# illumina®

## Phasing Analysis Service for Whole Human Genome Sequencing

Delivering whole-genome phase information for a comprehensive view of genomic complexity.

Advances in genome sequencing technologies have enabled efficient characterization of the human genome. With these advances, the importance of the relationship between genotype and phenotype has become more apparent. Recent findings suggest that this complex relationship can be better understood with phase information<sup>1</sup>. Genome phasing provides haplotype information about a given genome, distinguishing between alleles on homologous chromosomes. The Phasing Analysis Service, provided by Illumina FastTrack Services through the Illumina Genome Network, delivers complete genomic phase information to enable comprehensive analysis of the human genome.

## What is genome phasing and why is it important?

Phase information refers to the unique content of the two homologous chromosomes in diploid organisms (or multiple chromosomes in polyploid organisms). Phasing technology enables the generation of haplotype fragments, so that maternally and paternally inherited alleles can be distinguished from one another. By capturing gene information specific to each chromosome, phase information provides insight into allele-specific variation, which can inform studies in evolutionary history and genetic disease research<sup>2</sup>.

### How does phasing work?

Phasing technology addresses the unique distribution of variants across two chromosomes by spanning more than one heterozygous variant, enabling the generation of haplotype fragments. First, genomic DNA is fragmented into lengths of approximately 10 kilobases (Kb) each. These fragments are then clonally amplified, sheared, and labeled with a unique index. Fragments are then sequenced, and the short read sequences generated from each molecule are assembled separately. This process results in long sequence fragments indicating variants of known linkage.

## What does the Phasing Analysis Service include?

The cost-effective Phasing Analysis Service is available through Illumina FastTrack Services as an add-on service to Whole-Genome Analysis. The Phasing Analysis Service includes preparation of each DNA sample and sequencing on one lane of a HiSeq<sup>®</sup> system, generating approximately 30 Gb of sequence data. Raw phasing data and whole-genome sequencing data (in variant call format) from each sample are then processed using Illumina's proprietary software.

## How does Illumina's phasing algorithm work?

First, the phasing algorithm 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<sup>3</sup>.

### How does the Phasing Analysis Service provide phased data?

Phased genomic data is delivered in variant call format (VCF), which indicates linkage between identified variants and the confidence scores of identified variant linkage. Data sets also include quality-filtered and demultiplexed short-read sequence data in FASTQ format, as well as phasing performance statistics in a summary report PDF. Phasing data and the summary report are provided in the BaseSpace<sup>®</sup> analysis environment<sup>4</sup>, Illumina's cloud computing environment for next-generation sequencing data. Short-read sequence data are delivered on an external hard drive.

## If I have already sequenced my sample of interest, can I still phase the genome using the Phasing Analysis Service?

The Phasing Analysis Service is available without Whole-Genome Analysis for whole-genome sequencing projects completed by Illumina's FastTrack Services laboratory in 2013. Contact your local account manager or sales representative for more information.

## How much input DNA is required?

This service requires a total of 4 µg of human genomic DNA for whole-genome sequencing and phasing. If a sample was recently sequenced using Illumina Genome Network services, only 1 µg of additional DNA is required for phasing.

#### Which sample types can I submit for whole-genome phasing?

The Illumina FastTrack Services laboratory performs whole-genome phasing with unamplified DNA samples extracted from blood, saliva, or cell lines.

#### What are the requirements for sample submission?

At least 130 µL of unamplified DNA at a minimum concentration of 30 ng/µL must be submitted for whole-genome sequencing with phasing projects. DNA must be submitted in 10 mM Tris/1 mM EDTA solution buffer.

#### What is the sample submission process?

Each project is coordinated by a Project Manager in the Illumina FastTrack Services laboratory. A Project Manager will contact you within three days after your order is placed to coordinate the logistics of sample shipment. Your Project Manager will serve as your point of contact throughout the process.

#### When will I receive the genome phasing results?

The Illumina FastTrack Services laboratory is committed to providing industry-leading turnaround times. Whole-genome sequencing and phasing results are provided within twelve weeks of sample receipt.

#### Does the Phasing Analysis Service phase all SNPs present in a genome?

Depending on the sample type and quality, as well as diversity characteristics of each genome, this service phases approximately 95–98% of the SNPs in a given genome. A phased data set generated using a human trio sample is available in the BaseSpace analysis environment.

#### Can I phase non-human species using the Phasing Analysis Service?

No, the algorithm and process for this service are specific to human genome sequencing and phasing.

#### What is the price of this service?

Contact your local account manager or sales representative to receive a pricing quote.

#### How can I learn more about the Phasing Analysis Service?

Learn more about this service and the Illumina Genome Network at www.illumina.com/services or www.illumina.com/longreadsequencing.

#### References

- 1. Tewhey R, Bansal V, Torkamani A, Topol EJ, Schork NJ (2011) The importance of phase information for human genomics. Nat Rev Genet 12: 215–23.
- 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: 317–32.
- 3. www.1000genomes.org
- 4. www.illumina.com/basespace

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

#### FOR RESEARCH USE ONLY

© 2013 Illumina, Inc. All rights reserved. Illumina, illumina/Dx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIlx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 770-2013-026 Current as of 17 July 2013

