GeneVision application whitepaper

BC Platforms offers a cloud based solution empowering it's users to manage and clinically interpret their patients' genomic data, creating local competencies. Our highly automated solution allows hospitals to reduce time and costs while facilitating secondary use of data for internal and external research. Hospitals also benefit by optimizing the use of genotyping and NGS resources to incorporate into clinical data, driving evidence based practices in precision medicine within the hospital on a large scale.

Abstract

Use of genomic information in healthcare is currently mainly limited to cancer and rare diseases, and the genetic tests are typically outsourced to external service laboratories. Due to continuously decreasing price of genome data production, and increased interest for secondary use of patient data for research, there is a growing interest to produce patients' genome once, and use the same data for various different healthcare applications and for research. This approach involves multiple challenging steps including genome data production and quality control, scalable IT environment for genome data handling, and genomic result interpretation and delivery to the patient.

GeneVision provides end-to-end Software-as-a-Service solution from raw genome data to patient reports, implemented using Microsoft Azure technologies.

Introduction

Impact of genomics on healthcare

Since the Human Genome-project the world has been waiting for breakthroughs arising from the use of genetic data. Instead of an avalanche of cures to diseases with genetic origins there's been a trickle-feed of improved specificity, preciseness and safety in diagnosis and medical treatment of many diverse diseases, often without any specific genetic cause. These small advances have converged into mainstream healthcare guidelines and recommendations for the use of patient's genetic data in defining the safest and most efficient healthcare. As genetically assisted decision making is no longer only in the realm of large research hospitals, but part of routine healthcare on population level, it is creating new challenges due to the sheer enormity of the data volume and complexity.

Value of the generated genomic data in itself is of significant magnitude. It is not just the financial investments put into chemistry, instruments and laboratory work, but the reusability of the data for further healthcare applications, and research. Hospitals with the capacity to resource their own genomic data repositories are in a position to exploit the data generated in-house, but very few technical options of suitable quality and sufficient ROI exists. Hospital management also needs to be ensured of the safety and efficacy of their genomic services to the patients.

Current applications in healthcare

Clinical genomics

Utilization of Genomic information in healthcare is in most cases limited to cancer and rare diseases, and not applied at population level. Genetic tests and genotyping are outsourced to 3rd party service laboratories, which will then receive either a physical patient sample, or raw data coming from sequencing or genotyping device, and return typically one PDF report answering the question e.g if patient carries risk genotypes for breast cancer. Based on the report, a patient treatment plan is selected. Turnaround is often slow and services expensive, and if the medical provider would need to ask further question, they need to order another test.

The increased volume of patients being sequenced has contributed to the lowering of costs of whole exome data production. Affordable access to data has driven demand for secondary use of patient data for research and the promise of biobanks, inspiring many hospitals to having their own genome laboratories. However this increase in demand has required hospitals to outsource their sequencing to third parties and has also resulted in the following challenges.

1) Raw genome data consumes significant amount of storage, and requires scalable computing power to pre-process. Hospitals usually don't have calculation clusters or very large storage and archival system available.
2) Building genome data pre-processing pipelines requires specialised PhD level knowledge in bioinformatics which is hard to find and keep. Also, compared to academia, use of genomic data in healthcare is regulated and all software needs to meet certain quality and data security criteria, making implementation difficult.
3) Genome data quality must be monitored according to different standards (i.e. AMP/CAP/ EuroGenTest), requiring documented standard operating procedure and software tools for implementing it.
4) Genome data interpretation reports must be generated and interpreted to patient certain way, and annotation information updated regularly.

As a result, deployment may be delayed and initial investment for sequencing device and genome lab instrumentation are not producing added value to the patients or research.

GeneVision Solution

GeneVision enabled by Microsoft Azure technology can help hospitals in expanding the role of genomics beyond individual genomic tests, towards producing genome data at larger scale and expanding its applications for healthcare and research use. GeneVision implements scalable, easy-to-use workflows for both genotyping and next-generation sequencing, quality monitoring, and production of different patient reports for clinical genomics. The use of genotyping technology allows hospitals to plan NGS resource use more accurately, and focus more expensive NGS technology efforts to patients that benefit the most. In addition, the solution provides a secure genome data warehouse and archive,

and easy access via open Application Programming Interface (API) for accessing genome data from other applications.

GeneVision is comprised of multiple interlinked and interacting modules, each dedicated to performing part of the whole functioning workflow:

- Genotyping, imputation, and automated quality control
- NGS data secondary analysis and quality control
- Genome data warehouse
- Genome variant interpretation

NGS data secondary analysis and quality control

GeneVision provides automated workflow for NGS secondary analysis, monitoring data quality and ensuring data governance throughout the process. Secondary analysis requires large computing resources, and handles massive amounts of data. For data processing GeneVision uses Microsoft Genomics Service.

Sequence data production process includes laboratory work, robotics, instrumentation and data management software, and quality of the produced data must be monitored continuously. GeneVision collects and provides comprehensive quality control metrics for sequencing data, and provides tools for ongoing quality monitoring. Quality status for each sample is simply stated as either PASS or FAIL, according to chosen standard operating procedures. GeneVision visualises quality parameters over time and benchmarks quality with other laboratories.

Population scale genotyping, imputation and quality control

Genotyping is technology for extracting predefined genome variants (SNPs) from DNA. Process includes multiple steps, including DNA extraction from sample (typically blood or spit), production of raw data using genotyping device and processing raw data to extract SNP variants.

Compared to NGS, genotyping technology is older, but very cost-effective way of producing whole genome data. Using imputation techniques it is possible to ‘guess’ whole genome sequence with improved accuracy. While resolution of imputation is not enough for detecting very rare diseases, and can’t be used in oncology or identifying de-novo mutations, it works well for preventive medicine and pharmacogenomics applications, and for academic and commercial research work.

Standard software tools delivered together with Illumina genotyping device require manual data processing, taking up valuable time from laboratory’s bioinformaticians. GeneVision streamlines this whole process by implementing an automated workflow, eliminating nearly all manual work. Solution reads raw data directly from native Illumina binary files from Illumina genotyping devices, performs a comprehensive quality check on different levels (batch, plate, chip and sample) and runs imputation process using 1000 genome or population specific genomic references. This makes GeneVision to scale up to population level analytics and health planning utilizing genotype data.

Genome Data Warehouse

GeneVision implements scalable, high-throughput and cloud optimised software solutions for storing and managing genomic data, including genotypes, imputed genotypes, targeted sequencing, whole exome sequencing, and whole genome sequencing data. The warehouse is suitable for research and clinical work. The solution has various interfaces enabling wide range of workflow implementations from visual reporting and job control to programmatic access for external applications. Currently the warehouse is designed to handle whole genome sequence data from 1 million subjects, and design limits are expanding continuously. System is optimised for Microsoft Azure and supports Microsoft Azure Batch for distributed data analyses.

Genome Variant Interpretation

GeneVision includes multiple different reports, from pharmacogenomic reports (PGx), to genomic risk scores for preventive medicine (GRS) and well-being reports. In addition, 3rd party reports can be linked using an API. GeneVision sends relevant genomic information to 3rd party servers using a RESTful API and receives prediction reports in a PDF format and structured data format to be stored to the database.

Pharmacogenomics (PGx)

Efficacy and safety of many approved drugs depends on genomic profile of the patient. PGx reports show all the known incompatibilities with certain drugs and the patients genomic profile. Drugs are classified in groups based on significance of the effect the patient's genotype has on the efficacy or safety of the drug molecules. In practice there are four groups: 1) genetic variation has significant clinical relevance, 2) genetic variation has moderate clinical relevance, 3) genetic variation has minor clinical relevance, and 4) genetic variation has no clinical relevance. Based on this information, physicians feel more confident as they have more options to consider when prescribing therapies for their patients.

Conclusion

The GeneVision modular approach provides all the required components for a small to high-throughput sequencing lab in hospital and utilization of genome data for clinical genomics, and pharmacogenomics applications. The solution is powered by Microsoft Azure technology, Microsoft Genomics Service, and services relevant to patient genomic reporting.

Compared to other solutions, it implements full workflow from both genotyping & raw genome data to patient reports in one integrated solution, providing

- Secure and scalable genome data management and archiving solution (HIPAA and EU compliant)
- Scalable sequence data secondary analysis (alignment and variant calling)
- Sequence data production quality monitoring solution (AMP/CAP and EuroGenTest compliant)
- Support for population scale genotyping and imputation for preventive medicine
- Different reports provided for PGx, clinical use and preventive medicine.