febit

HybSelect: high-throughput access to genomic regions of interest for targeted next-generation sequencing

We present a technology for highly selective sequence capture using microarray hybridization for researchers who want to focus on relevant genomic regions in next-generation sequencing (NGS) experiments. The unique combination of the microfluidic Geniom Biochip® with the fully automated processing station Geniom RT Analyzer® thereby allows minimal hands-on time and high-throughput processing. Geniom Biochips® consist of eight channels containing individual capture probe arrays with complete sequence flexibility, ensuring applicability to any target sequence and scalability.

Next-generation sequencing platforms have amplified the information content of genetic variation studies by a massive reduction of cost and sequencing effort[1–4]. However, the full potential of these technologies has still not been reached because of a lack of suitable methods for target-sequence isolation at large scale[5]. Though untargeted whole-genome sequencing with limited sample throughput is one potential application, many future studies would greatly benefit from focusing on smaller but biologically more relevant genomic subsets in high sample numbers. The microfluidic Geniom Biochip® with the fully automated processing station Geniom RT Analyzer® from febit are tools to facilitate high-throughput analysis of desired genomic loci using NGS technologies.
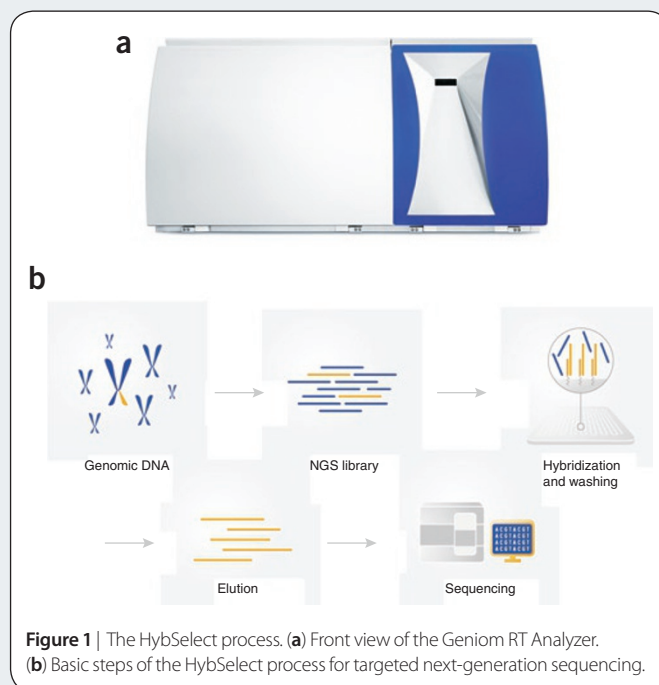
Analyzing exons or regulatory elements of whole gene sets related to specific diseases or drug responses would combine maximal content of customized information with economical usage of NGS instruments. Combined with multiplexing of samples within one NGS run, this strategy allows highly economical large-scale screening with massive sample numbers. The benefits of sequence capture thereby become even more evident for samples with increased complexity; for example, when applying NGS to microbial communities, host–pathogen mixtures or somatic variants.

Figure 1 | The HybSelect process. (**a**) Front view of the Geniom RT Analyzer. (**b**) Basic steps of the HybSelect process for targeted next-generation sequencing.

The Geniom RT Analyzer

The HybSelect™ process is conducted by the fully integrated hardware Geniom RT Analyzer (**Fig. 1a**) in three main steps: hybridization of a genomic NGS library to a Geniom Biochip, stringent washing, and elution of desired library fragments (**Fig. 1b**). febit's microfluidic technology thereby enables full flexibility for capture of different target regions owing to *in situ* oligonucleotide synthesis of capture probes on the Biochip.
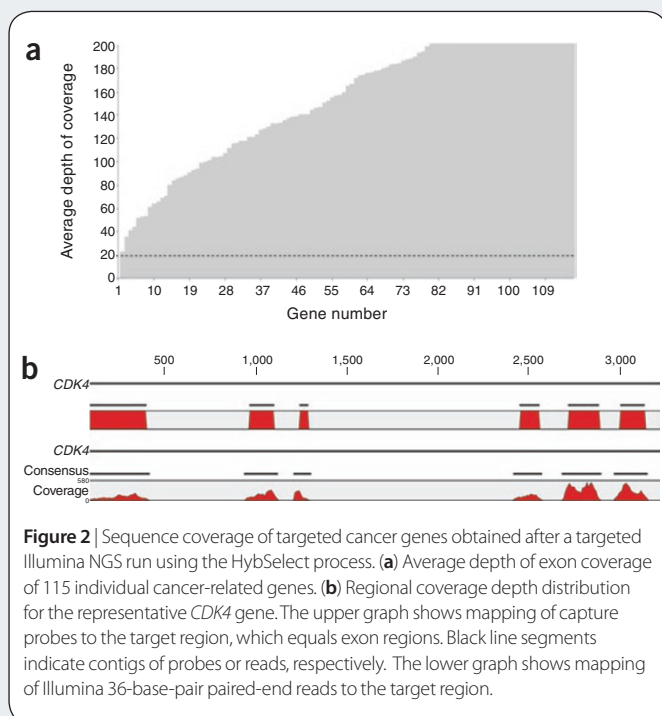
The integration of the microfluidic Geniom Biochip and the Geniom RT Analyzer as processing platform has several advantages. The hybridization steps used are very short, which results in shorter overall process duration. Furthermore, the process is highly automated, including automatic sample loading to the chip, on-chip denaturing, hybridization with controlled active motion and washing routines. This dramatically reduces workload and contamination risk, and increases reproducibility and standardization. Combined with scalability of sample numbers, these features enable true high-throughput sequence capture for large-scale NGS studies.

Daniel Summerer

febit biomed GmbH, Heidelberg, Germany. Correspondence should be addressed to D.S. (info@febit.de).

APPLICATION NOTES

Figure 2 | Sequence coverage of targeted cancer genes obtained after a targeted Illumina NGS run using the HybSelect process. (**a**) Average depth of exon coverage of 115 individual cancer-related genes. (**b**) Regional coverage depth distribution for the representative *CDK4* gene. The upper graph shows mapping of capture probes to the target region, which equals exon regions. Black line segments indicate contigs of probes or reads, respectively. The lower graph shows mapping of Illumina 36-base-pair paired-end reads to the target region.

Optimized process

We optimized and streamlined all aspects of our protocols for genomic DNA microarray hybridization to allow highest specificity and uniformity of capture during the HybSelect process.

We typically observe enrichment factors of several thousand–fold and uniform capture efficiencies for different target sequences. For example, we captured a full exonic set of 115 genes identified by the Wellcome Trust Sanger Institute as a set highly relevant to the onset of various cancer types. After the HybSelect process and sequencing with an Illumina GAII instrument, >97% of genes were covered, and 96% of all genes were in a range of coverage depth within <1 log. This demonstrates a well-balanced capture performance over the target region with low dependence on individual sequence context. In this experiment, the enrichment factor over the whole region was 1,600-fold and the region was covered >180-fold on average. All genes thereby received a coverage depth of 20-fold or deeper (**Fig. 2a**). A close look at the coverage distribution shows deep and uniform coverage of the capture probe–covered exonic regions of individual gene regions (**Fig. 2b**).

HybSelect allows accurate nucleotide calling

An important parameter for a sequence capture method is its accuracy of nucleotide calling. Especially in case of heterozygous positions, preferential hybridization of one allele could lead to biased representation and false nucleotide calls. We targeted 1,000 individual 500-base-pair regions in HapMap reference samples, each containing a central dbSNP position[6]. SNPs were chosen to have an increased heterozygous representation of ~50% in the samples.

After Illumina sequencing and mapping, comparison to HapMap reference data revealed an overall concordance of 98.6% for both homozygous and heterozygous SNPs with a minimum coverage depth of 20-fold. Notably, very similar concordances have previously been reported for untargeted Illumina whole-genome sequencing of HapMap samples. This indicates that the HybSelect process does not interfere with the accuracy of SNP calling and provides a powerful tool for targeted resequencing studies.

Summary

HybSelect allows researchers to tailor their NGS projects for the genomic regions that are really relevant to them. The capture process is highly flexible, customizable and applicable to any genome of interest. Typical enrichment factors and capture uniformities illustrate excellent specificity and low sequence bias. Notably, targeted NGS using HybSelect affords high quality data in terms of nucleotide calling.

The use of microfluidics enables the integration of Geniom Biochips with the Geniom RT Analyzer, resulting in a highly automated workflow with minimal manual interference.

Each Biochip contains eight microchannels with individual capture probe arrays and is thus scalable from one to eight samples. This scalability facilitates adjustment of an experiment to different target sizes and can substantially reduce per-sample cost for small targets. The current target capacity ranges from 250 kb to 2 Mb of actually captured sequence, for one and eight channels, respectively. On average, this corresponds to the coding sequence of ~224–1,800 human genes.

The multi-sample capacity of Geniom Biochips and the highly integrated process with low workload enabled by the Geniom RT Analyzer makes HybSelect an attractive option for researchers interested in high-throughput NGS studies involving large sample numbers. This is especially true when combined with multiplexing of several samples within one NGS run. For occasional users, HybSelect is also available as service, providing the same performance without the need to acquire additional equipment.

1. Bentley, D.R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**, 545–552 (2006).
2. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
3. Shendure, J. *et al.* Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* **5**, 335–344 (2004).
4. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
5. Garber, K. Fixing the front end. *Nat. Biotechnol.* **26**, 1101–1104 (2008).
6. Summerer, D. *et al.* Microarray-based multicycle-enrichment of genomic subsets for targeted next-generation sequencing. *Genome Res.* (in the press).