imec

An imec.icon research project | project results

SMART HEALTH

GENOME ANALYTICS PLATFORM

A European high-performance genome analytics platform for cost-effective DNA analysis

When the human genome sequence was published for the first time in early 2000s, the scientific community agreed that, in order to make genome sequencing available to every patient, the cost of Whole Genome Sequencing (WGS) had to decrease to 1,000 USD. Although sequencing itself has now become cost-effective, analyzing the avalanche of raw sequencing data in clinical use has become the economic bottleneck.

Within the genome analytics platform project, researchers and industry leaders in Belgium want to tackle this challenge. The aim is to develop a new genomic analytics platform that bridges the gap between the low cost of sequencing and the high cost of analyzing raw sequencing data. It generates and manages data for clinical and research genomic applications that analyze the genome of patients (such as Agilent Alissa OnePGT, Alissa Interpret, KU Leuven HiVA or KU Leuven WiNGS) and makes it easy for doctors and researchers to analyze and compare genome data for diagnosis, prognosis, or treatment selection.

THE OUTCOMES

1. A unique hybrid cloud platform for fast and cost-efficient whole genome sequencing and analysis

The genome analytics platform includes integrated storage technology from Western Digital, heavily optimized genomics software developed by imec and UGent, and works seamlessly with the BlueBee Genomics Cloud. Western Digital developed and optimized performance of its Unified Data Access interface to support the genome sequencing pipeline. Its advanced NVMe™ technology accelerates data analysis before enabling cost-effective and durable retention – distributing data over multiple clinical centers utilizing ActiveScale™, its scale-out object storage system.

BlueBee's intelligent software provides the ability to burst into a cloud-based sequencing service. This approach to sequencing enables the administrator to determine whether data is processed on-premise, in the cloud, or a combination of the two. Once in the cloud, the BlueBee platform will automatically provide the required hardware to execute the genomics workflows within the desired timeframe at a price per sample at any scale.

2. Whole genome mapping for each patient is an affordable reality

A key objective of the project was to process the full output of a Next Generation Sequencer (NGS), 48 human genomes, within 48 hours to ensure optimal levels of cluster utilisation and maximize the Return on Investment (ROI). The combination of Western Digital's innovative storage technologies enables full optimization of each sequencer run. Next to processing a steady data flow of 48 human genomes in 48 hours, BlueBee enables burst processing of 5000 samples linked to the Alissa OnePGT application in under 4 hours.

3. One workflow: from sequence data to therapeutics

Agilent has integrated the hybrid cloud platform into its Alissa OnePGT and Alissa Interpret solution, which is being used by the University Hospital of Leuven and the Human Genetics Group of KU Leuven to optimize patient diagnostic and screening.

The platform also encompasses the existing WiNGS platform from KU Leuven that compares data from patients from different hospitals without revealing the patient's identity. This is especially important to diagnose rare diseases.

NEXT STEPS

BlueBee has extended its cloud offering with a hybrid model whereby end users can decide to keep data on premise or in the cloud while keeping a consolidated overview. While operating in the cloud, BlueBee manages on-demand infrastructure to meet the turnaround times in an efficient manner at scale. Western Digital demonstrated how its device and system storage portfolio combines to provide the foundation for hybrid cloud deployments, to reduce costs and accelerate processes for petabyte-scale healthcare and research projects. UZ Leuven and the Human Genetics Group of KU Leuven will also use the GAP platform in patient diagnosis (rare diseases, pre- and neonatal). KU Leuven STADIUS is deploying its WiNGS platform across Belgian clinical genetic centers for rare genetic disease diagnosis, and will be further developing it in the framework of the Belgian node of the EU ELIXIR ESFRI infrastructure for life science data. Imec's ExaScience Life Lab released and published their elPrep 4.0 DNA software solution for efficient sequence analysis that produces identical results to established genome analysis programs such as SAMtools, Picard and GATK4. The elPrep software can be used stand-alone, or from within UGent's Halvade software. UGent developed Halvade on Spark, a software framework to process WGS data that runs on top of the well-known Spark framework and can distribute the genome processing over a cluster of machines very efficiently. The UZ Leuven, Imec's ExaScience Life Lab and UGent are considering a follow-up project, focusing on long reads. The software for handling long reads and transforming them into a genetic map and into useful information, is much less mature than it is for short reads.

The GAP project was co-funded by imec, with project support from Agentschap Innoveren & Ondernemen

AGENTSCHAP INNOVEREN & ONDERNEMEN
Vlaanderen is ondernemen

FACTS

NAME	GAP
OBJECTIVE	Bridging the gap between the low cost of sequencing and high cost of analysis to enable widespread clinical use of Whole Genome Sequencing
TECHNOLOGIES USED	DNA sequencing, Cloud computing, High Performance computing, Big data, Whole genome sequencing
TYPE	imec.icon project
DURATION	01/04/2017 – 31/03/2019
PROJECT LEAD	Kurt Florus, BlueBee Belgium
RESEARCH LEAD	Yves Moreau, STADIUS – KU Leuven
BUDGET	3,456,880 euro
PROJECT PARTNERS	Agilent Technologies Belgium, Amplidata, BlueBee Belgium
RESEARCH PARTNERS	STADIUS – KU Leuven, Human Genetics – KU Leuven
RESEARCH GROUPS	ExaScience Life Lab, IDLab, an imec research group at UGent

WHAT IS AN IMEC.ICON PROJECT?

The imec.icon research program equals demand-driven, cooperative research. The driving force behind imec.icon projects are multidisciplinary teams of imec researchers, industry partners and/or social-profit organizations. Together, they lay the foundation of digital solutions which find their way into the product portfolios of the participating partners.

GAP project partners:

Agilent · AMPLIDATA · BlueBee · GHENT UNIVERSITY · KU LEUVEN

CONTACT US
WWW.CONTACTIMEC.COM