Duality

Privacy Preserving Genome-Wide Association Studies Powered by Duality

In 2020, Professor Alexander Gusev of the Dana-Farber Cancer Institute, an affiliate of the Harvard Medical School, partnered with Duality to prove that privacy preserving Genome-Wide Association Studies (GWAS) are feasible and can be practically used.

GWAS studies typically include data from a large number of individuals – tens of thousands to millions – in which several factors are measured that may lead to the identification of specific genetic variants associated with a given medical condition. These factors typically include a genome-wide set of genetic variants (genotype), a trait and/or disease, as well as several other cofactors, such as age and gender. The goal of the GWAS is to identify if any genetic variants are associated with a specific trait and/or disease.

In this case study, we illustrate the project challenges, the reasons Dr. Gusev partnered with Duality, and the project outcomes.

Challenges: Privacy and Cross-Institutional Collaboration

Dr. Gusev noted that a typical GWAS study requires processing hundreds of thousands of patients' data from dozens of institutions - including large-scale clinical and genomic data sets as well as phenotypes.

The first challenge is a matter of logistics. In general, cross-institutional data collaboration is challenging due to both competing interests from participating institutions, as well as the challenge that different institutions may take different information from their patients and organize it differently, leading to possible disparate data inputs within the same pooled data set – and potential accuracy issues.

Logistics also present an issue in terms of data volume. While the statistics underlying the research are simple – analyzing the frequency of a specific genetic variant (or variants) in patients who may have underlying conditions – processing the data from hundreds of thousands or even millions of patients to make that discovery requires vast computing power. No one hospital contains enough patient data on its own to conduct this type of research.

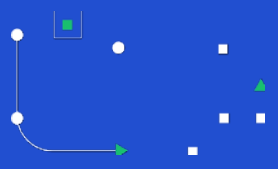
Another critical challenge is a matter of privacy, in particular the assurance that regulatory compliance factors will be met in accordance with sensitive patient data. Genetic data is extremely sensitive; so is clinical data from health records from various hospitals and other research institutions. Beyond the concern of how that data will be used, and what it will be used for - there are complexities, including the legal mandates regarding healthcare data and patient privacy.

In other words: as genomic data collection is becoming universal, the critical barrier to personalized medicine is no longer data availability but data security and privacy. At the same time as data collection has gotten faster, easier, and cheaper, the protocols for data sharing have become more cumbersome, fractured, and ad hoc.

Testimonial

"With Duality, we side-stepped the question of 'how do you share the underlying sensitive data' instead developed a technique to work on encrypted data without releasing any sensitive information at all."

Dr. Alexander Gusev, Dana-Farber Medical Center



Data Collaboration

HARVARD MEDICAL SCHOOL

Dana-Farber Cancer Institute

NIH National Institutes of Health

The Solution

Duality worked together with Prof. Gusev to conduct this research with the highest levels of accuracy, data discovery and analysis, while ensuring the preservation of patient privacy.

Duality uses data science, cryptography and leverages Privacy Enhancing Technologies (PETs) to derive insights from multi-sourced, encrypted data without ever decrypting the data – regardless of data size or scale. PETs are groundbreaking, they allow all the existing computational methods that researchers and clinicians are familiar with to operate directly on encrypted, secure datasets.

How it works

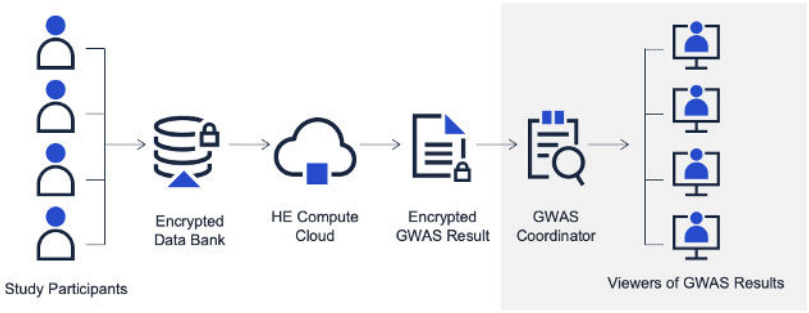
- Duality's researchers and data scientists first analyzed data from 25,000 individuals, using Fully Homomorphic Encryption (FHE). To maximize performance, the Duality team reformulated the GWAS testing methods to fully benefit from encrypted data packing and parallel computation; integrated highly efficient statistical computations; and developed multiple crypto engineering optimizations.
- Ultimately, the method was proven scalable to data sets from 100,000 individuals or more and can be performed in under 6 hours on a single server node, or in 11 minutes using 31 server nodes running in parallel.

Benefits

- Rather than contend with the question of cooperation between participating institutions, encrypting the data allows all participating researchers to take part in the study with trust, versus having to enter into the usual lengthy legal process.
- Patients can participate in research studies and receive personalized results knowing that their individual data remains well protected.
- Researchers can share full datasets or individual variables with collaborators without compromising the privacy of study participants or their own future analyses.
- Multiple institutions can securely collaborate on meta-analyses of tens of thousands of individuals while guaranteeing individual-level privacy and control over their valuable data.

Results

- Yielded results 30x faster than state-of-the-art methods available at the time the study was conducted
- Scalable genomic analyses on hundreds of thousands of individuals in a privacy-preserving way
- Maximized value from real-world patient data



Conclusions

In summary, Prof. Gusev confirms that the Duality Platform, can not only be leveraged in the GWAS context illustrated above, but also in other types of trials and research, such as:

- Clinical Trials - for secure collaboration across multiple clinical trials to identify genetic/clinical markers impacting preventative care, drug recommendations, etc.
- Trial Matching - allowing patients to share their genetic and clinical data with matching programs to inform them when an appropriate trial is available for them to join, but without compromising on their sensitive data
- Genetic Risk Prediction - thanks to recent advancements in machine learning, geneticists have discovered that millions of genetic markers can be aggregated into highly accurate genetic predictors of disease risk. Unfortunately, this aggregation has the potential to expose all of the genetic data of participants rather than just a handful of mutations. Using PETs, patients can participate in predictive studies and receive personalized risk scores without compromising their individual genetic data
- Rare Disease Studies - allowing carriers of rare diseases to securely submit their carrier status to a privacy-preserving database which would both match them to other carriers (for support group and community building) and allow researchers to quantify the frequencies of rare variants without revealing the identities of any of the individual carriers

For more information on this research and Duality, please read the full report in PNAS.

NEW JERSEY

5 Marine View Plaza,
Suite 301
Hoboken NJ 07030

CONTACT US

info@dualitytech.com
Duality Technologies
@Dualitytech

dualitytech.com