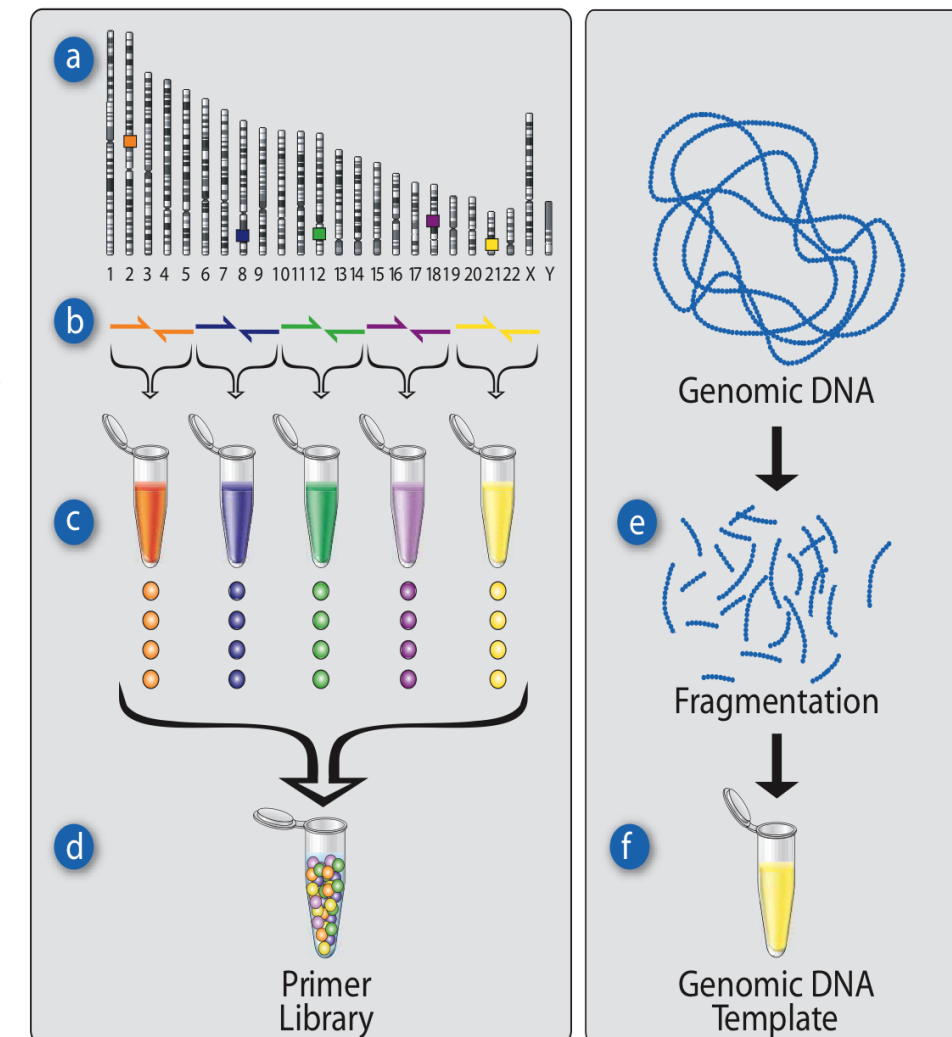
Genomic DNA Template Mix Preparation

- Genomic DNA is fragmented into 2 to 4 kb fragments and purified.
- Purified genomic DNA is mixed together with all of the components of the PCR reaction except the PCR primers.

Primer-template merge and PCR

- Primer Library droplets (~8pL) are dispensed to the microfluidic chip.
- 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.
- 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). Approximately 1 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.



Statistical Modeling

PCR Positive Droplets

Starting DNA	2 µg	5 µg
Fragment size	2-4 kb	2-4 kb
Yield after fragmentation	75%	75%
Input DNA	1.5 µg	3.75 µg
Fragments per sample	5x10 ¹¹	1.3x10 ¹²
Sample volume	25 µl	25 µl
Injection volume	20 µl	20 µl
Droplet volume	18.8 pl	18.8 pl
Droplets per sample	1x10 ⁶	1x10 ⁶
Fragments per droplet	3.7X10 ⁵	9.4X10 ⁵
Haploid genome fraction per droplet	34%	85%
Average amplicon length	400 bp	400 bp
gDNA fragment utilization	87%	87%
PCR probability per droplet	29%	74%
Merge efficiency with primer droplet	85%	85%
PCR positive droplets per sample	246,500	629,000

Primer Library Sampling

Starting DNA	2 µg	5 µg
# Primer pairs per Library	1536	1536
Average PCR positive droplets per primer pair	173	435
Minimum replicate PCR droplets per primer pair (99.73% CI)	133	371

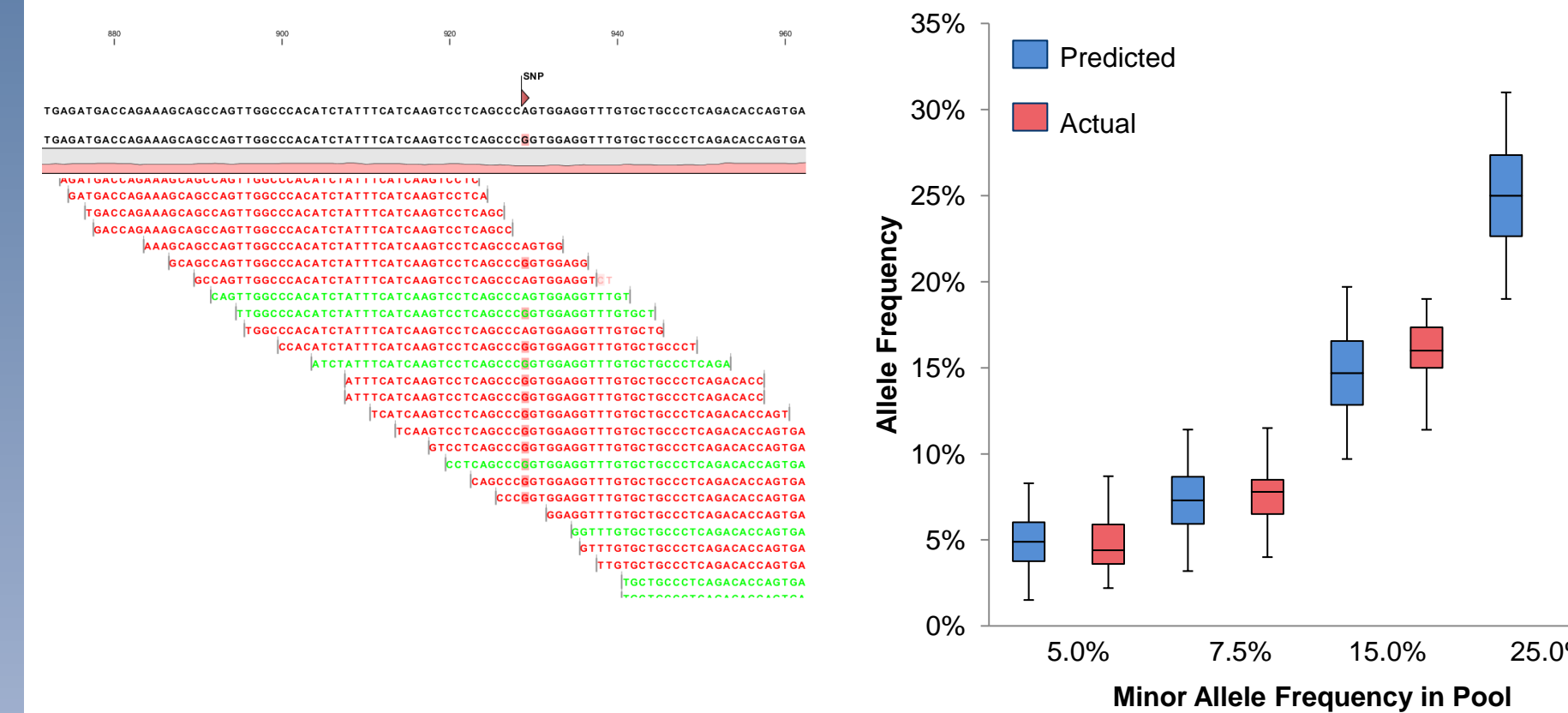
Allele Representation

Samples per Pool	1	4	6	8
Starting DNA	2	5	5	5
# Primer pairs per Library	1536	1536	1536	1536
Minor allele frequency in pool	50%	12.5%	8.3%	6.25%
Minimum PCR droplets per primer pair (99.73% CI)	133	371	371	371
Average PCR positive droplets per minor allele	66	46	30	23
Average percentage of reads per minor allele	50%	12.5%	8.3%	6.25%
Minimum PCR positive droplets per minor allele (99.73% CI)	48	26	14	9
Minimum percentage of reads per minor allele (99.73% CI)	36%	7%	3.8%	2.4%

Targeted Resequencing

Allele Representation

Two HapMap DNA samples were mixed together at varying ratios to represent sample pools of increasing complexity. 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 Genome Analyzer using 50 bp reads. Alignment and SNP calling was performed using the Genomics Workbench software package from CLC bio.



Sensitivity and Specificity

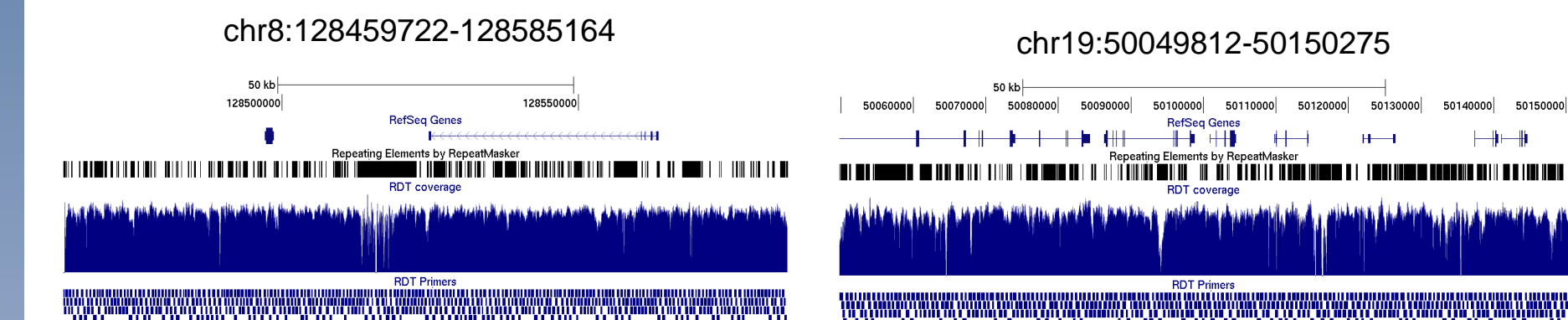
The sensitivity and specificity for detection of sequence variants within a complex pool with expected minor allele frequency of 7.5% was determined using a minimum of 10 high quality reads per allele. In order to accommodate sources of error introduced in upstream (sample prep) and downstream (sequencing error rate) processes outside of the RDT 1000 assay that are not included in the statistical models, a variable filter was applied to set a minimum threshold for observed allele frequency ranging from 1-4%.

Observed Allele Frequency Filter Cutoff	Expected Sensitivity	Observed Sensitivity	Observed Specificity	False Discovery Rate
1%	98.9%	98.7%	84.9%	15.1%
2%	98.9%	98.5%	90.4%	9.6%
3%	98.7%	97.7%	94.2%	5.8%
4%	97.0%	97.2%	96.5%	3.5%

Conclusions

The exquisite control and reproducibility of the droplet is the foundation for the statistical modeling. We will present targeted resequencing data to support the statistical models that can be generated using the RainDance Platform. By adjusting the amount of genomic DNA and the complexity of the Primer Library the researcher can tune the assay to the appropriate confidence interval for detecting sequence variants critical for the elucidation of the molecular basis of complex diseases.

The ability to accurately detect sequence variants in pooled DNA samples enables large-scale targeted resequencing studies for population genetics studies and characterization of allele frequencies as a follow-up to genome-wide association studies (GWAS).



The figures above demonstrate examples of targeted resequencing of two GWAS loci using the RainDance RDT 1000 microdroplet-based PCR assay. Coverage of uniquely mapped reads in each case was in excess of 98% enabling analysis of variants in coding and regulatory regions of the loci including regions containing repetitive sequence elements.

Other applications of the RainDance technology include discovery and validation of sequence variants in targeted gene networks in large cohort studies.

The combination of single-molecule PCR and large number of replicate independent PCR reactions maintains the allelic representation of complex samples for highly heterogeneous samples such as tumors.

Recent enhancements to the RDT 1000 sequence enrichment assay have increased the number of PCR droplets to two million droplets per sample (a 2X increase beyond the data presented in this poster) which will enable greater sensitivity to detect rare alleles as well as increasing the total number of primer pairs per library to 20,000.

The statistical power of the microdroplet-based PCR format will enable highly quantitative measurement of CpG methylation using targeted resequencing of bisulfite-converted template DNA.

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.