Dear Colleague

The genomics revolution is gaining momentum. In April, AstraZeneca announced it has launched a massive integrated genomics initiative, the most comprehensive bet to date by a pharmaceutical company on the potential of genomics to reshape drug development. The initiative will establish an in-house Center for Genomics Research and plans to collect and process up to 2 million genome sequences from samples donated by patients.

AstraZeneca is not alone in amassing vast quantities of omics and health data; omics data volumes all over the pharmaceutical industry are exploding. However, the industry lacks sophisticated informatics infrastructure and in particular enterprise level software that addresses the data management and analysis bottleneck which impacts the value to be gained from such big data initiatives.

In this edition of Insights we describe some of the challenges of building the informatics infrastructure to handle millions of genomes and associated data. Perhaps more importantly, we also describe the attributes that are key to delivering value from that infrastructure to drive the scientific insights needed to improve and make pharmaceutical R&D more efficient.

Marc Flesch, Ph.D.
Head, Scientific Software Engineering

On behalf of all the Genedata Profiler™ team, I hope you find *Insights - Genedata Profiler* informative and helpful. I welcome your comments and feedback.

Enjoy this second edition of *Insights - Genedata Profiler*!

With best regards,

Marc Flesch

HOT IN THE FIELD

Mark Collins, Ph.D.
Marketing & Business Development

Recent announcements by pharmaceutical companies and even governments regarding large scale "moon-shot" genomics initiatives to improve human health got us thinking here at Genedata. I discuss what we think it takes to leverage the full power of two million genomes. I will elaborate on seven key focus areas from the informatics perspective that we are working on to help organizations deliver more efficient pharmaceutical R&D.

Read Dr. Mark Collins' Commentary: The Bioinformatics Challenge of 2 Million Genomes to learn more about these seven imperatives to achieve the ambitious aims of the 2 Million Genomes Projects:

- Efficiency and scalability of data storage, management and processing
- Harmonizing disparate data
- Fast query performance
- Extensibility
- Security and Privacy
- Collaboration
- Scientific Insights

FROM OUR DEVELOPERS

Michael Remmert, Ph.D.
Software Engineering

Building a scalable system to handle millions of genomes and associated phenotype and other data requires vast amounts of data to be efficiently stored and processed in a reasonable time. Genedata Profiler has been designed from the ground up with such scalability in mind, bringing the following advantages to organizations who envisage leveraging massive quantities of complex omics data.

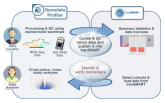
- Highly efficient storage and management of data using a hybrid data architecture that combines common, standard, extensible file formats (e.g. BAM or VCF), together with Genedata's open binary format (GBF). This approach stores encoded records within a genomic reference system reducing storage capacity by up to 20X compared with traditional methods and ensures rapid query and processing of data.
- Parallel processing of data via a grid scheduler that integrates with high-performance compute clusters. A unique feature is that third party algorithms, R-scripts, and other tools incorporated into Genedata Profiler via our APIs can take advantage of this parallelization, even if they were not originally developed with that capability.
- Scientific insights: users can see the big picture of their genomic data using advanced analytical algorithms and a highly scalable genome browser that allows for the real-time interactive visualization of thousands of genomes at a time.

LEARNING CENTER

Genedata produces scientific educational resources on the application of very large data sets to solve translational science challenges. We are pleased to make these available for download from our website. The latest additions to our Learning Center include:

Webinars

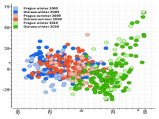
Ensuring patient privacy in multi-omic translational research
Genedata demonstrates how Genedata Profiler™ is able to overcome these challenges and optimize the process of translational research while ensuring compliance with ever-changing patient privacy regulations.



Leveraging the power of Genedata Profiler for more efficient translational research
Learn how the power of Genedata Profiler drives rapid and efficient biomarker identification and validation from raw data to results and how the platform integrates seamlessly with existing in-house platforms (e.g. the open source data warehouse tranSMART).

Molecular characterization of diseases using multidimensional big data
We address the major tasks performed by different members of an R&D team to dive deep into omics data to profile patients and compounds, applying a workflow to identify candidates for predicting response to an oncological drug.

Publications



Read about a collaborative project where Genedata helped to develop new bioinformatics workflows which deliver accurate measures for inter-laboratory comparison studies.
Herwig *et al*. (2015): Inter-laboratory study of human in vitro toxicogenomics-based tests as alternative methods for evaluating chemical carcinogenicity: a bioinformatics perspective

Videos

In this short video, Optimizing the process of omics based patient profiling in translational research, we show how the latest version of Genedata Profiler (v10.0) helps global, multidisciplinary teams perform translational research while ensuring patient data privacy.

Recent Posters

Epigenome-wide discovery of DNA methylation biomarkers for liquid biopsy-based ovarian cancer screening. As part of the EU FP7 EpiFemCare consortium, we report here the results of applying novel bioinformatics methods for analysis of Reduced Representation Bisulfite Sequencing (RRBS) data to detect cancer-specific methylation regions for use as clinical diagnostic assays.

A new translational research software platform for efficient patient and compound profiling. Genedata Profiler is a new translational research software platform developed in collaboration with leading pharmaceutical companies to process, manage, and analyze NGS and other omics data from patient samples, applying highest data quality and regulatory compliance standards.

Genedata Profiler - a collaborative, regulatory-compliant, integrated platform for omics-based patient and compound profiling. In this poster we discuss the key features of Genedata Profiler and illustrate the power of

the platform using a case study that also demonstrates how Genedata integrates with tranSMART.

Enterprise software for efficient translational research in an increasingly complex regulatory environment.In this poster we focus on lifecycle method and data management in Genedata Profiler - critical considerations in meeting quality and compliance goals for omics-based translational research.

Featured Articles

All ***Insights - Genedata Profiler*** **Featured Articles** are also available for download from the Genedata website.

Collins et al. (2016): Translational research in the era of heightened patient privacy concerns. In this paper we review the regulatory landscape and its impact on translational research, and define the key attributes of an "ideal" system to overcome the challenges of conducting translational research while respecting patient privacy.

MEET THE EXPERTS

Experts from Genedata will attend the following meetings over the next few months, where they will be delighted to discuss issues in genomic profiling of patients and characterization of diseases. Please visit our posters and ask for a demonstration of Genedata Profiler 10.0.

14th Annual Pharmaceutical IT Congress
September 28 - 29 | London, UK

BioData World Congress 2016
October 26 - 27 | Cambridge, UK

It was great to connect with many local Boston customers as well as new faces at the recent Festival of Genomics June 27-29, Boston, USA. Our team had many good discussions and interactions and thanks to those of you who dropped by our poster, *Lempiäinen et al, Epigenome-wide discovery of DNA methylation biomarkers for liquid biopsy based ovarian cancer screening*, which is now available for download

---

ABOUT GENEDATA PROFILER

Genedata Profiler optimizes the business process of translational research by combining high-performance raw omics data processing pipelines, sophisticated analytics, and innovative visualizations with an advanced distributed data management infrastructure. An open, interoperable enterprise software platform, Genedata Profiler empowers scientists to integrate NGS and other omics data from patients with clinical annotations, generating scientific insights while ensuring data privacy and security.

Basel | Boston | Munich | San Francisco | Tokyo
www.genedata.com