electronics

MDPI

Article

Blockchain-Based Privacy-Preserving System for Genomic Data Management Using Local Differential Privacy

Young-Hoon Park, Yejin Kim and Junho Shim *

Department of Computer Science, Sookmyung Women's University, Seoul 04310, Korea; yh.park@sookmyung.ac.kr (Y.-H.P.); yejinkim@sookmyung.ac.kr (Y.K.)
* Correspondence: jshim@sookmyung.ac.kr

Abstract: The advances made in genome technology have resulted in significant amounts of genomic data being generated at an increasing speed. As genomic data contain various privacy-sensitive information, security schemes that protect confidentiality and control access are essential. Many security techniques have been proposed to safeguard healthcare data. However, these techniques are inadequate for genomic data management because of their large size. Additionally, privacy problems due to the sharing of gene data are yet to be addressed. In this study, we propose a secure genomic data management system using blockchain and local differential privacy (LDP). The proposed system employs two types of storage: private storage for internal staff and semi-private storage for external users. In private storage, because encrypted gene data are stored, only internal employees can access the data. Meanwhile, in semi-private storage, gene data are irreversibly modified by LDP. Through LDP, different noises are added to each section of the genomic data. Therefore, even though the third party uses or exposes the shared data, the owner's privacy is guaranteed. Furthermore, the access control for each storage is ensured by the blockchain, and the gene owner can trace the usage and sharing status using a decentralized application in a mobile device.

Keywords: gene; blockchain; local differential privacy; DApp

Citation: Park, Y.-H.; Kim, Y.; Shim, J. Blockchain-Based Privacy-Preserving System for Genomic Data Management Using Local Differential Privacy. *Electronics* **2021**, *10*, 3019. https://doi.org/10.3390/electronics10233019

Academic Editor: Flavio Canavero

Received: 31 August 2021
Accepted: 29 November 2021
Published: 3 December 2021

1. Introduction

Since the human genome project, a significant amount of human genome data has been generated by researchers and healthcare employees [1]. Moreover, owing to the development of bioinformatics technology, genome data have been generated at an increasing speed. With regard to healthcare information, the security and privacy of gene data have received considerable attention because gene data contain privacy-sensitive information, such as the gene owner's characteristics and variant call formats [2]. However, security schemes that can preserve the properties of gene data are yet to be developed.

General healthcare data, including genome data, exhibit three common characteristics: large volume, sharing demand, and privacy sensitivity [3], which cannot be addressed simultaneously [2,4]. To ensure privacy, the healthcare data can be encrypted. Meanwhile, to address sharing issues, access control, key management protocols, or advanced cryptography schemes, such as attribute-based encryption, may be employed [5]. However, when these solutions are adopted in healthcare data management systems, efficiency may be affected due to the large volume of the data and the heavy computational costs associated with cryptography schemes [6]. With regard to genome data, the size of each datum is extremely large, i.e., 200–300 GB; therefore, efficient security and sharing schemes are necessary.

Various researchers have proposed using blockchain technology to overcome these problems. Blockchain is a type of data structure that guarantees the integrity of the stored data. It was first used for managing transaction data, known as cryptocurrency (e.g., Bitcoin [7] and Ethereum [8]). In addition, blockchain has been used not only in the finance domain but also in managing the history of stored data, including healthcare information.

Several blockchain-based schemes exist for the gene data. Dambrot et al. proposed ReGene to backup various types of information associated with the sharing of gene data [9], and many other researchers have attempted to adopt blockchain technology in healthcare data management systems [10–12]. Various blockchain-based gene data management solutions have also been proposed [13,14]. However, using only the blockchain cannot solve the aforementioned scalability and sharing problems.

Unlike general healthcare data, the size of each gene datum is extremely large; hence, a considerable amount of time is required to encrypt/decrypt or sign/verify the entire gene data. In addition, more than 99.9% of gene information is common to all people [15]; hence, the entire genome data need not be encrypted or signed. In other words, securing only the section that reveals the genome owner can ensure the confidentiality and integrity of the data while significantly reducing the computational cost.

Another issue regarding the gene data is that the demand for sharing gene data is expected to be high [16]. Although access control and key management schemes can be relied upon to securely share data, using only these existing solutions cannot entirely solve the sharing issue. Shared gene data can be accessed by a requesting user; however, once transmitted, the user can possess the data indefinitely. If the user leaks the gene data deliberately or by mistake, the privacy of the gene data owner will be invaded indefinitely.

To overcome the aforementioned problems in gene data, we herein propose a privacy-enhanced blockchain-based gene data management scheme using efficient encryption and differential privacy (DP). The main contributions of this study are as follows:

1. We propose a blockchain-based management system to supervise gene data during storage using a smart contract [17]. Using this smart contract, we developed an access control and integrity verification scheme for gene data. Thanks to the high level of security provided by the blockchain, the gene data can be delivered to only the authorized user, and the gene data can be authenticated.
2. To prevent privacy invasion while sharing gene data, we propose a two-storage model, one of which is a private storage for internal employees, and the other is a semi-private storage for external users who request to share the gene data. In semi-private storage, irreversibly modified gene data are stored; hence, privacy-sensitive information in the data is protected even when the external user exposes the data.
3. To irreversibly modify the gene data for concealing the gene owner's information, we propose a de-identification algorithm using local differential privacy (LDP). Using this scheme, the sections in the gene data that reveal the owner are modified randomly by adding noise generated using the RAPPOR scheme [18]. When partially modified gene data are assigned to a legal third party, the data can be utilized while maintaining the privacy of the data owner.

The remainder of this paper is organized as follows: We first discuss the related studies in Section 2, and we introduce the proposed system model in Section 3. Subsequently, we describe the blockchain-based gene data management system in detail in Section 4. In Section 5, we propose a methodology for introducing LDP to the system. We discuss the security improvement and performance of the proposed system in Section 6. Finally, in Section 7, we present our conclusions and directions for future research.

2. Related Works

2.1. Blockchain Schemes for Genome Data

Recently, research pertaining to genetic data has been actively pursued, particularly owing to the COVID-19 pandemic. The storage and management of a significant amount of gene data is challenging, and the security of these data remains dubious. Because a single human genome occupies 100 GB of storage space and more genomes will be sequenced in the future, by 2025, a storage capacity of approximately 40 EB will be required [19]. In addition, as genetic tests for which the results are delivered directly to consumers increase, people are becoming more interested in their personal genetic data. Therefore, whether these data are protected safely while they are used and stored is an important concern.

Currently, companies such as 23andMe [13] and Zenome [14] protect consumer information using blockchain technology. Experts expect blockchain to be vital to fields such as trust, transparency, and data-sharing [10]. Blockchain enables people to own their own data; however, the security and privacy issues of blockchain must be considered as well.

Based on a distributed system of blockchain, which is unlike previous centralized systems, blockchain has been widely used to manage security and privacy issues. Blockchain technology enables a decentralized environment and ensures integrity [12]. Using node authentication, hybrid signatures, and consensus mechanisms, privacy issues can be secured [20]. The authors in [21] proposed the use of blockchain to provide secure management and analyze big data regarding healthcare. The use of blockchain-based applications in the field of healthcare [22,23], as well as the development of IoT devices with blockchain have been attempted [24]; interest in applying blockchain technology in managing genomic data is currently increasing rapidly. Few researchers have attempted to use blockchain in this area and proposed a blockchain-based solution that provides not only genomic privacy, security, and anonymous data analysis, but also a method for reversing phenotypical expression errors [9]. The authors of [11] suggested using a blockchain algorithm to secure the medical data obtained. In addition, companies such as CrypDist, Nebula Genomics, and Gene-Chain use blockchain technology to enhance genomic data sharing.

2.2. Differential Privacy in Genome Data Managements

Gene data privacy has received considerable research attention because gene data contain privacy-sensitive gene owner identification data [3,25]. In several studies, the privacy invasion of gene owners has been prevented using membership inference attacks [26,27] or attribute inference attacks [28–30]. To protect against these attacks, various privacy-preserving solutions based on cryptographic techniques have been proposed [31–33]. However, using only cryptographic approaches may result in other privacy issues. As mentioned before, although an outside receiver is trustworthy, if the receiver keeps the decrypted gene data forever, and the scored gene data may be exposed because of several reasons such as receiver's carelessness. To address this limitation, we employed the DP, which have been adopted in several gene data management systems.

Owing to the relatively short history of DP, research and development of DP-based gene data management systems is scarce. In existing studies pertaining to DP for gene data, statistical values were obtained to protect private information in the gene data [34–36], and LDP has been employed to secure data in storage. However, to the best of our knowledge, only one LDP-based scheme exists for gene data [37], and it presents another efficiency problem when the LDP process is operated.

3. Proposed System Model

In this section, we introduce the entire model of the proposed system. We describe the roles of participants and the processes involved in protecting security. Our solution targets organizations that manage gene data, and we assume that they store the entire gene data in their internal storage.

The proposed system, as shown in Figure 1 employs two separate storages, that is, private and semi-private storage, and a blockchain system with a decentralized application (DApp). The gene data are stored in both storage systems, but to ensure security and privacy, the gene data are encrypted or noise is added to conceal sections that contain the owner information. The encrypted gene data are stored in the private storage, and the gene data with noise are stored in the semi-private storage. Moreover, the gene data owner can trace the usage and transmission information of gene data through DApp.

Let us first consider private storage. Private storage is for internal staff and highly trustworthy members, who can access and use the original gene data by decrypting the encrypted sections. We assume that such users do not deliver the original data to untrustworthy users. The original gene data would be delivered to users who investigate or test it among the trustworthy members.

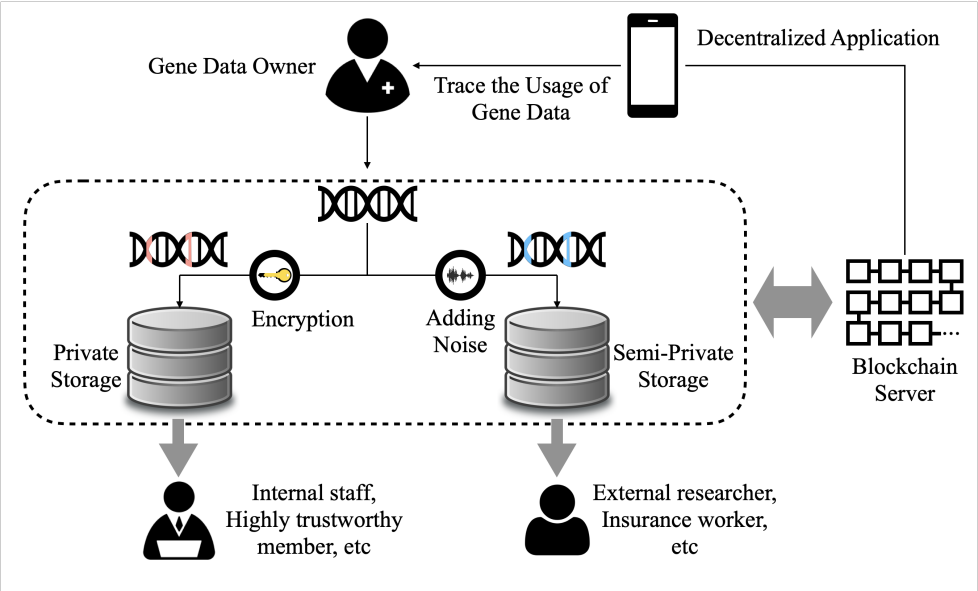


Figure 1. System Model.

Next, we considered semi-private storage. In semi-private storage, gene data with noise are stored and will not be retrieved from the original gene data by any other end users, except for relatively untrustworthy users, such as researchers who are employed by external organizations and insurance workers. Semi-private storage is not public; therefore, only permitted users can access the stored data. However, because such users may leak stored gene data to other users or organizations, irreversible gene data are available. In this regard, we adopted the LDP to generate noise and attach the noise data to the privacy-sensitive section of the gene data.

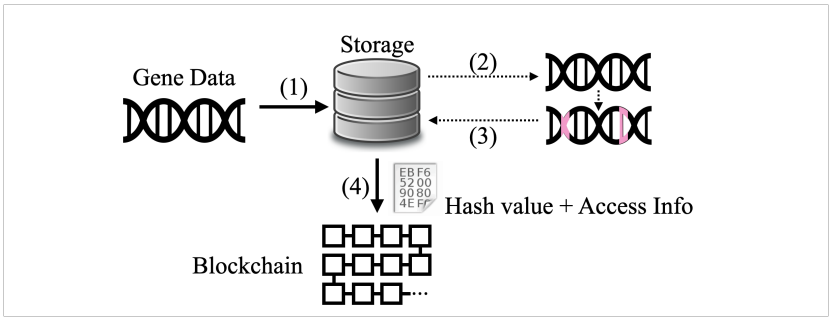
This two-storage-system model reduces the possibility of gene data leakage from private information. For trustworthy internal researchers, encrypted gene data in private storage are shared, and for relatively untrustworthy users or external organization workers, only gene data with noise in the semi-private storage are allowed to be accessed. Therefore, the original gene data can exist only in private storage and are accessible by permitted insiders; furthermore, they cannot exit the proposed management system. The data that can exit from the system are gene data with noise that cannot be transformed into the original data.

In addition to encryption and noise generation, the privacy of gene owners can be ensured using blockchain systems. In each block, various types of information pertaining to the gene data, such as the file size, creation/modification time, hash value of the data, and secured sections of the data, are stored. In addition, for access control, lists of users who can download encrypted data and gene data with noise are stored. Hence, using a smart contract, the blockchain can control the access requests. In other words, when the system receives a request for downloading the gene data, it verifies the availability of the blockchain system. If the smart contract in the blockchain returns "accept," the proposed system transmits the required data. Moreover, the DApp is connected to the blockchain system to trace the usage and transmission of gene data. Therefore, the gene data owner can verify the user that accesses their gene data, as well as the manner in which the data are utilized through the user's smartphone.

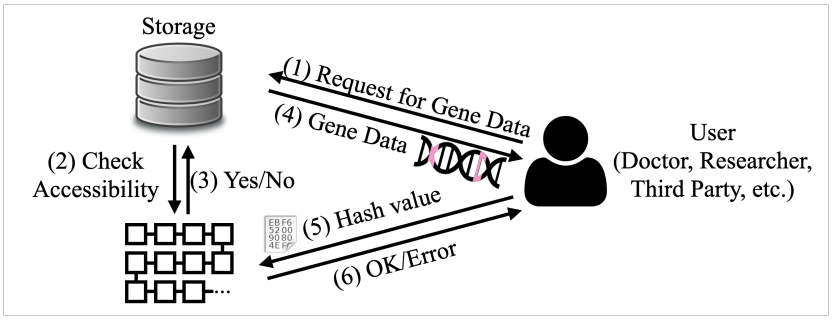
Using the proposed model, the sensitive sections of the gene data will not be shared by external members; hence, the privacy of the data owner is guaranteed. In addition, because of the smart contract of the blockchain, only authorized members can access the storage, whereas only trustworthy insiders can download the original data from private storage. The security of the proposed system is discussed in Section 6.

4. Blockchain-Based Gene Data Management

Figure 2 shows the interactions between the storage for the gene data and the blockchain. In this section, we address the blockchain used in the proposed gene data management system. The primary roles of blockchain are integrity verification of gene data and access control. Herein, we provide methods for managing gene data using a blockchain.



(**a**)



(**b**)

Figure 2. Interactions between storage and blockchain. (**a**) Saving of gene data; (**b**) Retrieval of gene data.

First, we consider the process by which gene data are stored in storage, as shown in Figure 2a. After the sensitive sections of the gene data are encrypted or noise is attached to those sections, the secured gene data are transmitted to the data storage. Subsequently, the hash function is used to calculate the digest of the gene data. This calculation was performed to ensure data integrity. In addition to the hash value, access information that reveals the member that can receive the data is delivered. After these data are transmitted, they are stored in the ledger; subsequently, a new block is created and connected to the blockchain.

Next, we explain the data retrieval process illustrated in Figure 2b. When the gene data are requested by an internal member or an external user, the storage first verifies the availability of access to the data using a smart contract in the blockchain. Because the blockchain retains the access information of the data and this information is not modified owing to the safety of the blockchain, the data are delivered only to authorized users. After data transmission, to verify the integrity of the gene data, the receiver first operates the hash function for the data and derives a message digest, and then verifies the integrity of the gene data by sending the hash value to the blockchain system. The data transmission process is terminated when the blockchain returns the results.

In this study, to realize a blockchain-based system, we employ Ethereum and a private network. It is noted that the aims of Ethereum are securely storing information about the gene data, conducting access control, and tracing the usage of the gene data. Hiding sensitive information by encrypting or adding noise to the gene data is not a role of Ethereum, but is realized by the encryption scheme and LDP, respectively.

5. Local Differential Privacy for De-Identification

DP [38,39] is widely used to protect user privacy in data stored in a database. DP is employed to prevent the leakage of personal information during the stage in a deep learning model [40–42]. Moreover, because healthcare data contain privacy-sensitive information, DP is adopted in various deep learning and artificial intelligence systems for healthcare system [43–47].

However, there are a few problems associated with DP. If the system employs a black box model, the database that stores the training data is not trustworthy, and the data that require privacy are exposed, and the DP is no longer usable. In addition, an attacker may refer to the training data by sending the same queries to the database and receiving the responses repeatedly. Hence, LDP [39], which conceals the stored data instead of the responses to the queries, is proposed. Specifically, the original DP is referred to as central differential privacy or general differential privacy. A randomized algorithm $\mathcal{A}$ provides $\varepsilon-$LDP if the following equation holds [18,48]:

Definition 1. *A randomized algorithm $\mathcal{A}$ satisfies ε-LDP if for all pairs of stored data v_1 and v_2, for all $Q \subseteq Range(\mathcal{A})$, and for $\varepsilon \geq 0$, the following equation holds:*

$$Pr[\mathcal{A}(v_1) \in Q] \leq e^{\varepsilon} \cdot Pr[\mathcal{A}(v_2) \in Q] \tag{1}$$

The logical meaning of Definition 1 is as follows: The probabilities that the modified versions of the two original data v_1 and v_2 by algorithm $\mathcal{A}$ are the same should be high. As ε increases, a greater difference between the two probabilities is allowed. This implies that the required level of privacy decreases. In contrast, as ε approaches 0, the gap between the two probabilities should be smaller; therefore, the privacy level increases.

The LDP modifies the data stored in the data; hence, the privacy of the data owner is protected even if the database is compromised or repetitive queries occur. Owing to this property, LDP can be utilized to secure healthcare data [49,50]. However, to the best of our knowledge, the application of LDP in gene data management systems has not been attempted.

As mentioned earlier, the size of each gene datum is extremely large; hence, a significant amount of time is required to apply the LDP to the entire dataset. However, because more than 99% of the gene data are common to humans, securing only sections that reveal each owner is sufficient. Therefore, in the proposed scheme, we provide a method for modifying the identification sections.

Figure 3 shows an example of adding noise using LDP. The gene data in the upper section of Figure 3 are those that have not been modified, and it is assumed that a group of the three bases surrounded by an over-brace of the genetic code is one of the sections that may reveal the owner. In our study, to conceal this section, we adopted LDP; hence, that section is modified to other bases, as shown in the lower section of Figure 3.

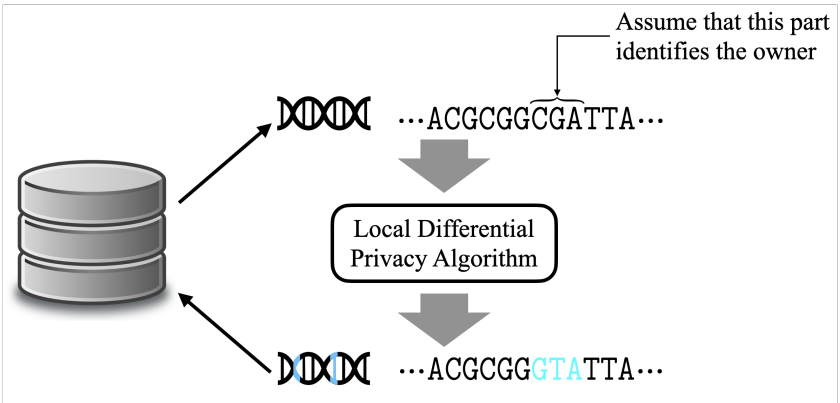


Figure 3. An example of attaching noise using local differential privacy.

Next, we discuss a method for modifying the privacy-sensitive section. In many studies, RAPPOR technology [18] was employed to generate noise. In our study, we modified the sensitive sections of the gene data using our own method based on RAPPOR. We assume that there are two independent unfair coins. In RAPPOR, the noise for a binary value b is changed by the following rule, where $b \in \{0, 1\}$.

- When the first coin is flipped and the head is shown, b is maintained.
- If the first coin shows the tail, then the second coin is flipped.
- If the second coin shows the head, b becomes 0. Otherwise, b becomes 1.

In the procedure above, if fair coins are used, we can conclude that the system yields $\log 3-$LDP. Based on this rule, we designed an algorithm to modify the privacy-sensitive sections of the gene data. The detailed algorithm is presented in Algorithm 1.

Algorithm 1 Gene Data Modification with LDP

Require: Gene data $\mathcal{G}$, set of sensitive sections in gene P, and privacy parameter ε

1: $\lambda \leftarrow \frac{4}{e^{\varepsilon}+3}$.
2: **for all** unmarked $g \in \mathcal{G}$ **do**
3: **if** $g \in P$ **then**
4: Choose a random positive real number r between 0 and 1
5: **if** $r < \lambda$ **then**
6: Let h be a pair of g.
7: Determine g among `A`, `C`, `G`, and `T` with the same probability.
8: **switch** g **do**
9: **case** `A`:
10: $h \leftarrow$ `T`; break;
11: **case** `C`:
12: $h \leftarrow$ `G`; break;
13: **case** `G`:
14: $h \leftarrow$ `C`; break;
15: **case** `T`:
16: $h \leftarrow$ `A`; break;
17: **end switch**
18: mark g and h.
19: **end if**
20: **end if**
21: **end for**

As previously mentioned, RAPPOR was used in Algorithm 1 to modify the bases of the privacy-sensitive sections. In Algorithm 1, note that `A`, `C`, `G`, and `T` are four bases of DNA code, and pairs of bases `A`, `C`, `G`, and `T` are `T`, `G`, `C`, and `A`, respectively; thus, we add lines 7–18. For a base g in the privacy-sensitive sections, the probability that g does not change despite the operation of this algorithm can be calculated as follows:

$$Pr[\mathcal{A}(g) = g] = (1 - \lambda) + \frac{\lambda}{4} = 1 - \frac{3\lambda}{4} \tag{2}$$

Subsequently, $Pr[\mathcal{A}(g) \neq g] = \frac{3\lambda}{4}$. Therefore, for $g, g' \in \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$ such that $g \neq g'$, $Pr[\mathcal{A}(g) = g'] = \frac{\lambda}{4}$ because the probabilities that $\mathcal{A}(g)$ is one of the three types of bases except g are the same. From these probabilities, we can derive the following theorem:

Theorem 1. *Algorithm 1 guarantees ε-LDP.*

A detailed proof of Theorem 1 is provided in the Appendix A. According to Theorem 1, we can conclude that if we employ Algorithm 1 to add noise to the sensitive sections of the gene data, then ε-LDP can be provided. In particular, when $\varepsilon = 0$, λ can only be 1; as such,

the IF statement is executed unconditionally. Hence, the probability that g becomes one of the four types of bases is $\frac{1}{4}$, and we claim that perfect privacy is guaranteed.

With LDP, the sections that identify the gene owner are modified. In contrast, the other sections of the gene data, which comprise more than 99.9% of the human genome data and represent the unique characteristics of humans, remain unchanged. Therefore, we expect that the modified method can be used for research and medical purposes. Moreover, we can adjust the level of modification in the situation by changing ε. However, to exactly demonstrate this, we need a further study about them.

6. Discussion

In this section, we discuss the proposed scheme provided in Sections 3–5. We first analyzed the security issues in blockchain and LDP, and subsequently measured the operation times for various cases, that is, gene data encryption and decryption, where we adopted LDP to demonstrate availability.

6.1. Security Analysis

The proposed system uses two schemes to provide security and privacy to the gene data: blockchain and LDP. First, we discuss the security effectiveness afforded by blockchain technology, and subsequently address privacy preservation afforded by LDP.

As mentioned above, the primary roles of blockchain are the protection of integrity of the gene data and access control. The blockchain securely stores the history of the data, and the stored data do not change. Therefore, blockchain can provide the two aforementioned functions.

We stored the hash values of the gene data in the proposed system. When an authorized user receives the gene data from the storage to verify the integrity of the data, the hash value is calculated and sent to the blockchain. It is noteworthy that the stored hash value is protected by the robust levels of security provided by blockchain. Using this method, the user can verify that the obtained gene data are not modified illegally.

The blockchain stores access control information, that is, not only the message digest of the gene data. In our system, when a user requests gene data, the storage server first verifies whether the user can access the data. In addition, two types of storage are employed: private storage and semi-private storage; the security methods for data in the two types of storages are different. The access control information in the blockchain oversees whether the user requests accessible storage.

However, another privacy infringment problems may occur due to the characteristic of blockchain. Although all of gene data are protected by the encryption and LDP schemes, data's owner can be traced from operations of the blockchain. For this, there are several countermeasures that overcome the privacy-related threats and keep the anonymity of the gene data such as mixing techniques [51–54], anonymous signatures [55,56], and secure multi-party computation [57], etc. Moreover, Zhang et al. [58] implies that the mentioned schemes can solve the deanonymization problems.

Next, we discuss the privacy-level improvement afforded by LDP. In our system, the LDP irreversibly modifies the privacy-sensitive sections in the gene data such that the receiver does not have any information regarding the data owner from the modified data. Therefore, the transformed gene data are stored in semi-private storage, and they can be delivered to external users, such as staff in other organizations or insurance workers. In addition, because the modified gene data are de-identified, the privacy of the gene data owner will not be invaded even if the receiver exposes the data to other users.

In ε-LDP, to adjust the privacy level, the ε value can be modified. If $\varepsilon = 0$, an attacker will not have any information regarding the original data from the modified gene data. In fact, the information that identifies the data owner vanishes completely. This appears to be the best solution for the gene data provider; however, it may not be suitable for gene data users. For example, the modified gene data are different from the human genome; therefore, the received data may become unsuitable for research. By contrast, increasing the privacy

parameter ε, implies that a lower privacy level is allowed; therefore, the modified gene data become similar to the original data, and the attacker might receive clues regarding the original data. However, the modified data may be similar to the human genome data; as such, researchers may be able to utilize the received data. Therefore, the system manager should select an appropriate ε to ensure both the security and availability of the modified gene data.

Now, let us analyze the privacy level owing to the ε-LDP mathematically. Let $g \in \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$ be the original base in the gene data, and g be modified into g'. Without loss of generality, let $g' = \mathtt{A}$. According to Equation 2, the probabilities that $g = \mathtt{A}$, $\mathtt{C}$, $\mathtt{G}$, and $\mathtt{T}$ are $1 - \frac{3\lambda}{4}$, $\frac{\lambda}{4}$, $\frac{\lambda}{4}$, and $\frac{\lambda}{4}$, respectively. Therefore, the probability that an attacker can guess the right gene base from the modified gene base is not greater than $1 - \frac{3\lambda}{4}$ because $\lambda \leq 1$ then $1 - \frac{3\lambda}{4} \geq \frac{\lambda}{4}$. When the number of modified gene bases is n, the probability that the attacker guesses the entire original gene data from the modified one is not greater than $\left(1 - \frac{3\lambda}{4}\right)^n$, which is close to 0 because n is sufficiently large. Therefore, we can conclude that it is difficult for an attacker to obtain the original gene data from the modified data.

6.2. Experimental Results

Next, we demonstrate that the proposed schemes can be used in real situations by presenting the experimental results. Because security and utility exhibit a tradeoff relationship, users may feel inconvenienced when the proposed security and privacy schemes are adopted. The size of gene data for one human is 200–300 GB; therefore, the operation times for encrypting/decrypting or the time required for adding noise may be long. In this subsection, we demonstrate that the operation times are not excessively long, and that the proposed schemes do not disrupt the management of gene data.

For the performance analysis, we used a computer with two Intel(R) Xeon(R) Gold 5118 CPUs @2.30 GHz with 12 cores, 128 GB of RAM, and Ubuntu Linux 18.04 LTS. These computers are used as proxy servers and client computers, respectively. In addition, GNU C++ and OpenSSL 1.1.1f libraries were used for LDP and AES cryptography, respectively. In this experimental analysis, we measured the operation times for data encryption and data modification using LDP.

Before providing the results, we explain the method of encrypting and modifying the privacy-sensitive parts of the gene data. Figure 4 shows an example of a line of the sequence alignment map (SAM), which contains a text version of the gene data. The text in the blue dotted box in Figure 4 shows the privacy-sensitive part of the gene data. The length of the gene data of this line, which is in the red dotted box in Figure 4, is 151. According to the blue box, the 136-th and 148-th bases of the array of the 151 gene bases in the reference gene data are `A` and `T`, respectively. Therefore, we can say that these two bases are privacy-sensitive parts of the gene data, which should be encrypted or modified by the LDP.

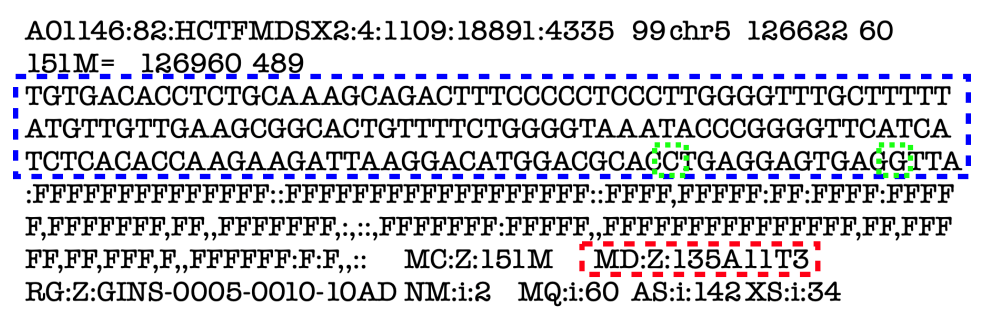


Figure 4. An example of a line in a gene data file.

To measure the operation times, we wrote a program for Algorithm 1 in C++ and compiled the source code using O3 optimization. Figure 5 shows the results of the time

durations for encrypting and modifying the gene data. To obtain experimental results, we measured the operation times for each chromosome because the size of the entire gene data was too large. The operation time for chromosome 1 (largest size) was the largest, and the time for chromosome Y (smallest size) was the smallest. From the graph, we can confirm that the operation time for transforming the sensitive parts in each chromosome does not exceed 3 min; therefore, we expect that the duration time for securing the whole gene data is not greater than 1 h in our experimental environment. Thus, we can conclude that Algorithm 1 guarantees a reasonable duration.

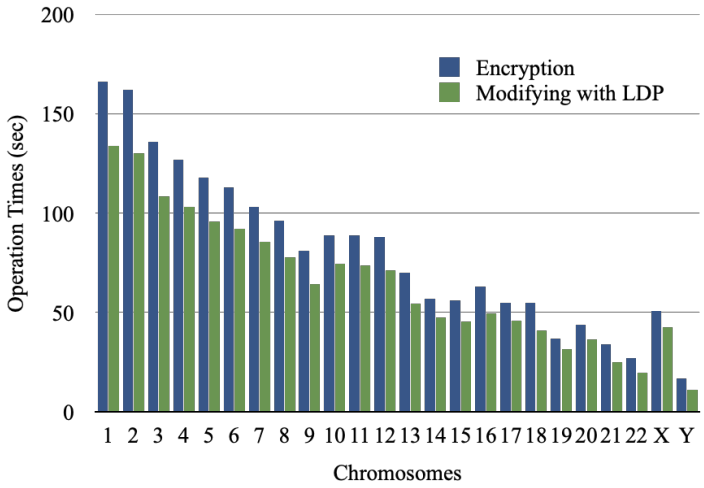


Figure 5. Results of operation times for encryption and modification with LDP.

7. Conclusions and Future Work

We proposed a blockchain-based gene data management system to preserve privacy using LDP. The provided model includes two types of storage: private and semi-private. Owing to this model with separated storage, the original gene data can exist only in the internal system, and only the irreversibly modified gene data can be shared with the external users. In terms of the modified data, only sections that include privacy-sensitive information are transformed using the LDP; hence, external users cannot obtain any information regarding the gene owner. Moreover, the blockchain enables perfect access control and guarantees integrity; hence, the proposed scheme provides the system manager with a dual-security solution for the gene data. The blockchain-based gene data management system with LDP enables volunteers to provide their gene data to healthcare institutes or research centers without anxiety. Furthermore, with additional schemes that prevent from deanonymization due to the characteristics of the blockchain, privacy of the gene data owner can be guaranteed definitely. In addition, such organizations can utilize gene data, although the data are modified because most parts of the gene data are still unchanged and have characteristics of the human genome.

A privacy-preserving system for managing gene data was proposed based on previously laid out theory. We demonstrated that the proposed algorithm is secure, and that the operation cost of the algorithm is somewhat expensive but reasonable. Heavy operation costs are unavoidable because of the original size of the gene data. In future studies, we will construct an entire system for a real gene data storage server. Through system operation and feedback, the proposed system will be upgraded and optimized to achieve more efficient and secure gene data management. In addition, as a next step, we intend to demonstrate that the gene data modified by the LDP can still be used for medical and research purposes to prove that our solution can be utilized in real environments.

Author Contributions: Supervision, J.S.; Writing—original draft, Y.-H.P.; Writing—review & editing, Y.K. All authors have read and agreed to the published version of the manuscript

Funding: This research was financially supported by Seoul R&BD Program (CT200018, Seoul Campus Town Technology R&D Project). This work was also supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning through the Basic Science Research Program under Grant 2020R1F1A1075952.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

Proof of Theorem 1. Let g be a gene base in the privacy-sensitive section of the gene data. According to Algorithm 1, g may change to another base. In addition, based on Equation (2), the probability that g is maintained is $1-\frac{3\lambda}{4}$, and the probability that g changes is $\frac{3\lambda}{4}$. Moreover, when g changes, the probabilities that g changes to one of the other three bases are the same. Therefore, for $g' \in \{\mathtt{A},\mathtt{C},\mathtt{G},\mathtt{T}\}$, $Pr[\mathcal{A}(g)=g']$ is either $1-\frac{3\lambda}{4}$ or $\frac{\lambda}{4}$.

Next, we prove that inequality (1) in Definition 1 holds for all subsets of $Range(\mathcal{A})$, Q. If $Q=\emptyset$, then $Pr[\mathcal{A}(g)\in Q]=0$; therefore, inequality (1) holds trivially. By contrast, when $Q=Range(\mathcal{A})=\{\mathtt{A},\mathtt{C},\mathtt{G},\mathtt{T}\}$, $Pr[\mathcal{A}(g)\in Q]=1$; hence, inequality (1) is established naturally because $\varepsilon \geq 0$ and $e^{\varepsilon} \geq 1$. Therefore, we next consider the cases where Q is neither $\emptyset$ nor $Range(\mathcal{A})$.

Let $|Q|=q$, where $1 \leq q \leq 3$. For a gene base g, we may define two cases, i.e., $g \in Q$ and the $g \notin Q$.

1. If $g \in Q$, among the q elements of Q, only one element is in Q and the other $q-1$ elements are not in Q. According to Equation (2), $Pr[\mathcal{A}(g)\in Q]=\left(1-\frac{3\lambda}{4}\right)+\frac{\lambda}{4}\times(q-1)=1-\frac{(4-q)\lambda}{4}$.
2. If $g \notin Q$, all elements in Q are not g. Therefore, according to Equation (2), $Pr[\mathcal{A}(g)\in Q]=\frac{q\lambda}{4}$.

Hence, these probabilities prove that inequality (1) in Definition 1 holds. Let g_1 and g_2 be two gene bases in the privacy-sensitive section. If $g_1 \in Q$ and $g_2 \in Q$, then $Pr[\mathcal{A}(g_1)\in Q]=Pr[\mathcal{A}(g_2)\in Q]$. In addition, if $g_1 \notin Q$ and $g_2 \notin Q$, then $Pr[\mathcal{A}(g_1)\in Q]=Pr[\mathcal{A}(g_2)\in Q]$ holds. Because $e^{\varepsilon}\geq 1$, inequality (1) is established.

Next, we consider the final two cases, where $g_1 \notin Q$ and $g_2 \in Q$, and $g_1 \in Q$ and $g_2 \notin Q$. In case $g_1 \notin Q$ and $g_2 \in Q$, $Pr[\mathcal{A}(g_1)\in Q]=\frac{q\lambda}{4}$ and $Pr[\mathcal{A}(g_2)\in Q]=1-\frac{(4-q)\lambda}{4}$, respectively. Because $\lambda=\frac{4}{e^{\varepsilon}+3}\leq\frac{4e^{\varepsilon}}{e^{\varepsilon}+3}$, we can derive that $e^{\varepsilon}\geq\frac{3\lambda}{4-\lambda}=\frac{\frac{3\lambda}{4}}{1-\frac{(4-3)\lambda}{4}}$. In addition, $\frac{\frac{3\lambda}{4}}{1-\frac{(4-3)\lambda}{4}}>\frac{\frac{2\lambda}{4}}{1-\frac{(4-2)\lambda}{4}}>\frac{\frac{\lambda}{4}}{1-\frac{(4-1)\lambda}{4}}$; hence, $\forall q\in\{1,2,3\}$, $e^{\varepsilon}\geq\frac{\frac{q\lambda}{4}}{1-\frac{(4-q)\lambda}{4}}$. $\therefore \frac{q\lambda}{4}\leq e^{\varepsilon}\cdot\left(1-\frac{(4-q)\lambda}{4}\right)$. In other words, $Pr[\mathcal{A}(g_1)\in Q]\leq e^{\varepsilon}\cdot Pr[\mathcal{A}(g_2)\in Q]$.

In the latter case where $g_1 \in Q$ and $g_2 \notin Q$, $Pr[\mathcal{A}(g_1)\in Q]=1-\frac{(4-q)\lambda}{4}$ and $Pr[\mathcal{A}(g_2)\in Q]=\frac{q\lambda}{4}$, respectively. Because $\lambda=\frac{4}{e^{\varepsilon}+3}$, we can derive that $e^{\varepsilon}=\frac{4-3\lambda}{\lambda}=\frac{1-\frac{(4-1)\lambda}{4}}{\frac{\lambda}{4}}$. In addition, $\frac{1-\frac{(4-1)\lambda}{4}}{\frac{\lambda}{4}}>\frac{1-\frac{(4-2)\lambda}{4}}{\frac{2\lambda}{4}}>\frac{1-\frac{(4-3)\lambda}{4}}{\frac{3\lambda}{4}}$; hence, $\forall q\in\{1,2,3\}$, $e^{\varepsilon}\geq\frac{1-\frac{(4-q)\lambda}{4}}{\frac{q\lambda}{4}}$. $\therefore 1-\frac{(4-q)\lambda}{4}\leq e^{\varepsilon}\cdot\frac{q\lambda}{4}$. In other words, $Pr[\mathcal{A}(g_1)\in Q]\leq e^{\varepsilon}\cdot Pr[\mathcal{A}(g_2)\in Q]$.

Hence, inequality (1) holds for all the cases. Therefore, we can conclude that Algorithm 1 guarantees ε-LDP. □

References

1. Gudodagi, R.; Venkata Siva Reddy, R.; Riyaz Ahmed, M. Investigations and Compression of Genomic Data. In Proceedings of the 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bengaluru, India, 11–12 December 2020; pp. 1–4. [CrossRef]
2. Pereira, S.; Gibbs, R.A.; McGuire, A.L. Open Access Data Sharing in Genomic Research. *Genes* **2014**, *5*, 739–747. [CrossRef]
3. Naveed, M.; Ayday, E.; Clayton, E.W.; Fellay, J.; Gunter, C.A.; Hubaux, J.P.; Malin, B.A.; Wang, X. Privacy in the Genomic Era. *ACM Comput. Surv.* **2015**, *48*, 1–44. [CrossRef]
4. Qin, S.; Zhou, F.; Zhang, Z.; Xu, Z. Privacy-Preserving Substring Search on Multi-Source Encrypted Gene Data. *IEEE Access* **2020**, *8*, 50472–50484. [CrossRef]
5. Yamamoto, Y.; Oguchi, M. A Decentralized System of Genome Secret Search Implemented with Fully Homomorphic Encryption. In Proceedings of the 2017 IEEE International Conference on Smart Computing (SMARTCOMP), Hong Kong, China, 29–31 May 2017; pp. 1–6. [CrossRef]
6. Sun, W.; Zhang, N.; Lou, W.; Hou, Y.T. When gene meets cloud: Enabling scalable and efficient range query on encrypted genomic data. In Proceedings of the IEEE INFOCOM 2017—IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9. [CrossRef]
7. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Technical Report. [Online]. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 30 October 2021).
8. Wood, G. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Proj. Yellow Pap.* **2014**, *151*, 1–32.
9. Dambrot, S.M. ReGene: Blockchain backup of genome data and restoration of pre-engineered expressed phenotype. In Proceedings of the 2018 9th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), New York, NY, USA, 8–10 November 2018; pp. 945–950. [CrossRef]
10. Bublitz, F.M.; Oetomo, A.; Sahu, K.S.; Kuang, A.; Fadrique, L.X.; Velmovitsky, P.E.; Nobrega, R.M.; Morita, P.P. Disruptive Technologies for Environment and Health Research: An Overview of Artificial Intelligence, Blockchain, and Internet of Things. *Int. J. Environ. Res. Public Health* **2019**, pp. 1–24. [CrossRef] [PubMed]
11. Sri Nidhi, P.V.; Akshayaa, S.; Vaisali, B.; Krishnan Namboori, P.K. DNA repair mutation detection using Deep learning strategy—A pharmacogenomic perspective. In Proceedings of the 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 22–23 March 2019; Volume 1, pp. 1–4. [CrossRef]
12. Ozercan, H.I.; Ileri, A.M.; Ayday, E.; Alkan, C. Realizing the potential of blockchain technologies in genomics. *Genome Res.* **2018**, *28*, 1255–1263. [CrossRef]
13. 23andMe. Available online: https://www.23andme.com/ (accessed on 30 October 2021).
14. Kulemin, N.; Popov, S.; Gorbachev, A. The Zenome Project: Whitepaper Blockchain-Based Genomic Ecosystem. 2017. Available online: https://zenome.io/download/whitepaper.pdf (accessed on 30 October 2021).
15. Whole Genome Association Studies. Available online: https://www.genome.gov/17516714/2006-release-about-whole-genome-association-studies (accessed on 30 October 2021).
16. Srinivasu, P.N.; Bhoi, A.K.; Nayak, S.R.; Bhutta, M.R.; Woźniak, M. Blockchain Technology for Secured Healthcare Data Communication among the Non-Terminal Nodes in IoT Architecture in 5G Network. *Electronics* **2021**, *10*, 1437. [CrossRef]
17. Putra, D.R.; Anggorojati, B.; Pratama Hartono, A.P. Blockchain and smart-contract for scalable access control in Internet of Things. In Proceedings of the 2019 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 19–20 November 2019; Volume 7, pp. 1–5. [CrossRef]
18. Erlingsson, U.; Pihur, V.; Korolova, A. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014; pp 1054–1067. [CrossRef]
19. Jin, H.; Luo, Y.; Li, P.; Mathew, J. A Review of Secure and Privacy-Preserving Medical Data Sharing. *IEEE Access* **2019**, *7*, 61656–61669. [CrossRef]
20. Wang, R.; Liu, H.; Wang, H.; Yang, Q.; Wu, D. Distributed Security Architecture Based on Blockchain for Connected Health: Architecture, Challenges, and Approaches. *IEEE Wirel. Commun.* **2019**, *26*, 30–36. [CrossRef]
21. Dwivedi, A.D.; Singh, R.; Srivastava, G.; Dhar, S. A decentralized privacy-preserving healthcare blockchain for IoT. *Sensors* **2019**, *19*, 326. [CrossRef] [PubMed]
22. Aileni, R.M.; Suciu, G. IoMT: A blockchain Perspective. In *Decentralised Internet of Things*; Springer: Cham, Switzerland, 2020. [CrossRef]
23. Liang, X.; Zhao, J.; Shetty, S.; Liu, J.; Li, D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017; pp. 1–5. [CrossRef]
24. Dey, T.; Jaiswal, S.; Sunderkrishnan, S.; Katre, N. HealthSense: A medical use case of Internet of Things and blockchain. In Proceedings of the 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 7–8 December 2017; pp. 486–491. [CrossRef]
25. Jin, X.L.; Zhang, M.; Zhou, Z.; Yu, X. Application of a blockchain platform to manage and secure personal genomic data: A case study of LifeCODE. ai in China. *J. Med. Internet Res.* **2019**, *21*, e13587. [CrossRef]

26. Homer, N.; Szelinger, S.; Redman, M.; Duggan, D.; Tembe, W.; Muehling, J.; Pearson, J.V.; Stephan, D.A.; Nelson, S.F.; Craig, D.W. Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. *PLoS Genet.* **2008**, *4*, 1–9. [CrossRef] [PubMed]
27. Wang, R.; Li, Y.F.; Wang, X.; Tang, H.; Zhou, X. Learning Your Identity and Disease from Research Papers: Information Leaks in Genome Wide Association Study. In Proceedings of the 16th ACM Conference on Computer and Communications Security, Chicago, IL, USA, 9–13 November 2009; pp. 534–544. [CrossRef]
28. Deznabi, I.; Mobayen, M.; Jafari, N.; Tastan, O.; Ayday, E. An Inference Attack on Genomic Data Using Kinship, Complex Correlations, and Phenotype Information. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *15*, 1333–1343. [CrossRef]
29. Humbert, M.; Ayday, E.; Hubaux, J.P.; Telenti, A. Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy. In Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, Berlin, Germany, 4–8 November 2013; pp. 1141–1152. [CrossRef]
30. Samani, S.S.; Huang, Z.; Ayday, E.; Elliot, M.; Fellay, J.; Hubaux, J.P.; Kutalik, Z. Quantifying Genomic Privacy via Inference Attack with High-Order SNV Correlations. In Proceedings of the 2015 IEEE Security and Privacy Workshops, San Jose, CA, USA, 21–22 May 2015; pp. 32–40. [CrossRef]
31. Ayday, E.; Raisaro, J.L.; Hubaux, J.P.; Rougemont, J. Protecting and Evaluating Genomic Privacy in Medical Tests and Personalized Medicine. In Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society, Berlin, Germany, 4 November 2013; pp. 95–106. [CrossRef]
32. Baldi, P.; Baronio, R.; De Cristofaro, E.; Gasti, P.; Tsudik, G. Countering GATTACA: Efficient and Secure Testing of Fully-Sequenced Human Genomes. In Proceedings of the 18th ACM Conference on Computer and Communications Security, Chicago, IL, USA, 17–21 October 2011; pp. 691–702. [CrossRef]
33. Wang, X.S.; Huang, Y.; Zhao, Y.; Tang, H.; Wang, X.; Bu, D. Efficient Genome-Wide, Privacy-Preserving Similar Patient Query Based on Private Edit Distance. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 492–503. [CrossRef]
34. Fienberg, S.E.; Slavkovic, A.; Uhler, C. Privacy Preserving GWAS Data Sharing. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011; pp. 628–635. [CrossRef]
35. Johnson, A.; Shmatikov, V. Privacy-Preserving Data Exploration in Genome-Wide Association Studies. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1079–1087. [CrossRef]
36. Yu, F.; Fienberg, S.E.; Slavković, A.B.; Uhler, C. Scalable privacy-preserving data sharing methodology for genome-wide association studies. *J. Biomed. Inform.* **2014**, *50*, 133–141. doi: 10.1016/j.jbi.2014.01.008. [CrossRef] [PubMed]
37. Yilmaz, E.; Ji, T.; Ayday, E.; Li, P. Genomic Data Sharing under Dependent Local Differential Privacy. *arXiv* **2021**, arXiv:cs.CR/2102.07357.
38. Dwork, C. Differential Privacy. In *Automata, Languages and Programming*; Bugliesi, M., Preneel, B., Sassone, V., Wegener, I., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–12.
39. Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
40. Fredrikson, M.; Jha, S.; Ristenpart, T. Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1322–1333. [CrossRef]
41. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership Inference Attacks Against Machine Learning Models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; pp. 3–18. [CrossRef]
42. Tramèr, F.; Zhang, F.; Juels, A.; Reiter, M.K.; Ristenpart, T. Stealing Machine Learning Models via Prediction APIs. In Proceedings of the 25th USENIX Conference on Security Symposium, Austin, TX, USA, 10–12 August 2016; pp. 601–618.
43. Sun, Z.; Wang, Y.; Shu, M.; Liu, R.; Zhao, H. Differential Privacy for Data and Model Publishing of Medical Data. *IEEE Access* **2019**, *7*, 152103–152114. [CrossRef]
44. Müftüoğlu, Z.; Kizrak, M.A.; Yildlnm, T. Differential Privacy Practice on Diagnosis of COVID-19 Radiology Imaging Using EfficientNet. In Proceedings of the 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Novi Sad, Serbia, 24–26 August 2020; pp. 1–6. [CrossRef]
45. Zia, M.T.; Khan, M.A.; El-Sayed, H. Application of Differential Privacy Approach in Healthcare Data—A Case Study. In Proceedings of the 2020 14th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 17–18 November 2020; pp. 35–39. [CrossRef]
46. Harris, D.R. Leveraging Differential Privacy in Geospatial Analyses of Standardized Healthcare Data. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 3119–3122. [CrossRef]
47. Chen, J.; Wang, W.H.; Shi, X. Differential Privacy Protection Against Membership Inference Attack on Machine Learning for Genomic Data. *bioRxiv* **2020**. [CrossRef]
48. Mahawaga Arachchige, P.C.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S.; Atiquzzaman, M. Local Differential Privacy for Deep Learning. *IEEE Internet Things J.* **2020**, *7*, 5827–5842. [CrossRef]

49. Wang, Z.; Ma, P.; Wang, R.; Zhang, J.; Chi, Y.; Ma, Y.; Yang, T. Secure Medical Data Collection via Local Differential Privacy. In Proceedings of the 2018 IEEE 4th International Conference on Computer and Communications (ICCC), Chengdu, China, 7–10 December 2018; pp. 2446–2450. [CrossRef]
50. Liu, X.; Zhou, P.; Qiu, T.; Wu, D.O. Blockchain-Enabled Contextual Online Learning Under Local Differential Privacy for Coronary Heart Disease Diagnosis in Mobile Edge Computing. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2177–2188. [CrossRef]
51. Bonneau, J.; Narayanan, A.; Miller, A.; Clark, J.; Kroll, J.A.; Felten, E.W. Mixcoin: Anonymity for Bitcoin with Accountable Mixes. In *Financial Cryptography and Data Security*; Christin, N., Safavi-Naini, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 486–504.
52. Corrigan-Gibbs, H.; Ford, B. Dissent: Accountable Anonymous Group Messaging. In Proceedings of the 17th ACM Conference on Computer and Communications Security, Chicago, IL, USA, 4–8 October 2010; pp. 340–350. [CrossRef]
53. Moniz, H.; Neves, N.F.; Correia, M.; Verissimo, P. Experimental Comparison of Local and Shared Coin Randomized Consensus Protocols. In Proceedings of the 2006 25th IEEE Symposium on Reliable Distributed Systems (SRDS'06), Leeds, UK, 2–4 October 2006; pp. 235–244. [CrossRef]
54. Ruffing, T.; Moreno-Sanchez, P.; Kate, A. CoinShuffle: Practical Decentralized Coin Mixing for Bitcoin. In Proceedings of the 19th European Symposium on Research in Computer Security, Wroclaw, Poland, 7–11 September 2014; Kutyłowski, M., Vaidya, J., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 345–364.
55. Boneh, D.; Boyen, X.; Shacham, H. Short Group Signatures. In Proceedings of the CRYPTO 2004, Santa Barbara, CA, USA, 15–19 August 2004; pp. 41–55.
56. Rivest, R.L.; Shamir, A.; Tauman, Y. How to Leak a Secret. In Proceedings of the ASIACRYPT 2001, Gold Coast, Australia, 9–13 December 2001; pp. 552–565.
57. Andrychowicz, M.; Dziembowski, S.; Malinowski, D.; Mazurek, L. Secure Multiparty Computations on Bitcoin. In Proceedings of the 2014 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, 18–21 May 2014; pp. 443–458. [CrossRef]
58. Zhang, R.; Xue, R.; Liu, L. Security and Privacy on Blockchain. *ACM Comput. Surv.* **2019**, *52*, 1–34. [CrossRef]