



**For more information, contact:**

Complete Genomics Inc.  
Jennifer Turcotte  
Vice President of Marketing  
(650) 943-2846  
[jturcotte@completegenomics.com](mailto:jturcotte@completegenomics.com)

Waggener Edstrom Worldwide  
Healthcare Practice  
Lisa Osborne  
Account Director  
(202) 261-7806  
[lisao@waggeneredstrom.com](mailto:lisao@waggeneredstrom.com)

**Complete Genomics Is Publicly Releasing a Large Sequencing Dataset — 60 Complete, High-Coverage Human Genomes — for Study by the Global Research Community**

*Samples come from NIGMS and NHGRI repositories and include three-generation CEPH/Utah Pedigree 1463*

**MOUNTAIN VIEW, Calif. – Feb. 3, 2011** – Complete Genomics Inc. (NASDAQ: GNOM), a complete human genome sequencing company that has developed and commercialized an innovative DNA sequencing platform, announced today that it is providing the research community with access to 60 complete, high-coverage human genome sequences. The company introduced its new public genomic repository at the annual Advances in Genome Biology and Technology (AGBT) meeting in Marco Island, Fla. These genomes have on average more than 55x mapped read coverage, and the sequencing of these 60 genomes generated more than 12.2 terabases (Tb) of total mapped reads. This dataset will complement other publicly available whole genome data sets, such as the 1000 Genomes Project's recent publication of six high-coverage and 179 low-coverage human genomes. Complete Genomics has data for 40 genomes currently available for download from its corporate website, and the remaining 20 genomes will be released by the end of March 2011.

To date, Complete Genomics has sequenced and analyzed more than 1,000 high-coverage genomes for its customers, generating more than 230 Tb of mapped reads.



The current capacity of its commercial operations is more than 400 complete genomes per month. Further, the company is working toward additional expansion of its genome sequencing capacity in the coming months.

The 40 genomes that have been analyzed and are being released now continue to display the high quality of Complete Genomics' sequencing results. On average, 97 percent of each genome and 96 percent of each exome is called with high confidence. On average, more than 98.6 percent of each genome had coverage of 10x or higher. Genome-wide single nucleotide polymorphisms (SNP) detection concordance with the high-quality Infinium subset of the International HapMap Project dataset averages 99.93 percent.

"We are building this extensive public genomic repository and are also providing the global research community with access to our downstream analysis tools as part of our ongoing commitment to make complete human genome analysis easier and more efficient," said Complete Genomics Chairman, President and CEO Dr. Clifford Reid. "Our ultimate goal is to enhance the research community's understanding of the basis, treatment and prevention of complex diseases."

The 60 genomes included in this public dataset were drawn from two resources housed at the Coriell Institute for Medical Research: the National Institute of General Medical Sciences (NIGMS) Human Genetic Repository and the NHGRI Sample Repository for Human Genetic Research. Included in this sample set is a 17-member, three-generation CEPH pedigree from the NIGMS Repository and ethnically diverse samples from the NHGRI Repository that represent nine different populations. The samples selected are unrelated, with the exception of the three-generation CEPH pedigree, a Yoruba trio and a Puerto Rican trio. The majority of these samples have been previously analyzed as part of the International HapMap Project or 1000 Genomes Project.

This data release builds upon the Yoruba trio dataset released by Complete Genomics on Jan. 6, 2011. The new 37-sample panel of genomes being released immediately includes 32 ethnically diverse samples from the NHGRI Repository and five Caucasian samples from the NIGMS Repository. Complete Genomics is posting its variant reports for each sample, which include SNPs, insertions/deletions, copy number variations and structural variations, together with the read alignments supporting those calls, coverage information and quality scores.

The 17-member CEPH pedigree dataset as well as three other genomes from the ethnically diverse panel of genomes will be publicly available by the end of March. Researchers will be able to use this new multi-generational dataset to develop family-based, inheritance or phasing analysis methods.

As terabyte-size genomic datasets are large enough to present challenges for the research community, the released data will also be available on a mirror site that is part of the Bionimbus Cloud. Bionimbus is a cloud-based system for managing, analyzing



and sharing genomic data. Bionimbus integrates technology for the high-performance analysis and efficient transport of large datasets over wide area networks.

“Scientists who are members of the Bionimbus Community can analyze the data without moving it, and researchers with access to high-performance research networks such as Internet2 or the National LambdaRail can download the data to do their own analysis,” said Robert Grossman, Ph.D., director of informatics at the Institute for Genomics and Systems Biology (IGSB) at the University of Chicago.

Kevin White, Ph.D., IGSB director and professor of human genetics at the University of Chicago, whose team was given early access to some of these data, said, “The Complete Genomics variant data we have examined so far benchmarks very favorably with the data from the NIH 1000 Genomes Project. Sensitivity and specificity rates for SNP calls appear to be greater than 99.9 percent and 99.8 percent, respectively, when comparing Complete Genomics data to high-confidence calls from the 1000 Genomes Project. The Complete Genomics platform seems to be a very competitive approach for generating whole genome data, which will enrich both gene identification efforts for Mendelian diseases and association studies between rare alleles and complex traits. We look forward to using these 60 genomes along with other public data in the Bionimbus Cloud to help mine datasets produced by IGSB and collaborators.”

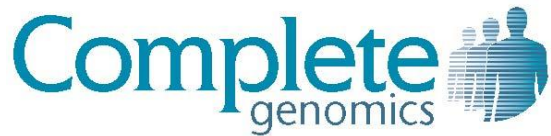
All available data can be accessed on the Complete Genomics website at <http://www.completegenomics.com/sequence-data/download-data> and at the Bionimbus mirror site at <http://www.bionimbus.org> under “public data.”

### **About Complete Genomics**

Complete Genomics is a complete human genome sequencing company that has developed and commercialized an innovative DNA sequencing platform. The Complete Genomics Analysis Platform (CGA™ Platform) combines Complete Genomics’ proprietary human genome sequencing technology with our advanced informatics and data management software. We offer this solution as an innovative, end-to-end, outsourced service, CGA™ Service, and provide customers with data that is immediately ready to be used for genome-based research. Additional information can be found at <http://www.completegenomics.com>.

### **Forward Looking Statements**

Certain statements in this press release, including statements relating to our expectations regarding the timing of the data release and public availability of complete genomes and data sets of complete genomes, as well as our expected monthly genome sequencing capacity and expectations regarding the timing of the expansion of our sequencing capacity, are forward looking statements that are subject to risks and uncertainties. Readers are cautioned that these forward looking statements are based on management’s current expectations, and actual results may differ materially from those projected. The following factors, without limitation, could cause actual results to



differ materially from those in the forward looking statements: our limited operating history, delays in production due to technical issues and our inability to increase yield. More information on potential factors that could affect our monthly genome sequencing capacity is included in our Securities and Exchange Commission filings and reports, including the risks identified under the section captioned "Risk Factors" in our Quarterly Report on Form 10-Q filed on December 22, 2010. We disclaim any obligation to update information contained in these forward looking statements, whether as a result of new information, future events or otherwise.