City, CA, USA. [5]*Reneo Capital Management LP, New York, NY, USA.* [6]*GSWC Group, Rancho Santa Fe, CA, USA.* [7]*ChromaCode, Carlsbad, CA, USA.* [8]*Genetic Alliance, Washington, DC, USA.* [9]*Illumina, San Diego, CA, USA.* [10]*Vascular Cures, Redwood City, CA, USA.* [11]*Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY, USA.*
**e-mail: bob@lunadna.com; dawn@lunadna.com*

Published online: 19 September 2019
https://doi.org/10.1038/s41587-019-0278-9

References
1. Yuan, J. et al. *Nat. Genet.* **50**, 160–165 (2018).
2. Nelson, S. C., Bowen, D. J. & Fullerton, S. M. *Am. J. Hum. Genet.* **105**, 122–131 (2019).
3. Zook, M. et al. *PLOS Comput. Biol.* **13**, e1005399 (2017).
4. Gibney, E. *Nature* https://doi.org/10.1038/d41586-018-03880-4 (2018).
5. Erlich, Y., Shor, T., Peèr, I. & Carmi, S. *Science* **362**, 690–694 (2018).
6. Ducharme, J. *Time* (26 July 2018).
7. Terry, S. F. *Sci. Transl. Med.* **9**, eaaf1001 (2017).
8. Roberts, J. L., Pereira, S. & McGuire, A. L. *Nat. Biotechnol.* **35**, 18–20 (2017).
9. Shabani, M. J. Am *Med. Inform. Assoc.* **26**, 76–80 (2019).
10. Largent, E. A. & Fernandez Lynch, H. *Yale J. Health Policy Law Ethics* **17**, 61–141 (2017).
11. Skiba, K. Older Americans targeted in DNA testing scams. *AARP* https://www.aarp.org/money/scams-fraud/info-2019/dna-testing-scam.html (2019).
12. Visscher, P. M. et al. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
13. Kish, L. J. & Topol, E. J. *Nat. Biotechnol.* **33**, 921–924 (2015).

Acknowledgements
We thank the LunaDNA member community for their early adoption of the platform. We are grateful for the participation of all of the LunaPBC advisors in this work.

Competing interests
All authors have a financial interest in LunaPBC, Inc., the management company of the LunaDNA, LLC platform, as employees, advisors or shareholders. Y.E. is also an employee of MyHeritage, a consumer genomics company.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0278-9.

Corrected: Author Correction

Data privacy in the age of personal genomics

To the Editor — The growing adoption of personal genomics has attracted attention to various issues concerning genomic data privacy. For example, some personal genomics companies sell data to pharma companies, a practice that has been found to lack transparency[1,2]. Rapidly growing genomic databases also have attracted the interest of law enforcement and helped solve criminal cases[3]. This has drawn criticism due to concerns over government access to genomic data of individuals who have not committed any crimes. To many people, the risk of discrimination is the most concerning issue. In the United States, the Genetic Information Nondiscrimination Act protects individuals from discrimination by employers and health insurance companies. However, it does not apply to life insurance and disability insurance, nor does it protect from discrimination in other areas, such as education and housing. In the future, additional, potentially concerning, uses for genomic data may be developed. For example, personal genomic data might become valuable for targeted advertising.

These concerns might be justified, as risks of privacy infringement and discrimination have already become reality in some parts of the world[4]. Furthermore, privacy concerns must be addressed because they increasingly deter people from genetic testing and data sharing with researchers[5]. Here we propose a privacy-focused model for direct-to-consumer (DTC) personal genomics and outline multiple complementary approaches that can be used to secure genomic data (Table 1).

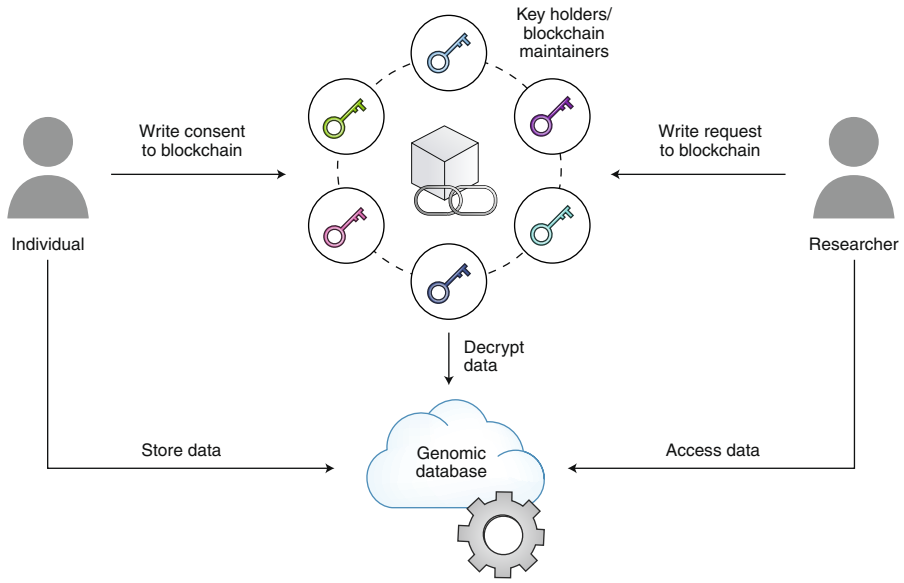

Fig. 1 | Multiparty data access control and record keeping on a blockchain. Access to genomic data is controlled by multiple independent parties that hold shares of a split encryption key. Additionally, these parties maintain a blockchain that immutably and transparently stores data access requests and users' consent.

User anonymity

Personal genomics companies may enable their customers to purchase genetic testing while remaining anonymous. To this end, cryptocurrencies, such as Bitcoin, can enable pseudo-anonymous transactions that rely on blockchain addresses, which have no connection to real-life identities, such as name, address or bank account[6]. In the past, this property has made cryptocurrencies attractive for black-market transactions and given them a questionable reputation. However, pseudo-anonymous payments can also facilitate the purchase of legal but potentially sensitive products and services, such as genetic testing.

Enabling pseudo-anonymous payments can eliminate several potential vulnerabilities individuals are exposed to when purchasing genetic testing services. Most importantly, enabling individuals to remain pseudo-anonymous would eliminate the dependence on data deidentification by personal genomics companies before data sharing. Additionally, because customer data such as name and credit card information do not need to be collected, individuals using such systems are at a reduced risk of being affected by security breaches.

However, as blockchain addresses are associated with real identities, transactions with cryptocurrencies are not fully anonymous. Furthermore, because

Table 1 | Approaches to protecting genomic data privacy

Feature	Possible implementation
User anonymity	Anonymous genetic testing enabled by payments with cryptocurrencies
Data access control	Multiparty access control to genomic data through the splitting of encryption keys
Record auditability	Immutable storage of data access requests and users' consent on a blockchain
Secure data analysis	Data analysis in controlled computing environments supplemented by privacy-preserving technologies

verification of blockchain transactions requires them to be publicly readable, user privacy can be compromised if the identity of an individual behind a blockchain address becomes known. Some cryptocurrencies aim to address this vulnerability. For example, one approach is to encrypt transactions, which are publicly stored on a blockchain, and validate them using so-called zero-knowledge proofs[7]. This cryptographic technique enables verifying that a statement is true without revealing any information about the statement itself. This allows verification of blockchain transactions without revealing the addresses of senders or receivers.

Although purchases of genetic testing services can be effectively anonymized, genomic data cannot because they contain unique, inheritable genetic markers. This information can be used for long-range familial searches that can identify anonymous subjects by linking them to distant relatives whose identity is known[8]. Thus, because genomic data anonymization alone is insufficient to protect privacy, genomic data sharing must occur in a controllable, transparent and privacy-preserving manner.

Data access control

Individuals should have full control over their personal genomic data. However, today DTC genomics companies effectively own and control all genomic data that they produce[1]. This introduces several risks. First, centralized genomic databases could be breached by hackers, as has already happened in the past[9]. Second, access to genomic databases can be enforced by government agencies[10]. And third, because there are no checks in place, personal genomics companies may deliberately or inadvertently infringe data privacy.

This issue can in principle be addressed by enabling individuals to manage the encryption keys for their personal genomic data. This approach has already been adopted for other products, such as e-mail services where the provider does not have the ability to decrypt and access user e-mails[11]. However, there are several drawbacks. First, it is not possible to recover encryption keys if there are no backup copies. If a user loses the key for his personal genomic data, they will become permanently inaccessible. Second, this approach might hinder data sharing with investigators because users would have to manually approve every data access request. In particular, this might make it practically infeasible to access large genomic datasets.

Delegated access control that relies not on a single party but multiple independent organizations can be a reasonable compromise between security and usability. To this end, keys that are used to encrypt genomic data can be split into shares that are distributed to multiple, independent parties; for example, research institutions (Fig. 1). The key splitting scheme can also incorporate some redundancy by making a subset of key shares sufficient to reconstruct the encryption key and decrypt the data[12]. Multiparty access control would provide better protection against breaches and misuse because it distributes data access control and thus does not rely on any single trusted party. Furthermore, if the organizations that received encryption key shares are located in different jurisdictions or are anonymous, this approach would also prevent governments from obtaining access to genomic databases without also obtaining the consent of the individuals to whom the data belong.

Record auditability

To establish trust and incentivize genomic data sharing, data access requests and users' consent must be communicated transparently and maintained immutably, which would ensure auditability and deter misuse. This can potentially be implemented using a blockchain—an immutable, public database that is maintained by a peer-to-peer network[6]. A network participant can propose to add a new entry to the blockchain by broadcasting a transaction to other participants in the network. The network accepts a new transaction only if it has been validated by a majority of participants. Transactions are bundled into timestamped blocks and each block references its preceding block, which creates a sequential ordering that prevents the deletion of data stored on a blockchain.

The blockchain can be maintained by the same network of organizations that hold encryption key shares and collectively control data access (Fig. 1). Thus, blockchains can supplement multiparty access control by enabling tamper-proof, auditable record keeping. Investigators who wish to access genomic data can write data access requests to the blockchain that include the investigator's identity, affiliation and study description. In turn, individuals can write their consent to share their genomic data to the blockchain. Holders of encryption key shares can read out these access permissions from the blockchain, collectively decrypt data and provide access to the authorized investigator.

There are several examples of utilization of blockchain technology for auditable record keeping. For instance, Google subsidiary DeepMind developed a blockchain-like database for a tamper-proof recording of computations on clinical data from the United Kingdom's National Health Service hospitals[13]. This ensures that consent for any data has been obtained from the patients. Another example is Estonia, where blockchain technology is being used to track when and how the health records of 1.3 million residents are being accessed[14]. Estonian citizens are able to log into their electronic profiles to express consent to different uses of their health data.

Secure data analysis

Although splitting of encryption keys and consent management on a blockchain can enable controlled and auditable data sharing, these technologies cannot protect shared genomic data against deliberate misuse. Privacy of shared genomic data can be preserved, however, by creating secure computing environments within which the data are analyzed. The idea of 'bringing algorithms to the data' rather than transferring data to external systems has already been adopted by several projects. For example, Blockstack is a company building a general-purpose, decentralized computing network that enables users to provide their own computing and storage resources to bring apps to wherever their data are located[15]. In genomics, the Global Alliance for Genomics and Health (GA4GH) Beacon Project has embraced a similar concept[16]. The Beacon Network is a federated

ecosystem that consists of connected genomic databases that are owned by different organizations. Investigators can submit queries—for example, for the presence of specific genetic variants—and those queries are then executed on the decentrally stored data. The results are sent back to the investigator.

By bringing computations to the genomic data, potentially sensitive information can be protected from disclosure to investigators conducting a study. However, when data are decrypted for analysis, privacy could also be infringed by providers of data storage and computing services. Fully homomorphic encryption and secure multiparty computations are privacy-preserving technologies that can help address this challenge. These technologies make it possible to encrypt data such that they can be analyzed as if they were in plaintext, yet remain encrypted and thus protected during analysis. Although the adoption of privacy-preserving technologies has been hindered by insufficient performance, recent advances enable increasingly practical execution times and scalability. For example, one recent study has demonstrated scalable, privacy-preserving genomic data exploration enabled by a combination of multiple privacy-preserving technologies[17]. Another study has presented a secure multiple-party computation protocol for genome-wide association studies (GWAS) with a computational complexity that scales linearly, rather than quadratically, with the number of genomes[18].Toward privacy-focused personal genomics

Above, we propose multiple mechanisms that can be adopted by DTC genomics companies to enhance the protection of personal genomic data. However, these mechanisms also constitute self-imposed restrictions that go against the business model of extracting maximum value from generated genomic data. Whether a privacy-focused personal genomics company can be successful will depend on whether consumers will reward a focus on data privacy protection.

The general trend is that consumers are increasingly paying more attention to how businesses handle their personal data, compelling even large tech companies to adopt stricter privacy policies and more sophisticated data protection mechanisms. For instance, after widely publicized misuses, Facebook was subject to significant scrutiny from both the public and regulators. As a result, the company has announced a pivot toward becoming a privacy-focused social network and implementing end-to-end encryption of all communication[19]. Apple is moving in a similar direction with its commitment to build hardware and software that protect customer privacy, rather than collecting and monetizing user data[20]. At the same time, Google's search engine business is experiencing growing competition from privacy-focused alternatives[21].

In contrast to large, profitable tech companies, most DTC personal genomics companies might not be able to adopt more privacy-focused business models. High custom acquisition costs and low margins have forced them to leverage data monetization as an additional revenue stream. It has become a crutch that is used to justify unsustainable unit economics. Development of more attractive genetic testing products that generate more revenue at lower customer acquisition costs will be required to alleviate the dependence on data monetization. As multiple personal genomics startups are exploring privacy-focused business models, the market will soon determine the value of genomic data privacy. ❐

Dennis Grishin[1,2*], Kamal Obbad[2] and George M. Church[1,2]

[1]*Harvard Medical School, Boston, MA, USA.*
[2]*Nebula Genomics, Inc., San Francisco, CA, USA.*
**e-mail: dgrishin@g.harvard.edu*

Published online: 19 September 2019
https://doi.org/10.1038/s41587-019-0271-3

References

1. Laestadius, L. I., Rich, J. R. & Auer, P. L. *Genet. Med.* **19**, 513–520 (2017).
2. Ducharme, J. A major drug company now has access to 23andMe's genetic data. Should you be concerned? *Time* (26 July 2018).
3. Kolata, G. & Murphy, H. The Golden State Killer is tracked through a thicket of DNA, and experts shudder. *The New York Times* (27 April 2018).
4. Wee S. L. China uses DNA to track its people, with the help of American expertise. *The New York Times* (21 February 2019).
5. Winkler, R. 23andMe's growth slows. *WSJ Online* (8 February 2019).
6. Nakamoto, S. Bitcoin: a peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf (2009).
7. Goldwasser, S., Micali, S. & Rackoff, C. The knowledge complexity of interactive proof-systems. in *Proc. Seventeenth Annual ACM Symposium on Theory of Computing* 291–304 (ACM, 1985).
8. Erlich, Y., Shor, T., Pe'er, I. & Carmi, S. *Science* **362**, 690–694 (2018).
9. Shaban, H. DNA testing service MyHeritage says 92 million customer email addresses were exposed. *Washington Post* (5 June 2018).
10. Haag, M. FamilyTreeDNA admits to sharing genetic data with F.B.I. *The New York Times* (4 February 2019).
11. Sawers, P. How ProtonMail is pushing email privacy standards. *VentureBeat* https://venturebeat.com/2018/05/13/how-protonmail-is-pushing-email-privacy-standards/ (2018).
12. Shamir, A. *Commun. ACM* **22**, 612–613 (1979).
13. Armstrong, S. *BMJ* **361**, k1996 (2018).
14. Heston, T. F. *Int. J. Curr. Res.* **9**, 60587–60588 (2017).
15. Ali, M., Nelson, J., Blankstein, A., Shea, R. & Freedman, M. J. Blockstack technical whitepaper v 2.0. https://blockstack.org/whitepaper.pdf (2019).
16. Global Alliance for Genomics and Health. *Science* **352**, 1278–1280 (2016).
17. Raisaro, J. L. et al. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **16**, 1328–1341 (2018).
18. Cho, H., Wu, D. J. & Berger, B. *Nat. Biotechnol.* **36**, 547–551 (2018).
19. Zuckerberg, M. A privacy-focused vision for social networking. *Facebook* https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/ (2019; accessed 27 June 2019).
20. Etherington, D. Apple is now the privacy-as-a-service company. *TechCrunch* (3 June 2019).
21. Evangelho, J. Why you should ditch Google search and use DuckDuckGo. *Forbes Magazine* (3 October 2018).

Acknowledgements

The authors thank J. Lunshof for critical discussion and feedback on the manuscript.

Competing interests

D.G. and K.O. are employees of Nebula Genomics, Inc., and G.M.C. is a cofounder.