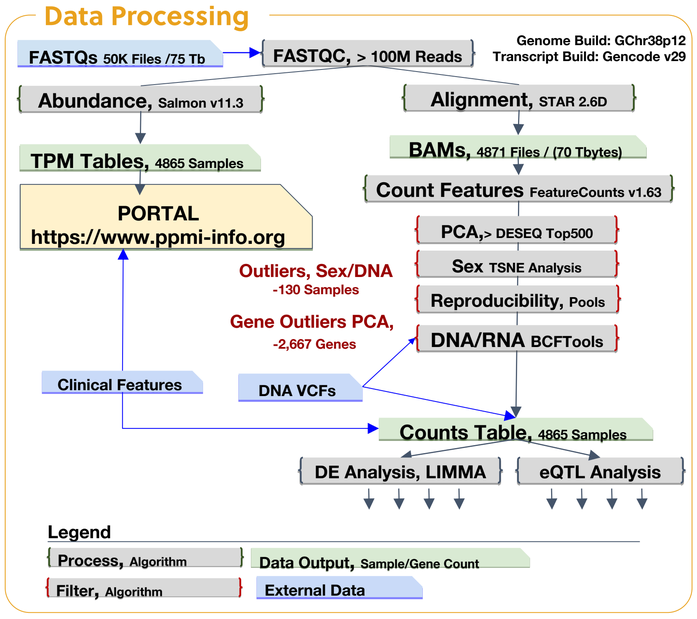
Longitudinal Analysis of 4500 whole blood transcriptomes across over 1500 whole-genome sequenced individuals within the Parkinson's Progression Markers Initiative

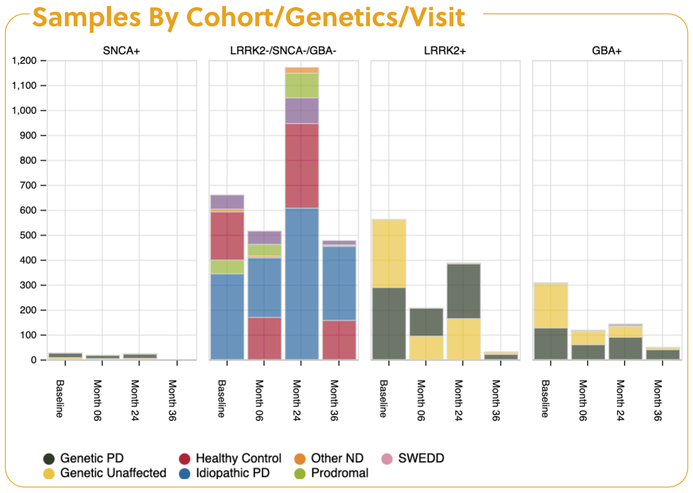
David Craig[1] , Ivo Violich[1], Elizabeth Hutchins[2], Eric Alsop[2], Cornelis Blauwendraat[4], Shawn Levy[3], Andy Singleton[4], Mark R Cookson[4], Raphael J. Gibbs[4], Kendall Van Keuren-Jensen[2]

[1]Department of Translational Genomics, University of Southern California [2]Neurogenomics, TGen [3]Genomic Services Laboratory, HudsonAlpha Institute for Biotechnology [4]Laboratory of Neurogenetics, National Institutes of Aging

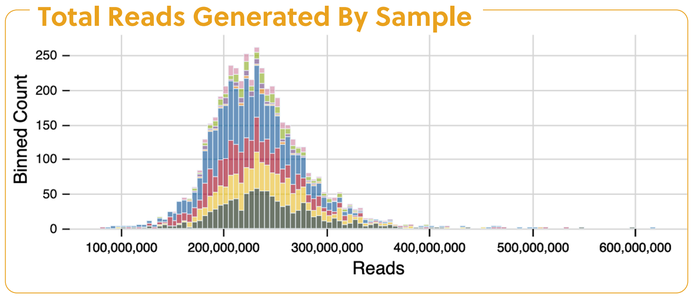
Data Processing



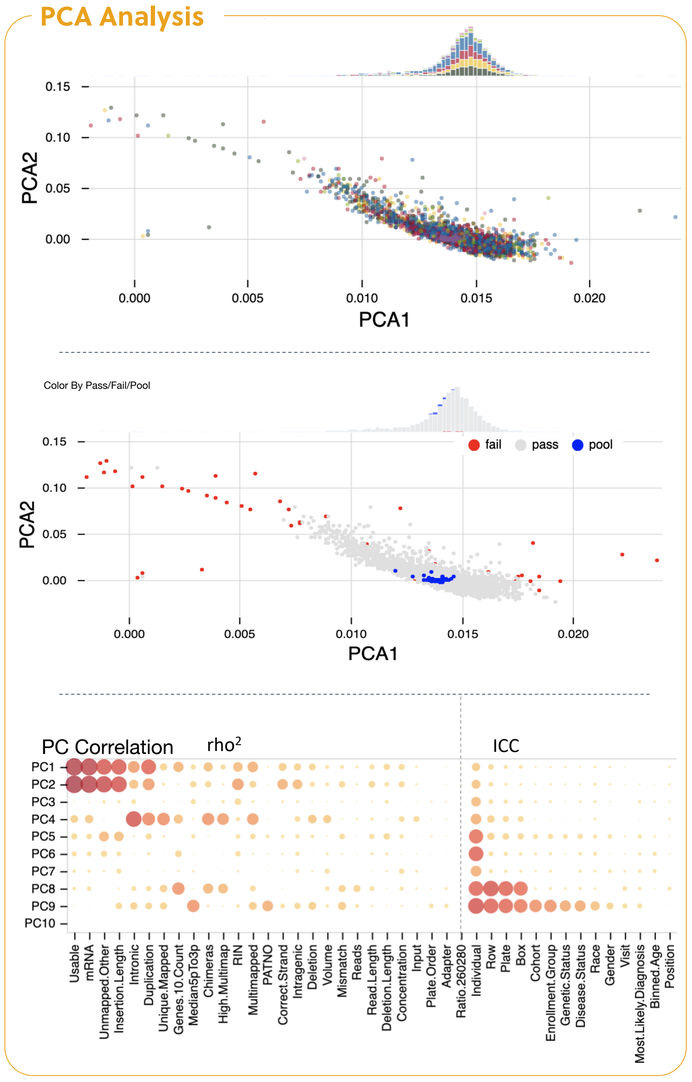
Samples By Cohort/Genetics/Visit



Total Reads Generated By Sample



PCA Analysis



Abstract

Transcriptome analyses of whole blood samples from normal healthy control subjects and patients with Parkinson's disease (PD) have the potential to uncover biological pathways disrupted by disease processes. Detectable changes in the transcriptome of a readily accessible biofluid, such as blood, have the potential to be valuable biomarkers. Moreover, integration of transcriptomic data with genomic whole-genome sequencing data has the potential to give insight to variants of unknown significance and previously identified genetic loci. In this study, we report transcriptome profiling of whole blood from 4,756 samples longitudinally collected from over 1,570 individuals who also have whole genome DNA sequence, as part of Michael J. Fox Foundation's Parkinson's Progression Markers Initiative (PPMI) cohort. Paxgene derived whole-blood samples were sequenced at baseline and at months 6, 12, 24, and 36 to a depth of over 200 million reads, with many individuals having over 1 billion reads over the time series. Longitudinal analysis was conducted using across a series of clinical variables, examining clinical phenotypes, carrier status, and therapeutic drug usage. We analyzed expression data with existing whole-genome sequencing data, reporting on the robustness of expression quantitative loci (eQTL). Finally, we describe a series of resources to enable other researchers, including a series of tools for querying and searching through the clinical/genomics data via an analysis portal.

Background

Parkinson's disease (PD) is the second most common neurodegenerative disease in the world. While several known factors contribute to PD and PD risk, mechanisms underlying pathology and disease progression are poorly understood. In order to increase our knowledge of gene expression changes associated with the disease, to provide insights about underlying pathology and potential targets for therapeutic development, whole blood samples were collected and sequenced.

PPMI is a longitudinal study with multiple study arms collecting both specimens and clinical data over time. Samples were obtained at different timepoints, predominately at Baseline, Month 6, Month 12, Month 24, and Month 36. The Sankey plot below provides an overview of patients at each timepoint. Many patients have 5 time points, and RNA-seq data is available at each time point. With 200 million reads per sample, some patients have over 1 billion reads over the 3 year course.

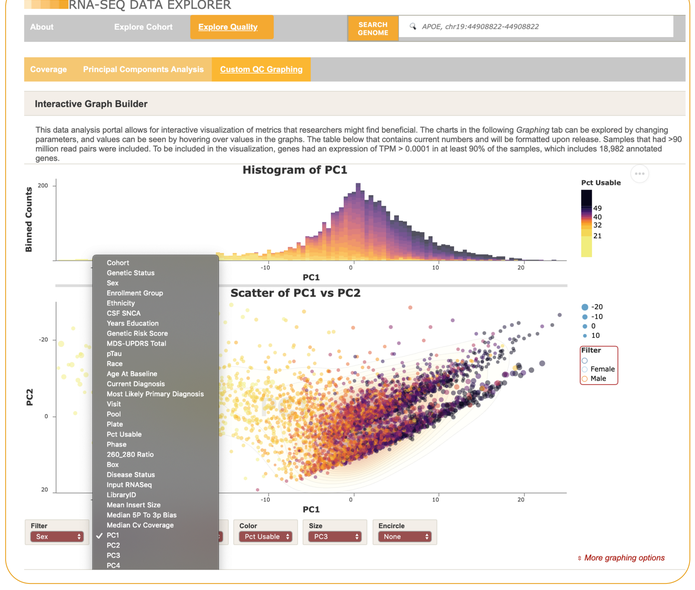
Resource

By many metrics, this is one of the largest and most significant released set of RNA-seq of disease. The depth of sequencing (~240M reads) and breadth of species (coding and non-coding RNA) is also significantly higher than most other studies. Beyond PD, there is a significant utility for studies and researchers seeking to determine which genes are accessible in the blood.

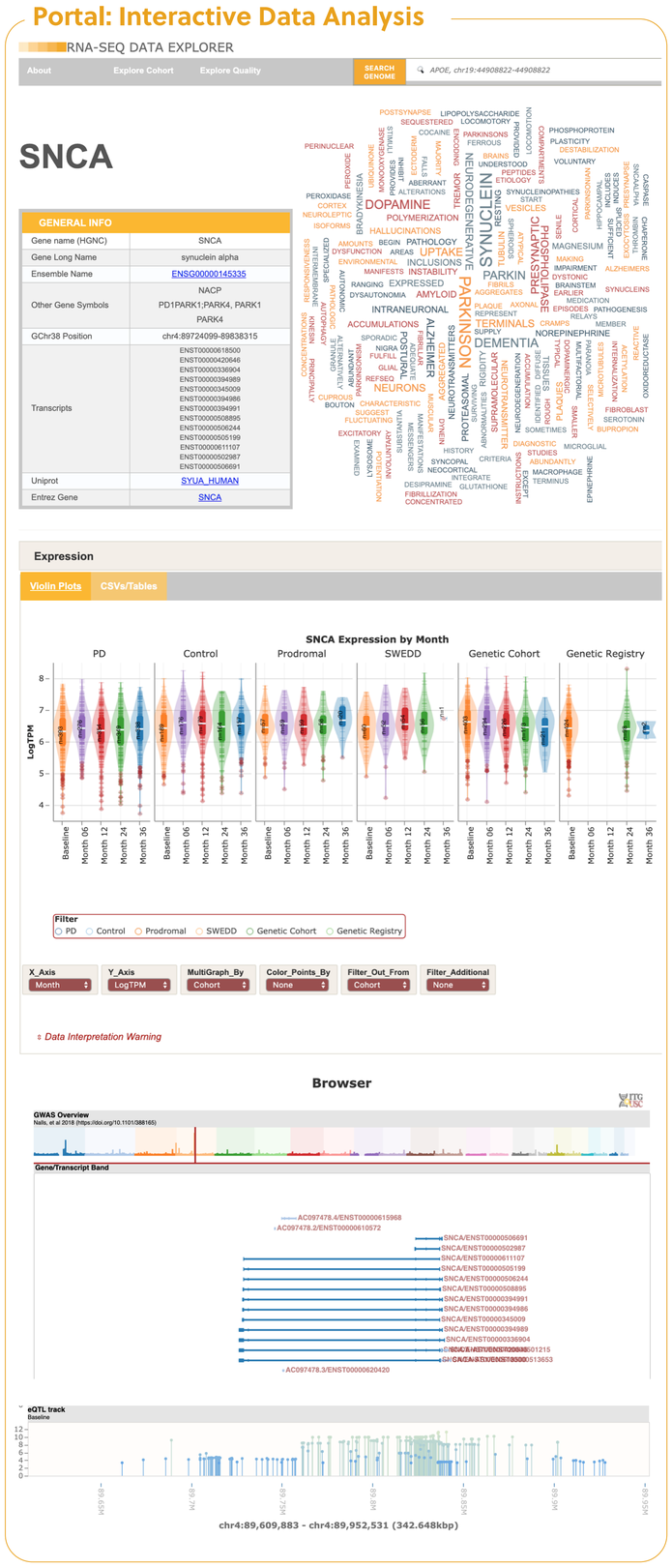
Samples: 4,756 **Individuals**: 1,570
Mean Reads Per Individual: 228 Million **Read Pairs Per Individual**: 114 Million
Total Reads Sequenced: 558 Billion **Total Bases Sequenced**: 111 Trillion
Total Disk Footprint: 108 Tbytes **CPU Core Hours**: 480 Thousand

A web portal provides a space for researchers to examine the data interactively. The raw data are also available for download through www.ppmi-info.org as TPMs, Counts or FASTQs/BAMs. Additional Data Available: WGS, Imaging, Clinical Correlates.

Portal: Interactive



Portal: Interactive Data Analysis



Funding

THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH

tgen AN AFFILIATE OF City of Hope

USC Institute Of Translational Genomics Keck Medicine of USC

HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY

NIH NIA LNG LABORATORY OF NEUROGENETICS