Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

# Genomic Data Analysis with Variant of Secure Multi-Party Computation Technique

**Manas Kumar Yogi[1], Yamuna Mundru[2]**

[1]Assistant Professor, Department of Computer Science and Engineering, Pragati Engineering College (A), Surampalem

[2]Assistant Professor, Department of Computer Science and Engineering (AI & ML), Pragati Engineering College (A), Surampalem

**Email:** [1]manas.yogi@gmail.com, [2]yamunamundru@gmail.com

## Abstract

The increasing availability of genomic data for research purposes necessitates innovative approaches to ensure privacy while facilitating collaborative analysis. This study explores the integration of a variant of Secure Multi-Party Computation (SMPC) techniques into genomic data analysis. The conventional challenges of sharing sensitive genetic information among multiple entities, such as research institutions or healthcare providers, are addressed by leveraging advanced cryptographic protocols. The research focuses on the development and implementation of a secure framework for collaborative genomic data analysis using an adapted SMPC variant. This variant is designed to efficiently handle the complexities of genetic data while ensuring robust privacy preservation. By encrypting individual genomic inputs and enabling computations without revealing the raw data, the proposed SMPC variant facilitates joint analyses, contributing to advancements in personalized medicine, disease research, and genetic epidemiology. The variants of SMPC, namely oblivious transfer protocol, is used, this allows the receiver to obtain one out of several pieces of information forwarded by the sender without revealing which one they obtained. It can be integrated into SMPC protocols for enhancing the privacy with less effort and cost. The proposed mechanism involves the validation of the SMPC variant through simulations using


*DOI: https://doi.org/10.36548/jtcsst.2023.4.006*
*Received: 27.11.2023, received in revised form: 21.12.2023, accepted: 07.01.2024, published: 20.01.2024*

real-world genomic datasets and assessing its performance in terms of computational efficiency and privacy preservation. Results from experiments demonstrate the feasibility and effectiveness of the proposed technique in enabling secure multi-party genomic data analysis. This research contributes to the evolving landscape of privacy-preserving techniques in genomics, offering a promising avenue for collaborative research without compromising the confidentiality of sensitive genetic information.

**Keywords:** Secure, Multi-Party, Privacy, Genetic, Attack, Security.

## 1. Introduction

Secure Multi-Party Computation (SMPC) is a cryptographic technique that allows multiple parties to jointly compute a function over their inputs while keeping those inputs private. The fundamental principles of SMPC are based on cryptographic protocols and techniques to ensure privacy, security, and correctness. Below mentioned are the key principles [1-2]:

- **Privacy-Preserving Protocols**

SMPC ensures that the inputs of each party remain confidential throughout the computation. Cryptographic techniques, such as encryption and secret sharing, are employed to achieve this. Parties jointly perform computations on encrypted or shared data without revealing the actual values to each other.

- **Cryptographic Primitives**

SMPC relies on various cryptographic primitives, including homomorphic encryption, secret sharing schemes, and oblivious transfer. Homomorphic encryption allows computations on encrypted data, secret sharing divides data among parties in a way that collaboration is required for reconstruction, and oblivious transfer ensures one party obtains information without the other knowing what was transferred.

- **Consistency and Correctness**

SMPC ensures that the final result of the computation is correct and consistent with what would be obtained if all parties revealed their inputs. Protocols are designed to guarantee accuracy despite the fact that parties do not have access to each other's private information.

- **Zero-Knowledge Proofs**

Zero-knowledge proofs are cryptographic tools that allow one party to prove to another that they know a particular piece of information without revealing the information itself. In SMPC, zero-knowledge proofs may be used to verify that parties are following the protocol correctly without disclosing their inputs.

- **Secure Channels**

Communication channels between parties need to be secure to prevent eavesdropping or tampering. The use of secure channels, often established through cryptographic protocols like TLS (Transport Layer Security), ensures the confidentiality and integrity of messages exchanged during the SMPC process.

- **Multiparty Trust Model**

SMPC assumes a multiparty trust model where no single party should be able to learn more than what is allowed by the protocol. Even if some parties collude, they should not gain additional information about the private inputs.

- **Efficiency Considerations**

While ensuring privacy and security, SMPC protocols also need to be designed with efficiency in mind. The goal is to minimize computational and communication overhead to make secure multi-party computations feasible for real-world applications.

Genomic data analysis plays a crucial role in various fields, contributing to advancements in medicine, biology, personalized healthcare, and scientific research. The several key aspects highlighting the importance of genomic data analysis is discussed below [3-4]:

- **Understanding Genetic Basis of Diseases**

Genomic data analysis allows researchers and healthcare professionals to identify genetic variations associated with various diseases. This knowledge is fundamental for understanding the genetic basis of diseases, enabling more accurate diagnoses, prognosis predictions, and targeted treatment strategies.

- **Personalized Medicine**

Genomic data analysis facilitates the development of personalized medicine by considering individual genetic variations. This approach tailors medical treatments to an individual's genetic makeup, increasing treatment efficacy and minimizing adverse effects. It allows for the identification of specific drug targets and the prediction of patient responses to different therapies.

- **Disease Risk Prediction and Prevention**

By analyzing genomic data, researchers can identify genetic markers associated with an increased risk of certain diseases. This information is valuable for preventive measures, early intervention, and lifestyle modifications to reduce the risk of disease development.

- **Biomedical Research and Drug Development**

Genomic data analysis is integral to biomedical research, aiding in the discovery of new genes, pathways, and potential therapeutic targets. It accelerates drug development by providing insights into the genetic basis of diseases and guiding the design of targeted therapies.

- **Genomic Epidemiology**

Genomic data analysis contributes to understanding the spread and evolution of infectious diseases. It helps trace the origin of outbreaks, identify transmission patterns, and develop strategies for disease control and prevention.

- **Population Genetics and Evolutionary Studies**

Studying genomic data across populations provides insights into genetic diversity, migration patterns, and evolutionary history. This information is crucial for understanding human evolution, migration, and adaptation to different environments.

- **Forensic Genetics**

Genomic data analysis is used in forensic investigations to establish identity, parentage, and kinship. DNA profiling and analysis contribute to solving criminal cases, identifying missing persons, and providing evidence in legal proceedings.

- **Reproductive Health and Genetic Counseling**

Genomic data analysis is employed in assessing reproductive risks, identifying genetic disorders, and offering genetic counseling to individuals and families. It helps in making informed decisions about family planning and reproductive choices.

- **Biomarker Discovery**

Genomic data analysis aids in the identification of biomarkers associated with diseases or treatment responses. These biomarkers can be used for diagnostic purposes, monitoring disease progression, and evaluating treatment efficacy.

- **Precision Agriculture**

In agriculture, genomic data analysis is utilized for crop improvement, disease resistance, and optimizing breeding programs. It contributes to the development of crops with improved yield, nutritional content, and resilience to environmental challenges.

### 1.1 Increasing Availability of Genomic Data

Advancements in genomic technologies, such as next-generation sequencing, have led to a rapid increase in the generation and availability of genomic data. Large-scale genomic projects, biobanks, and initiatives like the Human Genome Project have contributed to vast repositories of genetic information. This wealth of data holds great potential for scientific research, personalized medicine, and understanding the genetic basis of diseases.

### 1.2 Need for Privacy-Preserving Techniques

Genomic data is inherently sensitive, containing unique and personal information about an individual's genetic makeup. As the availability of genomic data grows, so does the need to address privacy concerns. Traditional approaches to data sharing and analysis might risk the exposure of identifiable information, leading to ethical and legal challenges. Preserving privacy while allowing collaborative analysis and data sharing becomes paramount in order to encourage participation in research and maintain public trust.

### 1.3 Concept of Secure Multi-Party Computation (SMPC)

Secure Multi-Party Computation (SMPC) is a cryptographic technique that allows multiple parties to jointly compute a function over their inputs while keeping those inputs private. In the context of genomic data analysis, SMPC provides a way for multiple entities or organizations to collaboratively analyze genomic information without sharing the raw data. It ensures that computations are performed in a secure and privacy-preserving manner, even when parties do not fully trust each other.

### 1.4 Relevance to Genomic Data Analysis

The application of SMPC in genomic data analysis addresses the dichotomy between the increasing demand for collaborative research and the necessity to protect individual privacy. By utilizing SMPC protocols, researchers and organizations can perform computations on shared genomic data without directly accessing the raw genetic information. This enables collaborative studies, such as identifying common genetic variants across populations or studying the genetic basis of diseases, while mitigating the risk of privacy breaches.

SMPC is particularly relevant in scenarios where different institutions or research groups want to combine their genomic datasets for a broader analysis without exposing individual-level data. This approach ensures that the benefits of large-scale genomic studies can be realized without compromising the privacy and confidentiality of the individuals contributing their genetic information.

## 2. Related Work

The motivation for employing a variant of SMPC in genomic data analysis arises from the imperative to balance the need for collaborative research and insights with the equally critical requirement to preserve the privacy and confidentiality of sensitive genetic information. Traditional approaches to data sharing often involve centralized databases or data pooling, which can raise significant privacy concerns. The motivation and suitability for utilizing a variant of SMPC in the context of genomic data analysis is stated below [5-7]:

- **Preserving Individual Privacy**

Genomic data is inherently sensitive and personally identifiable. Individuals may be hesitant to share their genetic information due to privacy concerns. A variant of SMPC allows parties to collectively analyze genomic data without revealing the raw genetic information, ensuring the privacy of individuals contributing to the dataset.

- **Collaborative Research across Institutions**

Genomic research often involves collaboration among multiple institutions, each possessing valuable datasets. SMPC facilitates secure collaboration by allowing these institutions to jointly analyze genomic data without the need to share the actual genetic information. This is particularly relevant in multi-center studies and international collaborations.

- **Compliance with Data Protection Regulations**

Data protection regulations, such as general data protection regulation (GDPR) in Europe, emphasize the importance of safeguarding individuals' privacy. SMPC aligns with these regulations by enabling collaborative genomic analysis while ensuring compliance with privacy standards and regulations.

- **Facilitating Cross-Organization Data Sharing**

In scenarios where different organizations or entities possess portions of a larger genomic dataset, a variant of SMPC allows them to collaboratively analyze the integrated data

without the need for a central repository. This decentralized approach enhances security and reduces the risk of data breaches.

- **Addressing Security Concerns in Cloud Environments**

Many organizations leverage cloud computing for data storage and processing. SMPC is suitable for secure genomic data analysis in cloud environments, addressing concerns related to data security and unauthorized access. The cryptographic protocols involved ensure that the data remains confidential during computations.

- **Enabling Patient-Centric Healthcare**

With the increasing emphasis on personalized and patient-centric healthcare, individuals may wish to contribute their genomic data for research purposes without compromising their privacy. SMPC allows for secure and privacy-preserving analysis, encouraging broader participation in genomic research.

- **Robustness against Insider Threats**

SMPC is designed to withstand collusion or malicious behavior by a subset of parties. This is crucial in scenarios where some collaborating entities may have conflicting interests or where there is a need to guard against insider threats.

- **Flexibility in Data Ownership and Control**

SMPC provides a flexible framework where data owners retain control over their information while still participating in collaborative analysis. This decentralized approach respects data ownership and allows for more equitable data sharing arrangements.

- **Scalability for Large-Scale Genomic Studies**

The variant of SMPC selected for genomic data analysis should be scalable to accommodate the complexities and large-scale nature of genomic datasets. This ensures that the method remains viable and efficient as the volume of genomic data grows.

**Table 1.** Existing Mechanisms for Genomic Data Analysis [5-9]

| Technique | Merits | Demerits |
|---|---|---|
| Traditional Centralized Approaches | Simplicity in implementation. | Privacy concerns due to centralized storage. |
| | Easy data sharing within the organization. | Risk of data breaches and unauthorized access. |
| | Faster computation in a centralized environment. | Limited collaboration in multi-institutional studies. |
| Differential Privacy | Strong privacy guarantees for individual data. | Potential loss of data utility due to noise injection. |
| | Allows for statistical analysis while protecting privacy. | Challenges in setting appropriate privacy parameters. |
| | Well-established theoretical foundations. | Complex implementation and potential performance impact. |
| Homomorphic Encryption | Allows computations on encrypted data. | Computational overhead, especially for complex operations. |
| | Strong security guarantees. | Limited support for certain types of computations. |
| | Protects data confidentiality during analysis. | Key management complexities. |
| Secure Multi-Party Computation | Joint computation without revealing raw data. | Computational and communication overhead. |
| | Suitable for collaborative analysis. | Complexity in protocol design and verification. |
| | Robust against collusion and insider threats. | Challenges in scaling to large datasets. |
| Blockchain-Based Solutions | Immutable and transparent record-keeping. | Scalability challenges for large genomic datasets. |
| | Enhanced security through decentralized consensus. | Latency in data validation and transaction processing. |
| | Potential for establishing trust in data provenance. | Resource-intensive consensus mechanisms. |

## 3. Proposed Mechanism

Secure multi-party computation (SMPC) is a cryptographic technique that allows multiple parties to jointly compute a function over their inputs while keeping those inputs private. In the context of genomic data analysis, it can be used to perform computations on sensitive genomic data without revealing the raw data to any party involved.

A variant of Secure Multi-Party Computation (SMPC) protocol tailored for genomic data analysis, addresses the challenges associated with sharing sensitive genetic information:

- **Privacy Preservation**

Challenge: Genetic information is highly personal, and individuals may be hesitant to share their genomic data due to concerns about privacy and the potential misuse of sensitive information.

Solution: The SMPC variant employs cryptographic techniques to allow collaborative analysis without revealing the raw genetic data. Through techniques like homomorphic encryption or secure function evaluation, computations can be performed on encrypted data, ensuring the privacy of individual genomic information.

- **Secure Collaboration**

Challenge: Traditional methods of sharing genomic data often involve centralizing the data, which raises security concerns and requires trust in a central authority.

Solution: The SMPC variant enables secure collaboration among multiple parties without the need for a trusted intermediary. Each party can contribute its data, and the computations are performed in a distributed manner, ensuring that no single entity has access to the complete set of raw genetic information.

- **Data Ownership and Control**

Challenge: Individuals may be reluctant to contribute their genomic data if they feel they lose control over how it's used or shared.

Solution: SMPC allows data owners to retain control over their raw genetic data. Since computations are performed on encrypted or private shares of the data, individuals can participate in collaborative research without relinquishing ownership or control over their genetic information.

The formulation of high-level design for a mathematical model using a variant of SMPC for genomic data analysis is shown below:

### 3.1 Problem Statement

Let's consider a scenario where multiple parties (hospitals, research institutions, etc.) each possess genomic data from patients. The parties want to collaboratively analyze the data to derive meaningful insights without sharing the raw genomic information.

### 3.2 Objectives

1. Genomic Data Processing: Enable joint computation of statistical measures (e.g., average, variance) and analytical operations (e.g., association analysis, genetic variant detection) on genomic data [10].

2. Privacy Preservation: Ensure that raw genomic data is not shared among parties, and the final results reveal only the necessary aggregated information.

### 3.3 Mathematical Model

#### a) Data Representation

Let G_i represent the genomic data of party i. Each genomic dataset is represented as a set of genomic features, such as single nucleotide polymorphisms (SNPs), gene expressions, or other relevant genetic markers.

#### b) Homomorphic Encryption

In the context of homomorphic encryption for genomic data analysis, below process is involved:

- **Encryption Function (Enc)**

The encryption function takes a plaintext message m and a public key pk as inputs and produces the ciphertext c.

$$c=Enc_{pk}(m) \quad (1)$$

- **Decryption Function (Dec)**

The decryption function takes a ciphertext c and the corresponding private key sk to recover the original plaintext message m.

$$m=Dec_{sk}(c) \quad (2)$$

- **Homomorphic Addition**

The homomorphic addition allows the addition of two cipher texts c1 and c2 to obtain a cipher text that decrypts to the sum of the corresponding plaintexts.

$$Enc_{pk}(m1)+Enc_{pk}(m2)=Enc_{pk}(m1+m2) \quad (3)$$

- **Homomorphic Multiplication**

The homomorphic multiplication enables the multiplication of a ciphertext c by a constant k to obtain a ciphertext that decrypts to the product of the original plaintext and the constant.

$$k \cdot Enc_{pk}(m)=Enc_{pk}(k \cdot m) \quad (4)$$

- **Homomorphic Evaluation of Functions**

Some homomorphic encryption schemes support the evaluation of specific mathematical functions on encrypted data. For example, given a function f(x), homomorphic encryption allows the computation of f(m) on encrypted data.

$$Enc_{pk}(f(m))=f(Dec_{sk}(Enc_{pk}(m))) \quad (5)$$

- **Bootstrapping**

In some fully homomorphic encryption schemes, a bootstrapping operation is used to refresh the ciphertext, reducing noise and allowing for additional computations without degradation in security.

$$\text{Bootstrapping}(\text{Enc}_{pk}(m)) \quad (6)$$

Utilize homomorphic encryption to enable computation on encrypted data. This allows parties to perform operations on encrypted genomic data without decrypting it. The result, when decrypted, is equivalent to the result of operations on the unencrypted data [11].

**c) Protocol Steps**

- **Key Generation**

Each party generates a pair of public and private keys for homomorphic encryption.

Model is as below:

$$(pk_i, sk_i) = \text{KeyGen}() \quad (7)$$

Where $pk_i$ is the public key and $sk_i$ is the private key for party i.

- **Data Encryption**

Each party encrypts its genomic data using the public keys of all parties involved.

Mathematical Model:

$$\text{EncryptedData}_i = \text{Enc}_{pk_1}(\text{GenomicData}_i) \quad (8)$$

Each party i encrypts its genomic data ($\text{GenomicData}_i$) using the public key ($pk_1$) of the first party. This process is repeated for all parties involved.

- **Secure Summation (Aggregation)**

Parties collaboratively perform encrypted addition to calculate the sum of their encrypted data.

**Mathematical Model**

$$\text{EncryptedSum}=\sum \text{EncryptedData i where i=1 to n.} \quad (9)$$

Parties perform an encrypted summation to obtain the sum (EncryptedSum) of all encrypted genomic data.

**d) Statistical Analysis**

Perform statistical computations (e.g., mean, variance) on the encrypted data. Utilize protocols for secure computation (e.g., secure addition, secure multiplication) [12-13].

Model (for Mean):

$$\text{EncryptedMean}=n\text{EncryptedSum} \quad (10)$$

Model (for Variance):

$$\text{Assuming EncryptedData}i-\text{EncryptedMean for each party i.} \quad (11)$$

$$\text{EncryptedVariance} = \frac{\sum_{i=1}^{n}(\text{EncryptedData i}-\text{EncryptedMean})}{n} \quad (12)$$

**e) Results Sharing**

The final result, derived from the encrypted computation, is shared among parties.

Parties securely share the encrypted statistical results among themselves.

**f) Decryption**

Parties use their private keys to collectively decrypt the final result, revealing the aggregated statistical measures without exposing individual genomic data.

Mathematical Model

$$(\text{EncryptedResult})\text{DecryptedResult}=\text{Decsk1}(\text{EncryptedResult}) \quad (13)$$

Each party uses its private key (sk1) to collectively decrypt the final result (Decrypted Result). The Figure.1 below shows the flow diagram of the proposed method.

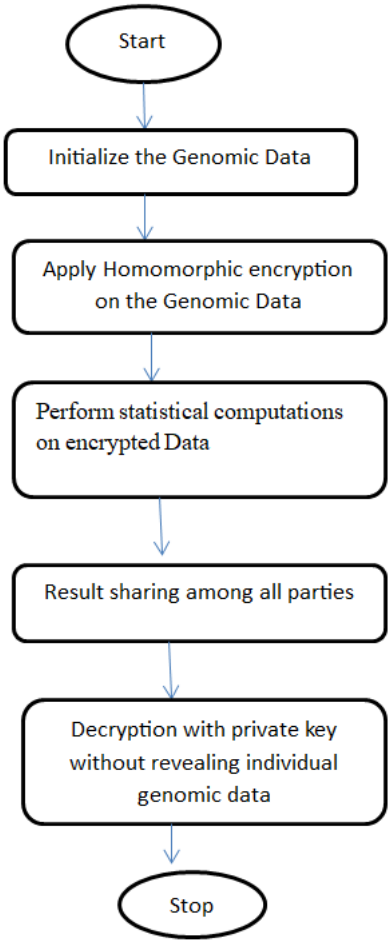


**Figure 1.** Flow Diagram of Proposed Method

## 4. Experimental Results

The benchmark dataset available in GitHub is used for verification of genomic sequence for a human or a worm [16]. The dataset used contains 1345 samples of genome sequence related to both human and a worm. The 70% of the dataset is used for training and 30% of the sample from this dataset is used for testing. The Python and Jupyter note book were used in implementing the proposed mechanism, The libraries PyTorch and TensorFlow were used to enable secure multi-party computations. The Python module Matplotlib was utilized to produce graphs. Below results prove the application of the proposed mechanism is prudent enough to save effort and time for a complex operation like genomic data analysis.

### 4.1 Validation Process

#### a) Data Preparation

Out of the 1345 samples present in the dataset we have divided into training and testing sets and then pre-processing has been performed to achieve consistency.

#### b) Training the Model

This training phase involved adjusting the model's parameters to learn patterns and relationships within the data which indicate the genomic data sequence is of humans or worms.

#### c) Validation Set

In addition to the training set, a separate validation set is used for fine-tuning the model and preventing overfitting. This set is distinct from the testing set and helps optimize the model's performance.

#### d) Testing the Model

The final evaluation is performed on a separate 30% of the rest of the samples in the dataset that the model has not seen during training or validation. This provides an unbiased assessment of the method's performance on new, unseen data.

### 4.2 Performance Metrics

#### a) Accuracy

The accuracy of the proposed method helps in advocating its application in genomic data analysis. The result in Figure.2 proves that the SMPC variant provides a robust trade-off when compared to other popular and traditional methods.

#### b) Computational Efficiency Metrics

Here we have used the amount of resource utilization and processing time to evaluate the performance of the proposed method with traditional method used in genomic data analysis. The results observed in terms of privacy preservation, accuracy, data utility and the resource utilization are depicted in the figures 3-5.

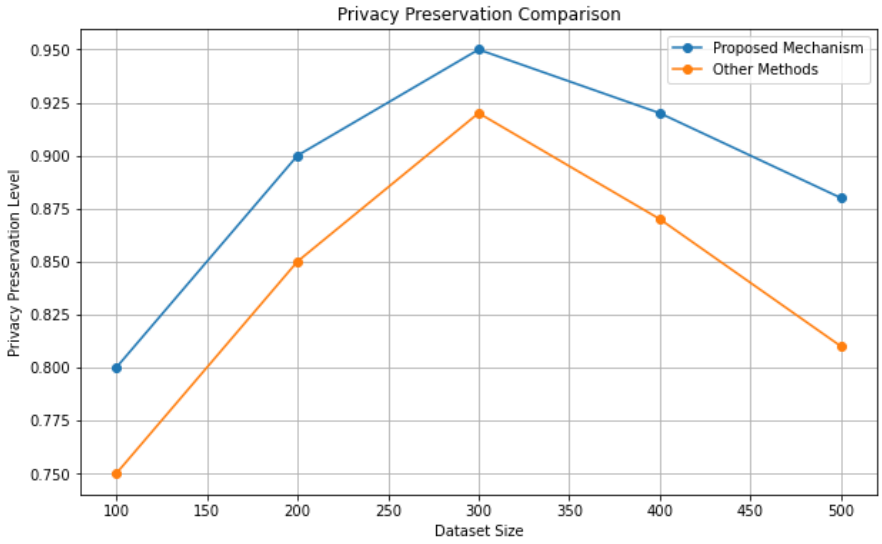


**Figure 2.** Degree of Privacy Preservation using Proposed Mechanism

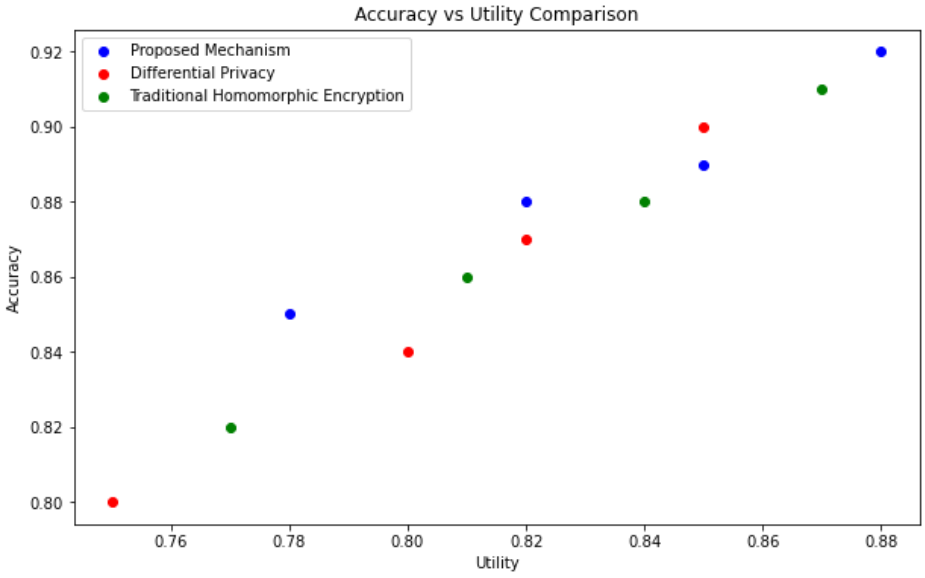


**Figure 3.** Accuracy versus Data Utility Comparison

The results obtained in Figure.3 assists in making decisions about the level of privacy preservation and its subsequent effect on the reliability. The accuracy versus data utility ratio is important because it helps determine a balance between preserving the privacy of sensitive genetic information and ensuring that the findings acquired are valuable.

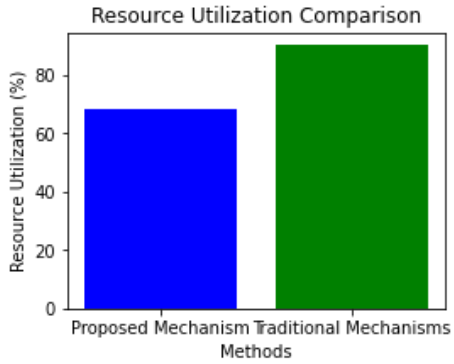


**Figure 4.** Comparison of Resource Utilization

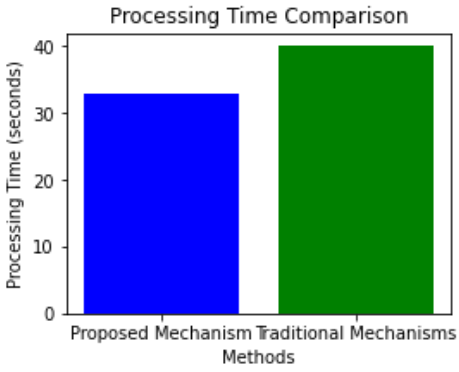


**Figure 5.** Comparison of Processing Time

The results in Figure 2 and 3 indicate the usefulness within the extent of any regulatory framework thereby proving the effectiveness of the proposed method. Figures 4 and 5 show the comparative performance with respect to the resource utilization and processing time. The resource utilization is nearly 15% to 18% less and the processing time is also nearly 8% to 10% less. This indicates that the application of the proposed method is also effective in real time scenario.

Sharing genomic data is very sensitive in nature and hence the proposed method can be applied to protect the sensitivity as observed from figure 2. The degree of privacy preservation is high when compared to other traditional methods. The complexity of existing methods in terms of resource utilization and processing time is also not so high as shown by the experimental results in figure 4and 5. The accuracy of the system along with its utility is also optimal as shown in figure 3. This proves the applicability of the proposed method is effective in real applications of genomic data analysis.

## 5. Future Scope

### a) Scalability and Efficiency Improvements

Future research could focus on enhancing the scalability and efficiency of secure multi-party computation (SMPC) mechanisms for genomic data analysis. This involves developing more efficient cryptographic protocols, exploring parallelization techniques, and optimizing algorithms to handle larger datasets and involve more parties. Addressing scalability challenges is crucial for the widespread adoption of privacy-preserving genomic data analysis in large-scale studies.

### b) Integration with Advanced Cryptographic Techniques

Explore the integration of advanced cryptographic techniques, such as homomorphic encryption, functional encryption, and zero-knowledge proofs, with SMPC for genomic data analysis. Combining these techniques may lead to more robust privacy-preserving solutions while providing additional security guarantees. Investigate the trade-offs and synergies between different cryptographic primitives to design more sophisticated and secure genomic data analysis frameworks.

### c) Dynamic and Adaptive Security Models

Develop dynamic and adaptive security models that can adjust to the evolving nature of genomic data and emerging privacy threats. Research could focus on mechanisms that dynamically adapt the level of privacy-preserving measures based on the sensitivity of the data, the trustworthiness of the involved parties, and the current security landscape. This could lead to more resilient and flexible solutions capable of responding to changing privacy and security requirements.

## 6. Conclusion

The proposed SMPC mechanisms offer a viable solution to the intricate challenge of conducting meaningful analyses on genomic datasets distributed across multiple entities. By encrypting the genomic data and enabling computations on the encrypted data, SMPC ensures that individual-level information remains confidential, addressing concerns related to data privacy and security. The cryptographic protocols employed in these variants, such as

homomorphic encryption and secure summation; provide a robust foundation for collaborative genomic research. Future research directions in this domain are poised to propel the field further. Enhancements in scalability and efficiency are crucial for accommodating the ever-expanding volumes of genomic data and involving a growing number of parties in analyses. Integrating advanced cryptographic techniques, exploring dynamic and adaptive security models, and fostering interdisciplinary collaboration are key avenues for further exploration. As the field evolves, standardization efforts will be imperative to ensure the seamless integration of SMPC mechanisms into genomics research workflows and healthcare applications. The potential impact of secure multi-party computation in genomics extends beyond the research realm, reaching into personalized medicine, drug discovery, and clinical diagnostics. As the broader scientific community continues to recognize the importance of privacy-preserving methodologies in genomics, the adoption of SMPC variants is likely to witness a surge.

## References

[1] Cho, Hyunghoon, David J. Wu, and Bonnie Berger. "Secure genome-wide association analysis using multiparty computation." Nature biotechnology 36.6 (2018): 547-551.

[2] Zhao, Chuan, et al. "Secure multi-party computation: theory, practice and applications." Information Sciences 476 (2019): 357-372.

[3] Blanton, Marina, and Fattaneh Bayatbabolghani. "Efficient server-aided secure two-party function evaluation with applications to genomic computation." Cryptology ePrint Archive (2015).

[4] Huang, Zhicong. On Secure Cloud Computing for Genomic Data: From Storage to Analysis. No. THESIS. EPFL, 2018.

[5] Sousa, João Sá, et al. "Efficient and secure outsourcing of genomic data storage." BMC medical genomics 10.2 (2017): 15-28.

[6] Blatt, Marcelo, et al. "Secure large-scale genome-wide association studies using homomorphic encryption." Proceedings of the National Academy of Sciences 117.21 (2020): 11608-11613.

[7] Evans, David, Vladimir Kolesnikov, and Mike Rosulek. "A pragmatic introduction to secure multi-party computation." Foundations and Trends® in Privacy and Security 2.2-3 (2018): 70-246.

[8] Evans, David, Vladimir Kolesnikov, and Mike Rosulek. "A pragmatic introduction to secure multi-party computation." Foundations and Trends® in Privacy and Security 2.2-3 (2018): 70-246.

[9] Asvadishirehjini, Aref, Murat Kantarcioglu, and Bradley Malin. "A Framework for Privacy-Preserving Genomic Data Analysis Using Trusted Execution Environments." In 2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), pp. 138-147. IEEE, 2020..

[10] Aziz, Md Momin Al, et al. "Privacy-preserving techniques of genomic data—a survey." Briefings in bioinformatics 20.3 (2019): 887-895.

[11] Froelicher, David, Juan R. Troncoso-Pastoriza, Jean Louis Raisaro, Michel A. Cuendet, Joao Sa Sousa, Hyunghoon Cho, Bonnie Berger, Jacques Fellay, and Jean-Pierre Hubaux. "Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption." Nature communications 12, no. 1 (2021): 5910.

[12] Saadeh, Angelo. Applications of secure multi-party computation in Machine Learning. Diss. Télécom Paris, 2023.

[13] Jain, Shreyan. Developing a cloud-based secure computation platform for genomics research. Diss. Massachusetts Institute of Technology, 2020.

[14] Abinaya, B., and S. Santhi. "A survey on genomic data by privacy-preserving techniques perspective." Computational Biology and Chemistry 93 (2021): 107538.

[15] Pascoal, Túlio. Secure, privacy-preserving and practical collaborative Genome-Wide Association Studies. Diss. University of Luxembourg, Luxembourg, 2022.

[16] https://github.com/ML-BioinfoCEITEC/genomic_benchmarks/tree/main/datasets/demo_human_or_worm on her name.