



## Fact Sheet - First Publication of an Individual Diploid Human Genome Sequence

### Results from *PLoS Biology* Publication

- The new diploid human genome published in *PLoS Biology* (called “HuRef”) represents the sequence and assembly of one individual—Dr. J. Craig Venter, in which his two sets of chromosomes (one inherited from his mother and one set from his father) are represented. It is this kind of genome sequencing and analysis that will usher in the true era of individualized medicine.
- Researchers at the JCVI have been sequencing and analyzing this version of Dr. Venter’s genome since 2003.
- Using whole genome shotgun sequencing and highly accurate DNA sequencing using Sanger-based chemistry, the team produced 32 million sequence reads or more than 20 billion base pairs of DNA from which they were able to assemble the majority of Dr. Venter’s genome.
- This diploid genome uniquely catalogues the contributions of each of the parental chromosomes, showing for the first time the amount of variation existing between the two.
- The HuRef analysis shows that human to human variation is 5 to 7 times greater compared to previous human genome analyses. This is between 15-30 million bases of sequence that is different between humans.
- There is far more non-Single Nucleotide Polymorphism (SNPs) genetic variation in each human than previously thought. The HuRef genome shows that an individual human has 4.1 million DNA variants and of these 22% are non-SNP, but the non-SNP variants account for 74% of all the variant DNA nucleotide bases.
- The HuRef analysis found 3.2 million SNPs and nearly one million non-SNP variants including insertion/deletions (“indels”), copy number variants, block substitutions and segmental duplications. There were also 1.2 million never before seen variants found in Dr. Venter’s genome.
- There are over 300 disease genes and over 4,000 genes overall exhibit different protein forms. With further study the JCVI team can begin to determine how these altered proteins impact all aspects of Dr. Venter’s health status.
- Another important feature that is made possible by having an individual, diploid genome is the ability to generate more informed haplotype assemblies. Haplotypes are groups of linked variations along the chromosomes. Other studies have generated many common haplotypes; however these are based on averages of large populations rather than individuals. Individual haplotypes enable scientists to study rare or ‘private’ variants that might explain and help predict traits and diseases in that particular person—allowing an individualized approach in genomic applications.

- In the HuRef analysis, the team used the heterozygous portion of the 4.1 million variant set and new algorithms to build haplotype assemblies. These haplotype assemblies were typically an order of magnitude larger than what can be achieved by genotyping a single individual, with over half the genome contained in segments greater than 200,000 base pairs in length.
- The HuRef analysis is helping researchers to begin to understand the impact of Dr. Venter's genome in relation to his traits. For genetic diseases, like Huntingtons's Disease and cystic fibrosis, the team can definitively show the absence of implicating mutations. For complex diseases, like heart disease and diabetes, whose explanation resides in DNA variants from multiple locations in the genome, the team can provide a probability of what Dr. Venter can expect. As more and more genomes are sequenced and more analysis is done involving those individuals' disease traits, it will be possible to assign a risk to the detected variants.

### Sequencing Methods

- For the HuRef project, the team at JCVI used a more traditional method of sequencing—whole genome shotgun assembly which is built upon Sanger dideoxy sequencing. Then, Applied Biosystems 3730xl high-throughput DNA sequencing machines were employed since these methods still produce the longest and most accurate lengths of DNA.
- This project was designed to produce an accurate and more complete version of a single individual's genome rather than producing a fast and potentially less expensive version. From the HuRef genome however the researchers believe that newer methods for sequencing can be used to enable more people to have their genome's sequenced and analyzed. It is clear that the HuRef version is likely the last time that these more traditional methods of sequencing will be employed.

### Previous Human Genome Sequencing Projects

- The first sequence and analysis of the human genome was published in Science in 2001 by Dr. Venter and colleagues at Celera Genomics. The publicly funded genome project also published their version of the human genome at the same time in the journal Nature.
- At Celera there were five individuals whose genomes were used for that consensus human genome assembly. One of those individuals was Dr. Venter whose DNA constituted the majority (60%) of the DNA for that genome. The publicly funded genome project used DNA from a variety of individuals and is a composite version.
- Dr. Venter and the team at JCVI have long been proponents of finding new and improved methods for sequencing genomes since it is only through cost-effective and accurate sequencing methods that millions of human genomes can be sequenced. In September 2003, the JCVI announced a \$500,000 prize for advances leading to the sequencing of one genome for \$1,000 or less. This JCVI prize was eventually joined with the \$10 million Archon X Prize for this enabling technology in DNA sequencing.