Watermarking Module for Shared Genetic Data

R. Bellafqira, M. Ben-Ghali, G. Coatrieux. Cyber-Health, LaTIM Inserm UMR 1101, IMT Atlantique.
E. Genin, T. Ludwig, G. Marenne. Inserm UMR 1078.
A. Amzert. Inserm, UMS 56 EMERGEN.

PRIVGEN follow-up action

- **Privacy-preserving sharing and processing of genetic data** - Labex CominLabs project 2017-2020
- **PRIVGEN Context** –

  Cloud Computing and data outsourcing - A successful paradigm to flexibility store, share and process large amount of data while minimizing costs

  Security needs of outsourced applications and data are worsened - Owners loss the control on their data and applications (confidentiality, integrity, availability?); Service providers may in turn transmit data to third-party service providers (traceability, intellectual/scientific ownership protection?);

- **PRIVGEN Challenges** –

Challenge 1 - Mechanisms for a continuous digital content protection - Merging different security mechanisms into one configurable digital content protection tool for multipurpose security services.

Challenge 2 - Composition of security and privacy-protection mechanisms - Provide a development approach for privacy-preserving distributed genetic applications

Challenge 3 - Distributed processing of genetic data - Develop a platform for: 1) sharing relevant genomic information while maintaining privacy; 2) supporting distributed execution of applications over shared genetic data.

- **PRIVGEN Results** -

Result 1 – 1) watermarking modulations for genomic data shared in genome wide association studies (GWAS) ; 2) dynamic joint homomorphic encryption-watermarking scheme able to detect and identify altered data under user data update constraints. ; 3) Processing of encrypted genetic data

Result 2 – 1) An architecture-based method for the definition of high-level definition of genetic analyses in terms of compositions of different types of cloud infrastructures as well as organization-owned computing and data storage infrastructures; 2) An object-oriented framework supporting compositions of a large range of security and privacy-preservation operators.

Result 3 - Platform PRIVAS: a tool to perform Privacy Preserving Association Studies
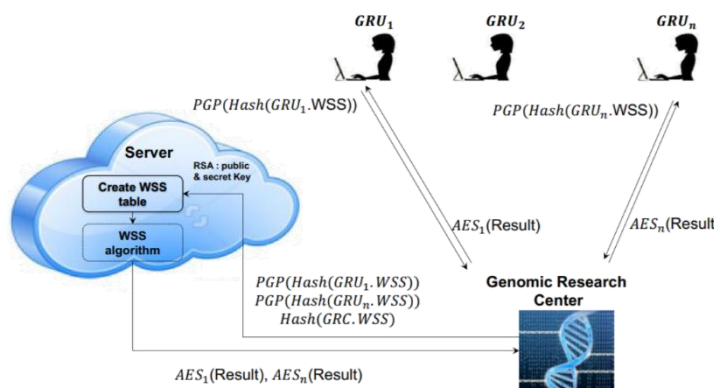
TADOP Project (11/22-04/25)

- **Project purpose** - Valorization of the PRIVGEN watermarking tools for genetic data with the development of a demonstrator to meet various security objectives (e.g., fight against data leakage, integrity control, IP Protection) as part of the prefigurative platform of the *Collecteur Analyseur de Données* (CAD) of the *plan France Médecine Génomique 2025*, currently being integrated into the Inserm HDS Cloud platform.

- **Project Challenges** – Scaling up of the PRIVGEN genetic data watermarking modulations (proof of concept for GWAS) to satisfy various watermarking based security services in the context of the CAD prefigurative platform. Industrial transfer at the issue of the project.

TADOP Use case

- Two GWAS types will be considered: at the variant level for frequent variants in the population and at the gene level for rare variants.
- Availability of experimental genomic dataset (Whole Genome Sequences – WGS) from GAZEL-ADN project to conduct the first tests.
- Availability of GWAS statistical analysis tools that will be deployed on the platform.

Current Status & Next Steps

- Development phase complete: Watermarking system successfully created
- Validation testing completed on multiple databases
- System packaged as Docker container for easy deployment
- **Current Phase**: Integration and testing of the watermarking module within Inserm's Health Data Hosting (HDS) certified cloud platform
- This integration will ensure secure, compliant handling of sensitive genetic data in a certified healthcare environment

Publications

- R. Bellafqira, M. Al-Ghadi, E. Génin, and G. Coatrieux, *Robust and Imperceptible Watermarking Scheme for GWAS Data Traceability*. International Workshop, IWDW 2022.
- R. Bellafqira, T.E. Ludwing, D. Niyitegeka, E. Génin, and G. Coatrieux. *Privacy-preserving genome-wide association study for rare mutations – a secure framework for externalized statistical analysis*. IEEE Access, 2021.

TADOP Integration platform