



# “ican publish in record time.”

“Working with one assistant, I was able to go from installation of an Illumina Genome Analyzer to publication in *Nature Methods* in just three months. The system’s automation and massive throughput have had a huge impact on my research.”

Dr. Yuan Gao  
Assistant Professor  
Center for the Study of Biological Complexity &  
Department of Computer Science  
Virginia Commonwealth University

Power. Speed. Simplicity. The Illumina Genome Analyzer puts publishable results into your hands. Quickly.

~~Next-gen Sequencing~~  
*now*

[www.illumina.com/sequencing?gr](http://www.illumina.com/sequencing?gr)

SEQUENCING  
GENOTYPING  
GENE EXPRESSION

illumina®

# BIOSUPPLYNET

A Current, Integrated Information Source  
for Life Science Laboratory Supplies



BIOSUPPLYNET

## Searching

Searching is very convenient. Info on over 20,000 products from over 6,500 suppliers is right at your finger tips.

## Find A Kit

Many suppliers offer time and effort saving kits to perform research techniques. Search our exclusive database of available research kits.

## BioToolKit

An annotated directory of over 1000 links to online molecular biology resources, including basic research tools and advanced applications for genome, transcriptome, and proteome data retrieval, analysis, and visualization.

## Featured Protocols

Download free protocols from Cold Spring Harbor Laboratory Press manuals, and find sources for the reagents and equipment needed.

## Career Center

Find or place listings for available positions in academic research, pharmaceutical companies, clinical research facilities, and the biotechnology industry.



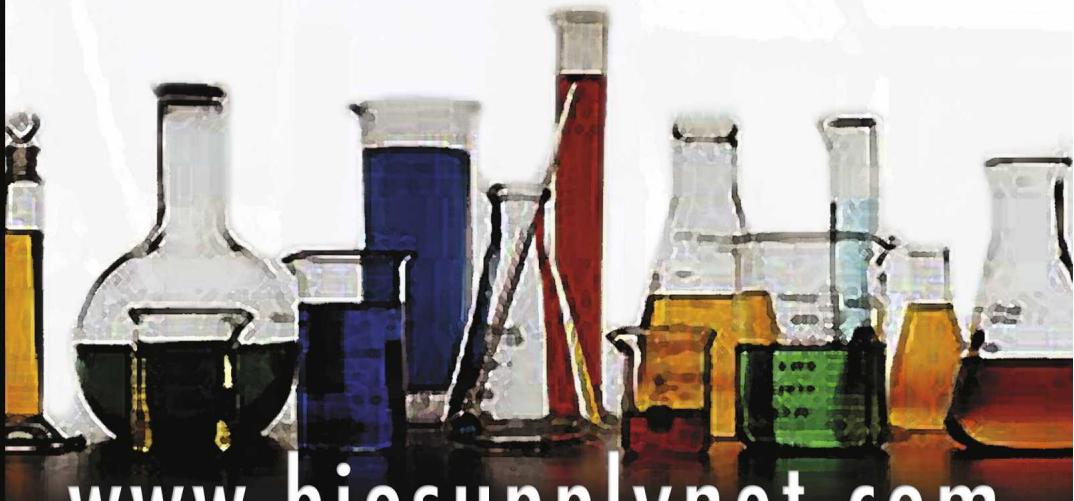
## CSH Protocols and BioSupplyNet Are Merging to Better Serve Your Laboratory Needs!

- New CSH Protocols functionality is being added to BioSupplyNet
- SAVE TIME! No need to search through stacks of laboratory manuals and numerous websites
- Over 20,000 products and supplies listed (indexed by type and/or by company)
- Over 6,500 prominent suppliers participating and the number continues to grow
- Searching is FAST and EASY
- NO REGISTRATION necessary

## In addition to searching for products, Scientists can quickly:

- Get immediate inside info about new products and special deals
- Find kits to perform research techniques
- Download free featured protocols from Cold Spring Harbor Laboratory Press manuals
- Visit our Career Center to search and post job listings
- Search CSH Protocols for up-to-date laboratory methods
- Order Catalogs
- Sign up for Newsletters

**Visit BioSupplyNet.com Today!**



# HudsonAlpha Institute for Biotechnology

*Where genome-scale technology addresses human diversity and disease*

HudsonAlpha is poised to increase the quality and health of human life by leveraging its unique model of genomic research, educational outreach and economic development to expedite the creation of tools, diagnostics and treatments for patients in need. Investigators at HudsonAlpha are studying human genetics and disease, particularly focusing on cancer, diseases of the nervous system, and infectious diseases, and large-scale genomic projects including ENCODE and TCGA. We welcome you to join our growing community.

*Resumes are currently being accepted for:*

**Investigators**

Senior Research Scientists  
Postdocs  
Research Associates and  
Assistants

**Current Investigators:**

Richard M. Myers, Ph.D.  
Director and Investigator  
Jian Han, M.D., Ph.D.  
Investigator

Devin M. Absher, Ph.D.  
Investigator  
Jeremy Schmutz  
Investigator

Jane Grimwood, Ph.D.  
Investigator  
Greg Barsh, M.D., Ph.D.  
Visiting Investigator

For descriptions of research areas see [hudsonalpha.org/pages/sr-researchareas.html](http://hudsonalpha.org/pages/sr-researchareas.html)



Please send resume and cover letter to:

Dr. Chris Gunter  
Director of Research Affairs  
HudsonAlpha Institute for Biotechnology  
601 Genome Way  
Huntsville, AL 35806  
[cgunter@hudsonalpha.org](mailto:cgunter@hudsonalpha.org)



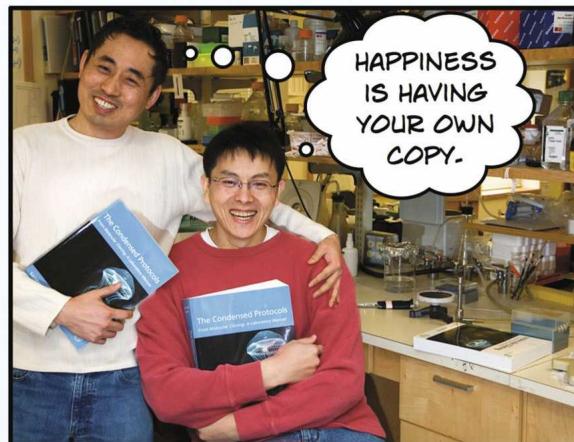
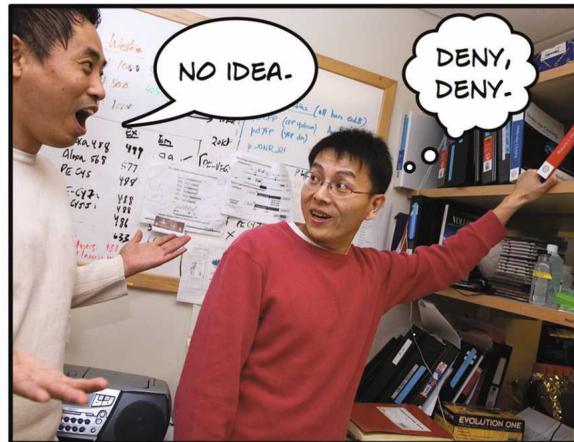
Huntsville, AL • [hudsonalpha.org](http://hudsonalpha.org)

*About HudsonAlpha*

From spirit to physical design, the institute's primary facility embodies and nurtures the sharing of ideas and information. Researchers employed by the not-for-profit HudsonAlpha Institute reside in one wing of the 270,000 square-ft. facility, while a separate wing houses 12 for-profit businesses. The wings are physically bridged with walkways spanning a soaring atrium that features inviting common areas. Proximity to the University of Alabama in Huntsville, the University of Alabama at Birmingham, Auburn University and Vanderbilt University adds to a rich intellectual environment for collaboration, discovery and innovation.

genomic research • educational outreach • economic development

# A Book in the Hand is Worth Three (somewhere) in the Lab

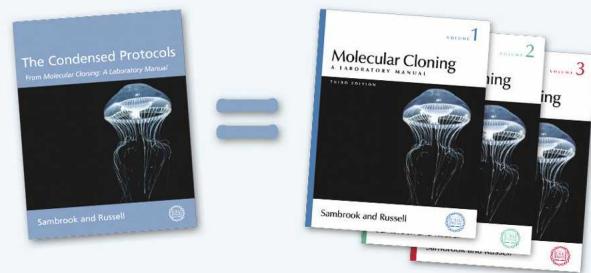


You're immersed in an important experiment, and the three volumes of your old standby, *Molecular Cloning*, are nowhere to be found. *Volume 2* walked home with the new graduate student, *Volume 3* was last seen caked in chemicals, and the lab downstairs borrowed *Volume 1*—again.

Sound familiar? What to do?

Purchase your very own, personal copy of all *Molecular Cloning* protocols! *The Condensed Protocols* contains ALL protocols from the third edition of *Molecular Cloning* in one affordable volume. With all of the information to perform essential biomolecular techniques in one spot, you'll never be left hanging again.

*The Condensed Protocols* is designed for bench use, and includes step-by-step instructions on how to perform each protocol, comprehensive lists of reagents and equipment, and recipes for preparing buffers and stock solutions. Each protocol is cross-referenced to the appropriate pages in *Molecular Cloning*, where you can learn about why the techniques work, how they were first developed, and how they have evolved.



Mention source code CPO8 for a 15% discount. Gold Members receive an additional 10% discount—sign up today!

# Speed Up Your Genome Research!

## Ensure Rapid and Efficient Blunt-End Repair of Genomic DNA

The *End-It™ DNA End-Repair Kit* rapidly and efficiently converts nebulized, sonicated, or sheared genomic DNA to 5'-phosphorylated, blunt-ended DNA for subsequent genomic cloning or next-gen sequencing.

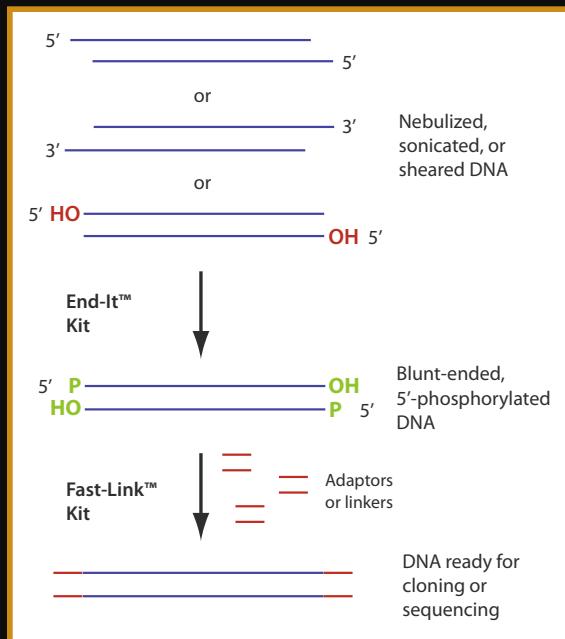
- Repaired DNA is blunt-ended and 5' phosphorylated for immediate blunt-end ligation.
- End-repair up to 5 µg of genomic DNA per reaction.

Then,

## Perform Blunt-End Ligations in as Little as 15 Minutes

The *Fast-Link™ DNA Ligation Kit* uses a high-quality, specially formulated ligase to provide extremely rapid, high-efficiency DNA ligations.

- Blunt-end ligations in as little as 15 minutes.
- Cohesive-end ligations in 5 minutes.



shades of crimson.

## Taq DNA Polymerase from New England Biolabs

### HIGH YIELD, ROBUST AND RELIABLE PCR REACTIONS IN CONVENIENT FORMATS

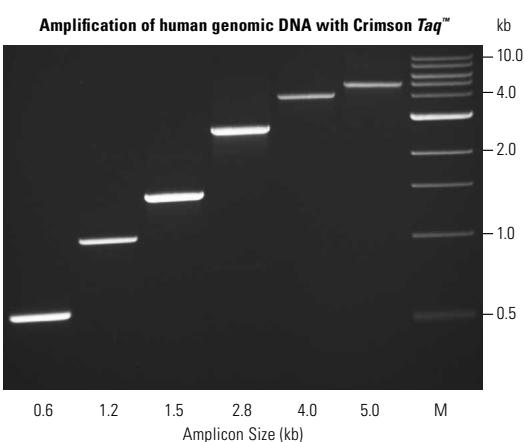
Looking for the right solution for your high yield PCR? Choose from an expanded selection of recombinant *Taq* DNA Polymerase based products from New England Biolabs. And for even greater convenience, NEB introduces Crimson *Taq™* DNA Polymerase for direct loading of samples onto a gel. Choose the *Taq* DNA Polymerase from NEB that's right for you and experience guaranteed Performance – Convenience – Results.

■ <b>NEW</b> Crimson <i>Taq™</i> DNA Polymerase	■	M0324S/L
■ LongAmp™ <i>Taq</i> 2X Master Mix	■	M0287S/L
■ LongAmp™ <i>Taq</i> DNA Polymerase	■	M0323S/L
■ LongAmp™ <i>Taq</i> PCR Kit	■	E5200S
■ Quick-Load™ <i>Taq</i> 2X Master Mix	■	M0271S/L
■ <i>Taq</i> 2X Master Mix	■	M0270S/L
■ <i>Taq</i> 5X Master Mix	■	M0285S/L
■ <i>Taq</i> PCR Kits	■	E5000S/E5100S
■ <i>Taq</i> with Standard <i>Taq</i> Buffer	■	M0273S/L/X
■ <i>Taq</i> with Standard <i>Taq</i> (Mg-free) Buffer	■	M0320S/L
■ <i>Taq</i> with ThermoPol Buffer	■	M0267S/L/X
■ <i>Taq</i> with ThermoPol II (Mg-free) Buffer	■	M0321S/L

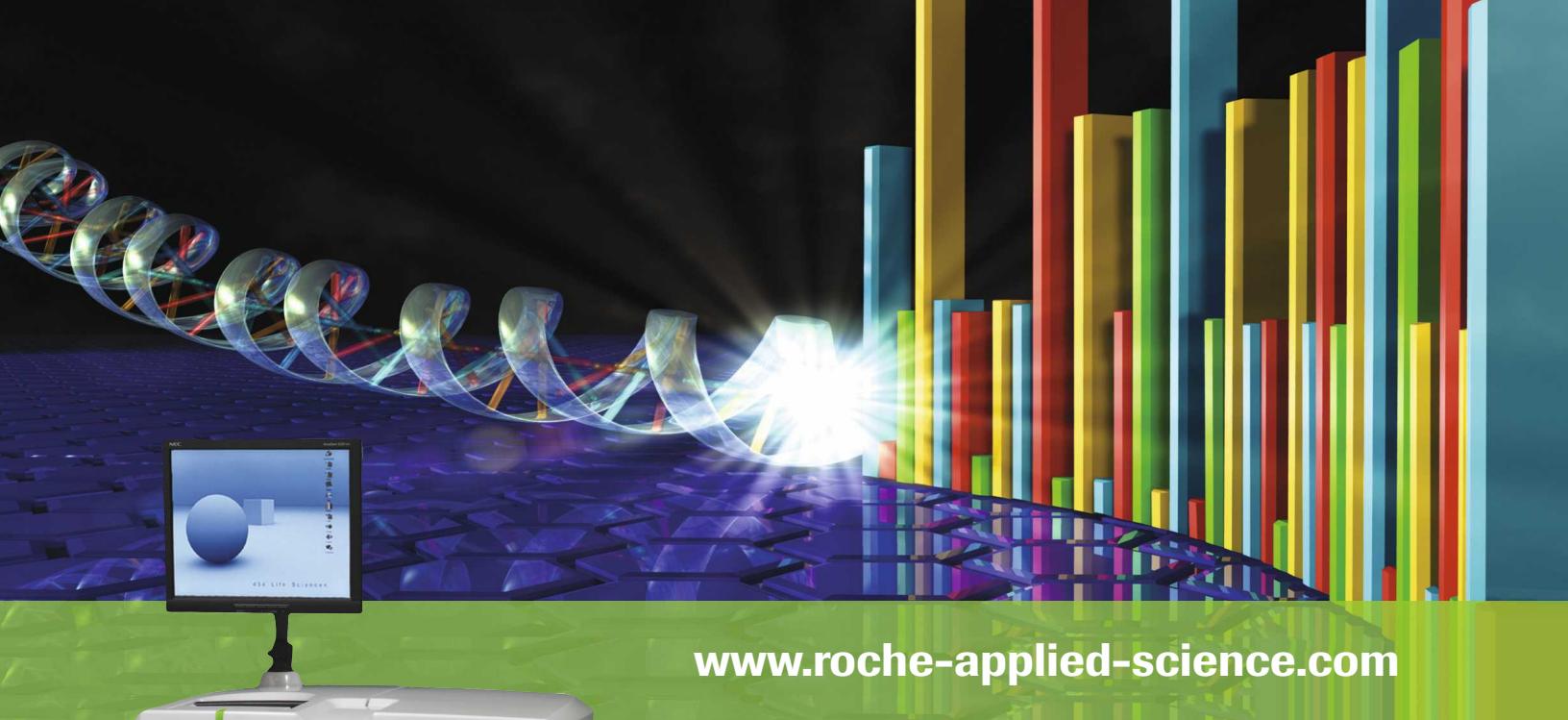
■ = Recombinant

For more information and international distribution network, please visit [www.neb.com](http://www.neb.com)

**New England Biolabs Inc.** 240 County Road, Ipswich, MA 01938 USA 1-800-NEB-LABS Tel. (978) 927-5054 Fax (978) 921-1350 [info@neb.com](mailto:info@neb.com)  
**Canada** Tel. (800) 387-1095 [info@ca.neb.com](mailto:info@ca.neb.com) • **China** Tel. 010-82378266 [beijing@neb-china.com](mailto:beijing@neb-china.com) • **Germany** Tel. 0800/246 5227 [info@de.neb.com](mailto:info@de.neb.com)  
**Japan** Tel. +81 (0)3 5669 6191 [info@neb-japan.com](mailto:info@neb-japan.com) • **UK** Tel. (0800) 318486 [info@uk.neb.com](mailto:info@uk.neb.com)



Choose Crimson *Taq™* DNA Polymerase for robust reactions in a convenient format. Amplicon sizes indicated below gel. Marker M is the 1 kb DNA Ladder (NEB #N3232).



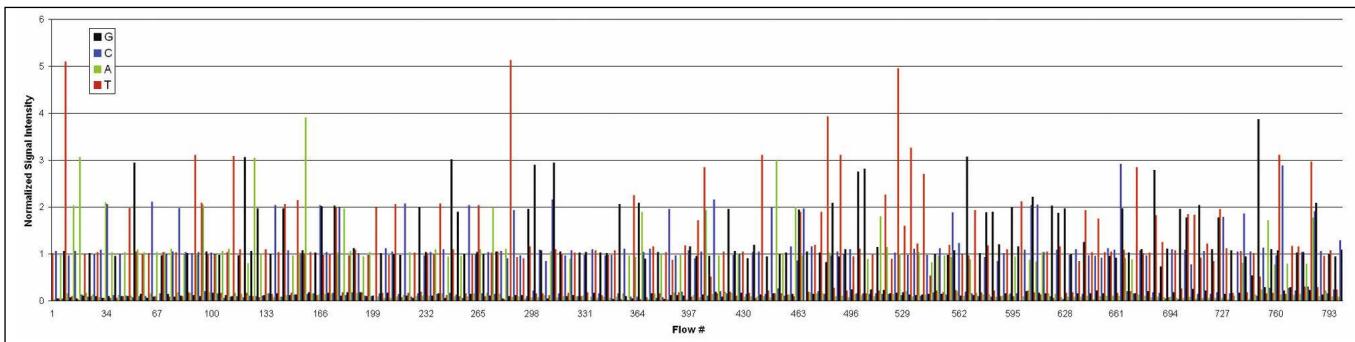
[www.roche-applied-science.com](http://www.roche-applied-science.com)



## Genome Sequencer FLX System

Really  
Length Matters

*Introducing the GS FLX Titanium Reagents*



**DNA Sequencing Flowgram:** Each bar within the flowgram represents a discrete nucleotide (A, T, C, or G) and the height of the bar corresponds to the number of nucleotides detected. The flowgram above represents a 458-base-pair sequencing read from *E. coli* K-12.

- Obtain sequencing read lengths of 400 to 500 bases.
- Generate more than 1 million sequencing reads per 10-hour instrument run.
- Improve performance by using GS FLX Titanium reagents — without instrument upgrades.
- Perform more applications with longer sequencing reads.

Learn more at [www.genome-sequencing.com](http://www.genome-sequencing.com)

**454**

SEQUENCING

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

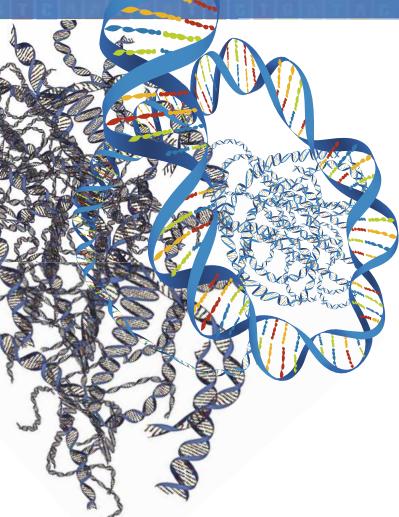
454, 454 LIFE SCIENCES, and 454 SEQUENCING are trademarks of 454 Life Sciences Corporation, Branford, CT, USA, a Roche company. GS FLX TITANIUM is a trademark of Roche. Other brands or product names are trademarks of their respective holders.  
© 2008 Roche Diagnostics. All rights reserved.

Roche Diagnostics  
Roche Applied Science  
Indianapolis, Indiana





[www.nimblegen.com](http://www.nimblegen.com)



# Seize the Genome

## NimbleGen Sequence Capture Arrays and Service

*Maximize the power of next-generation sequencing by capturing and enriching specific regions of interest for targeted resequencing.*

- **Target Specific Regions of Interest**

Capture up to 5 Mb total sequence on a single array with high coverage and specificity.

- **Reduce Cost**

Significantly reduce time and cost compared to laborious and limiting PCR-based methods.

- **Generate Data with Confidence**

Ensure system performance prior to sequencing with built-in QC probes.

- **Customize Each Capture Design**

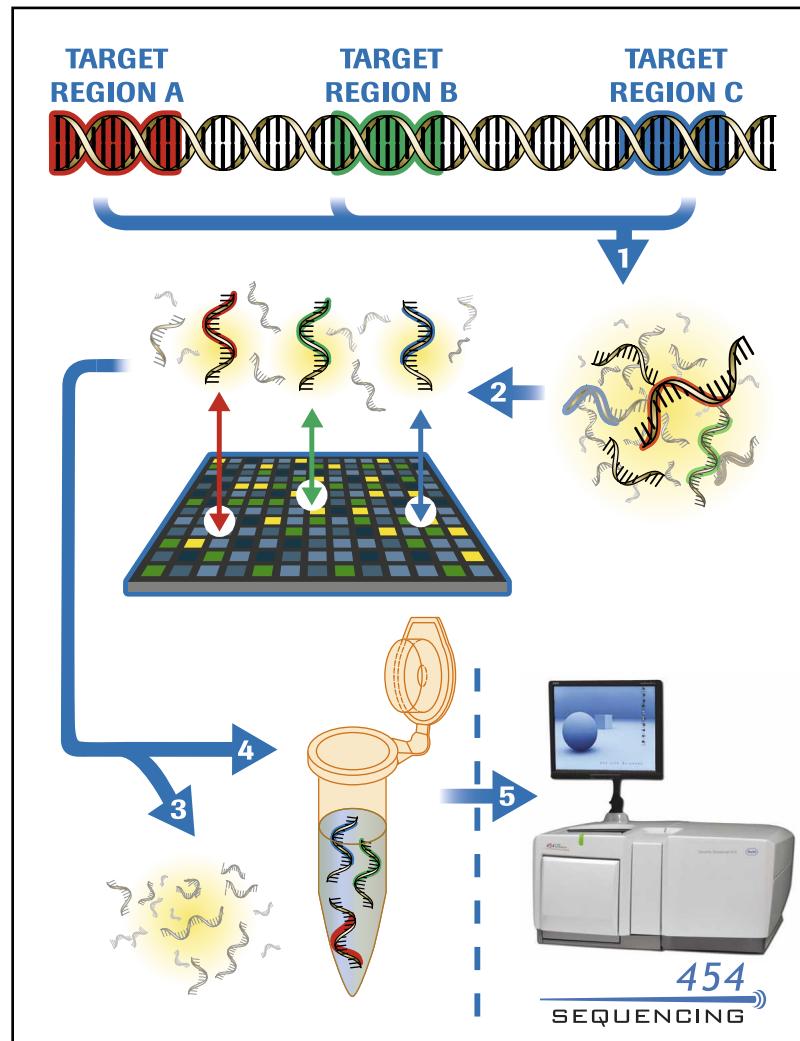
Specify the array design to capture contiguous genomic regions or thousands of exons in parallel.

*To seize command of your sequencing project,  
visit [www.nimblegen.com/seqcap](http://www.nimblegen.com/seqcap)  
or call (877) NimbleGen / (608) 218-7600*

*Roche has developed NimbleGen Sequence Capture technology that enables targeted sequencing of thousands of exons or contiguous genomic loci of up to 5 Mb in a single experiment. The microarray hybridization-based NimbleGen Sequence Capture technology has considerable cost, throughput, and quality advantages when compared to PCR.*

Recent technological advances in sequencing systems have rapidly increased raw sequence output. However, these next-generation sequencing systems do not currently have the throughput to sequence the whole human genome cost-effectively. Thus, they require that the complexity of genomic DNA samples be reduced to a manageable subset prior to sequencing. The prevailing method for complexity reduction has been the preparation of amplicons by parallel, multiplex, or long range PCR amplification. These PCR methods have severe cost and performance limitations when scaled to the level required to take full advantage of the capacity of currently available sequencing systems. As a result of these limitations, the bottleneck for sequencing projects has shifted to sample preparation.

To address this sample preparation bottleneck, Roche has developed the microarray hybridization-based NimbleGen Sequence Capture technology that utilizes high-density oligonucleotide microarrays as a programmable genomic selection device to allow targeted sequencing of subsets of the genome. These genome subsets can be exons, disease-associated regions, quantitative trait loci, promoters and enhancers, and other targeted regions. The revolutionary and simple workflow of NimbleGen Sequence Capture technology enables isolating megabase regions in as little as one week and eliminates the cost, labor, and infrastructure required for large-scale PCR experiments.



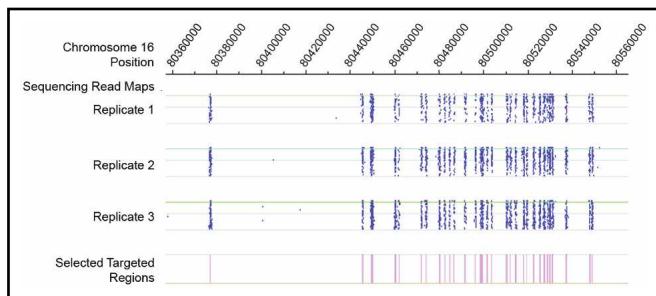
**Figure 1: The NimbleGen Sequence Capture Protocol.**

1. The genomic DNA sample is fragmented.
2. The sample is hybridized to a custom NimbleGen Sequence Capture array.
3. Unbound fragments are removed.
4. The target-enriched pool is eluted and amplified.
5. The enriched sample is ready for processing in the Genome Sequencer FLX sample processing workflow.

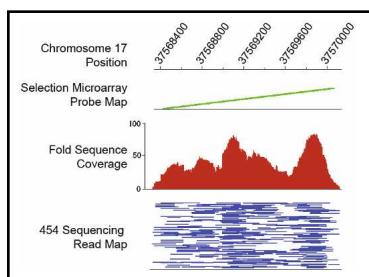
## Sequence Capture Performance

The performance data shown here are derived from the first peer-reviewed publication (1) of the technology, and the NimbleGen Sequence Capture Service offers data quality equal or higher than what has been published. The actual performance depends on the sizes and exact locations of the target regions, and pilot projects involving both Sequence Capture and sequencing is recommended to accurately determine how this technology works for your region of interest. To receive the latest updates on the technology's performance, we encourage you to subscribe to NimbleGen Sequence Capture news at [www.nimblegen.com/seqcap](http://www.nimblegen.com/seqcap).

■ With a single-array experiment covering 5 Mb target regions, followed by a single Genome Sequencer FLX run producing approximately 100 Mb total sequencing data, the majority of sequencing reads represented selected target regions (typically >70%). While the coverage typically depends on the composition of the target regions, approximately 8X median coverage can be achieved for exon-sized regions (Figure 2), and approximately 18X median coverage can be achieved for a single 5 Mb contiguous genomic region (Figure 3). The sequence coverage will increase for smaller cumulative target region sizes, if the same amount of sequencing runs is performed. For example, an experiment covering a 500 kb contiguous target region followed by a single Genome Sequencer FLX run yielded >90X median coverage. In research requiring small regions, NimbleGen Sequence Capture technology combined with 454 sequencing utilizing only a portion (1/2 to 1/16) of the picotiter plate will generate sufficient reads for sequencing applications.

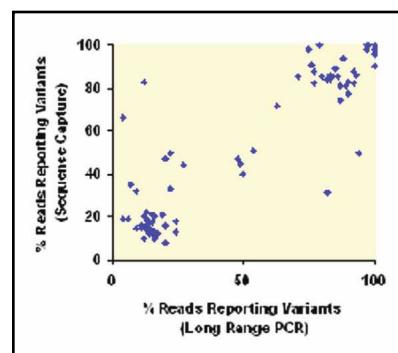


**Figure 2: Sequencing Read Maps of approximately 190 kb of Chromosome 16.** Sequencing capture read maps depict approximately 190 kb of chromosome 16 from three replicates of sequence capture experiments of human exons.



**Figure 3: Sequencing Read Map of 2 kb of Chromosome 17.** A sequencing read map shows 2 kb of chromosome 17 from a microarray selection of a 2 Mb contiguous region that contains the BRCA1 gene.

■ Although arrays are typically designed based on the reference genome, NimbleGen Sequence Capture technology does not bias against discovery of unknown variants. As shown in Figure 4, in an experiment designed to compare the performance of this technology versus long range PCR, almost all the variants were captured with the same fidelity as PCR. In the 70 kb region targeted by both methods, 98 SNPs were identified by both, and 9 and 5 rare variants were identified by NimbleGen Sequence Capture technology and long range PCR, respectively. In addition, 22 variants in repeats were detected only by long range PCR because no probes on the array were designed for these repetitive regions.



**Figure 4: Sequencing Read Map of 2 kb of Chromosome 17.**

In this experiment, genomic DNA from a mixed cell population was used, representing both common and rare SNP variants. A 200 kb region surrounding the EGFR gene was captured using NimbleGen Sequence Capture technology, and a 70 kb region out of the 200 kb region was amplified using long range PCR. DNA samples derived from target regions using both methods were sequenced separately with the Genome Sequencer FLX instrument, and SNP discovery was performed on both data sets. The percentages of 454 sequencer reads that report variants, either from PCR or NimbleGen Sequence Capture technology, were plotted for each of the SNPs detected by both methods.

## References

1. Albert TJ, et al. Direct selection of human genomic loci by microarray hybridization. *Nature Methods* 2007 Nov; 4(11):903-5.
2. Okou DT, et al. Microarray-based genomic selection for high-throughput resequencing. *Nature Methods* 2007 Nov; 4(11):907-9.
3. Hodges E, et al. Genome-wide in situ exon capture for selective resequencing. *Nature Genetics* 2007 Dec; 39(12):1522-7.

## No More Tedious and Costly PCR Reactions

NimbleGen Sequence Capture technology has clear advantages for large-scale, targeted sequencing when compared to traditional PCR-based methods, for example:

- **Select Large Genomic Regions:** For a 1 Mb disease-associated region identified by a genome-wide association study, a traditional long-range PCR approach requires 200 work days to optimize all of the PCR conditions. With NimbleGen Sequence Capture technology, this region can now be easily captured in a few weeks. In addition, you can target regions up to 5 Mb with a single array.
- **Capture Exonic Regions:** Sequencing a panel of 1,000 cancer genes with approximately 7,000 exons requires the design and synthesis of 14,000 PCR primers at a cost of >\$30,000. Plus, 7,000 PCR reactions are required at a cost of >\$10,000 per sample. Thus, a PCR approach is cost-prohibitive for most researchers. However, a single NimbleGen Sequence Capture array can target all 7,000 exons easily, in a single enrichment cycle, providing substantial cost and time savings.

## Advantages of NimbleGen Sequence Capture

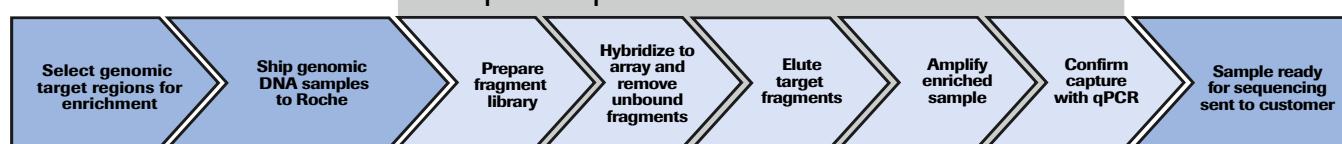
- **Target Specific Regions of Interest:** Capture up to 5 Mb total regions on a single array with high coverage and specificity.
- **Reduce Cost:** Significantly reduce time and cost compared to laborious and limiting PCR-based methods.
- **Generate Data with Confidence:** Ensure system performance prior to sequencing with built-in QC probes.
- **Customize Each Capture Design:** Specify the array design to capture contiguous genomic regions or thousands of exons in parallel.

## NimbleGen Sequence Capture Arrays and Service

Perform the NimbleGen Sequence Capture workflow in your own lab or utilize the NimbleGen Service Lab. See the workflow chart below for details of each option.

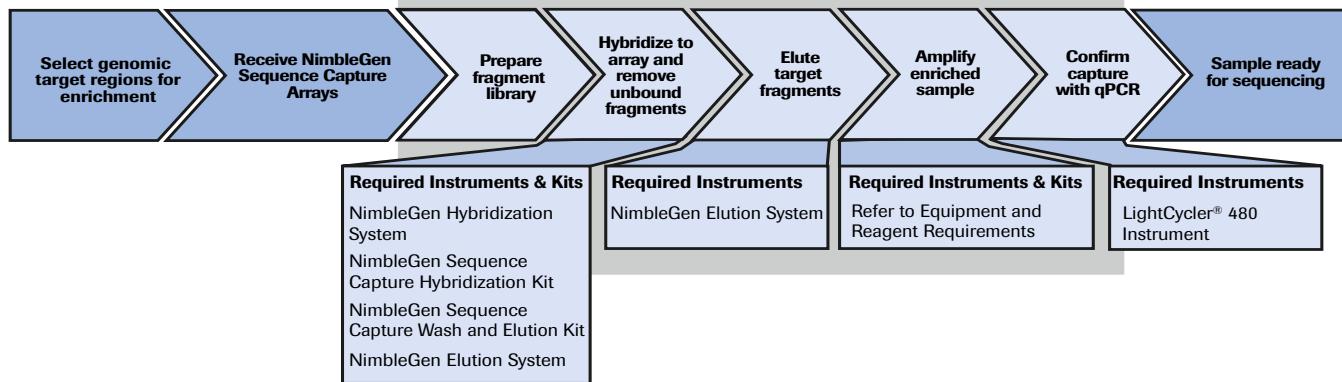
## Service

### Sample Preparation at Roche



## Delivery

### Sample Preparation at Customer Laboratory



*"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, Ph.D.**  
**Director, Human Genome Sequencing Center**  
**Baylor College of Medicine**

## For additional information

Toll-free in US: (877) NimbleGen / (877) 646-2534  
(608) 218-7600  
ngsales@nimblegen.com  
www.nimblegen.com

*Published by*

Roche Diagnostics  
Roche Applied Science  
Indianapolis, Indiana

© 2008 Roche Diagnostics. All rights reserved.



**Plant & Animal Genome XVII**  
**The International Conference on the**  
**Status of**  
**Plant & Animal Genome Research**  
**January 10-14, 2009**  
**Town & Country Hotel**  
**San Diego, California**

**Organizing Committee**

**Chairman:**

Stephen R. Heller, NIST, USA

**Plants:**

Michael Gale, John Innes Center, UK  
Ed Kaleikau, USDA/CSREES, USA

Dave Matthews, USDA, ARS Cornell University, USA  
Graham Moore, John Innes Center, UK

Jerome P. Miksche, Emeritus Director, USDA Plant  
Genome Program, USA

Rod Wing, University of Arizona, USA

**Animals:**

Cecilia Penedo, USA  
Shu-Hong, China  
Joan Lunney, USA  
Jim Reecy, USA

**Abstract Coordinators:**

Victoria Carollo, USDA, ARS, WRRC, USA  
Gerard Lazo, USDA/ARS/WRRC, Albany, CA, USA  
David Grant, USDA/ARS & Iowa State University, USA

**Sponsors**

**USDA, Agricultural Research Service**

**USDA, National Agricultural Library**

**USDA, NRI Competitive Grants Office**

**USDA, Cooperative State Research, Education, and  
Extension Service (CSREES)**

**John Innes Centre**

**NCGR, National Center for Genome Resou**

**Speakers**

Ajaz Hussain, Philip Morris  
Lee Hood, Institutes for Systems Biology  
Andrew Clark, Cornell University  
Ewan Birney, EBI  
Jonathan Wendel, Iowa State Univ.  
George Church, Harvard  
Steve Briggs, UC-San Diego

**Workshops**

Abiotic Stress / Allele Mining / Apomixis Aquaculture / Banana  
(Musa) Genomics / Barley / Bioinformatics / Brassicas /  
Brachypodium Distachyon / Cattle/Sheep / Challenge Program:  
Unlocking Crop Genetic Diversity for the Poor / Citrus  
Compositae / Computer Demonstrations / Cotton / Compositae /  
Connectrons Cool Season Legumes / Curcurbit / Equine / Forage &  
Turf Plants / Forest Trees Fruit and Nut Crops / Functional Genomics  
/ Host Pathogen Interactions / Insect Genetics / ICSB / ICGI /  
Int'l Grape Genome Project / Int'l Lolium Genome  
Initiative / IGGI / ITMI / Large-Insert DNA Libraries and Their  
Applications / Legumes/ Maize / Microarray Analysis /  
Molecular Markers for Plant Breeders/ Mutation Screening /  
NRSP-8 / NC 1010 Development and Implementation of  
Ontologies in the Database / Organellar Genetics / Plant  
Cytogenetics / Plant Development and Signal Networks /  
Interagency Working Group on Plant Genomics / Plant  
Interactions with Pests and Pathogens / Plant Transgene  
Genetics / Plant Reproductive Genomics / Polyploidy / Poultry /  
Proteomics / QTL Cloning / Reduced-representation/Sequencing  
Methods and Applications / Rice / Rice Blast /  
Root Genomics / Solanaceae / Sorghum and Millets / Soybean  
Genomics / Statistical Genomics / Sugar Beet / Swine /  
Swine Genome Sequencing / TAIR /  
Weedy and Invasive Plant Genomics

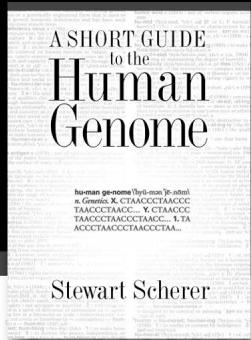
**Organizer**

Scherago International  
525 Washington Blvd., Ste. 3310  
Jersey City, NJ 07310  
201-653-4777 x20  
201-653-5705  
E-mail: pag@scherago.com

For complete details, including on-line registration, visit our website at

**www.intl-pag.org**

# A Short Guide to the Human Genome



Stewart Scherer

By Stewart Scherer

How many genes are in the human genome? Which genes are commonly associated with genetic diseases? How many mobile elements, simple sequence repeats, or protein kinases are encoded in the genome? What are the largest genes and proteins? How similar are human proteins to those of mouse, yeast, or bacteria?

Although the human genome has been sequenced, it often can be surprisingly difficult to find answers to seemingly simple questions about its characteristics. This convenient handbook, written in question-and-answer format, allows researchers and teachers alike access to basic facts about the human genome.

Using a recent assembly of the human genome sequence, Stewart Scherer has compiled answers to a broad range of questions about the structure and function of the human genome. Answers to each question are presented in a direct, straightforward style. Numerous figures and tables are included to illustrate and summarize the information.

2008, 173 pp., illus., index

Paperback \$29

ISBN 978-087969791-4

## CONTENTS

### 1. Introduction

#### 2. DNA and Chromosomes

What Is the Size of Each Chromosome?

What Is the Base Composition of the Nuclear Genome?

What Is the Base Composition of the Mitochondrial Genome?

What Is the Frequency of Each Dinucleotide?

Does Base Composition Vary Among the Chromosomes?

What Are the Frequencies and Sizes of Simple Sequence Repeats?

Which Sequences Are Present at the Centromeres?

What Are the Sequences at the Telomeres?

### 3. Genes

How Many Protein-coding Genes Are Present in the Genome?

Is Gene Density Uniform Across the Chromosomes?

How Common Are Pseudogenes?

What Is the Size of a Typical Gene?

What Are the Largest Genes?

Which Genes Have the Most Exons?

Does Exon Number Correlate with Gene Size or Protein Size?

What Is the Size of a Typical Exon?

What Is the Size of a Typical Intron?

Which Genes Have the Largest Introns?

Do Gene-rich Chromosomes Have Smaller Genes?

How Are CpG Islands Associated with Genes?

Do Large mRNAs Have Large UTRs?

Which Genes Are Located in the Introns of Other Genes?

Which Genes Are Present in the Mitochondrial Genome?

How Are Genes Organized in the Mitochondrial Genome?

### 4. RNA

How Are the Ribosomal RNA Genes Organized in the Genome?

Which Transfer RNAs Are Present in the Genome?

Which Genes Host Small Nuclear RNAs and Small Nucleolar RNAs?

How Are microRNA Genes Distributed in the Genome?

What Is the Size Distribution of snRNA Genes and Related Sequences in the Genome?

What Are the Sequences at Splice Junctions? Which Genes Use RNA Editing?

### 5. Proteins

What Is the Size of a Typical Protein?

What Is the Amino Acid Composition of a Typical Protein?

How Do Mitochondrial Proteins Differ in Composition from Typical Proteins?

Which Amino Acids Are Commonly Located in the Amino-terminal Region of a Protein?

Which Amino Acids Are Commonly Located in the Carboxy-terminal Region of a Protein?

What Are the Largest Proteins?

Which Proteins Have Large Homopolymer Tracts?

What Proteins Are Rich in Particular Amino Acids?

For Proteins That Are Rich in a Specific Amino Acid, How Are the Residues Distributed Across the Polypeptide Chains?

Which Proteins Lack Specific Amino Acids?

Which Proteins Are Post-translationally Cleaved into Multiple Hormones and Related Peptides

### 6. Translation and Protein Modification

How Does Codon Usage Vary Among Human Genes?

Which Proteins Contain Selenocysteine?

Which Amino Acids Are Introduced into Proteins by Post-translational Modifications?

Which Sequences Are Associated with the Formylglycine Modification?

Which Proteins Contain  $\tilde{\alpha}$ -Carboxyglutamate?

### 7. Gene Families

How Are DNA Polymerases Related to Each Other?

How Are the Histone Gene Families Organized?

How Are the Keratin Genes Organized?

How Many Protein Kinases Are Encoded in the Genome?

Which Types of Proteases Are Found in the Human Genome?

How Many Genes Are in the Major Transcription Factor Families?

Which Are the Important Residues in the Homeobox?

How Are the HOX Genes Organized?

Which Sequences Are Shared in Helix-loop-helix Transcription Factors?

Which Residues Define the DEXD Helicases?

Which Are the Conserved Features of the WD Domain?

How Are the Olfactory Receptor Genes and Pseudogenes Distributed in the Genome?

How Are the Taste Receptor Genes and Pseudogenes Distributed in the Genome?

### 8. Mobile Elements and Rearranging Genes

How Much of the Genome Is Composed of Mobile Elements?

What Types of L1 Elements Are Present in the Genome?

What Is the Number and Size Distribution of Alu Family Sequences in the Genome?

What Are the Common DNA Transposons in the Genome?

Which Endogenous Retroviral Genomes Are Largely Intact?

Are the Immunoglobulin and T-cell Receptor Loci Very Large?

How Are the Immunoglobulin Genes Organized?

How Are the T-cell Receptor Loci Organized?

What Sequences Direct DNA Joining at the Immunoglobulin and T-cell Receptor Loci?

### 9. Polymorphism

What Is the Frequency of Single Nucleotide

Polymorphisms in the Genome?

Which ABO Allele Is in the Reference Genome?

How Variable Are the HLA Proteins?

Which Genes Have Common Polymorphisms?

Which Genes and Alleles Are Associated with Common Genetic Diseases?

Which Genes Function as Tumor Suppressors?

### 10. Comparative Genomics

What Are the Major Differences Between the Human and Mouse Genomes?

How Do the Mitochondrial Genomes Vary Among Species?

How Does Codon Usage Differ in the Eukaryotes?

How Similar Are Human Proteins to Those in Other Species?

How Similar Are Human Genes to Those in Other Species at the Nucleotide Level?

How Are Human and Bacterial Ribosomes Related?

Are Human DNA Replication Proteins Related to Those of Bacteria?

Are Human Proteins for mRNA Processing Functions Conserved in Eukaryotes?

How Does Protein Sequence Conservation Differ Among the Histone Types?

How Conserved Are the Branches of the Ras Superfamily?

How Do the Sizes of Gene Families for Structural Proteins Vary Among the Eukaryotes?

How Large Are Gene Families for Developmental Signaling in Animals?

Which Species Have Genes Related to Those with Functions in Human Host Defense?

Which Human Proteins Are Related to Retroviral Proteins?

### Human Gene Index