© 2010 Nature America, Inc. All rights reserved.

Targeted resequencing for large-scale genomic studies: accelerating NGS throughput using molecular barcodes

Next-generation sequencing (NGS) is a powerful tool for the massively parallel investigation of specific genomic loci. However, sequence capture methodologies are required for time- and cost-effective sequence analysis in high-throughput genomic projects. HybSelect is a microarray-based sequence capture strategy that enables selective enrichment of a desired target region with increased sample numbers. Adding sample barcoding technology to the automated HybSelect approach further increases sample throughput and efficiently amends NGS technologies involved in large-scale discovery studies.

Massively parallel DNA sequencing technologies, combined with sequence-capture methodologies, have accelerated the investigation of specific genomic regions while reducing the time, cost and effort needed to interrogate multiple loci[1]. NGS is a powerful tool to study the sequence of many DNA bases in a sample, and sequence-capture methods allow for targeted next-generation sequencing (tNGS) in increasing sample numbers[2,3]. The possibility of sequencing relevant subsets of a genome with high sample throughput and at low cost has become of major interest to numerous researchers. Genomic studies require large sample sizes to ensure sufficient statistical power for the detection of specific mutations and alterations associated with a particular disease. Sample barcoding technology enables scientists to increase the throughput of samples on NGS instruments even more by processing many samples in parallel. To benefit from barcoding and to achieve high throughput, researchers need samples to be reformatted and scaled down from the whole genome to a selected region of interest. Here we present HybSelect, a highly scalable method for enriching biologically relevant sequence subsets. HybSelect is a microarray-based sequence-capture strategy that enables selective enrichment of a desired target region, making it the ideal front end for large-scale genomic studies on NGS platforms. Very high levels of multiplexing are achievable when HybSelect parallel sample processing capabilities are interlinked with barcoding technology (**Fig. 1**). Combining HybSelect with the SOLiD system in one streamlined, targeted NGS pipeline, we have established a powerful facility to resequence large genomic regions in hundreds of clinical samples per study in a very economical way and within a short time.

Julia Lange

febit holding gmbh, Heidelberg, Germany.
Correspondence should be addressed to J.L. (info@febit.de).

Technology

HybSelect sequence-capture microarrays are built within microfluidic biochips for robust handling and process automation. The precise microfluidic control ensures reproducible HybSelect sequence-capture results. The eight microfluidic arrays per biochip allow for the parallel processing of samples. Furthermore, owing to the compartmentalized, closed nature of the Biochips, risks of cross-contamination and evaporative sample loss are completely eliminated. Biochips are designed and manufactured using a proprietary production process, with the capability to go from tailor-made target sequence capture to sequencing results in days.

The HybSelect Biochips are processed on febit's HybSelector instrument, the first automated system for sequence capture on microarrays (**Fig. 2**). The HybSelector automates sample loading, hybridization, washing and precise temperature control. A

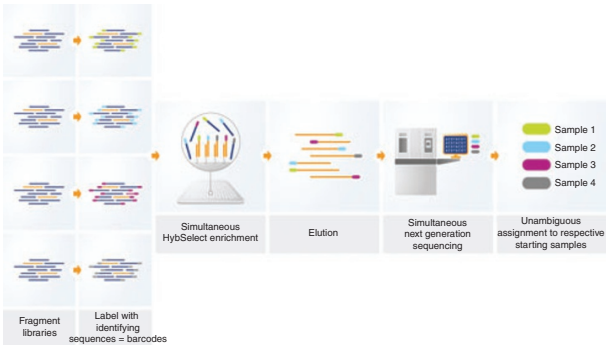

Figure 1 | High levels of multiplexing achieved by combining HybSelect™ parallel sample processing capabilities with bar-coding reagents. After automated sequence capture with HybSelect, targeted NGS was carried out on the SOLiD™ Sequencing System. The use of molecular barcodes allows the confident identification and tracking of samples

Tecan Ex. 2064, Page 1 of 2
IDT, Inc. v. Tecan Genomics - IPR2025-00015

APPLICATION NOTES

Figure 2 | The HybSelector instrument and a HybSelect Biochip for the selective enrichment of genomic target regions for next-generation sequencing.

pressurized system allows automated movement of samples through the microchannels of the Biochip with accurate timing. An active mixing process increases sensitivity and stringency while reducing hybridization time. Thus, the HybSelector enables convenient sequence capture on arrays with less hands-on time compared to other systems, and it fully supports barcoding approaches for the parallel capture of many samples on one microarray.

Barcoding of samples

Barcodes are short DNA tags added to each library before clonal amplification and sequencing of the fragments. Samples are barcoded before the HybSelect enrichment process. Molecular barcodes enable the parallel processing of 16 (**Fig. 3**) and soon 64 HybSelected samples per microarray. Barcoding works particularly well in microfluidic biochips owing to the efficient kinetics. Barcodes have been successfully applied to process 64 clinical samples in one sequencing run, and it may be possible to exceed 1,000 samples per run in the near future. The use of molecular barcodes allows confident identification and tracking of samples, and greatly strengthens laboratory information management systems in regulated production sequencing environments. Barcoded targets from hundreds of individuals can be sequenced each week using a SOLiD System in combination with febit's HybSelector instrument, and this capability can be applied for large-scale genomic studies. Sample throughput will continue to increase with the introduction of additional barcodes and higher sequencing output of even more powerful platforms.

Case studies

Many studies demonstrate the power of the targeted resequencing pipeline presented here. More than 350 samples from four large collaborative studies with clinical partners have been HybSelected as 4-plex pools and sequenced on the SOLiD System. These studies concentrated on clinically relevant genomic regions and yielded a rich set of data within a very short time. In a joint breast cancer study, targeted NGS was tested as a future platform for routine screening for hereditary forms of breast cancer, with high-throughput resequencing of thousands of samples. The aim of this and similar studies is to develop and validate an NGS workflow to take molecular diagnostics to the next level. Genetic heterogeneity makes it difficult to characterize patients in a fast and cost-effective way, and the establishment of a routinely applicable protocol for screening patients for known and novel disease-related genes remains a big challenge. The method of choice for future genetic diagnostics is still to be determined, but most likely will be one of the new NGS methods. Targeted resequencing has shown very good technical performance, and the ongoing data analysis will form the basis for further studies on the way to an innovative screening solution.

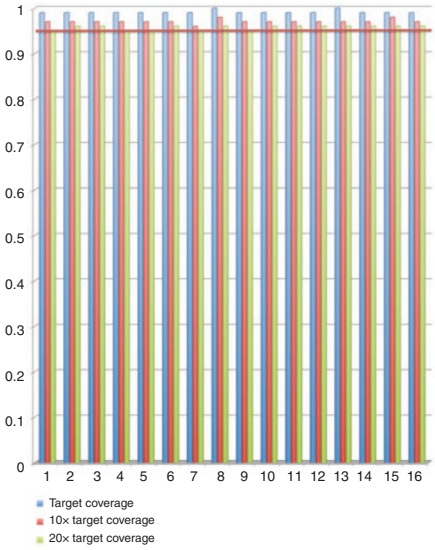

Figure 3 | Sixteen-plex parallel analysis of breast cancer genes reveals even distribution of coverage over all 16 barcoded samples. For every sample, a 20× target coverage of at least 95% was obtained.

Conclusion and outlook

HybSelect has successfully been established as a flexible and reliable method to enrich biologically relevant sequence subsets with increased sample numbers. It provides excellent completeness of coverage and coverage uniformity over multiple barcoded samples processed on a single microarray. The HybSelect method efficiently amends technologies involved in large-scale discovery studies such as whole-genome or whole-exome sequencing, enabling efficient follow-up projects such as disease association studies involving massive sample numbers. Targeted NGS facilities that combine the HybSelect technology with the SOLiD System benefit from outstanding sequencing capacity as well as superior accuracy. These facilities will take the lead through targeted sequencing of more than 1,000 samples per week, making extensive use of barcoding, serving large-scale genomic studies and clinical research, and taking NGS sample throughput to the next level.

1. Shendure, J., Mitra, R.D., Varma, C. & Church, G.M. Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* **5**, 335–344 (2004).
2. Summerer, D. Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* **94**, 363–368 (2009).
3. Summerer, D. *et al.* Targeted high throughput sequencing of a cancer-related exome subset by specific sequence capture with a fully automated microarray platform. *Genomics* published online (6 February 2010).

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.

© 2010 Nature America, Inc. All rights reserved.

Tecan Ex. 2064, Page 2 of 2
IDT, Inc. v. Tecan Genomics - IPR2025-00015