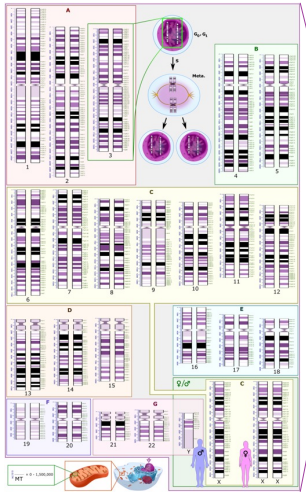
Towards the Deployment of Secure Computation Tools in Genomics

A Sociotechnical Perspective

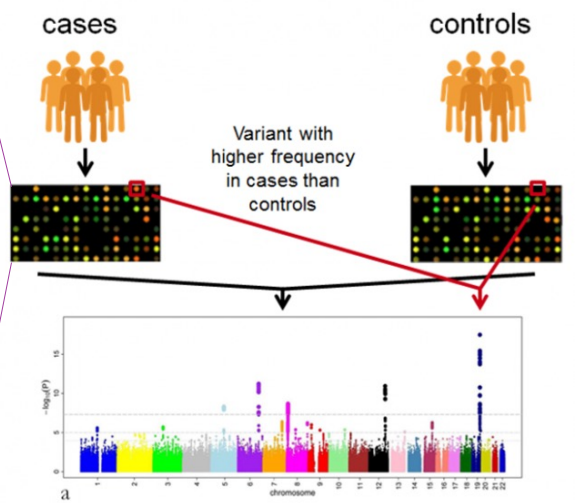
Natnatee “Ko” Dokmai, Ph.D.
Yale School of Medicine
Broad Institute of MIT and Harvard (formerly),
Indiana University Bloomington (formerly)

Privacy Challenges in Large-Scale Human Genome Research



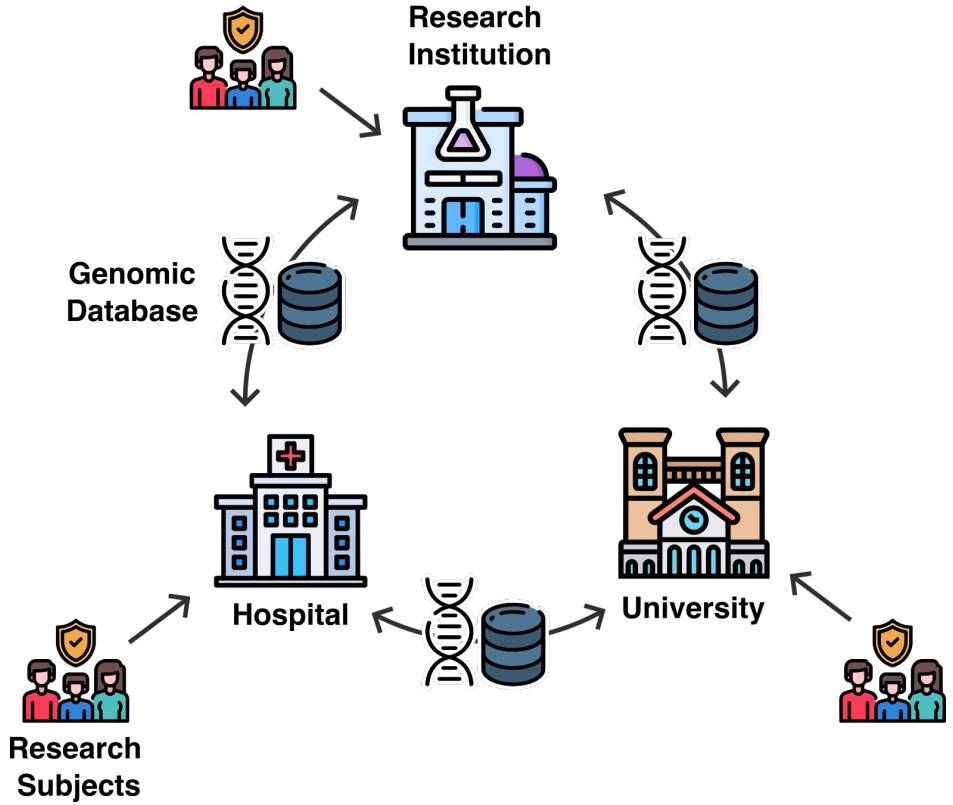
Human Genome

https://en.wikipedia.org/wiki/Human_genome#/media/File:Human_karyotype_with_bands_and_sub-bands.png



Genome-Wide Association Studies

https://www.ebi.ac.uk/training/online/courses/gwas-catalogue-exploring-snp-trait-associations/what-is-gwas-catalog/what-are-genome-wide-association-studies-gwas/



Privacy Challenges in Large-Scale Human Genome Research

Concerns regarding leakage of sensitive genomic data

↓

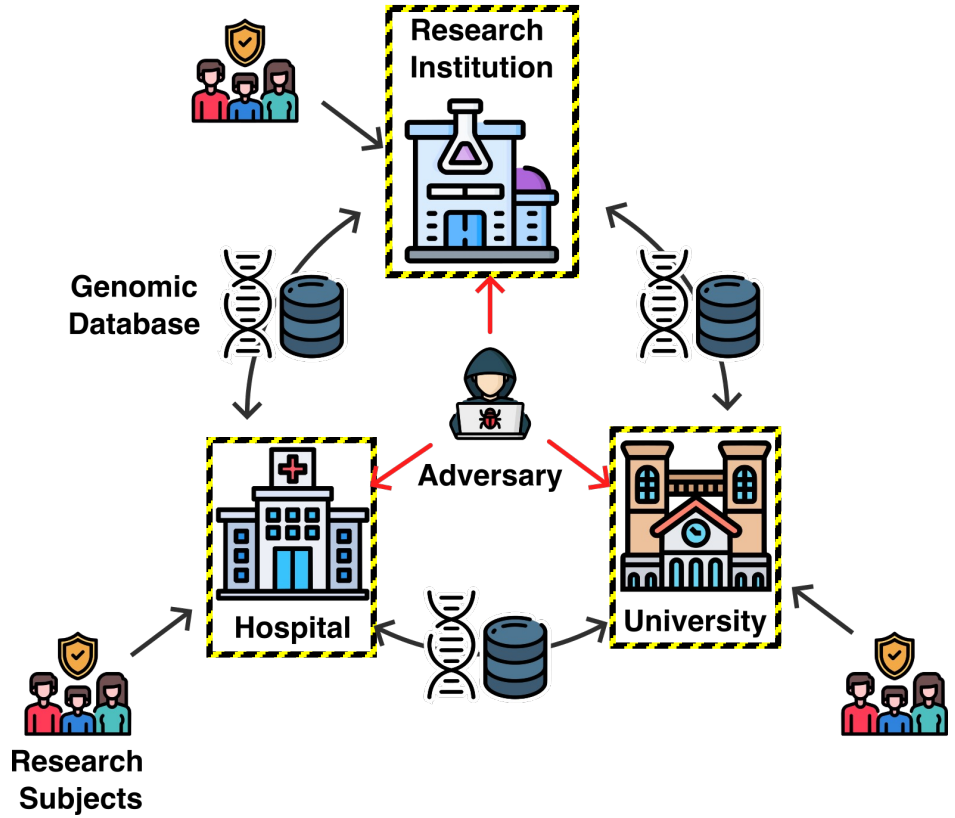
Data usage restriction

↓

Data mobility challenge in genomic research

Emerging Secure Computation Solutions

- Secure Multiparty Computation (MPC)
- Fully Homomorphic Encryption (FHE)
- Trusted Execution Environments (TEE)

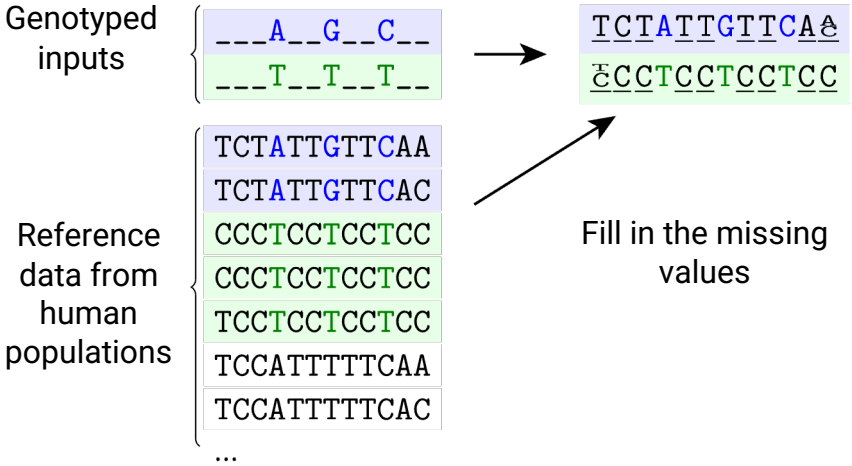


Our Technical Approach

Privacy-Preserving Genotype Imputation using Intel SGX*

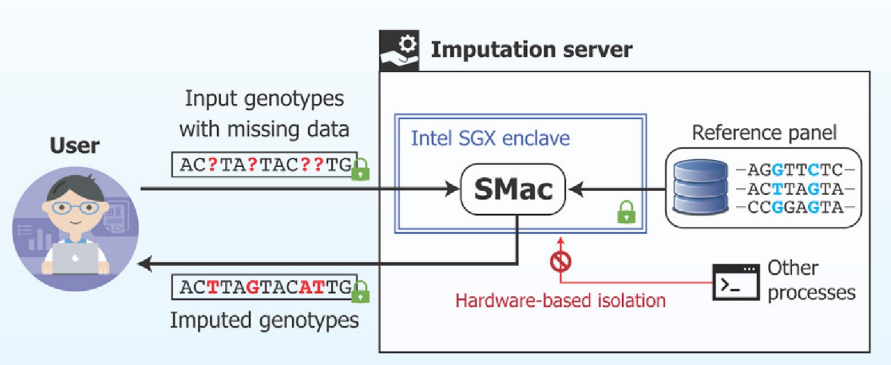
* Dokmai *et al.*, 2021, "Privacy-preserving genotype imputation in a Trusted Execution Environment", *Cell System* 12, 983–993

Genotype Imputation



Hoffmann and Witte, *Trends in Genetics*, 2015

Secure Genotype Imputation using SMac



Intel SGX (TEE Technology)

Runtime isolation and memory encryption against malicious OS

SMac

Novel side-channel resilient techniques while retaining efficiency and accuracy with state-of-the-art tools

Deployment and Adoption Challenges in Genomics

Current status

Secure computation tools are not currently in use in genomic research

Adoption barriers

Computational inefficiency, poor usability, organizational unawareness, and lack of economic incentives

Beyond existing assumptions

Is secure computation the right tool for collaborative genomic research?

Key idea: Contextual Integrity

Nissenbaum, *Privacy in Context*, 2009

Context matters

Privacy norms and expectations are context-dependent

Information flow

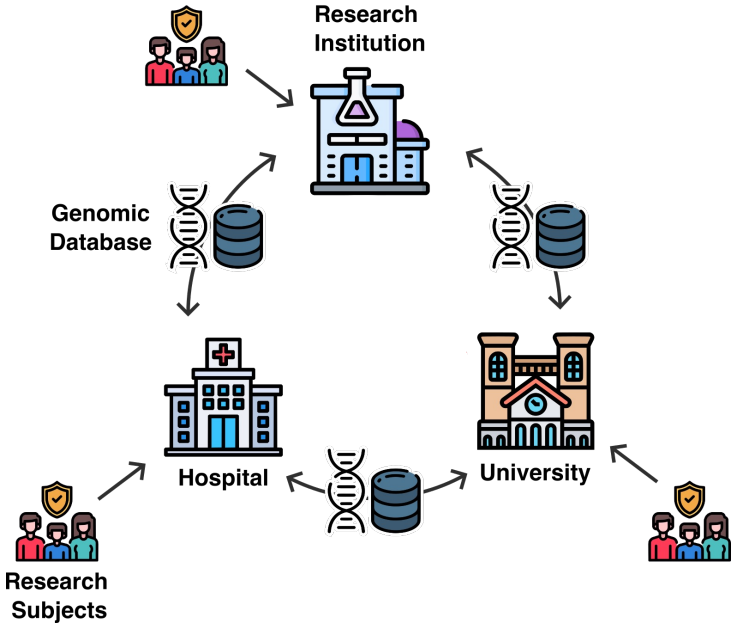
Focuses on how it aligns with established norms within a given context

Normative framework

An action is privacy-invasive if it disrupts the established norms for information flow

Analyzing Collaborative Genomic Research Context

Scientific Research or Healthcare?



Scientific Research

Purposes
Scientific results

Norms
Collaboration, transparency, reputation-based

Healthcare

Purposes
Diagnosis and treatments

Norms
Patient privacy and wellness

Shared norms
Trust, bioethics, accountability

Tension

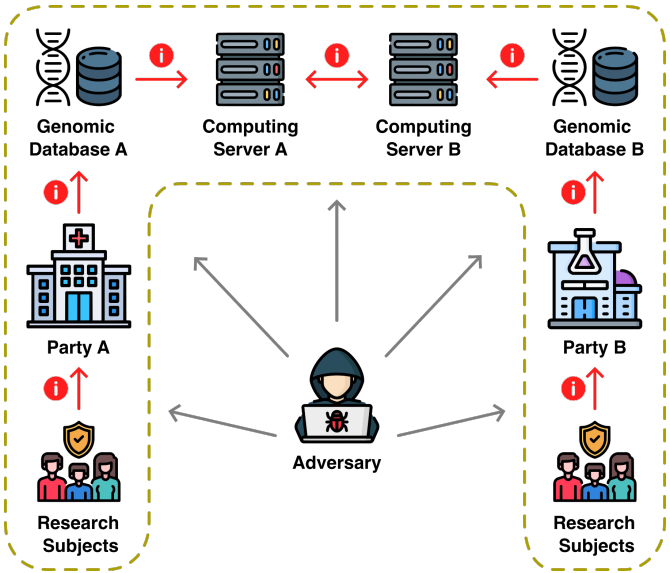
Institutions **vs** Research subjects

Publishable results **vs** Actionable healthcare

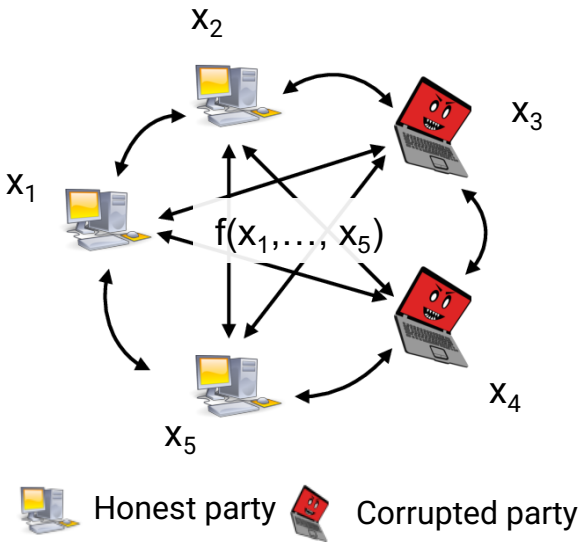
Transparency **vs** Privacy

How does applying secure computation tools for research disrupt collaborative genome analysis context?

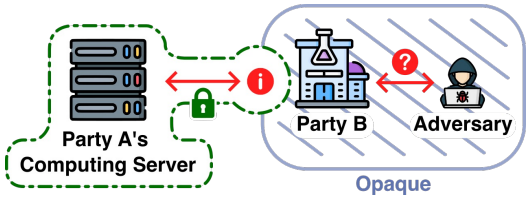
Secure Computation Threat Model Disrupts Genomic Context



An example of information flow and attack surfaces in collaborative genome research

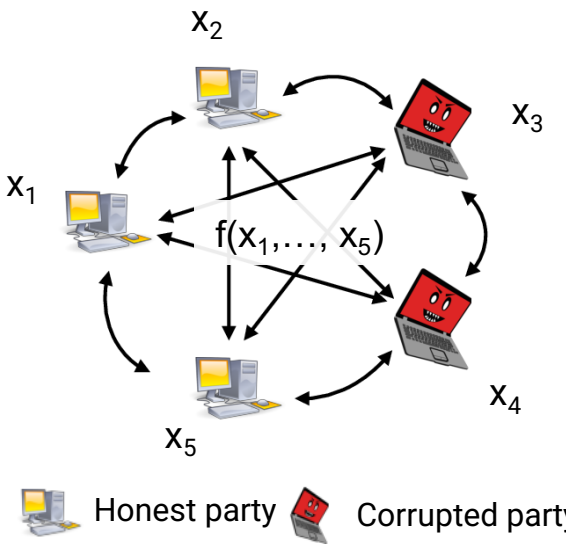


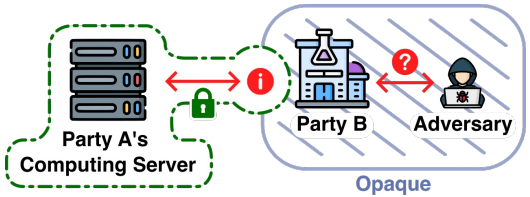
Secure computation threat model



Information flow interrupted by secure computation

Secure Computation Threat Model Disrupts Genomic Context





Institutional Trust → Distrust

Requires skepticism of other instutions' **scientific integrity** and **institutional capacity** to protect sensitive data

Patient Privacy → Institutional Data Security

Focuses on institutions as computing parties while **research subjects become invisible**

Collective → Individual Pursuit of Security and Privacy

Focuses on honest vs. corrupted parties instead of **security of the entire system**

Secure computation threat model

Information flow interrupted by secure computation

Contextual norms altered by secure computation

Our Goal

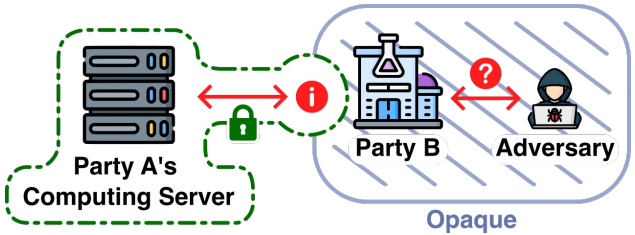
Reframing Secure Computation Framework in Genomics

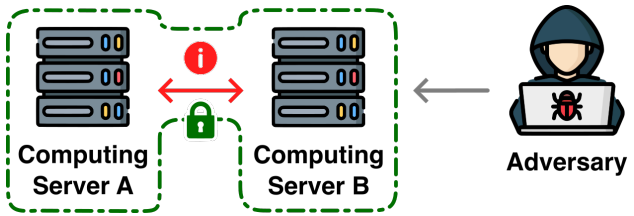
*We want a secure computation framework that **facilitates adoption** by **respecting contextual norms and purposes**, and enabling **risk analysis on the entire system***

that is,

*by aligning the tools with stakeholders' and practitioners' understanding of **institutional trust** and **risk management practices** and prioritizing **research subjects' privacy***

Move #1: A Trust-Based Secure Computation Framework





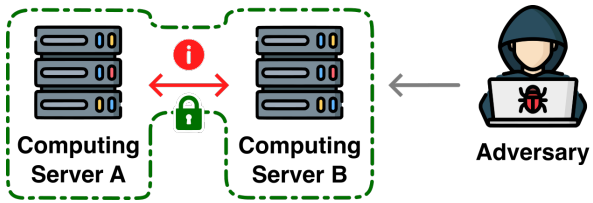
Distrust-based framework

- Institutions **distrusting** and working against each other
- Adversary and corrupted institutions are **not distinguished**
- **Convoluted** information flow
- **Security** preserved under certain assumptions
- **Vulnerable hardware:** adversary may have unrestrained access to performing attacks on hardware

Trust-based framework

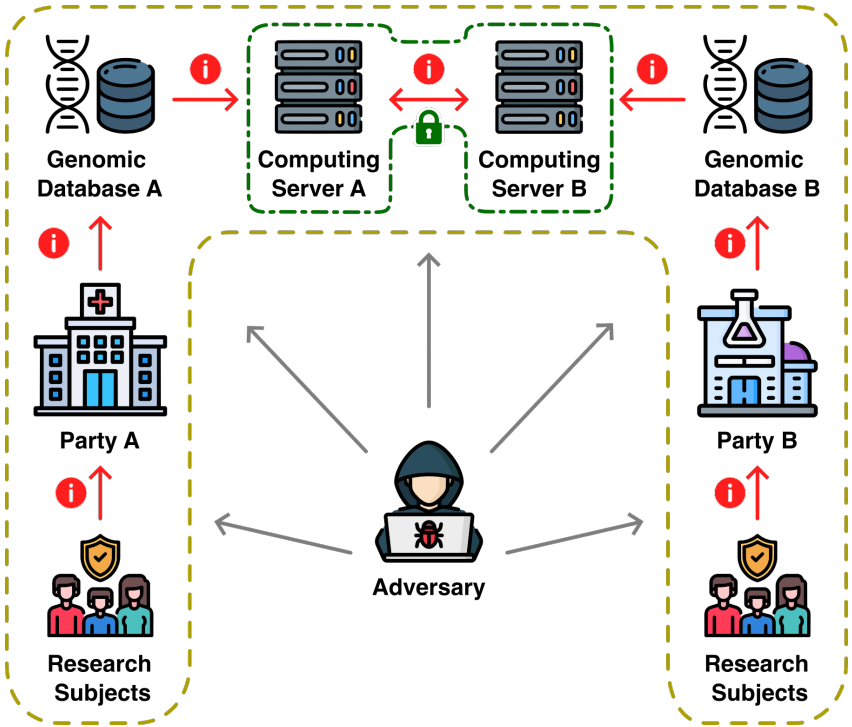
- Institutions **trusting** each other and working together to protect the system
- Adversary is **external** to the context and motived to compromise contextual integrity
- **Explicit** information flow (i.e., sender, receiver, information type, transmission principle)
- **Leakage control** in a security breach
- **Protected hardware:** institutions are willing to protect hardware against adversary

Move #2: Putting Secure Computation in the Information Flow



Trust-based framework

- Renders visible information flow and attack surfaces in a unified system
- Makes explicit the properties of secure computation as a risk mitigation tool
- Highlights research subjects as a data sender distinguished from the institutions
- Enables us to understand the consequence of secure computation on research subjects' privacy

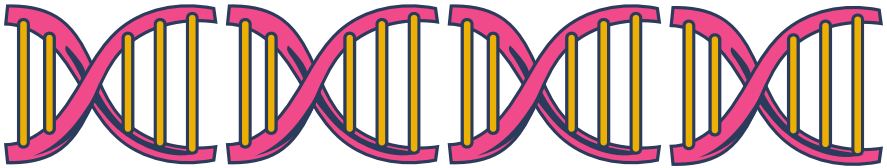


Secure computation as a risk mitigation tool

Ongoing Work

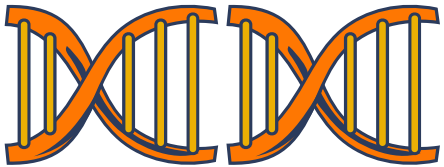
Reframing

a range of existing secure computation tools in genomics to align with practitioners' understanding of institutional trust and risk management practices



Deployment

of our privacy-preserving genotype imputation tool in Intel SGX



Designing

new privacy-preserving tools in genomics following the improved framework

Acknowledgement

Dr. Hoon Cho
Yale School of Medicine

Dr. Jean Camp
Indiana University Bloomington

Dr. Cenk Sahinalp
National Cancer Institute

Matthew Mosca
University of Edinburgh

Attribution

- Slide template: Slidesgo and Freepik
- Images: Flaticon.com

Questions?

Email: natnatee.dokmai@yale.edu