

**biomax**

biomax informatics ag

robert-koch-str. 2

82152 planegg

germany

tel.: +49 89 895574-0

fax: +49 89 895574-825

[www.biomax.com](http://www.biomax.com)

[info@biomax.com](mailto:info@biomax.com)

## Using the BioXM™ platform and KNIME to organize the next-generation sequencing data pipeline

“KNIME is the magic glue that allows me to integrate all my applications in a seamless way,” says one user of this open source exploration platform. At Biomax, we think it has some nice features — and have worked to make those benefits available to our clients by making KNIME seamlessly accessible in the BioXM™ Knowledge Management Environment.

Next-generation sequencing (NGS) covers a broad range of high-throughput sequencing platforms, an even broader range of experimental approaches, such as genome and exome sequencing or RNA-Seq, and a virtually unlimited spectrum of biomedical applications.

Correspondingly, there is no such thing as the one and only bioinformatics pipeline for NGS data. Instead, a rich and quickly expanding set of computational tools exists to implement various stages of NGS data analysis, such as quality filtering, read mapping, sequence assembly, SNP calling and so on. Depending on the particular data set to be analyzed and the scientific question to be addressed, these tools need to be connected to each other and parameterized accordingly. More often than not this is achieved by writing custom scripts and, as better tools appear and new data types emerge, these scripts need to be tweaked, which results in mountains of poorly maintained code.

---

Figure 1. Simple NGS workflow using KNIME.

We, and several of our clients, have discovered that creating easily maintainable and expandable bioinformatics workflows for NGS is greatly facilitated by the application of workflow management systems, such as KNIME. Each analysis step is encapsulated as a node model with clearly defined input and output ports as well as associated flow variables. A growing collection of such re-usable KNIME nodes is stored in a proprietary library, with virtually thousands of additional modules for various data analysis, manipulation and visualization tasks freely available from the KNIME community. Custom data flows can be easily created by plugging nodes into each other. Importantly, KNIME allows for easy creation of heavily branched processes, where data analysis nodes may have multiple inputs and outputs, and where several alternative or parallel analysis paths can be triggered as needed.

The diagram on the first page illustrates a simple NGS workflow using KNIME (the possibilities are virtually endless). After reading exome sequencing data (RunFastQC), generating read statistics (FastQC), and quality filtering (RawReadManipulator) the reads are mapped to the reference genome (BWA) and converted to an appropriate format (BAMSAMConverter). Subsequently, SNPs are called (SNPcall) and annotated (SnpEff, SnpSift, SnpEffGetDb). SNP calls are converted to a KNIME table (VCFtoTable) and passed to the BioXM Knowledge Management Environment.

Using the KNIME graphical user interface such a workflow can be crafted within minutes.

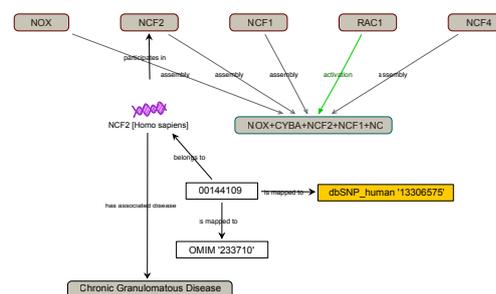
Biomax provides specialized KNIME nodes for smooth integration with the BioXM platform. For example, in the workflow shown on the first page, the BioXM importer node is used to load a variation data table into the internal BioXM data schema. Once variation data are imported into the BioXM system they can be accessed in the context of biological pathways and clinical information. Likewise, KNIME workflows have full access to powerful BioXM query and reporting capabilities. The BioXM reader node is capable of executing complex parametrized pre-defined BioXM queries and passing the corresponding result tables back to the KNIME workflows.

Perfect partners are seldom found. We think we may have hit on a possibility; seamless integration of KNIME with the BioXM Knowledge Management Environment can fast-track your research, analysis and solution development.

Learn more about KNIME here: [www.knime.org](http://www.knime.org).

Contact Biomax to learn more about the BioXM Knowledge Management Environment Platform or visit our website at [www.biomax.com/bioxm](http://www.biomax.com/bioxm).

**Figure 2. Direct access to KNIME from the BioXM Knowledge Management Environment**



**Figure 3. A BioXM graph view of a SNP identified in exome sequencing data.**

## About Biomax

Biomax Informatics AG is a leading provider of customized bioinformatics solutions. The company's flagship product, the BioXM™ Knowledge Management Environment, is a powerful enterprise platform enabling measurable productivity increases in biomedical R&D and clinical practice. The easy-to-use and highly configurable software allows domain experts, rather than programmers, to create, structure and visualize complex scientific models. The product is supplemented with sequence analysis, data integration, and text-mining software modules and domain-specific biomedical content solutions, for example in oncology.

Biomax creates value for clients by eliminating the coding process, slashing the implementation time of complex projects, and helping users to harness the information overload and distill critical insights from data. The Biomax worldwide customer community includes companies and research organizations that are successful in the areas of drug discovery, diagnostics, fine chemicals, food production and plant production.

*Biomax and BioXM are registered trademarks of Biomax Informatics AG in Germany and other countries. Registered names, trademarks, etc., used in this document, even when not specifically marked as such, are not to be considered unprotected by law.*  
BIOXMART1405