Selection of Cancer-Related Gene Exons for targeted resequencing with a Flexible and Fully Automated Microarray Platform

Summerer, D.; Schracke, N.; Wu, H.; Cheng, Y.; Haase, B.; Stähler, C.F., Stähler, P.F.; Beier, M.

febit biomed gmbh, Im Neuenheimer Feld 519, 69120 Heidelberg, Germany,
febit inc., 99 Hayden Ave, Suite 620, Lexington, MA 02421, USA

Next-generation sequencing studies are currently limited by an inability to enrich genomic DNA samples for specific regions of interest easily and effectively. This keeps sample throughput at the low-end and calls for focused and more inexpensive methods to analyze complex, eukaryotic genomes. Allowing for larger numbers of samples, the selection of relevant subsets for targeted resequencing greatly increases statistical power, while keeping datasets manageable. Here we present HybSelect™, a method to select human genomic loci of interest by hybridization on scalable, microfluidic DNA microarrays using the Geniom® RT Analyzer. We captured the complete coding sequence (1819 exons) of 115 cancer-related genes of a recently fully sequenced Yoruba individual representing a total region of interest (ROI) of 9.3 Mb on part of a biochip.

This corresponds to > 18.4 Mb capacity per biochip. Sequencing using Illumina technology revealed an average depth of coverage of 175.7-fold for all exons, and 97 % of targeted bases were covered at least once. Uniformity was such that 94 % of genes were in a range of < 1 log which indicates applicability to various sequence contexts. A comprehensive analysis of reference SNPs revealed a concordance to HapMap genotypes of 98.2 %. Using a fully automated microarray processing platform minimizes contamination risk, enhances reproducibility, and accelerates the experimental workflow.

Experimental setup

Probes for 115 cancer-related genes were distributed over the high complexity target sequence using a 9bp tiling design of 50 mer probes. The complete set representing a HybSelected Region of 0.48 Mb was synthesized with the Geniom® One System into four arrays of a geniom biochip. We used 1.5 µg of a Yoruban HapMap sample NA18507 sequencing library for hybridization to the biochip.

Upon automated hybridization, stringent washing and elution of the bound DNA in the Geniom® RT Analyzer, quality controls were performed.

Sequencing results

The sequencing results from a Solexa/Illumina Genome Analyzer II led to 33 million 36-bp paired-end reads. After filtering the data showed that the 115 genes were enriched up to an average coverage of 539 fold.

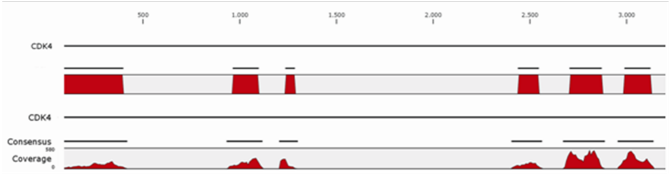


Figure1: Graphic overview of mapped sequencing reads. Part of the CDK4 gene sequence is shown. Upper graph: exon regions covered by capture probes (Hybselected Region). Lower graph: Mapping of the obtained reads to the Region of Interest.

Enriched genes	HybSelected Region[bp]	Enrichment [fold]	1-fold consensus [%]	20-fold consensus [%]	Average depth of coverage[fold]
all	482093	1599.7	97.1	83.2	183.4
Min	1098	232.4	74.7	33.2	23.5
Max	14982	4395.3	100	100	539.0
ABL1	5963	2307.09	100	99.1	404.9
AKT1	3249	717.5	91.9	87.6	119.1
AKT2	1702	1175.8	99.2	93.6	180.8
ALK	6220	1144.2	99.8	95.6	191.7
APC	10702	2312.6	99.3	91.1	409.4
ARHGAP26	9041	1074.2	99.1	90.7	186.5
ATM	13355	502.2	98.4	70.0	84.8
BCL6	3900	1264.4	99.3	88.1	218.7
...	...	...	...	...	...

Figure2: Statistics of mapping of sequencing reads obtained from Illumina paired end sequencing. Shown are the averages, minimal and maximal values and the first eight genes in alphabetical order.

Results

We analyzed the nucleotide representations of all HapMap reference SNP positions contained in the targeted exons of the Yoruban HapMap sample NA18507. 837 SNPs with reference data were present in the captured regions. These were used for further analysis. We first filtered the regions for SNP coverage depths of 20-fold or higher as a stringent and pre-established criterion for reliable base calling.
Nucleotide analysis and comparison to HapMap reference data (data from HapMap project phases 1 and 2) revealed an overall concordance of 98.2 % for all SNPs.

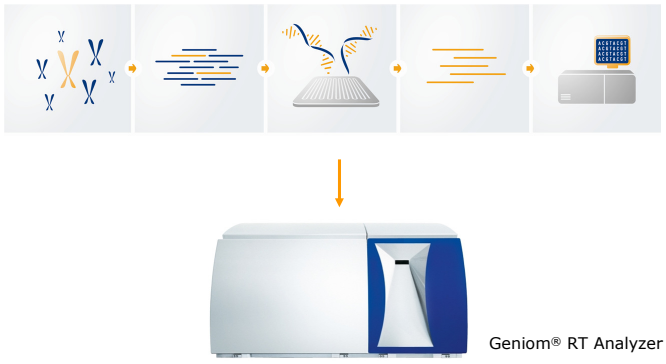
Conclusion

Subsequent analysis with next generation sequencers ensures maximal data output within very short time as compared to several weeks or even months with conventional Sanger sequencing.

Our method and platform provides a powerful tool with simple workflow and flexible probe design for reliable analysis of both homozygous and heterozygous mutations. We obtain high coverage depths for large fractions of the target sequence exceeding the minimal requirements for high quality SNP calls.

In addition, our method requires very little hands-on-time as hybridization and all washing steps are fully automated on febit's Geniom® RT Analyzer.

HybSelect™ is a powerful and cost-effective method to specifically enrich selected regions of any given chromosomal DNA.



Geniom® RT Analyzer

In Europe & Asia Pacific please contact:
febit biomed gmbh · Im Neuenheimer Feld 519 · 69120 Heidelberg, Germany
Phone +49 6221 6510-300 · www.febit.com · info@febit.eu

In America, please contact:
febit inc · 99 Hayden Ave, Suite 620 · Lexington, MA 02421, USA
Phone +1 781 - 391 - 4360 · www.febit.com · info@febit.com