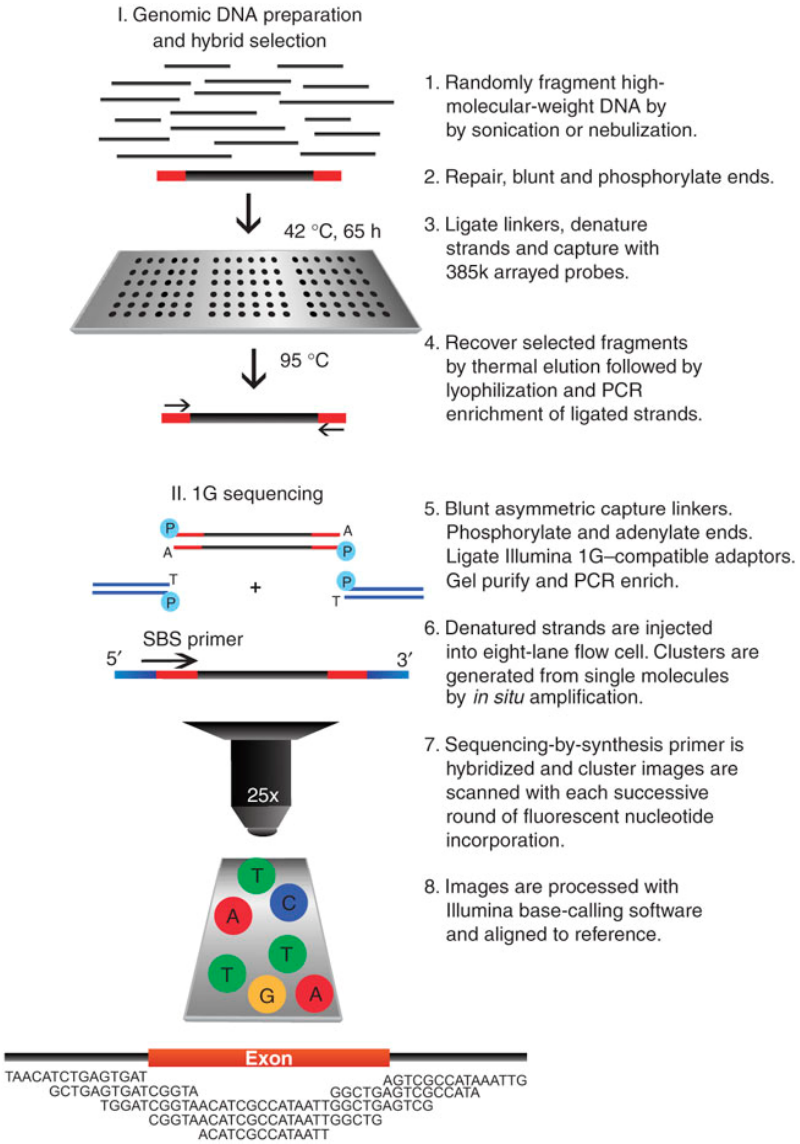


Genome Re-Sequencing

Genome Mapping
Single Nucleotide Polymorphism (SNP)
Deletion-Insertion Polymorphisms (DIP)

P. Tang (鄧致剛); RRC. Gan (甘瑞麒)
Bioinformatics Center, Chang Gung University.

Human Exome Sequencing



1. Randomly fragment high-molecular-weight DNA by by sonication or nebulization.
2. Repair, blunt and phosphorylate ends.
3. Ligate linkers, denature strands and capture with 385k arrayed probes.
4. Recover selected fragments by thermal elution followed by lyophilization and PCR enrichment of ligated strands.
5. Blunt asymmetric capture linkers. Phosphorylate and adenylate ends. Ligate Illumina 1G–compatible adaptors. Gel purify and PCR enrich.
6. Denatured strands are injected into eight-lane flow cell. Clusters are generated from single molecules by *in situ* amplification.
7. Sequencing-by-synthesis primer is hybridized and cluster images are scanned with each successive round of fluorescent nucleotide incorporation.
8. Images are processed with Illumina base-calling software and aligned to reference.

CSH PRESS GENOME RESEARCH

SNP detection for massively parallel whole-genome resequencing

Ruiqiang Li,[1,2,3] Yingrui Li,[1,3] Xiaodong Fang,[1] Huanming Yang,[1] Jian Wang,[1] Karsten Kristiansen,[1,2] and Jun Wang[1,2,4]

[1]*Beijing Genomics Institute at Shenzhen, Shenzhen 518000, China;* [2]*Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M DK-5230, Denmark*

Next-generation massively parallel sequencing technologies provide ultrahigh throughput at two orders of magnitude lower unit cost than capillary Sanger sequencing technology. One of the key applications of next-generation sequencing is studying genetic variation between individuals using whole-genome or target region resequencing. Here, we have developed a consensus-calling and SNP-detection method for sequencing-by-synthesis Illumina Genome Analyzer technology. We designed this method by carefully considering the data quality, alignment, and experimental errors common to this technology. All of this information was integrated into a single quality score for each base under Bayesian theory to measure the accuracy of consensus calling. We tested this methodology using a large-scale human resequencing data set of 36× coverage and assembled a high-quality nonrepetitive consensus sequence for 92.25% of the diploid autosomes and 88.07% of the haploid X chromosome. Comparison of the consensus sequence with Illumina human 1M BeadChip genotyped alleles from the same DNA sample showed that 98.6% of the 37,933 genotyped alleles on the X chromosome and 98% of 999,981 genotyped alleles on autosomes were covered at 99.97% and 99.84% consistency, respectively. At a low sequencing depth, we used prior probability of dbSNP alleles and were able to improve coverage of the dbSNP sites significantly as compared to that obtained using a nonimputation model. Our analyses demonstrate that our method has a very low false call rate at any sequencing depth and excellent genome coverage at a high sequencing depth.

[SOAPsnp is freely available from http://soap.genomics.org.cn under GPL license. The raw sequence data used in this report have been deposited in the EBI/NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) under accession no. ERA000005, and the SNP set has been deposited in dbSNP (release 130). These data are also available at http://yh.genomics.org.cn.]

nature International weekly journal of science

Letter

Nature **461**, 272-276 (10 September 2009) | doi:10.1038/nature08250; Receiv 2009; Published online 16 August 2009

Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[4], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

1. Department of Genome Sciences,
2. Department of Pediatrics, University of Washington,
3. Howard Hughes Medical Institute, Seattle, Washington 98195, USA
4. Agilent Technologies, Santa Clara, California 95051, USA

nature genetics

NATURE GENETICS | LETTER

Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants

Yingrui Li, Nicolas Vinckenbosch, Geng Tian, Emilia Huerta-Sanchez, Tao Jiang, Hui Jiang, Anders Albrechtsen, Gitte Andersen, Hongzhi Cao, Thorfinn Korneliussen, Niels Grarup, Yiran Guo, Ines Hellman, Xin Jin, Qibin Li, Jiangtao Liu, Xiao Liu, Thomas Sparsø, Meifang Tang, Honglong Wu, Renhua Wu, Chang Yu, Hancheng Zheng, Arne Astrup, Lars Bolund + ***et al.***

Affiliations | **Contributions** | **Corresponding authors**

Nature Genetics **42**, 969–972 (2010) | doi:10.1038/ng.680

nature genetics

Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng, Kati J Buckingham, Choli Lee, Abigail W Bigham, Holly K Tabor, Karin M Dent, Chad D Huff, Paul T Shannon, Ethylin Wang Jabs, Deborah A Nickerson, Jay Shendure & Michael J Bamshad

Affiliations | **Contributions** | **Corresponding authors**

Nature Genetics **42**, 30–35 (2010) | doi:10.1038/ng.499

Abstract

Abstract • **Change history** • **Author information** • **Supplementary information**

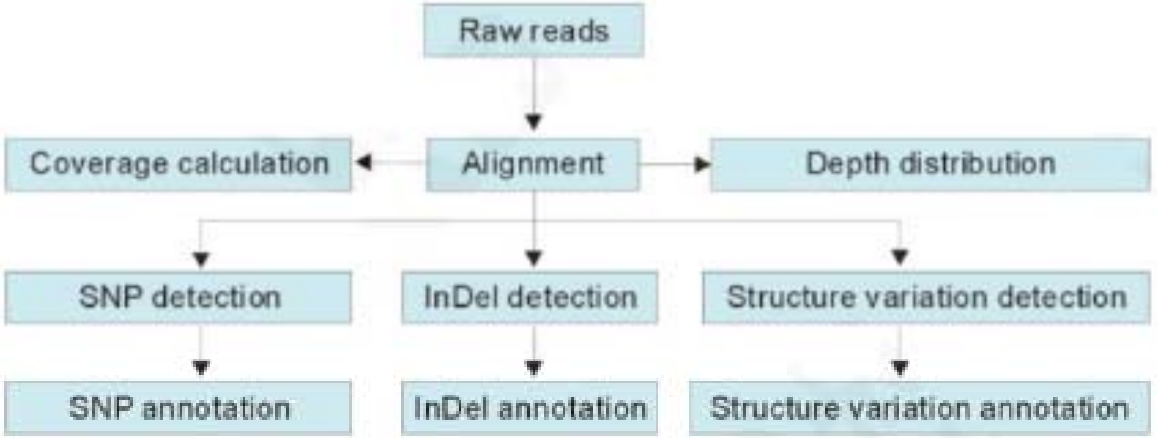
We demonstrate the first successful application of exome sequencing to discover the gene for a rare mendelian disorder of unknown cause, Miller syndrome (MIM% 263750). For four affected individuals in three independent kindreds, we captured and sequenced coding regions to a mean coverage of 40× and sufficient depth to call variants at ~97% of each targeted exome. Filtering against public SNP databases and eight HapMap exomes for genes with two previously unknown variants in each of the four individuals identified a single candidate gene, *DHODH*, which encodes a key enzyme in the pyrimidine *de novo* biosynthesis pathway. Sanger sequencing confirmed the presence of *DHODH* mutations in three additional families with Miller syndrome. Exome sequencing of a small number of unrelated affected individuals is a powerful, efficient strategy for identifying the genes underlying rare mendelian disorders and will likely transform the genetic analysis of monogenic traits.

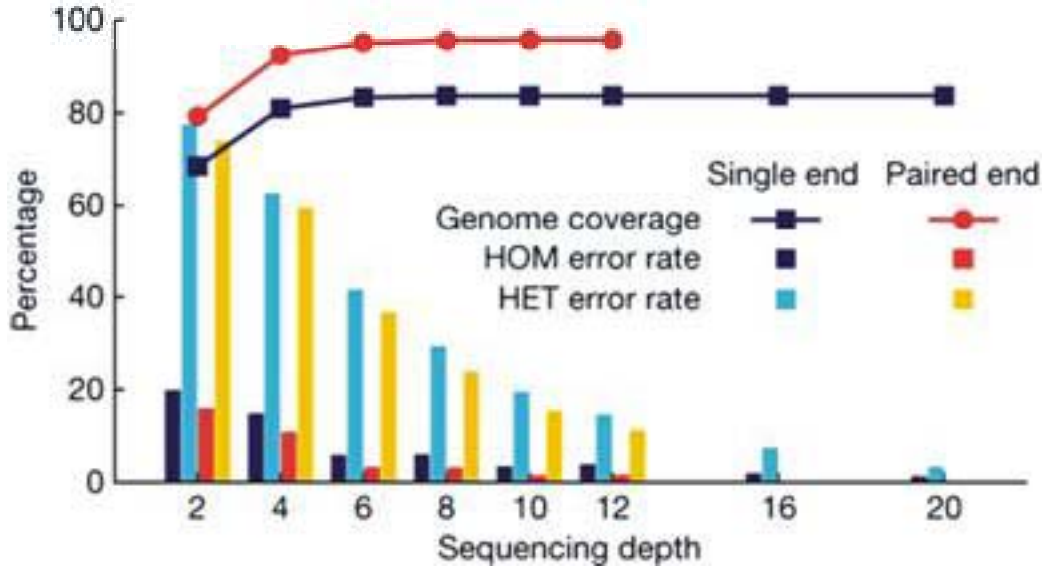
Bioinformatics of Genome Re-sequencing

Human whole genome re-sequencing

Human exome sequencing

Bacteria genome re-sequencing







Resequencing: Reference assembly followed by SNP and DIP detection

Analyzing a E. coli re-sequencing data set from 454 which contain more than 400,000 reads.