Phasing Analysis Service for Human WGS

Illumina Genome Network delivers comprehensive view of human genomic complexity.

Highlights

- **Inclusive View of Variation**
  Enables haplotyping and allele-specific analysis by identifying linkage between variants
- **Detailed Allele and Chromosomal Information**
  Supports studies in population reconstruction, evolutionary history, and genetic disease research
- **Rapid and Cost-Effective Genome Phasing**
  Delivers whole-genome sequencing and phasing within 12 weeks
- **Easily Interpretable Data**
  Data output in universal formats for seamless integration with analysis software

Phase Information Expands Insight into Human Genomes

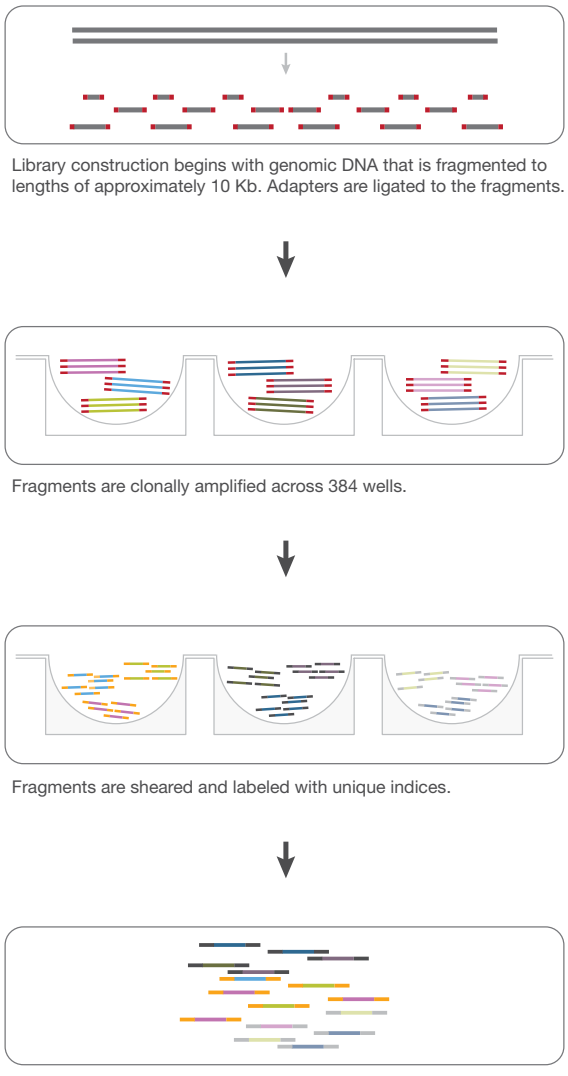
Advances in whole-genome sequencing (WGS) have enabled efficient characterization of the human genome. With the routine assembly and analysis of complex genomes, the importance of the relationship between genotype and phenotype has become more apparent. Recent findings suggest that this relationship can be better understood with phase information, which refers to the unique content of the two homologous chromosomes in diploid organisms[1]. Phasing technology addresses the unique distribution of variants across two chromosomes by spanning more than one heterozygous variant, enabling the generation of haplotype fragments. Examination of the unique variants on each chromosome contributes to greater understanding of the gene functions that influence specific phenotypes.

The Phasing Analysis Service provided by Illumina FastTrack Services through the Illumina Genome Network (IGN) assigns haplotype information to homologous chromosomes, empowering analysis with greater insight into the human genome (Figure 1).

Leveraging Proven Technologies

Powered by TruSeq® and Nextera® technologies, the Phasing Analysis Service delivers high-quality sequencing data generated using HiSeq® instruments (Figure 2). This cost-effective phasing solution is available as a complement to the Whole-Genome Analysis service, delivering results within 12 weeks, depending on laboratory queue. The unprecedented combination of long fragments (up to 10 Kb) and short-read accuracy enables complete human genome phasing, providing a more comprehensive view of genomic content.

Figure 1: Genome Sequencing and Phasing Workflow



Library construction begins with genomic DNA that is fragmented to lengths of approximately 10 Kb. Adapters are ligated to the fragments.

Fragments are clonally amplified across 384 wells.

Fragments are sheared and labeled with unique indices.

Fragments are sequenced using Illumina technology. Short read sequences are assembled into long sequence fragments indicating variants of known linkage.

The Phasing Analysis Service provides whole-genome sequencing and phasing, enabling the generation of haplotype sequences. This process identifies variant linkage and chromosome-specific variation for comprehensive genomic analysis.

Establishing Variant Linkage

When performing whole-genome phasing, Illumina's proprietary phasing algorithm first builds short sequence reads into the originally targeted long fragments. Next, it uses overlapping heterozygous regions to create long, molecularly phased sequence segments. The molecularly phased segments are then linked to each other in tandem and extended by leveraging known population information from the 1000 Genomes database[2]. By leveraging data from the 1000 Genomes database, molecularly phased segments can be extended to an N50 value of up to 500 Kb for samples from well-represented populations. This method provides phase information for approximately 95–98% of single nucleotide polymorphisms (SNPs) in a human genome.

Ensuring Accuracy in Gene Mapping Studies

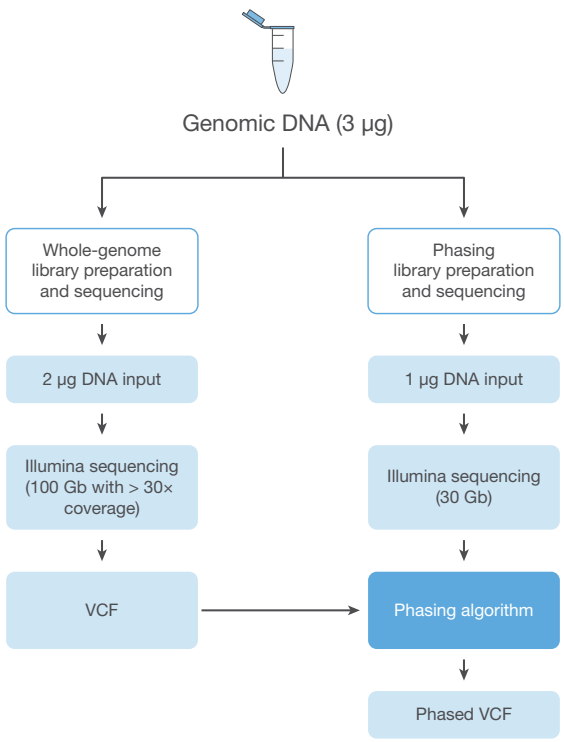
By capturing gene information from homologous chromosomes, phasing technology eliminates the traditional reliance on haplotype inference based solely on statistical information, which can be subject to error. Other traditional phasing methods include trio studies, which compare maternal and paternal sequences to the genome under study and consequently do not resolve the haplotypes of *de novo* variation. The Phasing Analysis Service eliminates the need to compare these samples, instead enabling thorough phasing analysis solely from the sample under study. The haplotype information provided by this service supports population and evolutionary studies as well as genetic disease research.

Enabling Comprehensive Analysis of Complex Traits

Because genome phasing technology provides haplotype information about each individual genome, it enables allele-specific and variant linkage analysis (Figure 3). This knowledge can inform studies of complex trait susceptibility, as allelic interactions among multiple genes influence these traits.

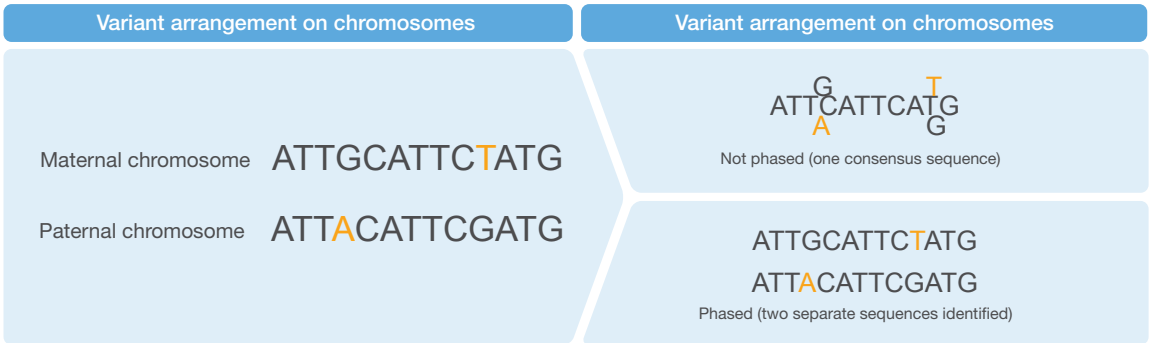
Haplotyping can also provide valuable information for genetic disease research, as disruptions to alleles in *cis* or *trans* positions on a chromosome or chromosomes[3] can cause some genetic disorders. The prevalence of compound heterozygosity—the presence of two different recessive alleles at a particular locus—in genetic diseases[1] supports the importance of phase information for relating genotype to phenotype. This distinction between variants from different chromosomes provides a valuable resource for investigating genetic causes of disease and their influence on disease phenotypes.

Figure 2: Phasing Analysis Service Workflow



The Phasing Analysis Service performs whole-genome sequencing and phasing from 3 µg of human genomic DNA. Sequencing and phasing are performed in parallel, and the sequence results are compared using proprietary algorithms to create molecularly phased sequence segments.

Figure 3: Phasing Resolves Haplotype Information



Genome phasing enables the generation of haplotype sequences, which indicate linkage between variants and can inform genetic disease studies.

Widely Compatible Data for Downstream Analysis

Data sets are provided in variant call format (VCF) for streamlined analysis, and include confidence calls for variant linkage. The Phasing Analysis Service also delivers filtered, demultiplexed reads in addition to the phased genome for further analysis (Table 1). These data sets are compatible with various downstream tools for subsequent analysis and visualization.

Phasing Analysis Service without Compromise

Powered by proven technology and comprehensive informatics, the Phasing Analysis Service delivers complete genomic phase information, enabling detailed and thorough analysis of human genomes. With industry-leading turnaround times, this cost-effective service eliminates the need to purchase additional instrumentation while providing an inclusive haplotyping solution. The Phasing Analysis Service provides researchers with the confidence to take the next step in their research. Learn more about this service at www.illumina.com/services.

References

1. Tewhey R, Bansal V, Torkamani A, Topol EJ, Schork NJ (2001) The importance of phase information for human genomics. Nat Rev Genet 12: 215–223.
2. www.1000genomes.org
3. Kleinjan DJ, Coutinho P (2009) *Cis*-ruption mechanisms: disruption of *cis*-regulatory control as a cause of human genetic disease. Brief Funct Genomic Proteomic 8: 317–332.

Table 1: IGN Phasing Analysis Service Details

Sample Input Requirements
3 µg genomic DNA, inclusive of WGS DNA requirements
Data Deliverables
Whole-Genome Sequence (VCF)
Single nucleotide polymorphisms (SNPs)
Indels (1–50 bp)
Copy number variations
Large deletions
Large insertions
Phased Genome Sequence
Variant linkage and confidence scores (VCF)
Filtered and demultiplexed short reads (FASTQ)
Additional Deliverables
Whole-genome concordance with genotyping arrays
Summary reports (whole-genome sequencing and phasing)

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 770-2013-024 Current as of 19 November 2014

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA