

Complete Genomics 69 Public Genome Data Set

Overview

Complete Genomics has released a set of public genome sequences on its FTP server (<ftp2.completegenomics.com>). These samples are from the CEPH/Utah Pedigree 1463 NIGMS Repository and from the NHGRI Repository (HapMap and 1000 Genomes Project) which represent nine different populations (Diversity). Both collections were drawn from the Coriell Institute for Medical Research. These data are also available on a mirror site that is part of the Bionimbus Cloud (<http://www.bionimbus.org/>).

By providing the research community with this genome data set, we encourage researchers to validate our performance and to further improve data analysis and interpretation methods.

For more information about this dataset including download and un-archiving instructions, please refer to the README document and the Quick Start Guide: Working with Complete Genomics Data available on our website (<http://www.completegenomics.com/sequence-data/download-data/>). For additional documentation and support, please refer to Complete Genomics Customer Support (support@completegenomics.com).

For each publicly available genome, Complete Genomics provides a master variation (masterVar) file. The masterVar file is an ideal starting point to get an overview of the types of information that Complete Genomics provides. For each location in the genome, the following information is reported:

- Variant calls (SNPs, small insertions, deletions, and substitutions, and complex small variants)
- Gene annotations
- Non-coding RNA annotations
- Known segmental duplication annotations
- Repeat masker annotation
- Read support information for the variation
- CNV information

Table 1: List of 69 Genomes.

Diversity Set				Puerto Rican Trio		CEPH/Utah Pedigree 1463		Yoruban Trio	
Coriell ID	Population	Coriell ID	Population	Coriell ID	Population	Coriell ID	Population	Coriell ID	Population
NA19700	ASW	NA19017	LWK	HG00731	PUR	NA12877	CEU	NA19238	YRI
NA19701	ASW	NA19020	LWK	HG00732	PUR	NA12878	CEU	NA19239	YRI
NA19703	ASW	NA19025	LWK	HG00733	PUR	NA12879	CEU	NA19240	YRI
NA19704	ASW	NA19026	LWK			NA12880	CEU		
NA19834	ASW	NA21732	MKK			NA12881	CEU		
NA06985	CEU	NA21733	MKK			NA12882	CEU		
NA06994	CEU	NA21737	MKK			NA12883	CEU		
NA07357	CEU	NA21767	MKK			NA12884	CEU		
NA10851	CEU	NA19648	MXL			NA12885	CEU		
NA12004	CEU	NA19649	MXL			NA12886	CEU		
NA18526	CHB	NA19669	MXL			NA12887	CEU		
NA18537	CHB	NA19670	MXL			NA12888	CEU		
NA18555	CHB	NA19735	MXL			NA12889	CEU		
NA18558	CHB	NA20502	TSI			NA12890	CEU		
NA20845	GIH	NA20509	TSI			NA12891	CEU		
NA20846	GIH	NA20510	TSI			NA12892	CEU		
NA20847	GIH	NA20511	TSI			NA12893	CEU		
NA20850	GIH	NA18501	YRI						
NA18940	JPT	NA18502	YRI						
NA18942	JPT	NA18504	YRI						
NA18947	JPT	NA18505	YRI						
NA18956	JPT	NA18508	YRI						
		NA18517	YRI						
		NA19129	YRI						