DOI: 10.24996/ijs.2023.64.9.35

ISSN: 0067-2904

Review of Challenges and Solutions for Genomic Data Privacy-Preserving

Hiba M. Yousif [1,2], Sarab M. Hameed[2]

[1]College of Engineering, University of Information Technology and Communications (UoITC), Baghdad, Iraq

[2] Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 16/4/2021 Accepted: 8/10/2022 Published: 30/9/2023

Abstract

The dramatic decrease in the cost of genome sequencing over the last two decades has led to an abundance of genomic data. This data has been used in research related to the discovery of genetic diseases and the production of medicines. At the same time, the huge space for storing the genome (2–3 GB) has led to it being considered one of the most important sources of big data, which has prompted research centers concerned with genetic research to take advantage of the cloud and its services in storing and managing this data. The cloud is a shared storage environment, which makes data stored in it vulnerable to unwanted tampering or disclosure. This leads to serious concerns about securing such data from tampering and unauthorized searches by those involved. In addition to securing inquiries, making calculations on this data, and generating differential privacy and garbled circuits, cryptography is considered one of the important solutions to this problem. This paper introduces most of the important challenges related to maintaining privacy and security and classifies each problem with appropriate, proposed, or applied solutions that will fuel researchers' future interest in developing more effective privacy-preserving methods for genomic data.

Keywords: security, privacy, genomic, homomorphic encryption

مراجعة التحديات والحلول السائدة لخصوصية البيانات الجينومية

هبة محمود يوسف[2,1] , سراب مجيد حميد[2]

[1]كلية الهندسة ، جامعة تكنولوجيا المعلومات والاتصالات (UoITC) ، بغداد ، العراق

[2] قسم علوم الحاسب . كلية العلوم . جامعة بغداد . بغداد . العراق

الخلاصة:

أدى الانخفاض الكبير في تكلفة تسلسل الجينوم خلال العقدين الماضيين إلى وفرة البيانات الجينومية. تم استعمال هذه البيانات في الأبحاث المتعلقة باكتشاف الأمراض الوراثية وإنتاج الأدوية. في الوقت نفسه ، أدت المساحة الضخمة لتخزين الجينوم (2–3 جيجا بايت) إلى اعتباره أحد أهم مصادر البيانات الضخمة ، مما دفع مراكز الأبحاث المعنية بالبحوث الجينية إلى الاستفادة من السحابة و خدماتها في تخزين وإدارة هذه البيانات. السحابة هي بيئة تخزين مشتركة ، الامر الذي يجعل البيانات المخزنة فيها عرضة للتلاعب أو الكشف غير المرغوب فيه ، مما أدى إلى مخاوف جدية بشأن تأمين هذه البيانات من التلاعب وعمليات البحث غير المصرح بها من قبل المعنيين. بالإضافة إلى تأمين الاستفسارات وإجراء الحسابات على هذه البيانات. يعتبر التشفير أحد الحلول المهمة لهذه المشكلة ، بالإضافة إلى الخصوصية التفاضلية والدوائر المشوشة. تقدم هذه الورقة معظم

*Email: heba.mahmoud1201a@sc.uobaghdad.edu.iq

التحديات المهمة المتعلقة بالحفاظ على الخصوصية والأمن وتصنف كل مشكلة بحلول مناسبة أو مقترحة أو تطبيقية من شأنها أن تغذي اهتمام الباحثين المستقبلي بتطوير طرق أكثر فعالية للحفاظ على الخصوصية للبيانات الجينية.

1. Introduction

The official announcement of the completion of the human genome project in 2003 drew attention to the importance and sensitivity of genomic data [1]. The tremendous development of gene sequencing technology has resulted in a massive amount of genomic information, which is considered the clue to many diseases' comprehension [2]. By using next-generation sequencing technologies, the growth of genomic data has become exponential in that the data volume is starting to reach petabytes [3]. It is appropriate for this huge amount of data to be stored in the cloud, provided that its security and privacy are guaranteed [4]. Cloud services are considered a wonderful and distinctive technology that provides dynamic and scalable services via the Internet. They are in increasing demand as technology develops [5]. However, it is vulnerable to attacks, which lead to data leaking to unwanted parties [6].

Storing, sharing, managing, and performing analysis on this data should be accomplished by insecure means, which ensure that the data is never lost or exposed to misuse [7]. Revealing the genome sequence of individuals leads to the possibility of genomic discrimination (even if it is prohibited), as well as the undesired revelation of sensitive information (biological family, medical history, or sensitive illness status) as a result of these breaches. Because they share most of their genetic DNA, the breadth of such injury might extend to offspring or relatives of the affected individuals. Furthermore, unlike the accounts of users and passwords (commonly hacked by information technology businesses), it is impossible to change the genetic information once it has been exposed [8]. However, disclosure of genomic data to an untrusted third party has substantial privacy implications [9].

In this paper, we present a detailed study on genomic data privacy preserving challenges and methods to preserve privacy through various techniques. The main contribution of the review can be precisely given as follows:

- Present an overview of the prominent challenges facing privacy-preserving for genomic data and discussed the important research in this area, with classification and a brief discussion of these challenges.
- This paper discusses different methods of solving the genomic data security and privacy problems using different perspectives of collaboration among Homomorphic encryption (HE), Garbled Circuit (GC), and Differential Privacy (DP).

In the rest of this review, the challenges of preserving the privacy of genomic data are described in section 2. Considering the need for more significant solutions is coupled with understanding the difficulties that should be addressed accurately. This is followed by reviewing different techniques for preserving genomic data privacy in Section 3. Finally, the major conclusions realized from this review are clarified.

2. Genomic data privacy-preserving: challenges

The processes of digital genomic data make it vulnerable to disclosure. The main operations which may violate the privacy of these data are sequence alignment, searching the database of genomics, and querying private genomic data. If necessary countermeasures are not taken, it is possible to violate the privacy of personal data [1]. The distribution of the related works published in various journals from 2014 to 2021 is summarized in Figure 1. Figure 2 shows a general taxonomy for different types of genomic data privacy challenges. The taxonomy aims to understand the challenges in protecting genetic data that represent different levels of difficulty**.** The following subsections present how these challenges add different levels of difficulties to the genomic data preservation challenge.

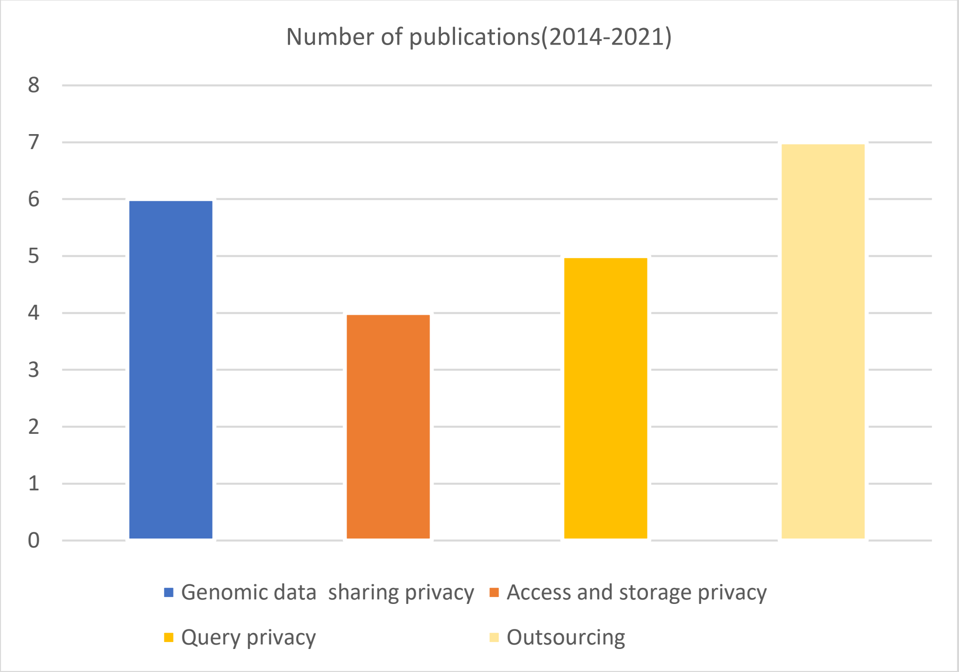

Figure 1: Distribution of the published papers in different journals between 2014–2021.

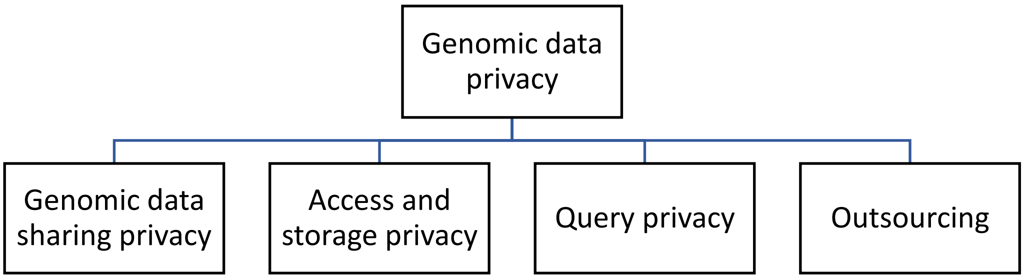

Figure 2: Genomic data challenges

2.1 Genomic data sharing privacy

Genomic data sharing processes can be categorized into public and private, each with its own set of access controls and rules. Various data identification attacks have increased the security and privacy of genomic data, which is not used for public sharing and has no privacy guarantee. These assurances, however, fall short for a variety of reasons (various adversary assumptions, different threat models/attacks) [6]. So, data security and privacy have grown to be a crucial necessity for many enterprises [10]. Bos et al. [11] discussed the available scenarios for the available applications in this field. This research paper sheds light on the issue of homomorphic encryption and demonstrates its importance and effectiveness by providing a practical application for the prediction service that works in the cloud on encrypted data. The application includes a cloud service to execute private predictive analysis jobs on health data that is encrypted using Some What Homomorphic Encryption (SWHE). The cloud service handles the encrypted data only and predicts without any knowledge of the secret medical data.

Lauter et al. [12] employed Genome-Wide Association Studies (GWAS) and its basic algorithms with HE to work with encrypted data. They discovered a variety of statistical algorithms that were evaluated through lower-order polynomials, such as the goodness of fit (Pearson) test or the Chi-Square statistical test, which were used to check for divergence from Hardy-Weinberg equilibrium. Cheon et al. [13] devised an approach for performing the edit distance technique on encrypted genomic sequences. The approach generates an encrypted value of their edit distance. They implement their proposed algorithm of edit distance over encrypted genomic data with lengths n and m by the SWHE scheme. The optimization in their algorithm was reducing the depth of computing edit distance for short sequences. On the other hand, Simmons et al. [14] made a discovery for the privacy-preserving technique of aggregate sharing of genomic data with low-cost accuracy from the traditional differentially private method in a way making the trade-off between accuracy and privacy by varying λ parameter in the Laplacian distribution.

To facilitate the privacy-preserving partnership of genomic data for GWAS, a decentralized network using the privacy-preserving sharing protocol (PPS) and the data fragmentation algorithm was proposed by Zhang et al. [15], which was restricted to a limited number of fragments. We noticed that Yang et al. [16] suggested a scheme to share medical data based on attribute cryptosystem and blockchain technology, which involved storing encrypted medical data in the cloud while storing the storage address and medical-related information in the blockchain, ensuring storage efficiency and removing the opportunity of data amendment irreversibly. The suggested technique integrates Attribute-Based Encryption (ABE) with Attribute-Based Signature (ABS), allowing medical data to be shared over many-to-many communications. Table 1 outlines the prominent methods for the privacy of genomic data sharing.

Table 1: The prominent methods for the privacy of genomic data sharing

Reference	Year	Technique(s)	Advantage(s)	Disadvantage(s)	Dataset
[11]	2014	Fully homomorphic encryption (FHE)	present an implementation for a cloud service that demonstrates an application of algorithms for outsourced prediction on encrypted, confidential medical data	For a low-degree computation, it might not be beneficial to implement the modulus switching technique that is required for deeper circuits	Polynomial function
[12]	2015	HE and the basic genomic algorithm used in GWAS	Efficient when applied to modest data	-Cannot be applied to big data. -not practical for different keys.	genotype and/or phenotype count tables
[13]	2015	SWHE and edit distance	the proposed algorithm performs analysis on the encrypted genomic sequence without privacy leakage	implementation over large parameters cannot be done due to large memory requirements	DNA sequences
[14]	2019	Perturbation, Bayesian, Markov chain, monte Carlo	A trade-off between accuracy and privacy by varying λ parameter in Laplacian distribution	The method depends on the addition of noise to calculate the risk	GWAS dataset

[15]	2019	Decentralized secure network and privacy-preserving protocol	-provide a protective method to solve the problem of data re-distribution that satisfies people's fundamental ethical interests in their data -prevent data re-identification problems by providing a high level of privacy, with the ability to perform GWAS analysis -fragmentation mechanism used provides scalability and decentralized analysis	- efficient when N (number of fragments) is low. when N increase, the accuracy of the method decreased	1000 Genomes Project
[16]	2020	Blockchain technology and attribute cryptosystem	High computational performance (lower computational overhead in encryption, signature, and decryption	/	Medical data

2.2 Access and storage privacy

Genomic data must be stored in a secure place, ensuring that there is no exposure, tampering, or disclosure from untrusted parties. Huang et al. [17] proposed a method called Selective Retrieval on Encryption and Compressing Reference-Oriented Alignment Map (SECRAM) that is used for compressing aligned data, storage and retrieval of encrypted data, and efficiency improvement during downstream analysis. Despite the high efficiency in saving space and preserving privacy provided by this method, when the coverage is low, we notice a decline in performance. Meanwhile, Liu et al. [18] suggested including the implementation of VA-Store, an approach that uses K-mers (i.e., a subsequence of length k) with various k values to address the substantial space required for repeated data in common genomic sequence analysis tasks. For one component of the input dataset, the VA-Store maintains a physical store while supporting several stores for other portions of the information.

VA-store has translated a given query on the virtual store into one or more queries on the physical store can be executed by utilizing essential linkage among repetitious data. Although it saves space and maintains privacy as well, it degrades in performance when a sequence of length less than k0 (the first subsequence length) comes along. On the other hand, Chen et al. [19] presented a framework for large-scale calculations on genomic data that was outsourced to a third-party (public cloud server) for greater scalability and security. Furthermore, the tree structure was used by them to exemplify arbitrary genomic data for computational competency and integrated homomorphic cryptography with the Garbled Circuit approach to ensure security. Although it provides a significant improvement in run time for the execution of queries, it requires an additional cost in the event of dishonest researchers and is exposed to security leaks during the search operation. Mehmood et al. [20] proposed an indexed-based method to answer queries of pathways scattered over various distributed datasets. They offered a heuristic-based source selection method for determining which datasets are appropriate for a given route query as well as a strategy for federating queries to select sources and assembling(merging) the paths obtained from those distant datasets. Table 2 illustrates the prominent methods for access and storage privacy.

Table 2: The prominent methods for access and storage privacy

Reference	Year	Technique(s)	Advantage(s)	Disadvantage(s)	Dataset
[17]	2016	SECRAM	-space saving -privacy-preserving	Compression is not efficient when the coverage is very low because of the storage overhead of encryption.	NA12878
[18]	2018	Virtual approximation store	mainly designed to save storage space for repetitive genome sequence data	the efficiency will suffer, when the chosen k length for an input user k-mer query is too far from the length k0 chosen for the physical store k0 of the underlying VA-store,	Streptomyces rapamycinicus NRRL 5491 genome
[19]	2018	HE, GC and, Tree indexing	-significant improvement in run time for the execution the of query	- It requires an additional cost in the case of dishonest researchers -security leakage	Neo4j
[20]	2019	Indexed based approach	- provide a heuristic-based source selection mechanism to select the relevant datasets - assembles the paths retrieved from those remote datasets	Datasets must be interconnected and contain a recourses that are relevant to each other	-Disease -hpoClass -doClass - phenotype - Protein -Variant -Gene -panther

2.3 Query privacy

Researchers querying on genomic data preferred to be secure and not to be disclosed to others (attackers/curious), in that the query and the output of it hold sensitive information about the individuals. It is a challenge to ensure the privacy of the query and the result [21]. Alaziz et al. [22] adopted the Paillier cryptosystem and order-preserving encryption to execute the count query and the ranked query securely. Despite the advantage of this method that the time of performing calculations on encrypted data is close to the time taken by the same operations on unencrypted data, it is expensive when decrypting. On the other hand, Sousa et al. [5] employed Private Information Retrieval (PIR) with HE to invent a hash-based solution. Some changes have been made to the standard PIR protocol to access specific variants while its identification parameters such as chromosome, position, and reference allele can be used instead of the usage of its relative position in the Variant Call Format (VCF) file.

Moreover, they used symmetric encryption to protect genomic data on the server side. The aforementioned method is characterized by an error rate associated with its hashing scheme and is slow if the database is large and multiple variants or files are queried. while Xu et al. [23] resorted to guaranteeing the integrity of the query result and preserving the confidentiality of the data through the proposed authenticated aggregate queries over a set of valued data. They suggested a privacy-preserving authentication framework for overall queries. Mahboubi et al.

[24] suggested a system called Secure Distributed TOPK (SD-TOP-K) in which the user data is encrypted and stored in a distributed system and can be evaluated by a top-k query processing algorithm which finds a set of encrypted data that is proven to contain top-k data items. This is done without having to decrypt the data in the nodes where they are stored. Moreover, they suggested a robust filter in the algorithm that strips the false positives as much as possible without decrypting the data. Meanwhile, Quan et al. [25] suggested a method to reduce top-k query privacy leakage when compared to order-preserving encryption (OPE). Top Order Preserving Encryption (TOPE), which allows top-k searches on encrypted data using partially ordered heap characteristics for balancing privacy and search capabilities, is the essential method. Table 3 summarizes the prominent methods for query privacy.

Table 3: The prominent methods for query privacy

Reference	Year	Technique(s)	Advantage(s)	Disadvantage(s)	Dataset
[22]	2016	Paillier and order-preserving encryption	The computation time of the secure computations is closer to the time of the corresponding regular computations over the plaintext	Decryption overhead	IDASH 2015 SNP
[6]	2017	PIR and HE	-optimal privacy -confidentiality -low storage complexity -low querying time -minimization of delivered data -generality	-error rate associated with a hashing scheme -slow down if the database is relatively large - suffer from scalability issues	iDash 2016
[23]	2017	Merkle Grid Tree (MG-tree)	-provide integrity and confidentiality for aggregate queries over set-valued data. - formal security analysis and cost models were provided for the proposed authentication protocols and algorithms	in max, top-k ,and FFQ queries, the performance breakdown severely due to dimensionality curse	-Personal Genome Project at Harvard Medical School - Foodmarket from Microsoft (FoodMarket) - TPC Benchmark H (TPC-H)
[24]	2018	Top k- query	-evaluate top k-queries over encrypted distributed data without decrypting -propose a new filtering algorithm	Need a lot of communications between cloud nodes	Gowalla database
[25]	2018	Order preserving encryption	- enable top-k queries on encrypted data with minimized privacy leakage - reduce the running time of generating a massive ciphertext significantly by supporting a dynamic dataset and supporting	Encryption time and generating hash table increasing linearly	Diabetes dataset

	batch encryption in the setup phase

2.4 Outsourcing

The growing interest in outsourcing to manage data is due to faster implementation, flexible scalability, reduction of costs, and improved latency and connectivity. Although consumers must trust cloud service providers, this raises privacy and security concerns when it is related to research data from patients or volunteers. Several solutions have been proposed to address the security challenges, particularly in the area of data processing in the cloud. [22]. Zhang et al. [26] introduced the Fully Outsourced secuRe gEnome Study basEd on homomorphic Encryption (FORESEE) architecture for computing Chi-square statistics on the public cloud in a safe and completely outsourced manner. The so-called semi-honest opponent model assumes that the cloud properly follows the protocol but is interested in information from the received data.

Secure division operations can be provided by the suggested FORESEE framework with homomorphically encrypted data and immediate release of research findings from the cloud. Although the cost is very high and the efficiency is reduced due to the large value of G, it is still efficient in supporting full cloud outsourcing while maintaining final result encryption. Meanwhile, Wang et al. [27] suggested a new HEALER framework for evaluating the P-value of accurate logistic regression parameters applied to homomorphically encrypted data. Secure outsourcing was facilitated and the danger of sensitive data analysis was reduced in untrustworthy cloud environments (e.g., Amazon EC2, or Microsoft Azure). A new rejection sampling technique, secure integer comparison method, and parallelizable mechanism were introduced to speed up the execution of this algorithm, making homomorphic encrypted precise logistic regression computing feasible. Furthermore, a compression strategy was used to lower the cost of storing and communicating homomorphically encrypted data. The cost of computation and storage is still significant with some limitations in the proposal and distribution availability in the encrypted domain, which might lead to a low acceptance rate. There is the challenge of a homomorphic division operation.

Then Ghasaemi et al. [28] suggested a model for outsourcing data by employing a paillier cryptosystem with permutation. The method provides count query and top-k operations with an outperformance technique, but there is vulnerability to Homer attack and de-identification attack. Ziegeldorf et al. [29] employed Fully and Partially Homomorphic Encryption with a bloom filter (FHE and PHE–BLOOM). These approaches are efficient in genetic disease tests, which securely outsource the storage that has been allowed by the data owner and computed to the untrusted cloud. FHE–BLOOM provided full security in the semi-honest model, while PHE-BLOOM provided little qualification in guarantees of security in a trade-off for enhancing performance improvement. It provides flexible and efficient management supporting the outsourced data and may be extended to support further query types, but still suffers from overhead in the setup of the patient's database.

Hassan et al. [30] introduced a new approach for outsourcing genome data that is both safe and efficient from the aggregate genome data. The suggested approach created an index tree, which was subsequently outsourced to a third-party cloud server. The nodes of the tree have been scanned by the cloud server and perform count query operations using a secure interactive interface during the data processing phase as well as the query execution phase. This approach does not expose any crucial genomic data, does not provide privacy against inference attacks, nor data access privacy as it reveals the tree traversal pattern. Raisaro et al. [31] suggested and implemented a safe and efficient privacy-preserving approach in a real-world setting for investigating genomic cohorts by employing HE and DF at Lausanne University Hospital. It enables the exploration of large genomic datasets.

Kim et al. [32] demonstrated a safe outsourced method for evaluating logistic regression models for quantitative characteristics and testing their genetic connections. They use a semi-parallel training strategy to create a logistic regression model for variables, then run a one-step parallelizable regression on all single nucleotide polymorphisms (SNRs). They increase the performance of the underlying approximation homomorphic encryption algorithm.

Table 4: The prominent methods in the privacy of outsourcing

Reference	Year	Technique(s)	Advantage	Disadvantage	Dataset
[26]	2015	FORSEE based on HE	support complete outsourcing to the cloud and outputting the final encrypted result	-a large value for G to ensure accuracy in computation led to degradation in efficiency of the - suggested framework cost of storage and computation is still very high	2015 iDASH
[27]	2015	HE and logistic regression	-supports secure outsourcing and alleviates analyzing sensitive data in the cloud -provide a new approach for rejection sampling and secure methods for integer comparison	if suitable proposal distribution is not available in the encrypted domain, the acceptance rate would be low, the computational and storage cost over encrypted data is still significant, need to improve in storage efficiency, and still a challenge to handle homomorphic division operation	Kawasaki Disease datasets.
[28]	2016	Paillier on permutation	provide count query and top-k operation with outperformance technique in terms of execution time	vulnerability to Homer attack and de-identification attack	1000 genomes
[29]	2017	FHE, PHE and BLOOM filter	provide flexible and efficient management supporting of the outsourced data , and may be extended to support further query types	overhead in the setup of patients database	iDASH
[30]	2018	Paillier cryptosystem and BLOOM filter	-Secure and efficient in outsource genomic data -scalable for large data set	-Venerable to inference attack -does not provide data access privacy	Single Nucleotide Polymorphism SNP sequences
[31]	2018	HE and DF	-preserve privacy	the added noise by i2b2 server to new queries of given user	real genomic data coming from the

			-enable outsourcing and exploration of large genomic data	linearly grows with number of queries already answered to the same user. this can degrade the utility of the system in mater that the later queries would be useless	exome sequencing of 392 samples giving a genotyping for 472,845 variants each
[32]	2020	HE and logistic regression	- achieved a very high level of accuracy in the final output	-cannot compute matrix inverse if the dimension is high	iDASH

3. Privacy-preserving techniques

There are several solutions to preserve privacy and security challenges for genomic data. Homomorphic encryption, Garbled Circuit, and Differential Privacy (DP) are considered the most significant privacy-preserving techniques [22]. For all works presented in this review, the distribution of the three existing privacy-preserving techniques is depicted in Figure 2.

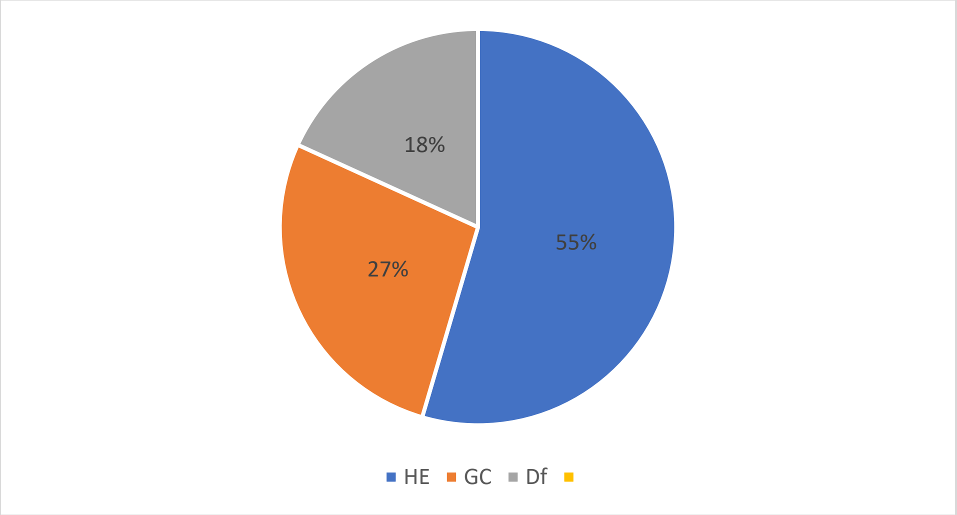

Figure 2: Distribution of preserving privacy for genomic during the period 2016-2021 using HE (blue), GC (brown), and DP (gray) techniques.

3.1 Homomorphic encryption

The performance of computation over encrypted data is allowed by homomorphic encryption with no need to decrypt it. HE can be classified into fully, partially, and somewhat homomorphic encryption. To preserve privacy during the computation of genomic data, different schemes were applied [33]. K. Shimizu et al. [34] suggested a method that combined efficient string data structures with cryptographic techniques constructed by additive HE. They produced an implementation of an efficient algorithm to search for sequences of SNPs in a large genome database. The server can not reveal the queried sequence.

On the other hand, Lain et al. [35] attempt to develop an Efficient Private Circular Query Protocol (EPCQP) with excellent accuracy, minimal computing, and transmission costs. The Moore curve was employed to transform two-dimensional spatial data to one-dimensional sequence and use Brakerski-Gentry Vaikuntanathan's (BGV) HE approach to protect the information about a point of interest (POIs). In order to reinforce the storage efficiency of the genome data sets, computation, and communication costs, Singh et al. [36] proposed a secure and efficient method for privacy-preserving personalized medicine by illustrating stream cipher-based homomorphic trans-ciphering on a cloud server.

Meanwhile, Wang et al. [37] proposed a novel scheme for healthcare queries on outsourced data called HeOC. Encrypted data is uploaded by trusted users into the cloud, and a perfect query is done on encrypted data about a particular disease. The operation is done by using a large number of sensors, which makes it expensive despite its efficiency. Then, Zheng et al. [38] employed an efficient k-NN query method to outsource encrypted data from e-healthcare. The encryption is done by the Paillier cryptosystem. This method provides efficient storing of encrypted data in the cloud and privacy-preserving k-NN query over encrypted data. This method is efficient in terms of privacy preservation and computational complexity.

Yan et al. [39] used edge computing based on the blockchain to construct a key solution that ensures the efficiency of the blockchain and reduces the computational overhead to clients by employing the paillier cryptosystem. They presented the advantages of blockchain and edge computing and constructed the key technological solutions of edge computing based on blockchain. They achieve the security protection and integrity check of cloud data and realize more extensive, secure multiparty computation. Blatt et al. [40] proposed a solution for GWAS security using (HE) to keep the encryption of all individual data during the association study. They presented a new Residue-Number-System(RNS) variant of the Cheon-Kim-Kim-Song (CKKS) HE scheme, new methods to switch between data encodings, and more than a dozen crypto-engineering optimizations. The solution can implement the full GWAS computation for 1000 individuals, 131,071 SNPs, and 3 covariates in about 10 minutes on a modern server computing node.

A novel encryption strategy based on HE was presented by Vizitiu et al. [41]. MORE (Matrix Operation for Randomization or Encryption) is proposed, which allows calculations within a neural network model to be directly conducted on floating-point data with reasonably little computational cost. At the same time, Blatt et al. [42] proposed statistical toolbox techniques that use HE to implement large-scale GWASs on encrypted genetic/phenotype data in an interactive manner where no decryption is required. The method presented a reformulation of GWAS tests to make use of packing of encrypted data and parallel processing, highly efficient statistical computation integration, and the development of a dozen crypto engineering optimizations.

Kuo et al. [43] invented three tracks for competition, which included genomic dataset access logging based on blockchain, securing HE in GWAS, and securing DNA segment searching. Kim et al. (2021) [44] utilized HE to introduce mathematical results and a warranty for security to protect genotype data at the time of imputation, which was implemented in a semi-trusted environment. Table 5 summarizes privacy preserving techniques for genomic data using homomorphic encryption.

Table 5: Privacy preserving techniques via homomorphic encryption

References	Year	Technique(s)	Advantage	disadvantage	dataset
[34]	2016	Homomorphic encryption and oblivious transfer	High advantage and powerful security	-the improvements in parallelization led to server run time may be reduced	1000 genome project
[35]	2017	Brakerski-Gentry-Vaikuntanathan homomorphic	-provides high accuracy query results while maintaining low	-the quality of results get worse if number	base stations datasets in China.

		encryption and Moore Curve	computation and communication costs.	of sub table gets large	
[36]	2018	FLIP scheme and BGV with FV comparison	Reinforcement efficiency of storage to the genome datasets, costs of communication and computation	-data exchange issues	Blood Group Antigen Gene Mutation Database (BGMUT)
[37]	2018	Health query scheme over the outsourced cloud (HeOC)	-Low overhead in computation and communication. -health service's provider model conditionality	Using large number of sensors	Personal and physiological Data
[38]	2019	Paillier and k d (dimension) tree	efficient data storage in the cloud and preserving the privacy of k NN queries on encrypted data	If the degree of the function f increased, the computational efficiency of the function will decrease	Synthetic dataset
[39]	2020	Paillier, Blockchain and edge computing	Enhance the performance of secure storage and computation	the negative value of the hash	Simulated data
[40]	2020	Cheon Kim Song (CKKS) based on Homomorphic scheme	Implement the full GWAS computation	needs to know the computation and parameter of semi-parallel procedure in advance and hand-tuned nature of many optimization applied on the solution	-iDash 2018 -Harvard Personal Genome Project
[41]	2020	Fully Homomorphic Encryption and MORE (Matrix Operation for Randomization or Encryption)	Ensured data security and perform data efficiently	Computational overhead	MNIST
[42]	2020	Ring Learning With Error (RLWE)	High efficiency	-The model assumed that encrypted data fully processed -GWAS can decrypt the encrypted data	AMD dataset
[43]	2020	Block chain, secure parallel GWAS by HE and secure search of DNA in large database	Enhance genomic security and privacy	High computation overhead	Simulated genomic dataset
[44]	2021	Fast Fully Homomorphic Encryption over Tours library	Effective -privacy preservin finger print authentication system	Suboptimal accuracy	1000 genome project

3.2 Garbled Circuit

Garbled is a two-party secure computation protocol that can be used for any general purpose. The usage of this protocol allows two parties to calculate the outcome of a function jointly without knowing anything concerning the inputs or intermediate results of the other party [45]. Yao's garbled circuit protocol is the most renowned of the Multi Parity Computation MPC techniques. It is commonly seen as the best-performing, and numerous of the protocols we cover build on Yao's GC [46]. The security of GC can be guaranteed by the equal participation of both parties communicating through the calculated functions.

Another benefit of GC is the secrecy of both parties' inputs, as the query frequently demands the same level of anonymity as the data. As a result, GC is typically utilized in sequence similarity situations when one party (researcher) has a data set of genomic sequences and the other party (data set) has a sensitive query sequence. The researcher wishes to locate sequences that are comparable to that specific query using any similarity metrics, such as Hamming and Levenshtein distances [33]. Al Aziz et al. [47] suggested approximation techniques for editing distance computation securely through genomic sequences and utilizing shingling specific set methods that include the algorithm of banded alignment intersected with garbled circuits to implement these methods. The method is considered to be accurate and time-efficient. On the other hand, the suggestion for a paradigm based on the basis of an indexed prefix tree for identical queries of patients by Mahdi et al. (2018) [48]. It ensures the privacy of data query requests and query responses. By employing the AES algorithm for preserving privacy, the encrypted and compressed tree is delivered to the cloud server to carry out query operations.

Researchers use GC to execute queries on accumulated data for semi-trusted models of opponents. Hasan et al. [30] proposed using distinct third parties to ensure secure exchange and execution of counter-question procedures on outsourced genomic data. The recommended method for creating an index tree from genetic data and then outsourcing it. The tree's nodes will be traversed by the cloud server and perform the count query using a secure interactive protocol. The checking will be done using Yao's GC over an interactive interface. Cheng et al. [2] proposed protocols to outsource the Similar Sequence Queries (SSQs) using an approximation of Edit Distance (ED), which depends on homomorphic encryption, and proposed a group of different security protocols to attain security efficiency and scalability depending on secret sharing, garbled circuit, and partial homomorphic encryption.

Mahdi et al. [49] suggested a technique to execute the count queries in a secure manner composed of genotype, phenotype, and numeric data by employing encryption and garbled circuits. Sotiraki et al. [50] developed a novel depth-optimized technique for computing set-maximal coincide between a database of aligned genetic sequences and an individual's DNA while preserving the database owner's individual privacy. Table 6 summarizes the privacy-preserving technique for genomic data via a garbled circuit.

Table 6: Privacy-preserving techniques for genomic data using a garbled circuit

Reference	Year	Technique(s)	Advantage(s)	Disadvantage(s)	Dataset
[47]	2017	Shingling private set intersection, banded alignment algorithm, and garbled circuit	Fast and achieve similar accuracy as traditional methods	-information leakage -weakness in case of brute force attack	1000 genome
[48]	2018	AES and garbled circuit	Preserve data, query and output privacy	-the sequences must be the same length -Query execution time depends on the dataset size	iDash 2016
[30]	2018	Indexed Tree and garbled circuit	Efficient and secure method for outsourcing data	It does not provide any privacy against heuristics attacks nor data access privacy because it exposes the tree traversal pattern.	Genomic dataset
[2]	2018	Partially homomorphic encryption, and Garbled Circuit	provide security, efficiency, and scalability under the semi-honest adversary model for small dataset	Not suitable for large dataset	Homo Sapiens Mitochondrion Complete Genome
[49]	2020	Encryption and garbled circuit	-provide data, query, and output privacy	Storage overhead cost	PGP
[50]	2020	Goldreich Micali Wigderson (GMW) and garbled circuit	Secure computing set maximal approach on semi-honest model	If n (length of sequence) is small, resulted in no improvement	Genomic database

3.3 Differential Privacy

Differential privacy is a model of privacy preservation that provides summary statistics about the dataset and ensures no one can learn anything about any record in the dataset [51]. It is considered widely accepted as a rigorous model for privacy protection. The present privacy-preserving algorithms are still problematic, such as k-anonymity. Before the appearance of differential privacy [52], it employed extra strict constraints and definitions by adding interference noise, as it conserves the potential privacy of users' information in the published data.

The attacker cannot conclude any information even if he has mastered specific information. Therefore, this completely excludes the possibility of disclosure of private information from the data source [53]. He et al. [54] suggested a differential privacy method that ensured genomic data release during belief propagation execution on a factor graph. This method is capable of factorizing the distribution of genomic data into a group of local distributions. Wei et al. [55] suggested differential privacy based on the genetic matching (DPGM) schema to attain efficient agreement and secure privacy in genetics. Park et al. [56] suggested a secure system for genomic data management by combining blockchain and local differential privacy. The suggested system uses two types of storage: private and semi-private, where genes are irreversibly modified by LDP in semi-private storage. While the data is stored in private storage accessible by internal employees only. Table 7 summarizes the privacy-preserving method using differential privacy.

Table 7: Privacy-preserving techniques using differential privacy

Reference	Year	Technique(s)	Advantage(s)	Disadvantage(s)	Dataset
[54]	2018	Differential privacy	guarantee the privacy	degrade the data utility	Genomic data
[55]	2019	Differential privacy based on genetic matching	-effective genetic matching	May cause privacy issues	-Simulated dataset -Real diabetic dataset -real diabetic DNA sequences
[56]	2021	Blockchain and differential privacy	-Develop access control and integrity verification -employed tow type of storage: private and semi-private	-data's owner can be traced from the operation of the blockchain -the operation cost is expensive	- genome sequence

3. Conclusions

During the past two decades, the importance of genomic sequencing and vital information has been demonstrated with the increase in genetic testing, analyses, and diagnostics and the spread of treatment based on individual genome sequencing. Because the cost of genomic sequencing has been dramatically reduced, people are being sequenced for a variety of reasons. Because it is important and sensitive information, unauthorized and undesirable access leads to a violation of the privacy of individuals.

Despite the benefits, the lack of protection and privacy-preserving methods creates risks and problems that outweigh the benefits. In this review paper, we present a clarification of the most important challenges facing maintaining the privacy of genomic data and a classification of the most important solutions used to meet these challenges. The use of the HE technique has produced remarkable results in terms of providing protection, privacy, and the ability to conduct operations without the use of data decryption. Furthermore, several forms of hybridization among HE, CG, and DP can be used to preserve genomic data privacy. It is hoped that in subsequent years, with the increase in genomic sequencing operations, the process of protecting this important and sensitive data generated will be essential in a manner appropriate to its rapid growth and will encourage researchers to focus on collaboration among HE, CG, and DP.

References

[1] M. Akgün, A. O. Bayrak, B. Ozer, and M. Ş. Sağıroğlu, "Privacy preserving processing of genomic data: A survey," *Journal of Biomedical Informatics*, vol. 56, no. 2015, pp. 103–111, Aug. 2015, doi: https://doi.org/10.1016/j.jbi.2015.05.022.

[2] K. Cheng, Y. Hou, and L. Wang, "Secure Similar Sequence Query on Outsourced Genomic Data," *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*, May 2018, doi: https://doi.org/10.1145/3196494.3196535.

[3] R. Mott, C. Fischer, P. Prins, and R. W. Davies, "Private Genomes and Public SNPs: Homomorphic Encryption of Genotypes and Phenotypes for Shared Quantitative Genetics," *Genetics*, vol. 215, no. 2, pp. 359–372, Apr. 2020, doi: https://doi.org/10.1534/ genetics.120. 303153.

[4] M. S. Rahman Mahdi, M. M. Al Aziz, N. Mohammed, and X. Jiang, "Privacy-preserving string search on encrypted genomic data using a generalized suffix tree," *Informatics in Medicine Unlocked*, vol. 23, p. 100525, 2021, doi: https://doi.org/10.1016/j.imu.2021.100525.

[5] None G.O. Ogunleye and S. E. Akinsanya, "Elliptic Curve Cryptography Performance Evaluation for Securing Multi-Factor Systems in a Cloud Computing Environment," *Iraqi Journal of Science*

, vol. 63, no. 7, pp. 3212–3224, Jul. 2022, doi: https://doi.org/10.24996/ijs.2022.63.7.40.

[6] J. S. Sousa *et al.*, "Efficient and secure outsourcing of genomic data storage," *BMC Medical Genomics*, vol. 10, no. S2, Jul. 2017, doi: https://doi.org/10.1186/s12920-017-0275-0.

[7] Z. Hasan, M. S. R. Mahdi, and N. Mohammed, "Secure Count Query on Encrypted Genomic Data: A Survey," *IEEE Internet Computing*, vol. 22, no. 2, pp. 71–82, Mar. 2018, doi: https://doi.org/10.1109/mic.2018.112102323.

[8] B. Berger and H. Cho, "Emerging technologies towards enhancing privacy in genomic data sharing," *Genome Biology*, vol. 20, no. 1, Jul. 2019, doi: https://doi.org/10.1186/s13059-019-1741-0.

[9] V. Popic and S. Batzoglou, "A hybrid cloud read aligner based on MinHash and kmer voting that preserves privacy," *Nature Communications*, vol. 8, no. 1, May 2017, doi: https://doi.org /10.1038/ncomms15311.

[10] M. Saad, "Designing a Secure Environment for IoT Networks Using Lightweight AES Algorithm," *Iraqi Journal of Science*, vol. 62, no. 8, pp. 2759–2770, Aug. 2021, doi: https://doi. org /10. /ijs.2021.62.8.29.

[11] J. W. Bos, K. Lauter, and M. Naehrig, "Private predictive analysis on encrypted medical data," *Journal of Biomedical Informatics*, vol. 50, pp. 234–243, Aug. 2014, doi: https://doi.org/10. 1016/j.jbi.2014.04.003.

[12] K. E. Lauter, A. López-Alt, and M. Naehrig, "Private Computation on Encrypted Genomic Data," *14th Privacy Enhancing Technologies Symposium, Workshop on Genome Privacy*, pp. 3–27, Sep. 2014, doi: https://doi.org/10.1007/978-3-319-16295-9_1.

[13] Jung Hee Cheon, M. Kim, and K. E. Lauter, "Homomorphic Computation of Edit Distance," 2015, pp. 194–212. doi: https://doi.org/10.1007/978-3-662-48051-9_15.

[14] S. Simmons, B. Berger, and C. Sahinalp, "Protecting Genomic Data Privacy with Probabilistic Modeling," *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, vol. 24, pp. 403–414, 2019, Accessed: Jul. 01, 2023. [Online]. Available: https://www.ncbi.nlm. nih.gov/pmc/articles/PMC6447070/

[15] Y. Zhang, X. Zhao, X. Li, M. Zhong, C. Curtis, and C. Chen, "Enabling Privacy-Preserving Sharing of Genomic Data for GWASs in Decentralized Networks," *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, Jan. 2019, doi: https://doi.org/10.1145/3289600.3290983.

[16] X. Yang, T. Li, X. Pei, L. Wen, and C. Wang, "Medical Data Sharing Scheme Based on Attribute Cryptosystem and Blockchain Technology," *IEEE Access*, vol. 8, pp. 45468–45476, 2020, doi: https://doi.org/10.1109/access.2020.2976894.

[17] Z. Huang *et al.*, "A privacy-preserving solution for compressed storage and selective retrieval of genomic data," *Genome Research*, vol. 26, no. 12, pp. 1687–1696, Oct. 2016, doi: https://doi.org /10. 1101 /gr.206870.116.

[18] X. Liu, Q. Zhu, Sakti Pramanik, C. M. Brown, and G. Qian, "VA-Store: A Virtual Approximate Store Approach to Supporting Repetitive Big Data in Genome Sequence Analyses," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 602–616, Mar. 2020, doi: https://doi.org/10.1109/tkde.2018.2885952.

[19] L. Chen, M. M. Aziz, N. Mohammed, and X. Jiang, "Secure large-scale genome data storage and query," *Computer Methods and Programs in Biomedicine*, vol. 165, pp. 129–137, Oct. 2018, doi: https://doi.org/10.1016/j.cmpb.2018.08.007.

[20] Q. Mehmood, M. Saleem, R. Sahay, A.-C. N. Ngomo, and M. D'Aquin, "QPPDs: Querying Property Paths Over Distributed RDF Datasets," *IEEE Access*, vol. 7, pp. 101031–101045, 2019, doi: https://doi.org/10.1109/access.2019.2930416.

[21] A. B. and S. S., "A survey on genomic data by privacy-preserving techniques perspective," *Computational Biology and Chemistry*, vol. 93, p. 107538, Aug. 2021, doi: https://doi.org /10.1016/j.compbiolchem.2021.107538.

[22] M. M. Al Aziz, M. Z. Hasan, N. Mohammed, and D. Alhadidi, "Secure and Efficient Multiparty Computation on Genomic Data," *Proceedings of the 20th International Database Engineering & Applications Symposium on - IDEAS '16*, 2016, doi: https://doi.org/10.1145/2938503.2938507.

[23] C. Xu, Q. Chen, H. Hu, J. Xu, and X. Hei, "Authenticating Aggregate Queries over Set-Valued Data with Confidentiality (Extended Abstract)," in *2018 IEEE 34th International Conference on*

Data Engineering (ICDE), Apr. 2018. doi: https://doi.org/10.1109/icde.2018.00256.

[24] Sakina Mahboubi, Reza Akbarinia, and P. Valduriez, "Privacy-Preserving Top-k Query Processing in Distributed Systems," in *Transactions on Large-Scale Data- and Knowledge-Centered Systems XLII*, Springer-Verlag, 2019, pp. 1–24. doi: https://doi.org/10.1007/978-3-662-60531-8_1.

[25] H. Quan, B. Wang, Y. Zhang, and G. Wu, "Efficient and Secure Top-k Queries With Top Order-Preserving Encryption," *IEEE Access*, vol. 6, pp. 31525–31540, 2018, doi: https://doi.org/10.1109/access.2018.2847307.

[26] Y. Zhang, W. Dai, X. Jiang, H. Xiong, and S. Wang, "FORESEE: Fully Outsourced secuRe gEnome Study basEd on homomorphic Encryption," *BMC Medical Informatics and Decision Making*, vol. 15, no. S5, Dec. 2015, doi: https://doi.org/10.1186/1472-6947-15-s5-s5.

[27] S. Wang *et al.*, "HEALER: homomorphic computation of ExAct Logistic rEgRession for secure rare disease variants analysis in GWAS," *Bioinformatics*, vol. 32, no. 2, p. btv563, Oct. 2015, doi: https://doi.org/10.1093/bioinformatics/btv563.

[28] R. Ghasemi, Md. M. Al Aziz, N. Mohammed, M. H. Dehkordi, and X. Jiang, "Private and Efficient Query Processing on Outsourced Genomic Databases," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 1466–1472, Sep. 2017, doi: https://doi.org/10.1109/jbhi.2016.2625299.

[29] J. H. Ziegeldorf *et al.*, "BLOOM: BLoom filter based oblivious outsourced matchings," *BMC Medical Genomics*, vol. 10, no. S2, Jul. 2017, doi: https://doi.org/10.1186/s12920-017-0277-y.

[30] M. Z. Hasan, M. S. R. Mahdi, M. N. Sadat, and N. Mohammed, "Secure count query on encrypted genomic data," *Journal of Biomedical Informatics*, vol. 81, pp. 41–52, May 2018, doi: https://doi.org/10.1016/j.jbi.2018.03.003.

[31] J. L. Raisaro *et al.*, "Protecting Privacy and Security of Genomic Data in i2b2 With Homomorphic Encryption and Differential Privacy," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 5, pp. 1–1, 2018, doi: https://doi.org/10.1109 /tcbb.2018.2854782.

[32] M. Kim, Y. Song, B. Li, and D. Micciancio, "Semi-Parallel logistic regression for GWAS on encrypted data," *BMC Medical Genomics*, vol. 13, no. S7, Jul. 2020, doi: https://doi.org/10.1186/s12920-020-0724-z.

[33] M. M. A. Aziz *et al.*, "Privacy-preserving techniques of genomic data-a survey," *Briefings in Bioinformatics*, vol. 20, no. 3, pp. 887–895, May 2019, doi: https://doi.org/10.1093/bib/bbx139.

[34] K. Shimizu, K. Nuida, and G. Rätsch, "Efficient privacy-preserving string search and an application in genomics," *Bioinformatics*, vol. 32, no. 11, pp. 1652–1661, Mar. 2016, doi: https://doi.org/10.1093/bioinformatics/btw050.

[35] H. Lian, W. Qiu, D. Yan, Z. Huang, and J. Guo, "Efficient Privacy-Preserving Protocol for *k*-NN Search over Encrypted Data in Location-Based Service," *Complexity*, vol. 2017, pp. 1–14, 2017, doi: https://doi.org/10.1155/2017/1490283.

[36] K. Singh, R. Sirdey, and S. Carpov, "Practical personalized genomics in the encrypted domain," in *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*, Apr. 2018. doi: https://doi.org/10.1109/fmec.2018.8364056.

[37] G. Wang, R. Lu, and Y. L. Guan, "Enabling Efficient and Privacy-Preserving Health Query Over Outsourced Cloud," *IEEE Access*, vol. 6, pp. 70831–70842, 2018, doi: https://doi.org/10.1109/access.2018.2880220.

[38] Y. Zheng, R. Lu, Y. Guan, S. Zhang, J. Shao, and H. Zhu, "Efficient and Privacy-Preserving Similarity Query With Access Control in eHealthcare," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 880–893, Jan. 2022, doi: https://doi.org/10. 1109/tifs.2022.3152395.

[39] X. Yan, Q. Wu, and Y. Sun, "A Homomorphic Encryption and Privacy Protection Method Based on Blockchain and Edge Computing," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–9, Aug. 2020, doi: https://doi.org/10.1155/2020/8832341.

[40] M. Blatt, A. Gusev, Y. Polyakov, K. Rohloff, and V. Vaikuntanathan, "Optimized homomorphic encryption solution for secure genome-wide association studies," *BMC Medical Genomics*, vol. 13, no. S7, Jul. 2020, doi: https://doi.org/10.1186/s12920-020-0719-9.

[41] A. Vizitiu, C. I. Niţă, A. Puiu, C. Suciu, and L. M. Itu, "Applying Deep Neural Networks over Homomorphic Encrypted Medical Data," *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–26, Apr. 2020, doi: https://doi.org/10.1155/2020/3910250.

[42] M. Blatt, A. Gusev, Y. Polyakov, and S. Goldwasser, "Secure large-scale genome-wide association studies using homomorphic encryption," *Proceedings of the National Academy of Sciences*, vol. 117, no. 21, pp. 11608–11613, May 2020, doi: https://doi.org/10.1073 /pnas.1918257117.

[43] T.-T. Kuo *et al.*, "iDASH secure genome analysis competition 2018: blockchain genomic data access logging, homomorphic encryption on GWAS, and DNA segment searching," *BMC Medical Genomics*, vol. 13, no. S7, Jul. 2020, doi: https://doi.org/10.1186/s12920-020-0715-0.

[44] M. Kim *et al.*, "Ultrafast homomorphic encryption models enable secure outsourcing of genotype imputation," *Cell Systems*, vol. 12, no. 11, pp. 1108-1120.e4, Nov. 2021, doi: https://doi.org /10.1016/j.cels.2021.07.010.

[45] J. Sancho, J. García, and Á. Alesanco, "Oblivious Inspection: On the Confrontation between System Security and Data Privacy at Domain Boundaries," *Security and Communication Networks*, vol. 2020, pp. 1–9, Sep. 2020, doi: https://doi.org/10.1155/2020/8856379.

[46] D. Evans, V. Kolesnikov, and M. Rosulek, "A Pragmatic Introduction to Secure Multi-Party Computation," *Foundations and Trends® in Privacy and Security*, vol. 2, no. 2–3, pp. 70–246, 2018, doi: https://doi.org/10.1561/3300000019.

[47] M. M. A. Aziz, D. Alhadidi, and N. Mohammed, "Secure approximation of edit distance on genomic data," *BMC Medical Genomics*, vol. 10, no. S2, Jul. 2017, doi: https://doi.org/10. 1186/s12920-017-0279-9.

[48] M. S. R. Mahdi, M. M. Al Aziz, D. Alhadidi, and N. Mohammed, "Secure Similar Patients Query on Encrypted Genomic Data," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2611–2618, Nov. 2019, doi: https://doi.org/10.1109/jbhi.2018.2881086.

[49] Safiur Rahman Mahdi, N. Sadat, N. Mohammed, and X. Jiang, "Secure Count Query on Encrypted Heterogeneous Data," Aug. 2020. doi: https://doi.org/10.1109/dasc-picom-cbdcom-cyberscitech49142.2020.00098.

[50] K. Sotiraki, E. Ghosh, and H. Chen, "Privately computing set-maximal matches in genomic data," *BMC Medical Genomics*, vol. 13, no. S7, Jul. 2020, doi: https://doi.org/10.1186/s12920-020-0718-x.

[51] Z. Wan, J. W. Hazel, E. W. Clayton, Y. Vorobeychik, M. Kantarcioglu, and B. A. Malin, "Sociotechnical safeguards for genomic data privacy," *Nature Reviews Genetics*, vol. 23, Mar. 2022, doi: https://doi.org/10.1038/s41576-022-00455-y.

[52] J. Wang, Z. Cai, and J. Yu, "Achieving Personalized k-Anonymity-Based Content Privacy for Autonomous Vehicles in CPS," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4242–4251, Jun. 2020, doi: https://doi.org/10.1109/TII.2019.2950057.

[53] B. Jiang, J. Li, G. Yue, and H. Song, "Differential Privacy for Industrial Internet of Things: Opportunities, Applications and Challenges," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10430–10451, Jul. 2021, doi: https://doi.org/10.1109/jiot.2021.3057419.

[54] Z. He, Y. Li, J. Li, K. Li, Q. Cai, and Y. Liang, "Achieving differential privacy of genomic data releasing via belief propagation," *Tsinghua Science and Technology*, vol. 23, no. 4, pp. 389–395, Aug. 2018, doi: https://doi.org/10.26599/tst.2018.9010037.

[55] J. Wei, Y. Lin, X. Yao, J. Z. Zhang, and X. Liu, "Differential Privacy-Based Genetic Matching in Personalized Medicine," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1109–1125, Jul. 2021, doi: https://doi.org/10.1109/tetc.2020.2970094.

[56] Y.-H. Park, Y. Kim, and J. Shim, "Blockchain-Based Privacy-Preserving System for Genomic Data Management Using Local Differential Privacy," *Electronics*, vol. 10, no. 23, p. 3019, Dec. 2021, doi: https://doi.org/10.3390/electronics10233019