# Hurdles for Genomic Data Usage Management

(Position Paper)

Muhammad Naveed
University of Illinois at Urbana-Champaign
www.cryptoonline.com
naveed2@illinois.edu

***Abstract*—Our genome determines our appearance, gender, diseases, reaction to drugs, and much more. It not only contains information about us but also about our relatives, past generations, and future generations. This creates many policy and technology challenges to protect privacy and manage usage of genomic data. In this paper, we identify various features of genomic data that make its usage management very challenging and different from other types of data. We also describe some ideas about potential solutions and propose some recommendations for the usage of genomic data.**

## I. Introduction

Our genome[1] is the complete blueprint of our body. It determines how we look, diseases we are susceptible to, our ancestry and much more. It helps in identifying criminals, early diagnosis of diseases, and enable personalized medicine and prenatal testing. A lot of effort has been put into correlating genomic data to human traits, e.g., cystic fibrosis, obesity, eye color, etc. Genome-Wide Association Studies (GWAS) are being conducted to learn genetic basis of diseases [1].

Genomic based medicine promises to revolutionize healthcare, but at the same time genomic data is susceptible to abuse. It can lead to employment and insurance denial, discrimination, and social stigma. Human genome is very stable and once leaked it can have a lifelong impact on one's life. A lot of work has been done to identify the risk of publishing genomic data. It has been shown that even aggregate genomic data can be used for re-identification [2], [3].

Genomic data usage management is very challenging. Genome is obtained from human in chemical form and then digitized. A reasonable solution to the problem of genomic data usage should prevent abuse of both chemical and digital form of the genome. Any privacy breach due to improper management of genomic data can have a lasting impact. Genomic data should be handled with great care; policies and technology used to manage and secure genomic data should be chosen according to the features of genomic data.

**Contributions** We summarize our contributions as follows:

- We identify features of genomic data (with quantification from biology research, where applicable) that make genomic data usage management and privacy difficult.
- We present policy and technology challenges created by these features for the management of genomic data usage.
- We explain how policy and technology together can give a meaningful solution to the management of genomic data usage.
- We propose some recommendations for the management of genomic data.

Features, challenges, solutions, and recommendations discussed in this paper are by no means complete or exhaustive. The goal of the paper is to be a starting point for more comprehensive studies and future work on the topic.

Rest of the paper is organized as follows: Section II describes the special features of genomic data, Section III presents the technology and policy challenges caused by the special features of genomic data described in Section II, Section IV describes some preliminary ideas to address the problem, Section V presents some data usage recommendations for genomic data, and finally we conclude the paper in Section VI.

## II. Special features of genomic data

This section discusses several features of genomic data which make genomic data usage management particularly challenging. These features impact the suitability of state of the art technology and policy solutions for the management of genomic data. Section III discusses why these features are hurdles for genomic data usage management.

**Feature 1** (Cyber-physical nature)**:** *Genomic information has both physical and digital existence. It can be easily digitized from its chemical form.*

A biospecimen is obtained from human, and DNA is extracted from it. Usually, biospecimens are stored in biobanks. The extracted DNA is then used to digitize the genome through a process called sequencing. Sequencing is mostly done by third-party companies such as Illumina and Roche and they require the biospecimen or extracted DNA to do so. Genotyping is an economical alternative to digitize a small part of the genome that varies among population. Direct-to-consumer companies, such as 23andme, use genotyping technology.

**Feature 2** (Sensitive information)**:** *Genomic data contain sensitive information about health, paternity, kinship, etc., which can lead to discrimination, employment and insurance denial, and social stigma.*

A lot of effort has been put into determining the genetic basis of diseases [4]. Cystic fibrosis and sickle cell anemia status can be determined by single mutation [5], [6]. BRCA1

[1]For the definitions of biological terms, please visit http://web.ornl.gov/sci/techresources/Human_Genome/glossary.shtml

  

and BRCA2 genes can be used for breast cancer diagnosis [7]. ApoE gene reveals one's chances of getting Alzheimer [8]. Our current knowledge about genomic data is just the tip of the iceberg; as research progresses, we expect to know more and more.

Paternity test is used to determine if a male individual is the father of a child [9]. It is accepted by courts as evidence of paternity [10]. Paternity test along with online genealogy databases make anonymization of sperm donors impossible [11], [12], [13]. Genomic data can also reveal information about other blood-relations such as sibling, cousin, uncle, etc [14].

**Feature 3** (Stablility over lifetime)**:** *DNA is very stable over the lifetime of human beings.*

**Feature 4** (Negligible intergenerational change)**:** *DNA changes very little from one generation to the next. Only* ***seventy*** *(out of six billion) nucleotides change per diploid human genome per generation [15].*

Human genome changes at a surprisingly slow rate over generations. A study done in 2000 reported the change to be only 175 nucleotides per diploid human genome per generation (i.e. from parent to child) [17]. A recent study showed that the change is only 70 nucleotides per diploid human genome per generation [15]. This is an extremely small change compared to the size of the human genome, i.e., six billion nucleotides. Essentially, this means that human genome is very stable over generations and contains large amount of information about past and future generations of a person.

Henrietta Lacks died in 1951 of cancer and at that time her cells were extracted from her body without permission. These cells led to many developments in medicine [16]. Recently, researchers published her genome online violating the privacy of Lacks' family. Grandchildren of Henrietta Lacks were concerned about their privacy. NIH acknowledged their right and now two out of six members of NIH committee that decide on usage of Henrietta Lacks genome are from her family. The story of Henrietta Lacks shows that genomic data leak can impact future generations.

**Feature 5** (Similarity with blood-relatives)**:** *Genome of a single human contains a lot of sensitive information about his blood-relatives.*

Closely related individuals have very similar genomes. Research has shown that to genotype a family, we only need to genotype few people in the family and use computational tools to infer other family members' genomes [18]. Genotyping is a technique used to obtain SNPs (Single Nucleotide Polymorphisms) in a human genome. Usually, one million SNPs are genotyped. It was found that 83%[2] of these SNPs can be inferred from those of few family members in practical experiments, while, in theory 97% can be inferred [18].

**Feature 6** (Increasing value with time)**:** *We will know more about our genomes as time passes; currently it is hard to estimate how much information is there in the human genome.*

[2]This is considering only SNPs, not the entire genome as in Feature 4 (Negligible intergenerational change).


**Fig. 1:** What's in genome?

First human genome was sequenced in 2003. Since then, there has been a lot of progress on finding genetic basis of disease and other phenotypes. Figure 1 shows examples of different kind of information present in genomic data. As research progresses, we will learn more and more about the genome. As an example, it was believed that 98% of the genome (corresponding to non-coding region) was junk, but now we know that 80% of human genome serves some biochemical purpose [19].

**Feature 7** (Diverse usage)**:** *Genomic data has very diverse applications including (but not limited to) healthcare, biomedical research, law enforcement and forensics, and direct-to-consumer (DTC) genomics (or recreational genomics).*

**Feature 8** (Very high dimensionality)**:** *Genomic data is inherently very high dimensional.*

Humans share 99.5% of their genome. Nonetheless, each of us has 4 million variations called SNPs (Single Nucleotide Polymorphisms)[3] [20]. Considering only SNPs, each individual genome has 4-million dimensions, which makes genomic data very high dimensional.

**Feature 9** (Less can be more)**:** *Completion attacks based on LD (linkage disequilibrium) can be used to infer unpublished genomic data from partial genomic data.*

Partial genomic data can be used to infer unavailable genomic data, which means that even data that is not directly related to some trait e.g., a disease, can leak information about the disease. James Watson (co-discoverer of DNA) donated his genome without publishing his ApoE gene, that shows increased susceptibility to Alzheimer's disease, but researchers have shown how to find out Watson's ApoE status without having corresponding genomic data [21].

**Feature 10** (Very large size)**:** *The size of human genomic data is very large.*

Human genome is made up of six billion nucleotides. Most of our genome is common with other people, but the part of

[3]There are many other types of variations, e.g., copy number variation.

genome that differs from person to person is also large (4 million nucleotides).

## III. Challenges

This section describes how the features of genomic data described in Section II make genomic data usage management challenging for both policy and technology solutions.

### A. Policy challenges

Feature 1 (Cyber-physical nature) makes complete genomic privacy at least partially a policy question. As genome sequencing agencies need to access biospecimen to sequence the genome and there is no way to encrypt chemicals; some type of legislation is required to prevent the abuse of genome (in its chemical form). Moreover, leftover DNA (e.g., from hairs and saliva) is considered abandoned property under US law and can be used for arbitrary purposes. This might not be a scalable approach but can be used for targeted attacks. A famous singer, Madonna hires a DNA-sterilizing team to clean all leftover DNA after she leaves hotel to prevent abuse [22].

Feature 3 (Stability over lifetime), Feature 4 (Negligible intergenerational change) and Feature 5 (Similarity with blood-relatives) complicate the policy issues related to genomic data as these features are unique to genomic data. Usually, data belongs to a single person and he/she can share it as he/she wishes. Genomic data is very different in this regard. It reveals a lot about one's relatives and about all of his/her past and future generations. As people are considered autonomous and they can do whatever they want with their data, it is not clear whether one should be allowed to share his/her genomic data without the consent of his/her relatives. Getting consent from relatives is very complicated, as it leaks information about very large number of relatives and getting consent from all of them might not be possible. We believe this is a good topic for policy makers to debate. Today, genomic data has no special laws when it comes to publishing it online. Many people have posted their genomes at openSNP.org.

Feature 6 (Increasing value with time) is a big challenge to create meaningful policies for protecting genomic privacy. Before defining a policy, we should know the actual information content of genomic data. As research progresses, genomic data will reveal more information, the policies and laws created today might not be appropriate in future. This requires policies and laws to evolve as we know more about genomic data.

Feature 7 (Diverse Usage) further hardens to police genomic data usage management. We need to answer questions such as, should genomic data collected by a hospital or DTC company be available to law enforcement for investigation?, should patient data be used for research?, how to regulate DTC businesses?, etc.

Feature 9 (Less can be more) requires some policy and legislation in place for appropriate use of genomic data and to guarantee that people and companies who abuse partial data would be held accountable.

### B. Technological challenges

Section III-A discusses why genomic data usage management cannot be guaranteed by technology alone. Nonetheless, technology is necessary to provide meaningful privacy guarantees while keeping laws, legislation and policy requirements to bare minimum.

Cryptography is one of the fundamental tool that provides confidentiality and integrity. History of cryptographic algorithms prove that they are secure for a limited period of time before they are broken. Since, symmetric cryptographic algorithms such as DES, AES or any efficient symmetric algorithm are not provably secure, lifetime of these algorithms is limited compared to the lifetime of genomic data. Public key algorithms are provably secure under some assumptions. For example, RSA is based on the assumption that factoring of very large numbers is computationally hard. We know that factoring can be solved in sub-exponential time using index-calculus [17], [23]. Recently, the asymptotic complexity of index-calculus algorithm was improved [24]. These index-calculus algorithms does not break RSA, but eventually we might be able to develop efficient algorithms for factoring.

Usually, data is considered confidential if it is encrypted in a manner that would cost the adversary more than the value of the data to decrypt or decryption by the adversary will take so much time that the data will be useless after a successful decryption. Both of these assumptions are not valid for genomic data, Feature 6 (Increasing value with time) shows that the value of genomic data will increase with time and Feature 3 (Stability over lifetime), Feature 4 (Negligible intergenerational change), and Feature 5 (Similarity with blood-relatives) show that lifetime of human genomic data is much larger than the lifetime of a typical cryptographic algorithm.

Genomic data is inherently high dimensional, as discussed in Feature 8 (Very high dimensionality). No statistical disclosure control methods (e.g., differential privacy or k-anonymity) exists to publish any useful genomic data. There are few papers published on differential privacy for genomic data [25], [26], but none of the approaches can publish large amount of aggregate data without destroying its utility. Publishing genomic data in privacy-preserving fashion is a very hard problem.

Completion attacks [27] can be used to infer unpublished part of the genome from partially published genome, as described in Feature 9 (Less can be more). We need to develop privacy-preserving techniques to publish partial genomic data such that unpublished data cannot be inferred from it.

## IV. Solutions: What can we do?

### A. On policy side

Genomic usage management and privacy requires appropriate legislation in place. Recently, presidential report on genomic privacy was published, which shows that US government is concerned about the privacy of genomic data [28]. The legislation should at least guarantee all aspects of data usage management and privacy that technology cannot offer, as discussed in Section III-A. There is already a law in place in US called Genetic Information Non-discrimination Act (GINA) that prevents employment and insurance discrimination based on genomic data. In the following, we describe how legislation would help provide a meaningful genomic data usage management strategy:

- ⋆ As discussed in Feature 1 (Cyber-physical nature), genomic data has cyber-physical nature. There should be laws that regulates the sequencing agencies and biobanks, to govern the usage and prevent abuse of genomic data. For example, laws can require the sequencing agencies to use the chemical DNA only once and destroy it after sequencing has been done.
- ⋆ Sequencing machines are highly sophisticated machines manufactured by big companies. It is easy to regulate sequencing machines to use trusted hardware such that the output of sequencing machines is encrypted. There should be legislation in place for tampering with the sequencing machines and trusted hardware in it. FDA regulates all medical equipment, so regulating sequencing machines should not be hard. Health privacy policies such as HIPAA show that government is concerned about the privacy of public and is willing to legislate when necessary.
- ⋆ Feature 3 (Stability over lifetime), Feature 4 (Negligible intergenerational change), Feature 5 (Similarity with blood-relatives), and Feature 10 (Very large size) show that the state of the art cryptographic tools might not be suitable for genomic data. However, with appropriate legislation in place, we can use existing cryptographic tools. For example, it can be regulated through laws that the genomic data of a person should be "securely deleted" when he/she dies or after some predefined life of encrypted genome.
- ⋆ Feature 8 (Very high dimensionality) creates the issue of inference control. It's really hard to publish any type of useful genomic data in anonymized form e.g., aggregate genomic data. However, we can create some data use agreement policies to prevent inference control as done by dbGap to provide access to the aggregate GWAS data.
- ⋆ People should be prohibited from posting their genome online as they do on openSNP.org, violating not only their's but their relatives' privacy as well [29].
- ⋆ A very important part of any policy or legislation would be to provide security against any leakage of genomic data, at least when it's not the victim's fault. People whose genomic data is made public due to their participation in medical studies, projects like 1000 genome project or any kind of data breach, should be protected by law. Recently, there was conflict between NIH and Lacks family after Henrietta Lacks genome was published online. NIH and Lacks family resolved the conflict by forming a six member committee (two of them being Henrietta Lacks' family members) responsible to handle usage of Henrietta Lacks' genome.

### B. *On technology side*

With appropriate legislation in place, current technology would be in much better position to manage genomic data usage and privacy. Many technical solutions attempt to solve genomic privacy issues [30], [31], [32], [33], [34], [32]. In the following we describe, how technology can be used to manage genomic data usage and privacy:

- ⋆ Today sequencing machines output genomic sequences in plaintext. This makes genomic data very vulnerable to abuse by insiders and potential hackers. As discussed in the previous section, sequencing machines can have trusted hardware e.g., Trusted Platform Module (TPM). This trusted hardware in conjunction with a key management authority can be used to output encrypted data instead of plaintext data. Trusted hardware can also be used for software attestation to verify that correct software is running on the machine.
- ⋆ Usually, encrypted data is considered useless if the corresponding decryption key is destroyed. Due to Feature 3 (Stability over lifetime) and Feature 4 (Negligible intergenerational change), genomic data have very long lifetime, so even encrypted data should be considered sensitive and should be handled with care. Stored genomic data should be reencrypted with state of the art ciphers available, and previous version of the encrypted data should be securely deleted as explained below.
- ⋆ Secure data deletion [35] should be used to delete data after time specified by a policy or law. However, standard secure data deletion techniques might not be suitable for genomic data because even encrypted data is sensitive and just destroying the deletion key is not enough.
- ⋆ Genomic tests should be performed in a privacy-preserving manner. There is some recent work on privacy-preserving genomic computation [30], [31], [32], [33], [34], [32]. These protocols have some drawbacks such as being computationally intensive, leaking more than necessary and being unscalable; mainly due to the very large size of genomic data as explained in Feature 10 (Very large size). Highly scalable cryptographic tools are required to handle data at the human genome scale.
- ⋆ Most importantly we can learn from other institutions such as Federal Census Bureau that has been publishing anonymized population data using hierarchical anonymization. It has been publishing this data for decades and there is no reported privacy breach from such data. Same techniques may not be directly applicable to genomic data, but we can learn from those techniques and try to develop similar techniques for genomic data.

## V. Recommendations for Genomic Data Usage Management

In this section, we present some data usage recommendations for genomic data, keeping in mind the features of genomic data. We use the term *data* to mean *genomic data* and the term *data owner* to mean the person whose cells are used to sequence the genome. We assume that genomic data is stored by an external party such as hospital, because we believe that the data owners should not store genomic data on their personal devices.

- ⋆ Do not use the data to identify the data owner.
- ⋆ Do not use the data to infer unpublished genomic data or any other information either about the data owner, his/her relatives or any other individual.
- ⋆ Do not share the data or any information computed or inferred from it with anyone without explicit permission from the data owner.
- ⋆ In case, the data storage is outsourced, appropriate technological measures and legal agreements should be signed with the cloud provider to prevent any abuse.
- ⋆ Data should be stored in the same legal jurisdiction as that of the data owner, e.g., data of a United States' citizen should not be stored outside of United States.
- ⋆ **Right NOT to know:** Do not report incidental findings if the data owner have opted for not knowing them.

  

- ⋆ Delete data if the encryption scheme used to encrypt the data is no longer secure and if required obtain the new copy of the data encrypted with a secure cipher.
- ⋆ Always store data in encrypted fashion on persistent storage.
- ⋆ If it is necessary to send data over a network, secure communication should be used.
- ⋆ When required, decrypt the data only in Random Access Memory (RAM) and delete it as soon as the job is done.
- ⋆ When deleting the data from persistent storage, encrypted data as well as corresponding decryption keys both should be securely deleted, such that neither can be recovered.
- ⋆ Data should be securely deleted on the legally valid request of the data owner such that it cannot be recovered.

## VI. Conclusion

Managing usage of genomic data is very challenging problem. It requires novel policy and technology solutions. State of the art technology is not ready to handle genomic data due to inherent features of genomic data. However, with proper policies and appropriate technological solutions in place, we can create a system that can manage the usage of genomic data. Nonetheless, a lot of research is required to improve the data usage management and privacy of genomic data.

## Acknowledgment

The author would like to thank anonymous reviewers for their useful comments. We also thank Carl A. Gunter, Vincent C. Bindschadler, Fahad Ullah, and Syed Abbas Bukhari for their useful suggestions.

This work was supported by NSF CNS 13-30491 (ThaW) and HHS 90TR0003-01 (SHARPS). The views expressed are those of the author only.

## References

[1] "Genome-wide association studies." [Online]. Available: http://www.genome.gov/20019523

[2] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig, "Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays," *PLoS Genetics*, vol. 4, no. 8, p. e1000167, 2008.

[3] R. Wang, Y. F. Li, X. Wang, H. Tang, and X. Zhou, "Learning your identity and disease from research papers: information leaks in genome wide association study," in *CCS*, 2009, pp. 534–544.

[4] "A catalog of published genome-wide association studies." [Online]. Available: http://www.genome.gov/gwastudies/

[5] "Cystic fibrosis." [Online]. Available: http://ghr.nlm.nih.gov/condition/cystic-fibrosis

[6] "What is sickle cell anemia?" [Online]. Available: http://www.nhlbi.nih.gov/health/health-topics/topics/sca/

[7] "Breast cancer." [Online]. Available: http://ghr.nlm.nih.gov/condition/breast-cancer

[8] G. Utermann, U. Langenbeck, U. Beisiegel, and W. Weber, "Genetics of the apolipoprotein e-system in man," *American journal of human genetics*, vol. 32, no. 3, p. 339, 1980.

[9] "Paternity testing." [Online]. Available: http://americanpregnancy.org/prenataltesting/paternitytesting.html

[10] "Legal dna paternity testing." [Online]. Available: http://www.genetica.com/GeneticaWebV2.nsf/XPaternityDNATesting-legal.xsp

[11] A. Motluk, "Anonymous sperm donor traced on internet," *New Scientist*, 2005. [Online]. Available: http://goo.gl/3FOpfc

[12] R. Stein, "Found on the web, with DNA: a boy's father," *The Washington Post*, 2005. [Online]. Available: http://goo.gl/q6X16E

[13] G. Naik, "Family secrets: An adopted man's 26-year quest for his father," *The Wall Street Journal*, 2009. [Online]. Available: http://goo.gl/z9xGBc

[14] A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W.-M. Chen, "Robust relationship inference in genome-wide association studies," *Bioinformatics*, vol. 26, no. 22, pp. 2867–2873, 2010.

[15] J. C. Roach, G. Glusman, A. F. Smit, C. D. Huff, R. Hubley, P. T. Shannon, L. Rowen, K. P. Pant, N. Goodman, M. Bamshad *et al.*, "Analysis of genetic inheritance in a family quartet by whole-genome sequencing," *Science*, vol. 328, no. 5978, pp. 636–639, 2010.

[16] "Cancer cells killed henrietta lacks - then made her immortal." [Online]. Available: http://hamptonroads.com/2010/05/cancer-cells-killed-her-then-they-made-her-immortal

[17] M. W. Nachman and S. L. Crowell, "Estimate of the mutation rate per nucleotide in humans," *Genetics*, vol. 156, no. 1, pp. 297–304, 2000.

[18] J. T. Burdick, W.-M. Chen, G. R. Abecasis, and V. G. Cheung, "In silico method for inferring genotypes in pedigrees," *Nature genetics*, vol. 38, no. 9, pp. 1002–1004, 2006.

[19] E. Pennisi, "Encode project writes eulogy for junk dna," *Science*, vol. 337, no. 6099, pp. 1159–1161, 2012. [Online]. Available: http://www.sciencemag.org/content/337/6099/1159.short

[20] "Whole genome sequencing: How many snps remain?" [Online]. Available: http://massgenomics.org/2009/06/whole-genome-sequencing-how-many-snps-remain.html

[21] D. R. Nyholt, C.-E. Yu, and P. M. Visscher, "On jim watson's apoe status: genetic information is hard to hide," *European Journal of Human Genetics*, vol. 17, no. 2, pp. 147–149, 2008.

[22] B. R. Villalva, "Madonna sterilization, star hires DNA team on tour," in *The Christian Post*, 2012. [Online]. Available: http://goo.gl/yj9p4v

[23] A. Joux, "Faster index calculus for the medium prime case application to 1175-bit and 1425-bit finite fields," in *Advances in Cryptology–EUROCRYPT 2013*. Springer, 2013, pp. 177–193.

[24] ——, "A new index calculus algorithm with complexity l (1/4+ o (1)) in very small characteristic." *IACR Cryptology ePrint Archive*, vol. 2013, p. 95, 2013.

[25] S. Fienberg, A. Slavkovic, and C. Uhler, "Privacy preserving GWAS data sharing," in *ICDM*, 2011, pp. 628–635.

[26] A. Johnson and V. Shmatikov, "Privacy-preserving data exploration in genome-wide association studies," in *KDD*, 2013, pp. 1079–1087.

[27] Y. Erlich and A. Narayanan, "Routes for breaching and protecting genetic privacy," vol. abs/1310.3197v1, 2013.

[28] P. C. for the Study of Bioethical Issues, "Privacy and progress in whole genome sequencing," 2012.

[29] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti, "Addressing the concerns of the lacks family: Quantification of kin genomic privacy," in *CCS*, 2013.

[30] J. R. Troncoso-Pastoriza, S. Katzenbeisser, and M. Celik, "Privacy preserving error resilient DNA searching through oblivious automata," in *CCS*, 2007, pp. 519–528.

[31] F. Bruekers, S. Katzenbeisser, K. Kursawe, and P. Tuyls, "Privacy-preserving matching of DNA profiles," *IACR Cryptology ePrint Archive*, vol. 2008, p. 203, 2008.

[32] E. Ayday, J. L. Raisaro, U. Hengartner, A. Molyneaux, and J.-P. Hubaux, "Privacy-preserving processing of raw genomic data," in *DPM*, 2013.

[33] E. Ayday, J. L. Raisaro, J. Rougemont, and J.-P. Hubaux, "Protecting and evaluating genomic privacy in medical tests and personalized medicine," in *WPES 2013*, 2013.

[34] P. Baldi, R. Baronio, E. De Cristofaro, P. Gasti, and G. Tsudik, "Countering gattaca: efficient and secure testing of fully-sequenced human genomes," in *CCS*, 2011, pp. 691–702.

[35] J. Reardon, H. Ritzdorf, D. Basin, and S. Capkun, "Secure data deletion from persistent media," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 271–284.

  