

**Contact Information:**

John Fosnacht  
SoftGenetics LLC  
The Oakwood Centre  
Suite 350  
State College PA 16803

**FOR IMMEDIATE RELEASE**

**New merge function creates Sanger Quality Sequence from NGS paired end reads**

**February 23, 2010, State College PA** announced the availability of a new unique function in its NextGENe software that provides Sanger quality sequence from short-reads from Next Generation Sequencing platforms such as the Illumina GA.

Short read lengths produced by Next Generation sequencers such as the Illumina Genome Analyzer can create difficulty for accurate analysis of data. Additionally, relatively high error rates (compared to other technologies such as Sanger sequencing) further complicate the analysis of next-gen sequencing data. For these reasons, lengthening the short reads prior to analysis and statistically correcting or removing errors is a valuable tool to improve alignment accuracy, Indel detection and assembly. A novel, highly accurate method to elongate reads, utilizing paired end reads, has been developed by SoftGenetics for its NextGENe software.

“Sequencing Paired End reads” indicates Megan Manion, NextGENe Product Specialist “is a useful technique which produces reads in pairs such that each pair of reads are a known distance from each other in the genome. This is accomplished by preparing DNA fragments of a certain length (200 bp, for example). This fragment size, or library size, is the distance between each pair of reads. Sequencing is then done from each end of the fragment, producing two paired reads. NextGENe’s paired end merging technique takes advantage of paired end information, along with the additional coverage from sequenced overlapping DNA fragments, to produce long reads spanning the entire library size with an extremely low error rate.”

Manion continued “Paired end reads can be merged by elongating the paired reads to the point that there is overlap between the two reads. This allows the paired reads to be joined together to form one continuous, longer read. The number of elongation cycles required depends on the read lengths and the library size. Each cycle of Condensation will generally increase the average read length to 1.6 the original length for shorter ( $\leq 36$  bp) reads and to 6 bases less than twice the original length for longer ( $> 36$  bp) reads. These values may be reduced with an average depth of coverage less than 30x. A single cycle of elongation of 75 bp reads from a 200 bp library, for example, allows the paired reads to overlap and be linked together. For 35 bp reads from a 200 bp library, three cycles of elongation allows for the linking of the paired reads. Reads should be extended until a significant portion of the paired reads (roughly 15% of the elongated read length) will be expected to overlap.”

The company offers 30-day trials and no cost web-based training on its genetic analysis software packages. Interested parties may request the software on the company website: [www.softgenetics.com](http://www.softgenetics.com) or via email: [info@softgenetics.com](mailto:info@softgenetics.com).

# #

*SoftGenetics, LLC specializes in the development of genetic analysis tools for both research and diagnostic applications. Hallmarks of SoftGenetics software tools are advanced technologies, providing exceptional accuracy, and sensitivity in an easy-to use Windows® user interface.*

*Trademarks are property of their respective owners.*