SQUiD: Ultra-Secure Storage and Analysis of Genetic Data for the Advancement of Precision Medicine

Jacob Blindenbach, Jiayi Kang, Seungwan Hong, Caline Karam, Thomas Lehner, Gamze Gürsoy

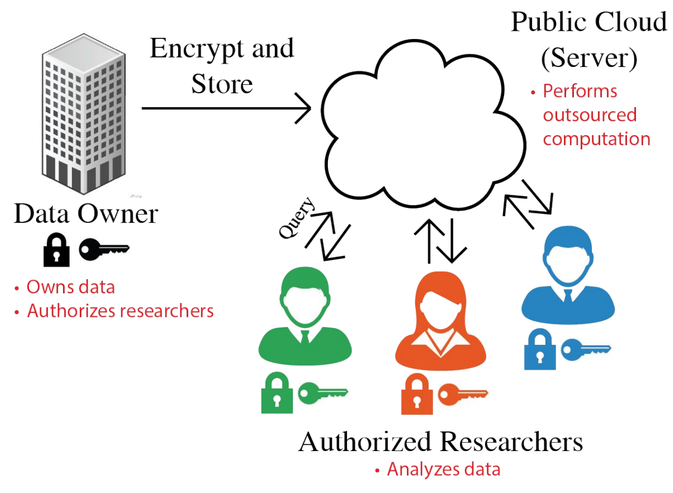
Problem Statement

- Databases are needed for storing extensive sensitive patient disease and genetic information.
- The large amount of data necessitates the use of cloud storage, coupled with strong security measures due to its sensitive nature.

Goal: Ensure the data is secure from both cloud vulnerabilities and unauthorized users, yet accessible for authorized researchers to safely perform queries and analyses.

Approach / Scenario

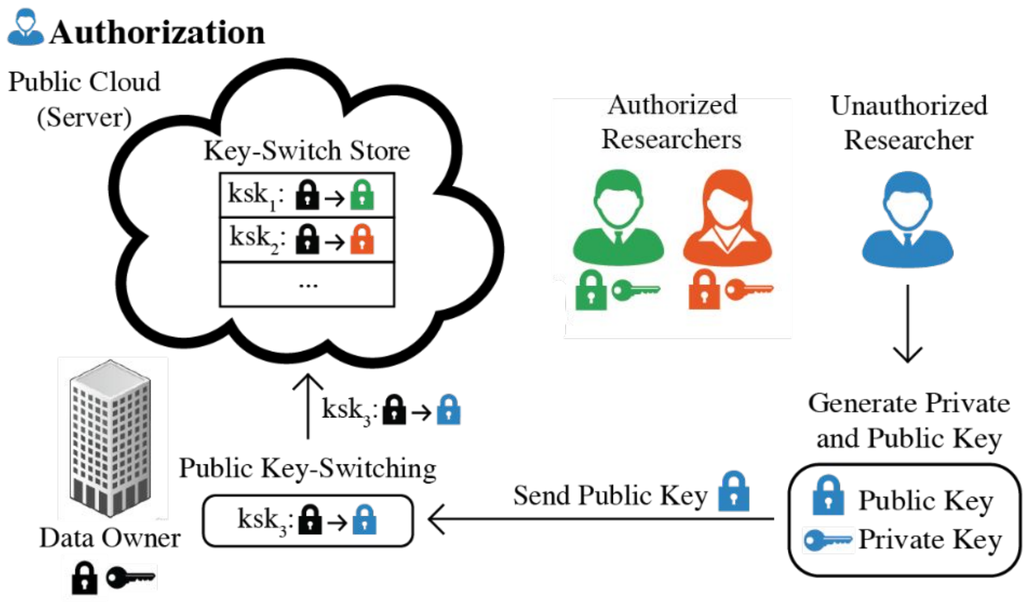
SQUiD (**S**ecure **QU**eryable **D**atabase) is designed for a multiparty setting with a data owner, a public cloud, and multiple researchers.

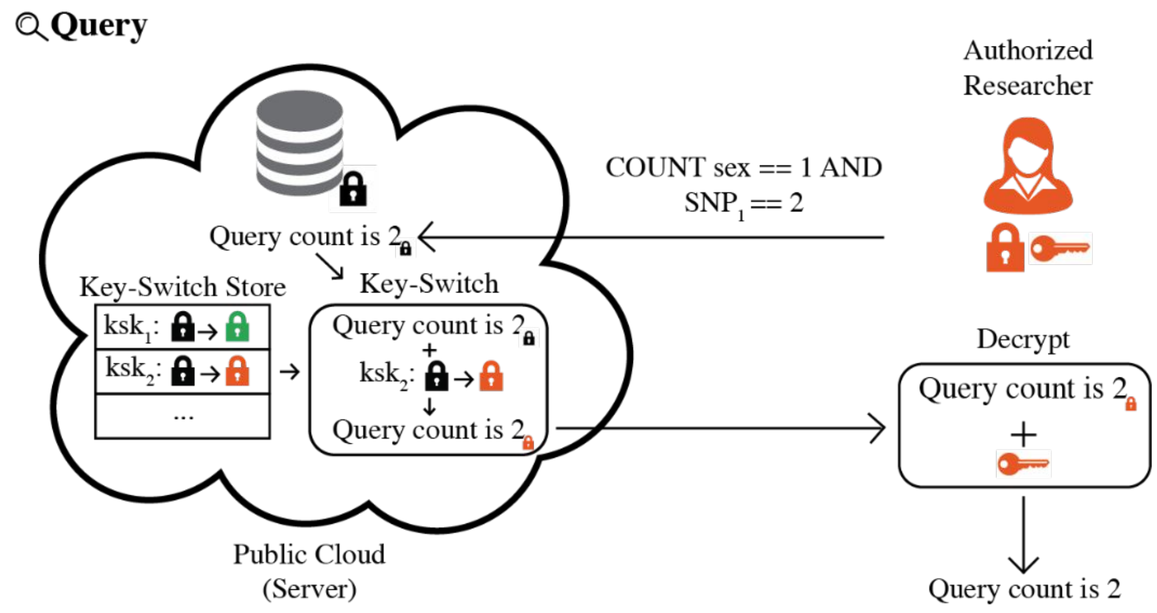


Results

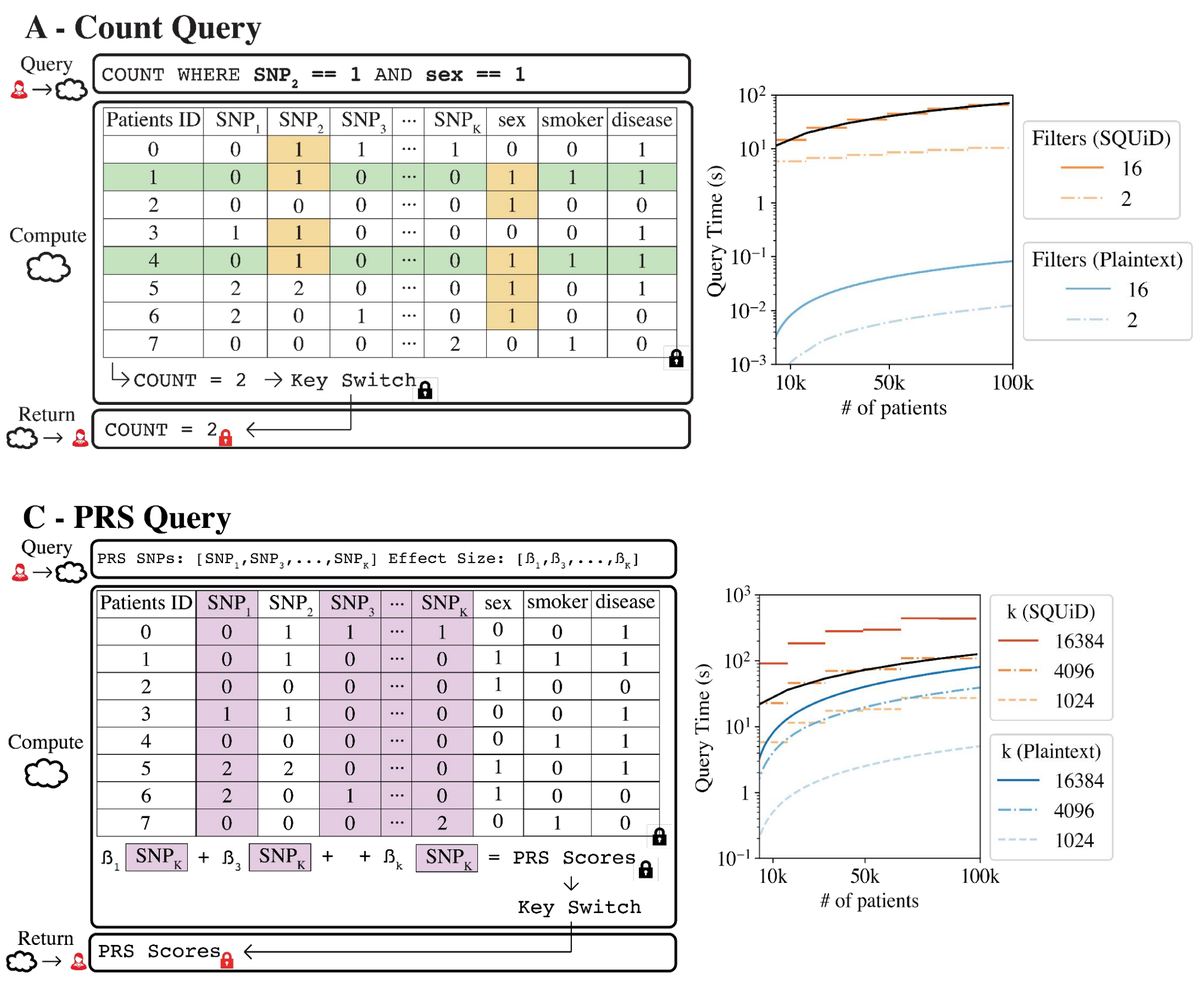
To implement an encrypted database using homomorphic encryption (HE), several challenges need to be addressed.

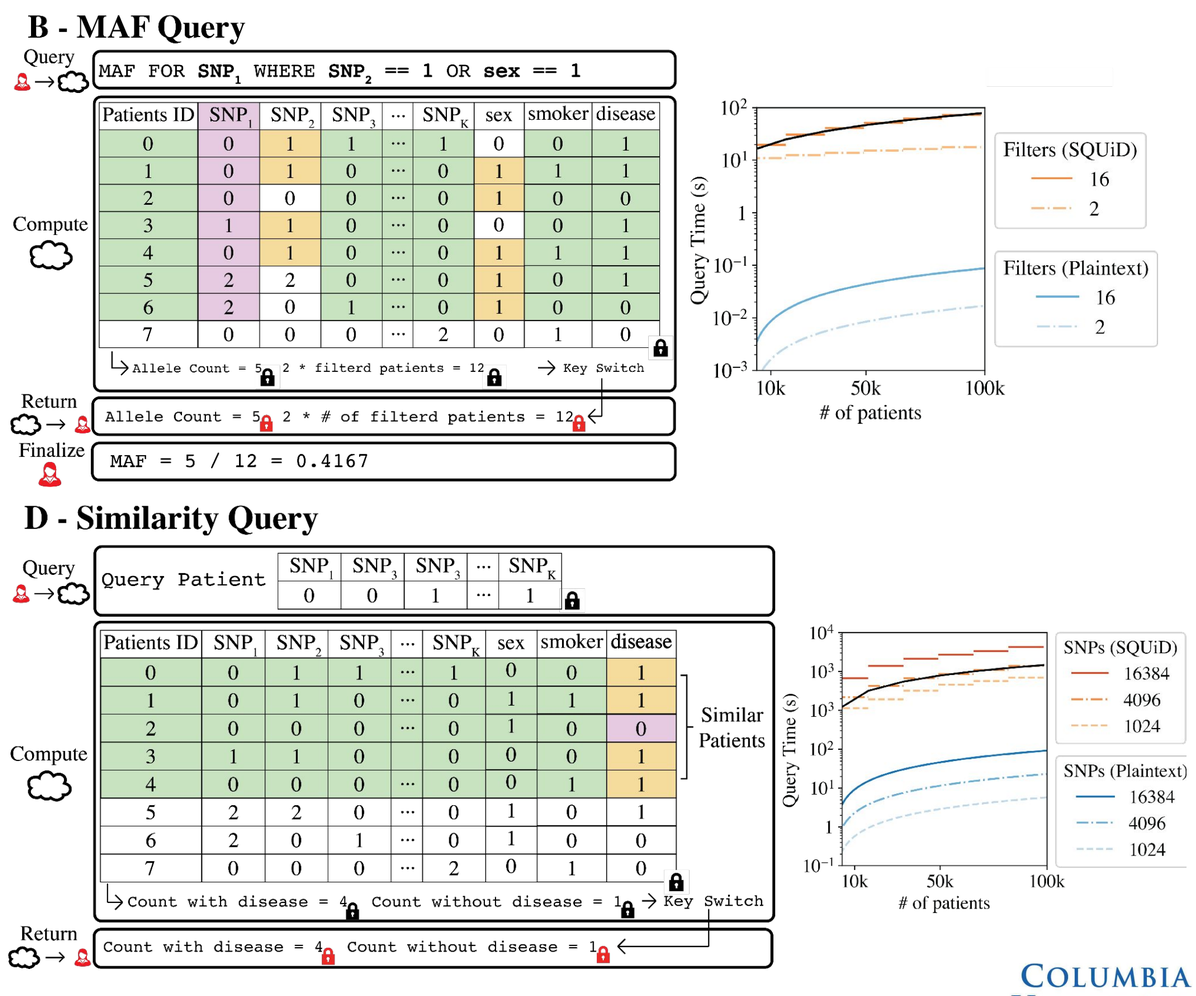
HE is designed for two parties: Using public key-switching, multiple researchers can access the database without sharing private keys. Furthermore, the data owner can control researcher access.





HE has limited practical functionality: SQUiD provides four useful queries that analyze genotype phenotype data.





Jacob Blindenbach
Columbia University

COLUMBIA UNIVERSITY

SQUiD: Ultra-Secure Storage and Analysis of Genetic Data for the Advancement of Precision Medicine

Jacob Blindenbach, Jiayi Kang, Seungwan Hong, Caline Karam, Thomas Lehner, Gamze Gürsoy

Abstract

Cloud computing provides the opportunity to store the ever-growing genotype-phenotype data sets needed to achieve the full potential of precision medicine. However, due to the sensitive nature of this data and the patchwork of data privacy laws, additional security protections are proving necessary to ensure data privacy and security. Here we present SQUiD, a Secure QUeryable Database for storing and analyzing genotype-phenotype data. With SQUiD, genotype-phenotype data can be stored in a low-security, low-cost public cloud in the encrypted form, which researchers can securely query without the public cloud ever being able to decrypt the data.

SQUiD utilizes homomorphic encryption to enable direct computations on encrypted data, introducing improvements that bolster security while also increasing practicality. Firstly, we developed a new cryptographic primitive, public key-switching, to allow for multi-client interactions with the database without ever sharing sensitive private keys. Secondly, we used ciphertext packing to decrease the storage and query runtime by orders of magnitude. Thirdly, we developed practical protocols to enable four queries commonly used with genotype-phenotype datasets. Specifically, SQUiD can count the number of patients that pass a filter, compute the minor allele frequency (MAF) for SNPs in a filtered patient cohort, compute polygenic risk scores for patients in the database, and perform patient similarity analysis. We demonstrate the usability of SQUiD by replicating various commonly used calculations such as cohort creation for GWAS, MAF filtering, and patient similarity analysis both on synthetic and UK Biobank data.

Jacob Blindenbach
Columbia University

COLUMBIA UNIVERSITY