NimbleGen Sequence Capture

Genetic Discovery Made Easy

Targeted Sequencing Enables Genetic Discovery

In life science research, understanding the human genome and the diseases associated with genetic mutations are critical. The deeper our understanding of genetic diseases becomes, the more biological complexity is revealed. Research has shown the majority of disease causing mutations can occur in the Exome (coding exons) or in other disease associated regions.

Identifying and targeting specific genomic regions of interest are key to advancing our knowledge of disease and the underlying mutations present in the genome. Researchers use targeted sequencing to investigate everything from causative mutations in disease associated regions, to novel mutation present in Mendelian disease, to detecting coding variants in the Exome. Targeted sequencing enables user specified enrichment to identify and 'capture' only the precise regions of interest that are responsible for disease.

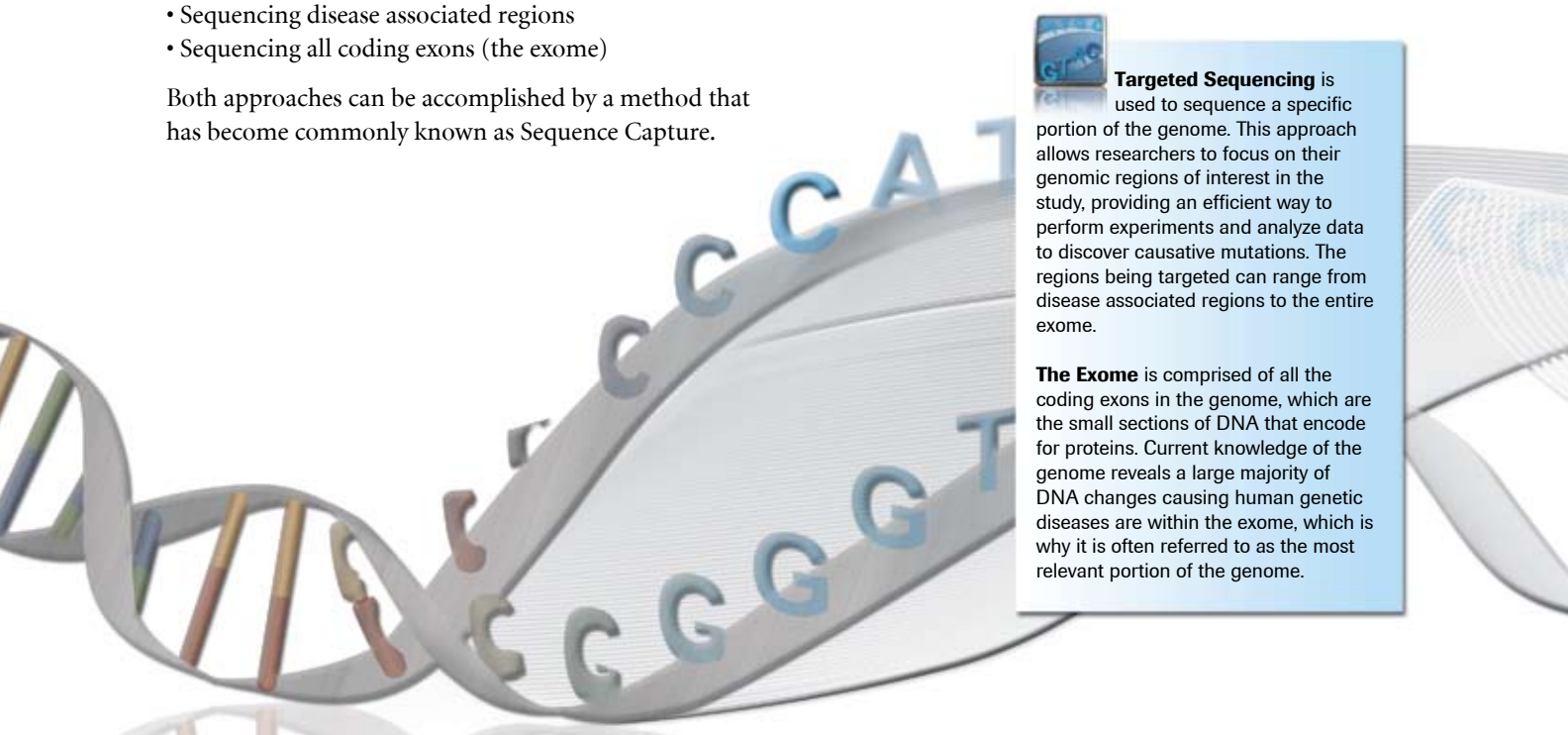
Two common approaches for targeted sequencing are:
• Sequencing disease associated regions
• Sequencing all coding exons (the exome)

Both approaches can be accomplished by a method that has become commonly known as Sequence Capture.

Sequence Capture Research Methods
Roche NimbleGen has pioneered targeted enrichment with the development of NimbleGen Sequence Capture technologies. Our unique experience developing Sequence Capture for genetic disease study has advanced genomic enrichment methods for next-generation sequencing platforms.

Two options are available for targeted enrichment of genomic regions:

- NimbleGen Sequence Capture arrays enable researchers to target custom regions of interest in the human genome.
- SeqCap EZ Exome is a solution-based capture method that enables enrichment of the whole exome in a single test tube.



Targeted Sequencing is used to sequence a specific portion of the genome. This approach allows researchers to focus on their genomic regions of interest in the study, providing an efficient way to perform experiments and analyze data to discover causative mutations. The regions being targeted can range from disease associated regions to the entire exome.

The Exome is comprised of all the coding exons in the genome, which are the small sections of DNA that encode for proteins. Current knowledge of the genome reveals a large majority of DNA changes causing human genetic diseases are within the exome, which is why it is often referred to as the most relevant portion of the genome.

The Power of Sequence Capture in Revealing Disease Causing Mutations

NimbleGen Sequence Capture technology is a revolutionary process for the enrichment of selected genomic regions from full complexity human genomic DNA in a single step. Sequence Capture was developed to eliminate the necessity of setting up thousands of PCR reactions, instead allowing for parallel enrichment of target regions in a single experiment.

Roche NimbleGen created, further refined, and optimized large-scale genomic enrichment technology. Our company is at the forefront of targeted sequencing, driving development of Sequence Capture technologies. Capitalizing on the efficiencies inherent with parallel enrichment, researchers can now design economical, higher throughput, and time-saving next-generation sequencing experiments. In combination with high throughput sequencing (short- or long-read), Sequence Capture has made targeted sequencing possible and accessible to more life science researchers.

Let Roche NimbleGen provide you with the flexible enrichment technologies, and built-in controls that enable high quality data that your sequencing experiments demand.

Delivering better coverage with less sequencing

- Achieve greater coverage and uniformity using our 2.1 million optimized DNA oligonucleotides
- Save on sequencing time and cost compared to other methods

Optimized and scalable workflow

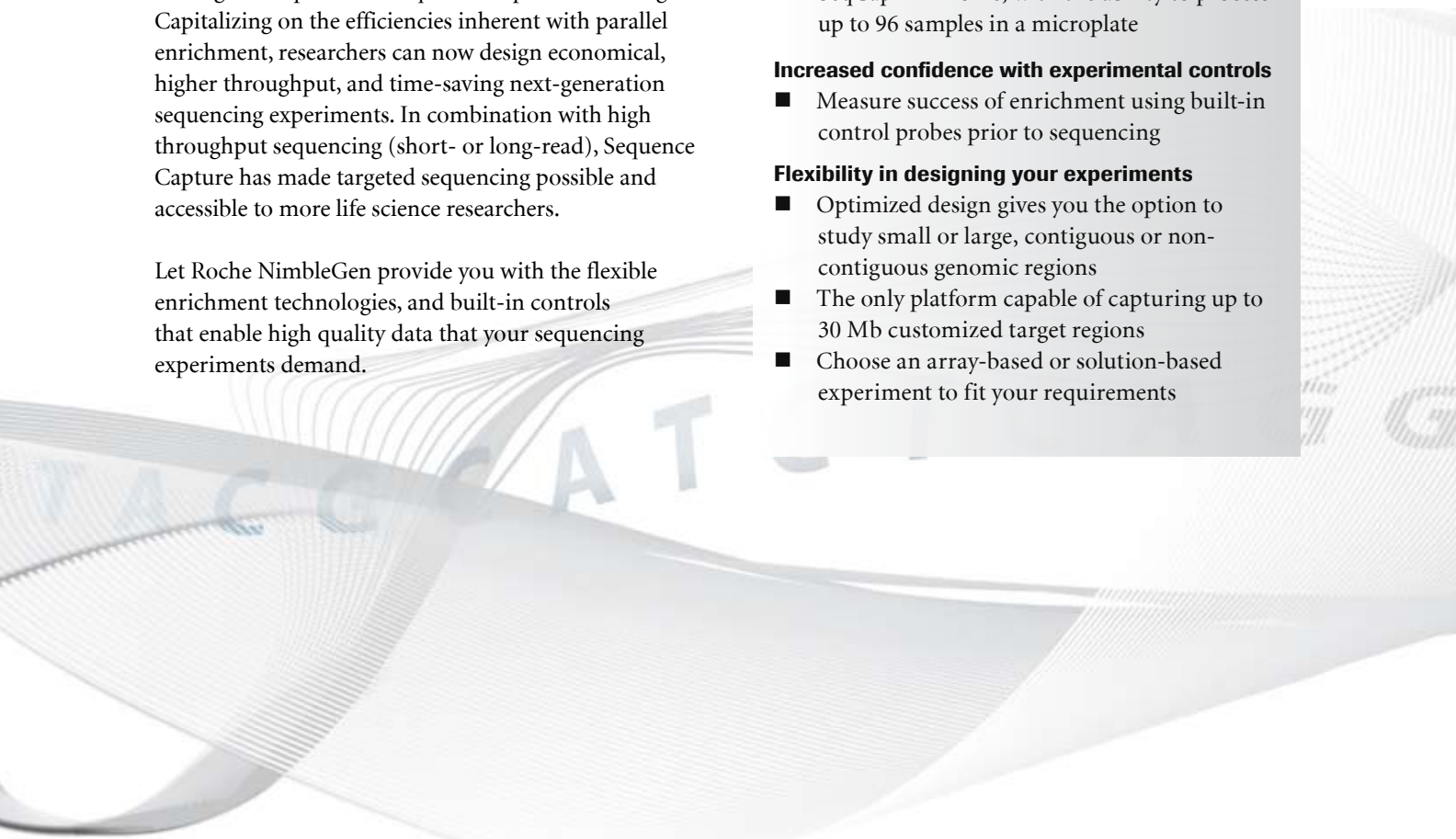
- Optimized protocols for next-generation sequencing platforms (short- and long-read)
- A truly scalable in-solution workflow using SeqCap EZ Exome, with the ability to process up to 96 samples in a microplate

Increased confidence with experimental controls

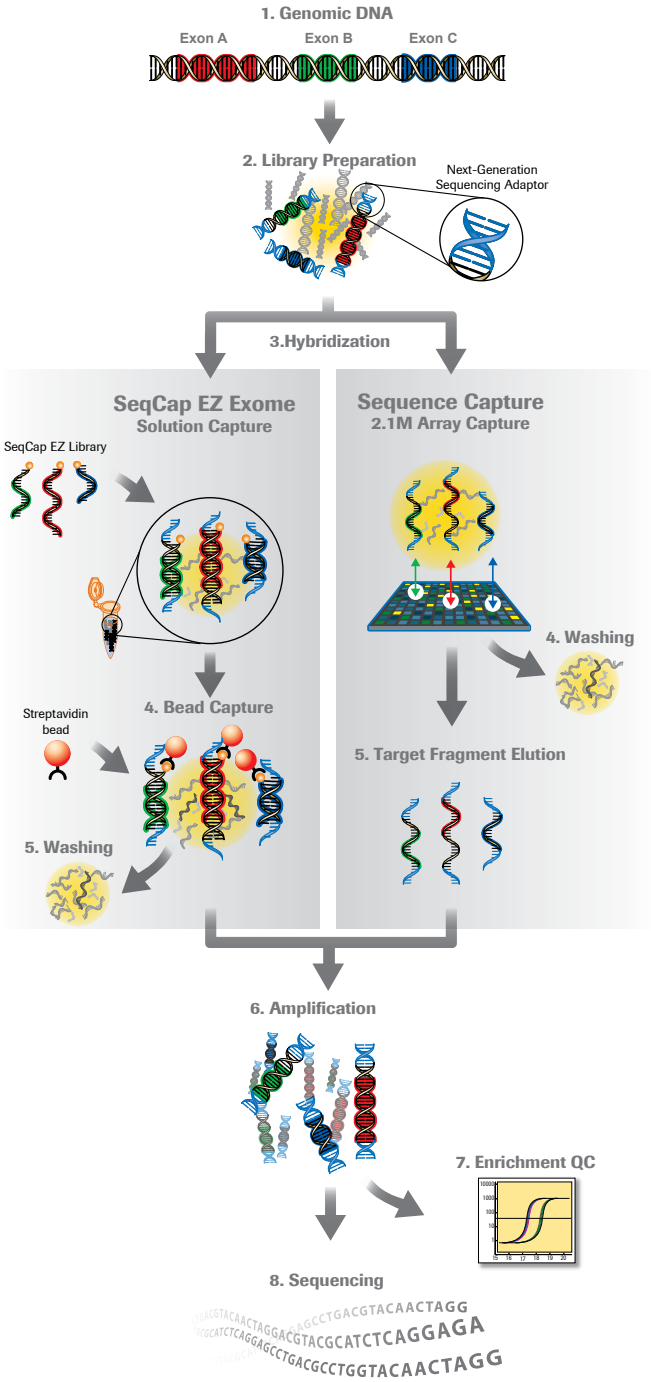
- Measure success of enrichment using built-in control probes prior to sequencing

Flexibility in designing your experiments

- Optimized design gives you the option to study small or large, contiguous or non-contiguous genomic regions
- The only platform capable of capturing up to 30 Mb customized target regions
- Choose an array-based or solution-based experiment to fit your requirements



Robust Protocols for High Quality Targeted Enrichment



Roche NimbleGen offers two types of capture methods: SeqCap EZ Exome, a solution-based method and Sequence Capture, an array-based capture method.

◀ **Figure 1: Sequence Capture protocols.**

1. **Genomic DNA:** SeqCap EZ Oligo pool or an array is made against target regions in the genome.
2. **Library Preparation**: Standard shot-gun sequencing library is made from genomic DNA.
3. **Hybridization:** The sequencing library is hybridized to the SeqCap EZ oligo pool or to the Sequence Capture array.

Steps 4 and 5 are different for each protocol:

SeqCap EZ Exome, biotinylated DNA oligos in solution

4. **Bead Capture:** Streptavidin beads are used to pull down the complex of capture oligos and genomic DNA fragments.
5. **Washing:** Unbound fragments are removed by washing.

Sequence Capture, capture probes synthesized on array:

4. **Washing:** Unbound fragments are removed by washing.
5. **Target Fragment Elution:** The enriched fragment pool is eluted and recovered from the array.

6. **Amplification:** Enriched fragment pool is amplified by PCR.
7. **Enrichment QC:** The success of enrichment is measured by qPCR at control loci.
8. **Sequencing-Ready DNA:** The end product is a sequencing library enriched for target regions, ready for high throughput sequencing.

Optimized Design for Improved Capture Performance

Probe selection plays a critical role in the performance of enrichment technologies. The most commonly used method is a standard tiling design, where probes are laid out at even spacing across the target regions. However, real world sequencing data have shown that a standard tiling design does not always give the best performance, especially for exonic regions where simple tiling tends to result in bias in coverage profiles between different exons.

Roche NimbleGen has developed a probe selection algorithm based on our extensive collection of sequencing data on multiple capture designs. The result is an empirically optimized design algorithm that improves capture uniformity over standard tiling methods and gives better coverage of target regions.

To discover more about NimbleGen Sequence Capture and a list of publications featuring our technologies, go to
www.nimblegen.com/seqcap

1000 Genomes Exon Capture Array*	Standard Tiling Design	Sequence Capture Algorithm
Targets Covered by Sequence	97.5%	98.3%
Bases with 5X Minimum Coverage	70.0%	90.1%
Bases with 10X Minimum Coverage	50.5%	75.0%

* 1000 genes, ~3 Mb target sequence, 1 Genome Sequencer FLX run/sample

▲ **Table 1: NimbleGen Sequence Capture design algorithm increases effective sequence coverage.** Both designs used similar amount of sequencing, however the Sequence Capture design algorithm produces a higher percentage of bases with ≥ 10X coverage compared to the standard tiling design. Experimental data courtesy of Dr. Richard Gibbs, Baylor College of Medicine.

NimbleGen Sequence Capture

"We are extremely pleased with the capabilities and efficiencies the NimbleGen Sequence Capture technology has brought to our sequencing research efforts. There are huge advantages when this technology is compared to PCR-based methods. This is the most exciting next phase in bringing genetic discovery to medicine."

Richard Gibbs, PhD
Director, Human Genome Sequencing Center
Baylor College of Medicine

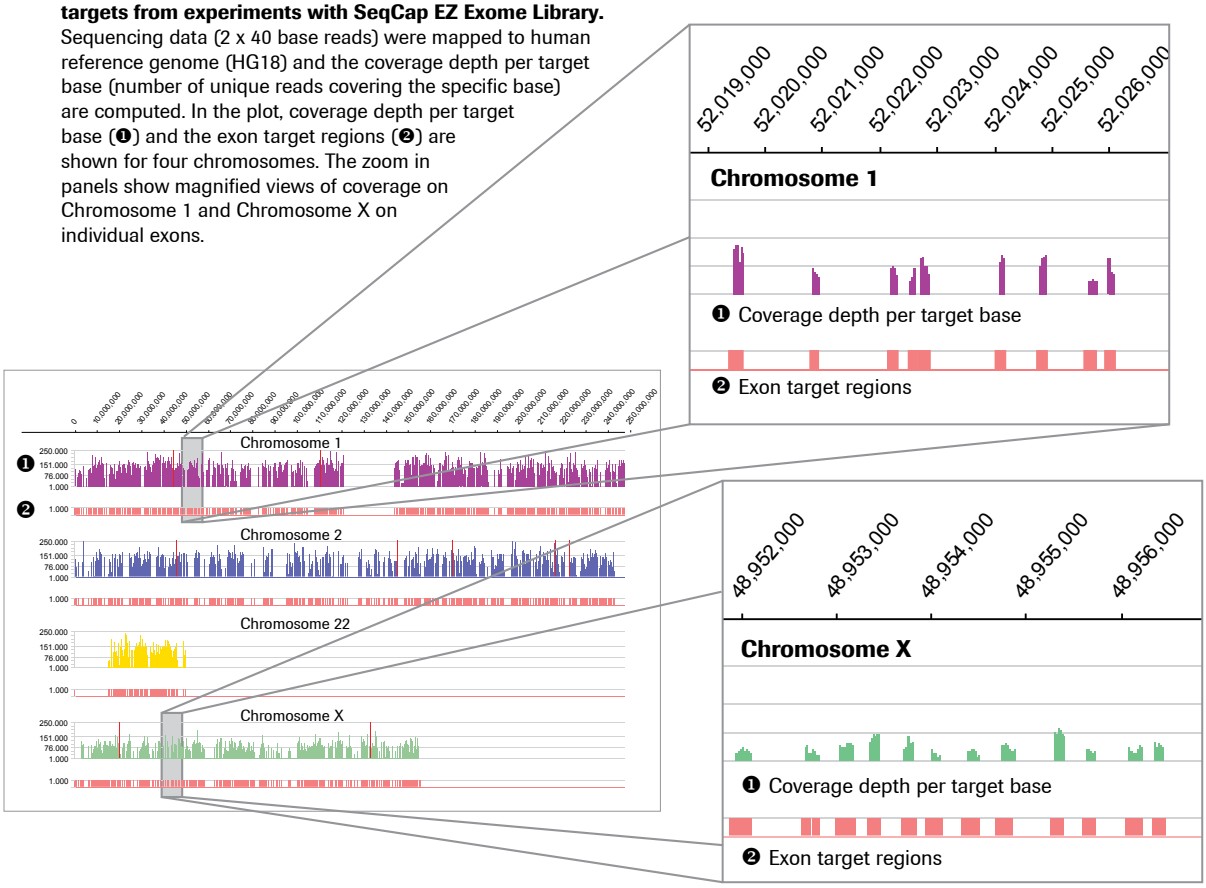
SeqCap EZ Exome - Focusing on the Most Relevant 1% of the Genome

The majority of known disease-causing mutations reside in coding regions in the genome, which comprise the most functionally relevant and best understood 1% of the human genome, or the exome. Targeted sequencing of the exome is an innovative technique to identify causative mutations for genetic disorders (both Mendelian and complex diseases) quickly and cost-effectively.

From our ongoing commitment to advancing targeted sequencing, SeqCap EZ Exome was developed as a solution-based capture method. Built upon an optimized design algorithm, SeqCap EZ Exome sets a new standard for a simple single-step enrichment method:

- Reduces the potential for DNA sample evaporation; no high-temperature liquid handling necessary.
- Built-in control probes measure enrichment success prior to sequencing, improving coverage uniformity.
- In-solution capture alleviates the need for additional equipment.
- Scalable and optimized workflow allows for large-scale experimental studies.
- Target and readily detect known, novel, and rare SNPs.

▼ **Figure 2: Example of sequencing coverage on exon targets from experiments with SeqCap EZ Exome Library.** Sequencing data (2 x 40 base reads) were mapped to human reference genome (HG18) and the coverage depth per target base (number of unique reads covering the specific base) are computed. In the plot, coverage depth per target base (❶) and the exon target regions (❷) are shown for four chromosomes. The zoom in panels show magnified views of coverage on Chromosome 1 and Chromosome X on individual exons.



Better Uniformity and More Coverage For Increased SNP Detection

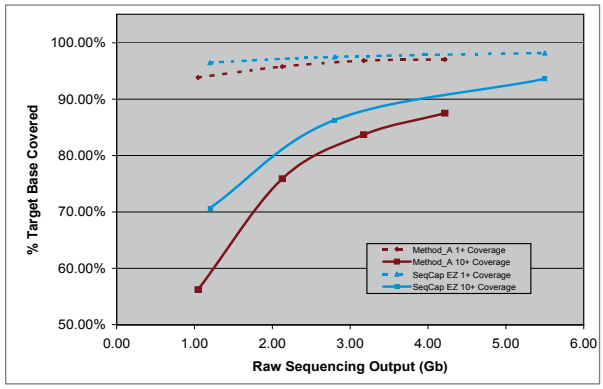
High performance data and results can be achieved by combining SeqCap EZ Exome with next-generation sequencing technologies. Targeted exons across the genome are efficiently captured and sequenced, as demonstrated by the example shown in Figure 2. For detection of single nucleotide variations, greater than 95% of all known SNPs in the exome are readily detected with about 3 Gb of raw sequencing (Table 2). This level of performance delivers about 30 fold reduction of sequencing comparing to whole genome sequencing, where typically ~ 90 GB raw sequencing is required to detect SNPs reliably in the genome.

The unique empirically optimized probe selection process from Roche NimbleGen delivers high levels of uniformity for exome capture, leading to better coverage with less sequencing. All enrichment technologies will introduce some level of coverage bias, shown as uneven coverage between different exons. This bias requires more sequencing to compensate for the unevenness, therefore increasing the cost of sequencing. By employing an intelligent design with 2.1 million oligos (average 10 oligos/exon) and an optimized protocol, SeqCap EZ Exome sets the new standard for uniform capture of exons. Compared with other methods, significantly less sequencing is required to achieve the same coverage. For example, in order to cover 80% of targets with 10 or more reads, about 2Gb of sequencing is needed for SeqCap EZ Exome, while greater than 2.5Gb is required for a competitors technology (Figure 3).

Sequencing Output	1 Lane	2 Lanes	4 Lanes
Total Sequence (Gb)	1.2	2.8	5.5
Total Number of Reads (million)*	15.2	35.5	69.2
% Read on Target	60.8%	60.2%	61.9%
Mean Coverage	20.0	38.2	67.2
% Target Base Covered by 1+ reads	96.4%	97.5%	98.2%
% Target Base Covered by 5+ reads	86.0%	92.9%	96.3%
% Target Base Covered by 10+ reads	70.7%	86.3%	93.6%
Detection Rate for Known Heterozygous SNPs in Exon Targets (6318)	88.8%	95.4%	97.9%
Detection Rate for Known Homozygous SNPs in Exon Targets (4787)	94.3%	96.3%	97.3%

* This is total number of raw sequence reads (paired-end 2x40 base reads). The reads are filtered for uniquely mapped reads for downstream SNP analysis.

▲ **Table 2: SNPs are readily detected by targeted sequencing on samples prepared by SeqCap EZ Exome.** HapMap sample NA12762 was used in this experiment, and the sequencing data was analyzed with Short Oligonucleotide Alignment Package (SOAP).



▲ **Figure 3: SeqCap EZ Exome provides better coverage with less sequencing compared to another exome enrichment method.** The 10+ coverage plots are more relevant for experiments aimed at discovering novel and rare SNPs because multiple reads (at least 8 or 10 reads) are needed to detect a heterozygous SNP reliably.

Discover More in Disease Associated Regions

Discover the Difference with Sequence Capture arrays

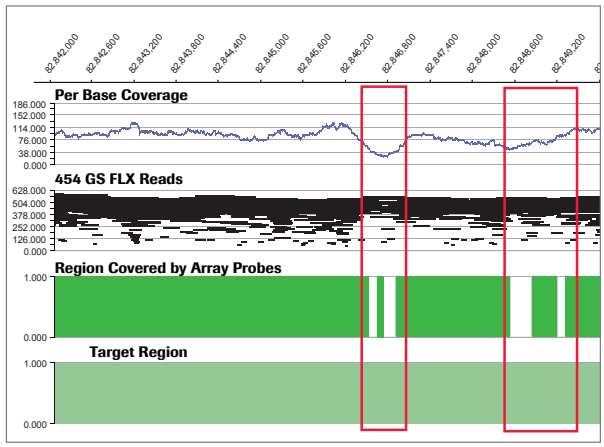
Genome-wide association studies (GWAS) have identified hundreds of loci, with regions ranging from a few hundred kilobases to a few megabases, that are associated with human diseases and traits. A major challenge for follow-up studies is to sequence disease-associated regions to identify the causative mutations for these diseases. Due to the large sizes of these regions, traditional technologies, such as PCR and capillary sequencing are time and cost prohibitive.

NimbleGen Sequence Capture 2.1M and 385K arrays provide an ideal solution for targeted enrichment of these disease associated regions. These customizable arrays can be made to target any region or sequence in the human genome, and offer a powerful solution for large-scale targeted resequencing studies to significantly reduce time, labor, and cost while improving data quality.

NimbleGen Sequence Capture arrays are suitable for targeted sequencing of any size, from small target regions like 250kb (Figure 4) to large regions as large as 30Mb (Table 3). All human designs utilize the empirically optimized Sequence Capture design algorithm to ensure highly uniform capture. For example, a 250kb contiguous region, representing a typical GWAS locus, is captured with high specificity and uniformity (Figure 4). Note that small repetitive regions where no probes were selected can still be covered by sequencing, due to efficient capture from neighboring probes and with the advantage of long reads from the Genome Sequencer FLX Titanium Series (red boxes in Figure 4).

Capture of Large Contiguous Regions using 2.1M Arrays		
Experiment	A	B
Total Reads (millions)	1.2	1.3
Total Bases	347 Mb	380 Mb
% Reads Mapped Uniquely	87.6	86.7
% Bases Mapped Uniquely	93.1	92.6
% Mapped Reads on Target	79.1	70.8
Average/Median Coverage	10.3/9	10.1/8

▲ **Table 3: The ENCODE pilot regions (~30Mb) are captured using 2.1M arrays and sequenced.** The target regions consist of ~ 50 individual contigs of ~ 500kb each



▲ **Figure 4: High-Performance Targeted Resequencing in a 250kb Target Region.**

Focus on Genomic Regions that Matter in Your Research

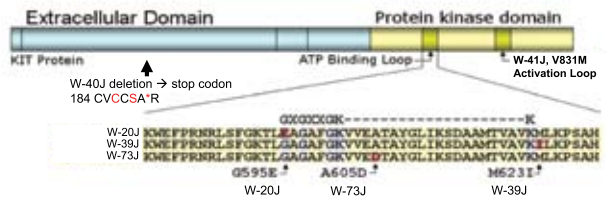
Roche offers a seamless workflow combining NimbleGen Sequence Capture Arrays and the high throughput sequencing of the Genome Sequencer FLX System from 454 Life Sciences. This complete solution of kits, arrays and instruments are specifically designed to optimize the workflow, reduce processing time, minimize costs, and enhance data quality. Furthermore, the GS Reference Mapper software from 454 Life Sciences enables researchers to easily identify variants like SNPs and indels from the final data output without complicated bioinformatics infrastructure (Table 4).

An example of discovering causative mutations, the mouse Kit locus (~200 kb) from 5 non-complementing Kit mutants is shown in Figure 5. These alleles include one known allele W-41J, and four unknown alleles, W-20J, W-39J, W-40J and W-73J. The known mutation from W-41J was confirmed in this experiment, and the data analysis successfully identified a non-synonymous coding mutation for each of the 4 unknown alleles. (D'Ascenzo et al, *Mamm. Genome*, 2009, 20:424–436)

454 Optimized Sequence Capture: Resequencing of HapMap Research Sample

Experiment	250 kb - 1	1 Mb - 1
Total Reads	70,190	140,374
Total Bases	27,646,394	55,453,593
On-Target Reads	75.2%	87.3%
Median Coverage	85	49
Target Bases with 1+ Coverage	98.6%	96.9%
Target Bases with 10+ Coverage	97.3%	92.8%
Known SNP Detection Rate	97.4%	96.5%

▲ **Table 4: Sequence Capture Performance on a 250 kb contiguous region and a 1 Mb contiguous region in the human genome.** Data shown are from 1 of the 4 independent experiments for each region. A HapMap sample is used in the study and SNP calls were generated by the GS Reference Mapper software.



▲ **Figure 5: Mutation discovery in the mouse *KIT* Locus using Sequence Capture and 454 Sequencing.**

SeqCap EZ Exome- Genetic Discovery Made Easy

SeqCap EZ Exome System, Short Read

Product	Delivery Cat. No.	Pack Size	Genome Build	Target Region (total coverage)
SeqCap EZ Human Exome Library SR, 4 reactions	05 977 215 001	4	HG18	Coding exons and miRNA exons
SeqCap EZ Human Exome Library SR, 48 reactions	05 977 223 001	48	HG18	Coding exons and miRNA exons

SeqCap EZ Exome System, Long Read

Product	Delivery Cat. No.	Pack Size	Genome Build	Target Region (total coverage)
SeqCap EZ Human Exome Library LR, 4 reactions	05 933 374 001	4	HG18	Coding exons and miRNA exons
SeqCap EZ Human Exome Library LR, 48 reactions	05 933 382 001	48	HG18	Coding exons and miRNA exons

The design targets 180,000 coding exons (from CCDS database) and ~550 miRNA exons (from miRBase). SeqCap EZ Human Exome Library SR is optimized for short read sequencing. SeqCap EZ Human Exome Library LR is optimized for long read sequencing with the 454 Sequencing System. Annotation files and a list of accessories required to process the SeqCap EZ Exome Library are available at *www.nimblegen.com/seqcapEZ.*

Reagents	Cat. No.	No. of Reactions (SeqCap EZ)
NimbleGen Sequence Capture Hybridization Kit	05 340 721 001	60
NimbleGen Sequence Capture Wash and Elution Kit	05 340 730 001	1200

For more information on NimbleGen products, contact Roche Microarray Technical Support:
www.nimblegen.com/arraysupport

Optimized Array Workflow Solutions - Targeted Enrichment in Your Laboratory

Roche NimbleGen offers an optimized workflow for NimbleGen Sequence Capture. The workflow includes arrays, reagents, and instruments necessary to prepare targeted resequencing samples in your laboratory.

Sequence Capture 2.1M Array

Product	D Delivery Cat. No.	Pack Size	Genome Build	Target Region (total coverage)
Sequence Capture Custom Human 2.1M Array	05 329 841 001	1 slide	HG18 or HG19	Up to 30 Mb
Sequence Capture Developer 2.1M Array	05 451 965 001	1 slide	customer specified	Up to 30 Mb

Sequence Capture 385K Array

Product	D Delivery Cat. No.	Pack Size	Genome Build	Target Region (total coverage)
454 Optimized Sequence Capture Human 385K Array	05 478 731 001	1 slide	HG18 or HG19	Up to 5 Mb
Sequence Capture Custom Human 385K Array	05 394 538 001	1 slide	HG18 or HG19	Up to 5 Mb
Sequence Capture Developer 385K Array	05 394 589 001	1 slide	customer specified	Up to 5 Mb
Additional array information is available at *www.nimblegen.com/seqcap.*				

Reagents	Cat. No.	No. of Reactions
NimbleGen Sequence Capture Hybridization Kit	05 340 721 001	30 per 2.1M Array 60 per 385K Array
NimbleGen Sequence Capture Wash and Elution Kit	05 340 730 001	15 per 2.1M Array 15 per 385K Array

Equipment	Cat No.
NimbleGen Hybridization System 4 (110V)	05 223 652 001
NimbleGen Hybridization System 12 (110V)	05 223 679 001
NimbleGen Hybridization System 4 (220V)	05 223 687 001
NimbleGen Hybridization System 12 (220V)	05 223 695 001
NimbleGen Elution System	05 329 752 001

For life science research only.
Not for use in diagnostic procedures.

Published by:
Roche Diagnostics GmbH
Roche Applied Science
Werk Penzberg
82372 Penzberg
Germany

www.roche-applied-science.com

05227887001 ● 05/10