SECURING DNA VERIFICATION PROCEDURES USING BLOCKCHAIN

Ramprashath R[1], Keerthana Devi E[2], Sangavi M[3], Gokul K[4], Karthik R[5]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract:

This project aims to encrypt DNA verification processes utilizing blockchain and AES (Advanced Encryption Standard). Several circumstances might call for DNA verification. For instance, they are conducting a study and testing the DNA, and they will also need to conduct some criminal and illness analyses. The secure method of procedure is therefore required in all circumstances. We've implemented this task using a combination of AES and blockchain technologies. Initially, the collected DNA pattern information as well as the person's private information are handled securely using AES (Advanced Encryption Standard)[2]. In contrast to a Feistel cipher, AES is iterative. It is completely based on the "substitution-permutation network" at its core. Blockchain is a mechanism for storing data in a way that makes system changes, hacking, and cheating difficult or impossible. Decentralized database administration is what is done in the situation. Individuals upload their genetic data to a web-based genomics research facility. The person stores his genetic statistics using a blockchain to protect or test their secrecy. By using hash values in the blockchain technique, analyzing those data, and preserving the statistics. It can be understood that blockchain is nothing more than a never-ending avalanche of blocks that may be linked together like a chain in a certain cryptography sequence[25]. Whenever it comes to securely storing transactional information, blockchains typically move as the technique of choice. Because of the security, decentralization, and transparency that blockchains have associated with them, we may assume that the dawn of a stable organizational environment isn't too far away.

Keywords: Blockchain Security, AES (Advanced Encryption Standard) , Block Storage, DNA Computing, Data privacy .

[1]Assistant Professor, Department of MCA, Karpagam College of Engineering , Coimbatore India.
[2,3,4,5]Scholar, Department of MCA, Karpagam College of Engineering , Coimbatore India.

Email: [1]ramprashath.r@kce.ac.in , [2]keerthanadevi656@gmail.com , [3]msangavi591@gmail.com , [4]gokulking9486@gmail.com , [5]karthickramamoorthy97@gmail.com

DOI: 10.31838/ecb/2023.12.s3.155

1. Introduction

The mission is that when hackers need to gain access to a system, they may target the weakest link. A system's encryption is no longer commonly used. The verification of DNA is a crucial process that requires high-level security to protect sensitive private information. The use of blockchain technology offers a reliable solution to secure DNA verification procedures[26]. Blockchain technology is a decentralized system that provides a secure and immutable way to record transactions. It is resistant to tampering and hacking, ensuring that the data is protected from unauthorized access[25]. The immutable nature of blockchain makes it difficult for anyone to tamper with the information stored on it.

Widely used encryption techniques like the Advanced Encryption Standard (AES) algorithm offers robust security for sensitive data. AES encryption employs a symmetric key scheme, which adds a degree of security by allowing the use of the same key for both data encryption and decryption[12] . The integrity and security of DNA verification data may be guaranteed by blockchain technology since it offers a decentralized, tamper-proof framework for storing and managing data. It can develop a safe solution for DNA verification operations by fusing blockchain technology with AES encryption. With the help of this technology, Genetic data might be verified using cryptographic techniques and shared and stored safely.

Users must make certain the software program below attention does what they need it to do, that it protects consumer information in the manner it is anticipated to, and that the general technique has no susceptible points. The DNA testing technique is necessary for many different circumstances. They include crime analysis, research purposes, disease analysis, and so on. Thus, security is required in most scenarios. We must achieve security during the DNA checking procedure as part of our project. Furthermore, there must be no grey areas or uncertainty regarding information storage and handling. Customers, for example, must recognize the location of the cloud if the information is stored there.

Every web page in a news ledger is a block in Blockchain technology. Through cryptographic hashing, this block influences the following block or web page. In other words, when a block is completed, it generates a one-of-a-kind consistent code that is linked to the next web page or block, thus growing a chain of blocks or a blockchain. When a blockchain is delivered to a new blockchain transaction or a new block is to be added to the blockchain, several nodes in the same blockchain implementation are required to execute algorithms to evaluate, confirm, and process the blockchain's records. In summary, using AES encryption on blockchain technology can provide a secure and reliable system for DNA verification procedures[13]. This can increase trust in the accuracy and integrity of DNA data, as well as protect the privacy and security of individuals' sensitive information.

Literature Review

There are no security measures to handle the DNA samples. Also, at the time of sample collection user details collection, no security measures were followed. Not only for the testing purpose of the DNA sample collected but also for the research purpose. So incorrect genome data may lead to a disaster for research results. In the existing system, the malicious user can easily access the data. In such cases, the malicious user can access and change the data, which will cause a major defect. Results can take as few working days to come back from the receipt of samples. A large amount of storage space to store the keys and other information is required for encryption and decryption. The test reports are not stored on a security basis.

DISADVANTAGES: The current approach makes it impossible to oversee and preserve data integrity, and third-party auditors may engage in malevolent behaviour. Management issues with the majority of the current public verification techniques. The current techniques are unable to fend off an entity that engages in hostile activity on behalf of the process.

[17]By G. Yan et al., "Blockchain-based secure data storage and sharing method for genomic data" (2018)

The authors suggest a secure data storage and sharing system for genetic data based on blockchain technology. The proposed method secures the confidentiality of genetic data by combining blockchain technology with the Advanced Encryption Standard (AES) algorithm. The study's findings demonstrate that the suggested plan offers efficient and secure data sharing and storage for genomic data.

James Bornholt et al [2] examined the exponentially increasing demand for data storage and the unsatisfying capacity of existing storage media. To keep up with the demand, using DNA to archive data is an attractive possibility. In this paper, they presented an architecture for a DNA-based archival storage system which is designed as a key-value store leveraging common biochemical techniques to provide random access.

[9]S. K. Park et al" Blockchain-based secure genomic data sharing and analysis framework for collaborative research" (2020)

The authors propose a blockchain-based secure genomic data sharing and analysis framework for collaborative research. The proposed framework uses a combination of blockchain technology, AES algorithm, and secure multi-party computation to protect the privacy of genomic data. The results of the study show that the proposed framework provides secure and efficient data sharing and analysis for collaborative research.

M. Fritz et al [12] stated that the data storage costs have become a large proportion of total cost in the creation and analysis of DNA sequence data. In this paper, they presented a new compression technique that efficiently compresses DNA sequences for storage.

[5]By A. El-Kaliouby et al., "Secure storage and exchange of genetic data using blockchain technology" (2020)

The authors propose a secure storage and sharing of genomic data using blockchain technology. The proposed scheme uses a combination of blockchain technology and AES algorithm to protect the privacy of genomic data. The results of the study show that the proposed scheme provides secure and efficient data storage and sharing for genomic data.

Andy Extance [14] discussed how they can afford to store the genome sequences and other data the world was creating a very fast rate. They mentioned that DNA storage would be slow and it would take hours to store data by synthesizing DNA strings with a particular pattern of bases. And also if one wants to recover that information, it will take more hours as it will require a sequencing machine.

[11]Homomorphic encryption-based blockchain-based secure genomic data sharing and analysis system, S. Choi et al (2020)

The authors propose a blockchain-based secure genomic data sharing and analysis system using homomorphic encryption. The proposed system uses a combination of blockchain technology, AES algorithm, and homomorphic encryption to protect the privacy of genomic data. The results of the study show that the proposed system provides secure and efficient data sharing and analysis for genomic data.

[6]Blockchain technology is used in "A Secure and Efficient System for Sharing Genomic Data" by J. Kim et al (2019)

The authors propose a secure and efficient system for sharing genomic data using blockchain technology. The proposed system uses a combination of blockchain technology and AES algorithm to protect the privacy of genomic data. The results of the study show that the proposed system provides secure and efficient data sharing for genomic data.

2. Proposed Methodology

The proposed system has been implemented so that each user gets their ID. Based on the ID, the upload and retrieval of the data are very convenient for the user. This system enhances security by implementing an AES encryption algorithm which may never allow cyber-attacks to happen. The encrypted values are controlled by an administrative view, which makes it authorized to access the file. The user can get the report at a reasonable price, which was fixed by the administration. In this system, results can take as few days to come back from the collection of samples. The overall process is monitored with the aid of using the administration. This will assist in saving from statistical leakage. The collection of samples and retrieval of the report is primarily based totally on the id. Which will assist in cleanly getting the information.

The AES encrypted algorithm is used in this data which allows privacy to see the important data. With Blockchain technology, each page in the ledger of reports forms a block. This block will have an impact on the next block or page through cryptographic hashing. In other words, when a block is completed, it creates a unique secure code, which ties into the next page or block, creating a chain of blocks or a blockchain.

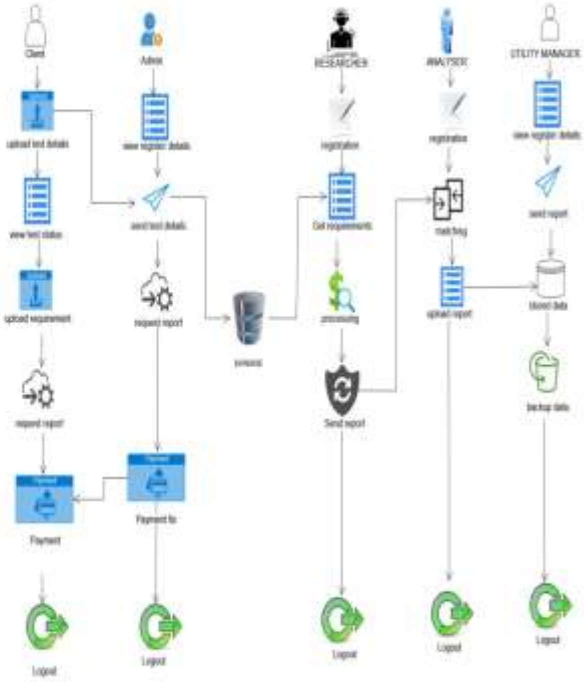

Fig 1: Architecture for Proposed System

Advantages:

- The overall process is monitored with the aid of using the administration. It will assist to save you from statistics leakage.
- The collection of samples and retrieval of the report is primarily based totally on the id. Which will assist to get the information clean way.
- The encrypted algorithm is used in this data which makes it privacy to see the important data.
- Advanced techniques used to achieve good accuracy in test reports.
- These new techniques help to quick these processes as much as possible.

Purpose of the System

- The overall process is monitored with the aid of using the administration. It will assist to save you from statistics leakage.
- The collection of samples and retrieval of the report is primarily based totally on the
- id. Which will assist to get the information clean way.
- The encrypted algorithm is used in this data which makes privacy to see the important data.

Implementation

Phase 1: User

If a user wishes to register and log in to the user page in this module, the user will be sent to the user home page, which displays options for the test information, test status, request report, payment, and download. The user page then uploads the DNA sample information after a successful login. It typically takes three to five days to receive the findings once the user uploads the DNA sample data. because DNA testing goes through a lot of steps. The admin then changed the test status, which will thereafter be updated.

Fig 2: User Dashboard

The customers were grateful for this since it made it simple for them to determine if the findings were ready or not. The user will ask the administrator for the report once the findings are ready. The admin will react to the user after receiving the request. The administrator will next request storage on the blockchain, obtain the report data, and establish the payment amount for the report data. The user will pay the amount in the payment option when the admin sets the payment amount to the report data. The website will immediately redirect to the download page after the payment has been made. However using a blockchain approach, the data was saved as hash values. Then, after obtaining the key from the administrator, download the report's data.

Phase 2: Admin

The view research team information, view analyzing team details, request, and payment update options are presented on the admin home page when the admin wishes to log in to the admin page in this module. The administrator will then review the study team's registration information. After the registration information is accurate, only the admin will give the go-ahead to continue; otherwise, it is forbidden. After that, the administrator will inform the study team of the DNA sample details.

Fig 3: Admin Dashboard

The administrator will deliver the DNA sample information in an encrypted manner by the security rules. The administrator will then review the team registration information. After the registration information is accurate, only the admin will give the go-ahead to continue; otherwise, it is forbidden. The administrator will then keep an eye on each process and update the test status often. After receipt of the user's report request, the admin will make a blockchain storage request to get the report data and establish the payment amount for that data.

Phase 3: Researcher

The registration, see status, decryption, process, and transmit options are all presented on the research home page when the research team wishes to register and log in with their credentials for this module. The first research team wishes to fill out the registration form with their information. Following registration, the research team must wait for admin clearance. Once the admin has verified that the registration information is accurate, the research team will be approved.

Fig 4: Decryption Dashboard

After receiving approval, obtain the sample information, which was in an encrypted format. The DNA sample information was then decrypted using the resulting decryption key. The study team performs two procedures, the first of which is the extraction of DNA molecules from the supplied material. Moreover, the limited sample amount results in a low quantity of extracted DNA molecules. so that DNA molecule sequencing and the process of replicating may happen. Upload the sequencing information last.

Phase 4: Analyzer

The registration, view status, matching details, and report menus are all presented on the analyzing home page when the analyzing team wishes to register and log in with their details in this module. The first analyzing team wishes to fill out the registration form with their information. Once registration is complete, the admin will review the registration data while you wait for their approval. The administrator will authorize the analyzing team if it is accurate.

After receiving approval, obtain the sequencing information. Afterward, it will take a few seconds to match the sequencing information. Following a successful match, obtaining the report will take a little while. Finally, in the blockchain technique, upload the previously created report to the blockchain storage.

Phase 5: Utility Manager

They want users to sign up and log in with their credentials. When this happens, the user is redirected to the blockchain storage home page, where the menus for sending reports, storing data, and backing up data are all displayed. Registration information for the initial research team and the analysis team are both kept separately. then verify the user's or the administrator's report request.

Fig 5: Block Chain Storage Dashboard

Blockchain storage will transmit report data to the administrator if there is any request. The report data is then hashed using the private key and saved using the blockchain mechanism. The storage system also offers a backup.

The process of using data encryption using the AES algorithm involves the following steps:

Step 1: Key Generation : The first step in AES encryption is key generation. A secret key is required to encrypt and decrypt data. The key must be kept secret and not shared with unauthorized parties[21]. The key is generated using a secure random number generator or a key derivation function.

Step 2: Data Padding : AES operates on fixed-size blocks of data, typically 128 bits[19]. If the data to be encrypted is not an exact multiple of 128 bits, it must be padded to bring it to the required size. Padding can be done using standard padding schemes such as PKCS#5 or PKCS#7.

Step 3: Encryption : [8]Once the key and the data have been prepared, encryption can begin. AES operates on a block-by-block basis, with each block being encrypted separately. The encryption process involves a series of rounds, with each round performing a series of operations on the data.

Key Expansion: The original key is expanded into a set of round keys, one for each round of encryption.

Initial Round: The first round involves an XOR operation between the block of data and the first round key.

Rounds: Each subsequent round consists of four operations: SubBytes, ShiftRows, MixColumns, and AddRoundKey. These operations combine to provide confusion and diffusion of the data, making it difficult for attackers to decrypt the data.

Final Round: The final round skips the MixColumns operation.

I/p Msg.SIZE in(KB)	AES (ms)	AES WITH DNA	Time Diff in ms
128	182	220	38
256	210	242	32

512	350	372	22
960	370	389	19
1028	380	395	15

Table 1: AES and DNA based AES Encryption

Step 4: Data Transmission - Once the data has been encrypted, it can be transmitted over an insecure network or stored on an insecure medium[17]. The encrypted data and the key should be transmitted separately to ensure security.

Step 5: Decryption :[10] To decrypt the data, the recipient uses the same key that was used to encrypt it. The decryption process involves the inverse of the encryption process, with each operation being performed in reverse order. The decryption process is identical to the encryption process, except that the order of the round keys is reversed.

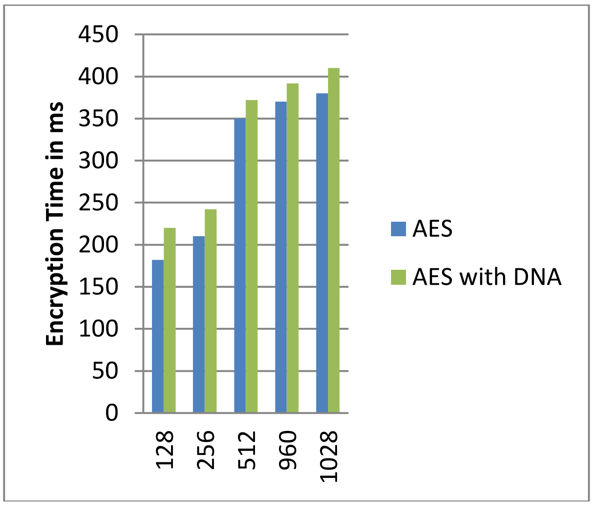

Figure 6: comparing Encryption using Time Difference Graph

The following steps are involved in storing DNA verification procedures using hash values in blockchain storage:

Step 1: DNA Verification Procedure Data Preparation : The DNA verification procedure data is prepared for storage in the blockchain. This may involve formatting the data into a standardized format, such as FASTA or FASTQ.

Step 2: Hash Function Application : A hash function is applied to the DNA verification procedure data to generate a fixed-length hash value. [25]The hash function used should be secure and irreversible. Some popular hash functions used for this purpose include SHA-256 and SHA-3.

Step 3: Blockchain Selection : The next step is to select a blockchain platform that is suitable for storing the DNA verification procedure data using hash values. The selected blockchain platform should provide high security and privacy, scalability, and efficiency. Some popular blockchain platforms that could be used for this purpose include Ethereum, Hyperledger Fabric, and Corda.

Step 4: Data Storage in the Blockchain : The final step is to store the hash values of the DNA verification procedure data in the blockchain using the developed smart contract[5]. The hash values are stored on-chain, which involves storing the data directly in the blockchain.

When the DNA verification procedure data needs to be verified, the data is retrieved from its source and a hash function is applied to it. The resulting hash value is then compared to the hash value stored in the blockchain. If the hash values match, the DNA verification procedure data is considered authentic and verified[3].

This process ensures the privacy and security of the sensitive DNA verification procedure data while also providing a tamper-proof and decentralized storage solution.

3. Result and Discussion

In recent years, there has been a lot of interest in the application of blockchain technology for the secure storage and exchange of sensitive data, including data used for DNA verification. In this situation, the blockchain's adoption of powerful encryption algorithms can greatly improve the security of such data.

The Advanced Encryption Standard (AES), a symmetric-key encryption method, is one of the most extensively used encryption techniques. AES is a dependable option for protecting DNA verification data since it uses a secret key for encryption and decryption.

When implementing AES on the blockchain, the encryption and decryption keys are stored in a decentralized manner, making it difficult for attackers to gain access to them. Additionally, each block in the blockchain is linked to the previous block using a cryptographic hash function, ensuring the integrity and immutability of the data.

However, it is important to ensure that proper key management practices are followed to prevent unauthorized access to the encryption keys. Additionally, regular security audits and updates to the encryption algorithms and key management practices should be performed to ensure the continued security of the data.

Input Data

INPUT DATA	**DESCRIPTION**
DNA Sample Data	The DNA sample to be verified in digital format
AES Encrpytion Key	The Secret key used to encrypt the DNA sample data
Blockchain	The network where the encrypted DNA sample and verification results will be stored

Output Data

OUTPUT DATA	DESCRIPTION
Encrypted DNA sample	The DNA sample data encrypted using the AES algorithm with the encryption key
Verification Results	The result of the DNA sample verification either positive or negative
Access Control Mechanism	The mechanism that allows only authorized parties to access the encrypted DNA Sample data for further analysis

Limitation

- Data integrity is difficult to manage and maintain in the existing system.
- Third party auditors can be able perform malicious activities.
- Management problem existing in most of the existing public verification schemes.

- The existing schemes cannot resist an entity that performs malicious activities for the process.

4. Conclusion

To store the report, our suggested version used blockchain technology. Each page in a ledger of reports constitutes a block using blockchain technology. As a result of several circumstances, DNA testing techniques are necessary for a wide variety of scenarios. It may be utilized for research, medical analysis, criminal analysis, and other applications. Security is typically required in certain circumstances. We are attempting to ensure security during the DNA testing procedure because it is required by our project. Via cryptographic hashing, the following block or page may be affected by this block.

To put it another way, each time a block is finished, it generates a special security code that connects to the following page or block, forming a chain of blocks or a blockchain. Also, the use of the AES encryption procedure makes the acquisition of user personal information and DNA samples very safe. Because hackers will attempt to infiltrate a system through its weakest spot. Regardless of whether a system uses a 128-bit key or a 256-bit key, the encryption of that system is often not done in this way. Users should confirm that the program meets their needs, secures user data as intended, and has no flaws in the whole procedure before deciding whether to use it. It has since been improved and implemented via experimentation for instances where it is truly necessary.

5. References

Balasubramaniam, S., Botvich, D., Suda, T., Nakano, S. F. Bush, and M. Foghl (2008, September). Address encoding, link switching, and error correction in molecular communication using hybrid DNA and enzyme-based computing.

Bornholt, J., Lopez, R., Carmean, D. M., Ceze, L., Seelig, G., & Strauss, K. (2016). A DNA-based archival storage system. ACM SIGARCH Computer Architecture News, 44(2), 637-649.

Wang, H., Chen, X., Dai, H. N., & Zheng, Z. (2018). Blockchain potential and problems: a survey. 14(4), 352-375, International Journal of Web and Grid Services.

Techapanupreeda, C., & Tasatanattakool, P. (2018, January). Blockchain: Applications and difficulties. 2018 saw the ICOIN (International Conference on Information Networking) (pp. 473-475). IEEE.

Ghoneimy, S.; Mohamed, R.; and El-Seoud, S. A. (2017). Problems and Applications of DNA Computing. International Journal of Mobile Interactive Technology.

The authors are Moritani, Y., Hiyama, S., Nomura, S. M., Akiyoshi, and Suda (2007, December). a molecular communication interface that uses vesicles with channel-forming proteins as the communication medium. 2007 saw the second iteration of the Bio-Inspired Models of Network, Information, and Computing Systems (pp. 147-149). IEEE.

Hoek, J. B., Markevich, N. I., & Kholodenko, B. N. (2004). Protein kinase cascades with multisite phosphorylation exhibit bistability and signalling switches. Cell Biology Journal, 164(3), 353-359.

V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based en-cryption for fine-grained access control of encrypted data," in ACM Conference on Computer and Communications Security, 2006, pp. 89–98.

S. Liu and J. Gaudiot (2006, May). Deadly Automaton powered by DNA: Revolutionary Nanomedicine. NSTI Nanotech Conference and Trade Expo Proceedings, Boston, Massachusetts.

Kholodenko, B. N., and Sauro, H. M. (2004). signalling network quantitative analysis. 86(1), 5-43. Advances in Biophysics and Molecular Biology.

A. P. De Silva and S. Uchiyama (2007). molecular computation and logic. 2(7), 399; Nature Nanotechnology.

The authors are Niazov, T., Baron, R., Katz, E., Lioubashevski, O., and Willner (2006). Four connected biocatalysts working in a succession to create concatenated logic gates. National Academy of Sciences Proceedings, 103(46), 17160–17163.

Komarova, N. L., Zou, X., Nie, Q., & Bardwell, L. (2005). a conceptual framework for cell signalling specialisation. Biology of Molecular Systems, 1 (1). H. Zimmermann, in 20

Extance, A. (2016). How DNA could store all the world's data. Nature News, 537(7618), 22.

Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, & E. Shapiro (2004). a self-contained molecular computer for the rational regulation of gene expression. 423. Nature, 429(6990).

K. Emura, A. Miyaji, A. Nomura, K. Omote, and M. Soshi, "A ciphertext-policy attribute-based encryption scheme with constant ciphertext length," in International Conference on Information Security Practice and Experience. Springer, 2009, pp. 13–23.

J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in IEEE Symposium on Security and Privacy, 2007,pp. 321–334

S. Hohenberger and B. Waters, "Attribute-based encryption with fast decryption," in

International Workshop on Public Key Cryptography.Springer, 2013, pp. 162–179.

J. Lai, R. H. Deng, C. Guan, and J. Weng, “Attribute-based encryption with verifiable outsourced decryption,” IEEE Transactions on information forensics and security, vol. 8, no. 8, pp. 1343–1354, 2013

Deng, H., Qin, Z., Wu, Q., Guan, Z., Zhou, Y.: Flexible attribute-based proxy re-encryption for efficient data sharing. Information Sciences 511, 94–113 (2020)

A. Lewko and B. Waters, “New proof methods for attribute-based encryption: Achieving full security through selective techniques,” in Annual Cryptology Conference. Springer, 2012, pp. 180–198.

G. Ateniese, K. Fu, M. Green, and S. Hohenberger, “Improved proxy re-encryption schemes with applications to secure distributed storage,” Acm Transactions on Information and System Security, vol. 9, no. 1, pp. 1–30, 2006

B. Berger, H. Cho, Emerging technologies towards enhancing privacy in genomic data sharing, 2019.

X.L. Jin, M. Zhang, Z. Zhou, X. Yu

Application of a blockchain platform to manage and secure personal genomic data: a case study of lifecode. ai in china J. Medical Internet Res., 21 (2019), p. e13587

M. Shabani, Blockchain-based platforms for genomic data sharing: a de-centralized approach in response to the governance problems? J. Am. Med. Inform. Assoc., 26 (2019), pp. 76-80

Fritz, M. H. Y., Leinonen, R., Cochrane, G., & Birney, E. (2011). Efficient storage of high throughput DNA sequencing data using reference-based compression. Genome Research, 21(5), 734-740.