PARKINSON'S PROGRESSION MARKERS INITIATIVE

Play a Part in Parkinson's Research

Exome-sequencing of PPMI samples

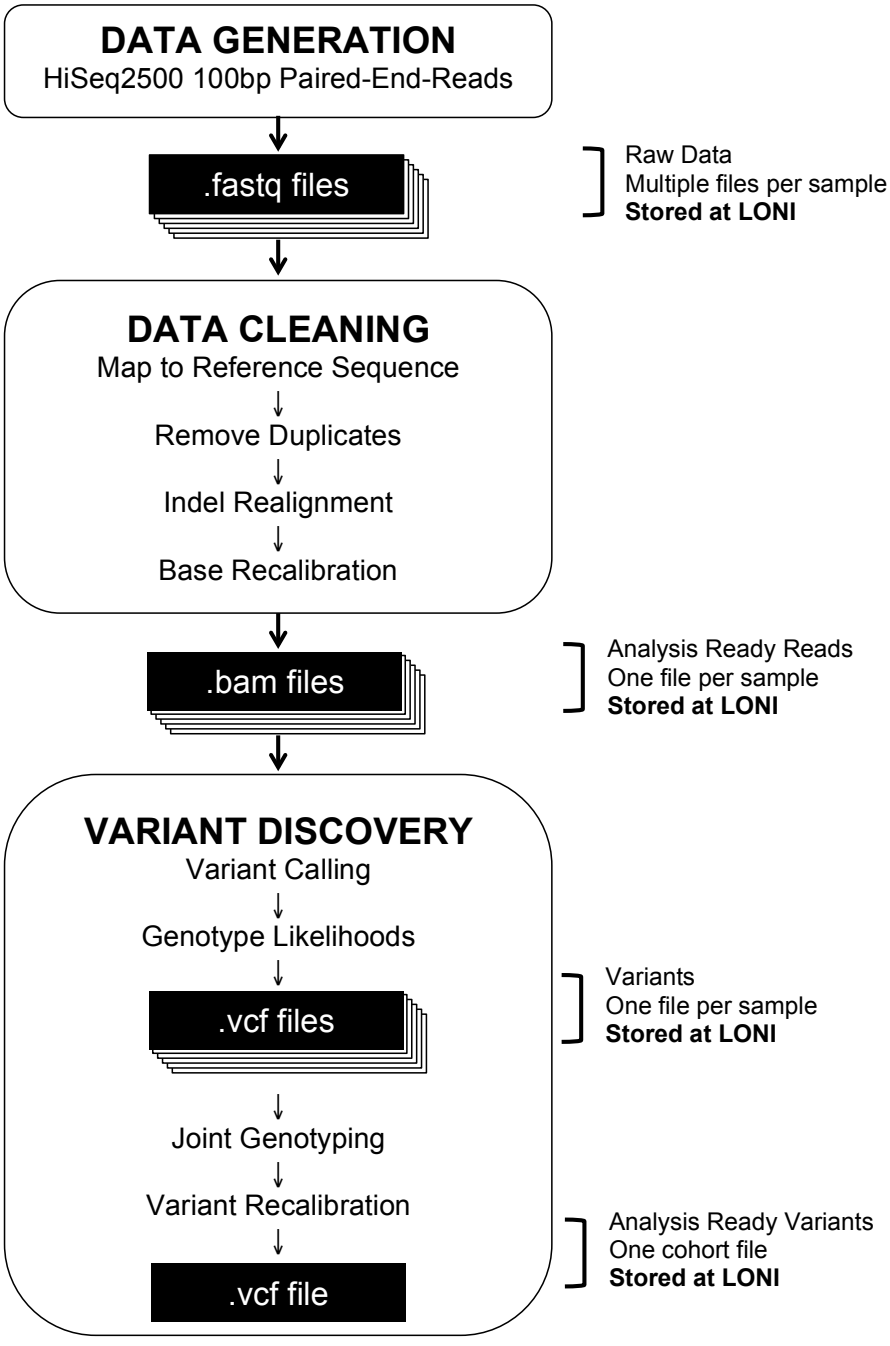
Dena G Hernandez,
J Raphael Gibbs
National Institute on Aging,
Laboratory of Neurogenetics

Summary

Exome sequencing was performed on whole-blood extracted DNA samples collected according to the PPMI Research Biomarkers Laboratory Manual using Illumina Nextera Rapid Capture Expanded Exome Kit. Nextera Expanded Exome targets 201,121 Exons, UTRs and miRNA and covers 95.3% of Refseq exome. >340,000 probes are constructed against the human NCBI37/hg19 reference genome. Targeted genomic footprint is 62Mb.

Methods

Library preparation for next-generation sequencing using Nextera Rapid Capture Expanded Exome Kit was performed per manufacturer's protocol (Illumina, Inc. San Diego). Exome-enriched libraries (multiplexed sets of 12 samples) were sequenced on the Illumina HiSeq 2500 sequencing platform using 2 x 100 bp paired-end read cycles. An overview of the data processing and available files is shown in the figure below. Briefly, paired-end sequence reads (fastq files) were aligned using BWA [1] against the reference human genome (UCSC hg19). Duplicate read removal; format conversion and indexing were performed with Picard (picard.sourceforge.net/index.shtml). The Genome Analysis Toolkit (GATK) [2–4] was used to recalibrate base quality scores and perform local re-alignments around indels for the aligned sequencing reads (per subject bam file). Variant calling and genotype likelihoods were generated per subject using the GATK HaplotypeCaller (per subject genomic vcf file). GATK CombineGVCFs and GenotypeGVCFs were used to perform joint genotyping for the cohort from the set of per subject genomic vcf files. Variant filtering was then applied using the GATK Variant Quality Score Recalibration tools (cohort vcf file). Subject quality control was performed based on variant call-rate, heterzygosity rate, gender check, relatedness/duplicates and population outliers using PLINK [5].



References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009 Jul 15;25(14):1754–1760. PMID: 19451168

2. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491–498. PMID: 21478889

3. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–1303. PMID: 20644199

4. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al. 2013 Oct 15;11(1110):11.10.1–11.10.33. PMCID: PMC4243306

5. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. PMCID: PMC4342193

About the Authors

This document was prepared by Dena G Hernandez and J. Raphael Gibbs, National Institute on Aging, Laboratory of Neurogenetics. For more information please contact Dena Hernandez by email at hernand@mail.nih.gov or Raphael Gibbs gibbsr@mail.nih.gov.

Notice: This document is presented by the author(s) as a service to PPMI data users. However, users should be aware that no formal review process has vetted this document and that PPMI cannot guarantee the accuracy or utility of this document.