The Genome Sequencer FLX™ Software - A complete solution for sequencing, *de novo* assembly, reference mapping, and amplicon variant analysis.

Olaf Kaiser[1], Marcus Dröge[1]

[1]Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany

Abstract

Data analysis and bioinformatics are rate limiting steps in many laboratories. For helping researchers to make discoveries faster, as well as to minimize bioinformatics efforts, the Genome Sequencer FLX™ system is bundled with a suite of state-of-the-art analysis tools, that integrate seamlessly with the instrument and are optimized for 454 Sequence data analysis [1].

This software suite comprises four tools: Instrument Software, *GS Reference Mapper, GS de novo Assembler*, and *GS Amplicon Variant Analyzer*. The Instrument Software controls the Genome Sequencer FLX™ system during a sequencing run and performs the post-sequencing steps of image- and signal-processing. The *GS Reference Mapper* application maps shotgun reads, versus a given reference sequence (up to a length of 3 GB) and assembles the mapped reads into consensus sequences (contigs). In addition, it assists the user by detecting high-confidence mutations (*for e.g.* SNPs etc.) automatically. The *GS de novo Assembler* is an assembler tool, which assembles shotgun sequencing reads de novo into contigs. In addition, the *GS de novo Assembler* is capable of ordering contigs into larger scaffolds, by using 454 paired-end sequencing reads and it also enables *de novo* assemblies of genomes up to 120 MB. Moreover, the *GS Reference Mapper* and *GS de novo Assembler* are capabale of co-assembling traditional Sanger and 454 sequencing reads. The fourth software tool, the *Amplicon Variant Analyzer* (AVA), performs an alignment of amplicon sequencing reads, sequenced on the Genome Sequencer FLX™ system, versus a given reference sequence and identifies differences between the reads and the reference sequence. The frequency of automatically detected variants is calculated and reported by the AVA software tool. Furthermore, the AVA software offers researchers manual detection of variants, by examination of the read alignments. Subsequently, a quantification of the manually detected variants can be performed automatically by the AVA tool. It has been shown that the AVA software is perfectly suited to the identification and quantification of somatic mutations in cancer samples [2] or the detection of mutations conferring resistance in HIV quasi species [3].

References

[1] Droege, M., and Hill, B., The Genome Sequencer FLX™ System—Longer reads, more applications, straight forward bioinformatics and more complete data sets. J. Biotechnol., doi:10.1016/j.jbiotec. 2008.03.021.

[2] Thomas, R.K., et al., Sensitive mutation detection in hetereogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nature Medicine 2005, 12:852-855.

[3] Hoffmann, C., et al., DNA barcoding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res. 2007, 35:e91.