Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020

Beyond Data Markets: Opportunities and Challenges for Distributed Ledger Technology in Genomics

Scott Thiebes
Karlsruhe Institute of Technology
scott.thiebes@kit.edu

Niclas Kannengießer
Karlsruhe Institute of Technology
niclas.kannengiesser@kit.edu

Manuel Schmidt-Kraepelin
Karlsruhe Institute of Technology
manuel.schmidt-kraepelin@kit.edu

Ali Sunyaev
Karlsruhe Institute of Technology
sunyaev@kit.edu

Abstract

During the past decade, distributed ledger technology (DLT) has found its way into application areas outside finance, such as supply chain management, the Internet of Things, or health care. To this end, this novel technology phenomenon has recently also caught the attention of researchers and practitioners in genomics. Although various DLT-based data markets for genome data already exist or are in development, the potential of DLT in this context is far from exhausted, whereas the possible risks related to the application of DLT in genomics are not yet sufficiently known. In this work, we investigate the potential opportunities and challenges for the application of DLT in the field of genomics. Thus, we make an important contribution to the safe and socially acceptable use of DLT in this unique and highly relevant use context.

1. Introduction

Distributed Ledger Technology (DLT) is one of the most hyped information technology innovations of the last decade that is probably best known for the cryptocurrency Bitcoin and its underlying DLT concept Blockchain [1]. Yet, the hype around DLT nowadays extends far beyond applications in the financial sector, with potential benefits of DLT being discussed in diverse industries and application scenarios, including supply chain management, the Internet of Things, and especially health care [2]. Often cited benefits of DLT pertain to enabling secure transactions between untrustworthy parties through automated, algorithm-based consensus building mechanisms, which eliminate the need for third-party trust enforcement [1], high availability of DLT-based systems [3], or the ability to automate and enforce processes by means of smart contracts [4].

Recently, DLT has also caught the interest of practitioners and researchers within the field of genomics [4-6]. Thereby, the diffusion of DLT in genomics is currently mainly driven by a small but rapidly increasing number of businesses such as Nebula Genomics, EncrypGen, or LunaDNA (see section 2.2 for a more detailed overview of DLT genomics businesses). In most cases, the objective of these businesses is to operate data markets, where users can share their genome data with third parties in exchange for tokens. Accordingly, DLT quickly gained the interest of the genomics community, as it promises to facilitate the exchange of genomic data and offers opportunities to reward data providers (e.g. through the use of tokens). However, to date, the majority of these initiatives is still in a pre-market phase. Furthermore, and despite the promising potential of DLT, applications of DLT have so far only been successfully implemented and operated in a limited number of cases. For example, only 8% of the DLT-related projects on GitHub are actively maintained [7].

Researchers within the field of genomics and associated disciplines have only recently begun to investigate the full potential of DLT for the genomics research community and other relevant stakeholders in genomics (e.g., genome data donors, medical professionals, pharma industry) [4, 6]. Due to the novelty of DLT for the genomics community, we still lack a profound understanding of what specific opportunities the application of DLT can bring forth for genomics, beyond the mere creation of genome data markets. Even more so, extant literature and, unsurprisingly, also white papers of involved businesses paint a rather positive picture of the prospective opportunities for DLT in genomics. To this end, the potential challenges that the application of DLT in genomics can bring forth are barely considered in the ongoing discussions about the utility of DLT for the genomics community. Overall, we as a community still lack knowledge on what are the most promising opportunities as well as the most demanding challenges

URI: https://hdl.handle.net/10125/64142
978-0-9981331-3-3

related to the application of DLT in genomics. We therefore ask the following research questions:

RQ1: What are opportunities and challenges for the application of DLT in genomics?

RQ2: What is the relative importance of the identified opportunities and challenges?

To answer our research questions, we conduct a ranking-type Delphi study with an international panel of experts on genomics and/or DLT and augment the results of the Delphi process with insights from a review of white papers and scientific publications on DLT in genomics. In doing so, the opportunities and challenges for the application of DLT in genomics that are described in this work present an important contribution to research and practice as they lay the foundation for directing adoption efforts towards the most promising opportunities as well as the most demanding challenges.

The remainder of this paper is structured as follows. In section two, we provide a brief introduction to DLT and an overview of the current state concerning the application of DLT in genomics. In section three, we detail on our research approach, including panel selection and the employed ranking-type Delphi method. Section four presents our results in terms of identified opportunities and challenges as well as their relative rankings, before we discuss our results in section five. We conclude the paper in section six.

2. Related research

2.1 Distributed Ledger Technology

DLT is an emerging technology that enables the operation of a distributed ledger, which is a special type of an append-only, distributed database that is particularly suited to the peculiarities of an untrustworthy environment [8, 9]. Inherent to DLT is the creation of a ledger that is replicated in a network of storage devices, which are referred to as nodes. Compared to traditional distributed databases, such structure allows for the presence of Byzantine failures. Byzantine failures include the presence of crashed or unreachable nodes, network delays, or malicious behavior of nodes [8]. In DLT, new data is added to the ledger using transactions that are committed on each node's replication [e.g., 10]. The data is then stored in a well-defined structure. Through the use of cryptographic techniques (e.g., hashing), data stored in the distributed ledger can hardly be removed or modified, resulting in near immutability. Each DLT design (e.g., Bitcoin [10]) employs a consensus mechanism to reach consistency between the replications stored on nodes of the distributed ledger. A consensus mechanism is an algorithm used to negotiate the valid state between nodes of the distributed ledger. Consensus mechanisms employ trust models, which consider threats and uncertainties in the process of consensus finding (e.g., Byzantine failures [8]). DLT can be employed to operate a distributed infrastructure across multiple parties, who can develop applications on DLT. Such applications on DLT make use of so-called smart contracts. Smart contracts are computer programs, which contain formalized business processes (e.g., conditional withdrawals) and are deployed and executed on a distributed ledger [e.g., 11]. Smart contracts can also retrieve data from the respective distributed ledger itself (on-chain) or from the external world (off-chain). Off-chain data feeds, which can be called by a smart contract to retrieve data are called oracles.

2.2. Distributed Ledger Technology in genomics

Genomics is the scientific discipline concerned with the sequencing, mapping, and analysis of genomes [12]. It is an integral part of contemporary (bio-)medical research and the ongoing shift towards a precision medicine treatment paradigm in health care [13]. Following the completion of the Human Genome Project, advances in genome sequencing, mapping, and analysis technologies have led to plummeting costs for the acquisition of genome data, from costs of around USD 95 million per whole genome sequence in 2001 to costs of around USD 1,300 per whole genome sequence today [14]. However, with the availability of relatively inexpensive means for acquiring and analyzing genome data on the one hand and the inherent characteristics of genome data (e.g., uniqueness, kinship, staticity) on the other hand [15], a variety of ethical, legal, and social challenges have emerged [16, 17]. Such challenges include, for example, trade-offs between individuals' right to privacy and the overall benefit of freely sharing genome data [16], interdependent privacy due to kinship [18], individuals' ability to comprehend and interpret the results of genetic testing [19], or the handling of inadvertent findings [20].

Employing DLT within genomics promises to address many of the aforementioned ethical, legal, social challenges [4, 6]. However, since the application of DLT in genomics is still in its infancy, DLT's diffusion within the community is currently to a large degree driven by a small but rapidly increasing number of young businesses. One of the pioneering and most prominent businesses in this area being Nebula Genomics. Like most players in the commercial DLT-

Table 1. Excerpt of DLT genomics businesses.

DLT Genomics Business	Objective
EncrypGen (https://encrypgen.com/)	DLT-based genome data market
Gene Blockchain (http://www.geneblockchain.org)	DLT-based genome data market
Genecoin (http://genecoin.me/)	Permanent storage of one's DNA using Bitcoin
Genomes.io (https://genomes.io/)	DLT-based genome data market
Nebula Genomics (https://nebula.org/)	DLT-based genome data market
Shivom (https://www.shivom.io/)	DLT-based genome data market
Zenome (https://zenome.io/about/)	DLT-based genome data market

in-genomics space, Nebula Genomics aims to offer direct-to-consumer genetic testing services, while at the same time creating a genome data market where the very same consumers can trade their genome data with interested third-parties (e.g., researchers, pharmaceutical companies) in exchange for tokens of Nebula Genomics' own crypto currency [21]. Similarly, EncrypGen, Gene Blockchain, Genomes.io, Shivom, and Zenome are also operating or planning to operate their own DLT-based genome data market together with their own cryptocurrency. A notable exception to the aforementioned businesses is Genecoin. Compared to the majority of businesses whose aim is to build DLT-based genome data markets, Genecoin aims at creating a permanent storage for individuals' DNA by storing their genome data in the Bitcoin network. Instead of giving out a Genecoin cryptocurrency, consumers are encouraged to create their own personal cryptocurrency off of their Bitcoin seed. Table 1 provides a non-exhaustive summary of the main players in the commercial DLT-in-genomics space.

Looking at scientific literature related to the application of DLT in genomics, we see that this stream of research is in a state of emergence (see Table 2). While an increasing number of research articles on DLT in health care is published, some of which also address the case of genomics [e.g., 22, 23, 24], we are aware of only four publications that explicitly focus on the peculiarities of bringing DLT into genomics. The viewpoint articles of Ozercan, et al. [6] and Shabani [4] and Thiebes, et al. [25] all provide a general discussion of potential that DLT holds for the genomics community. In addition, Ozercan, et al. [6] also describe the prototype of a DLT-based research infrastructure for genomics, called Coinami. In this infrastructure, research institutes may place genomics-related research jobs (e.g., genome sequence alignment) on the distributed ledger, which are then executed by nodes (so-called miners) that are rewarded with tokens via a coinbase transaction in return. Lee, et al. [5] on the other hand develop a DLT-based lossless compression platform for genome data. Here DLT is used as a means for handling data exchange requests. Interestingly, although a variety of different DLT-concepts exists (e.g., Blockchain, TDAG, BlockDAG), current discourses in research and practice on the application of DLT in genomics center around Blockchain only.

3. Methods

In order to identify salient opportunities and challenges related to the application of DLT in genomics, as well as their relative importance, we conduct a ranking-type Delphi study. We provide details on the Delphi panel selection and each of the three Delphi phases below.

3.1. Panel selection

Delphi studies typically rely in the input of experts with profound knowledge on the phenomenon under consideration (here DLT in genomics) [26]. Towards this end, we identified three groups of experts that could aid in the identification of challenges and opportunities for the application of DLT in genomics: (1) experts on DLT with at least basic knowledge on genomics; (2) experts on genomics with at least basic knowledge on DLT; and (3) individuals with high expertise in both, DLT and genomics.

Literature provides thresholds of 7 to 30 knowledgeable experts that are necessary to generate meaningful insights using Delphi [27]. To recruit an adequate number of experts from the three relevant groups, we employed a purposeful sampling strategy [28] and used different techniques to contact potential participants. First, we contacted the founders and employees (whom we assumed to have the required expertise) of the several DLT-in-genomics businesses

Table 2. Literature on DLT in genomics.

Article	Type	Research Focus
Lee, et al. [5]	Original article	Development of DLT-based lossless compression platform for genome data
Mackey, et al. [22]	Viewpoint	Challenges and opportunities for DLT in health care (genome data discussed as an example)
Roman-Belmonte, et al. [23]	Viewpoint	Discussion of the transformatory potential of DLT for health care (genome data discussed as an example)
Ozercan, et al. [6]	Viewpoint	Discussion of the transformatory potential of DLT for genomics / proposal of a DLS for genomics
Shabani [4]	Viewpoint	Discussion of the transformatory potential of DLT for genomics
Talukder, et al. [24]	Original article	Development of a consensus protocol for DLT-based electronic health records (genome data discussed as an example)
Thiebes, et al.	Viewpoint	Discussion of the transformatory potential of DLT for genomics

(see Table 1) via email, contact forms on their websites, and social networks. We also reached out to authors of white papers on this topic (many of which overlapped with the founders and employees of the previously named startups), researchers who had previously published articles on DLT genomics (see Table 2), and industry and international research consortia (i.e. The Global Alliance for Genomics and Health, and FORCE11). Lastly, we used personal contacts to researchers of whom we knew had some relevant experience and snowballing to recruit additional experts. Overall, this sampling strategy resulted in 12 actual participants for the brainstorming phase. Table A1 in the appendix provides and overview of relevant demographics for these participants.

3.2. Data collection and analysis methods

We conducted the entire study over the Internet by means of an online survey software for the questionnaire rounds and email to communicate with our participants after the brainstorming phase.

3.2.1. Brainstorming phase. At the beginning of the brainstorming phase we introduced participants to the Delphi procedure as well as to DLT and genomics in order to establish a common understanding for both concepts. Drawing on previous ranking-type Delphi studies [16, 26], we next asked participants to name and briefly describe three to ten opportunities and challenges for the application of DLT in genomics. We also gave participants the possibility to proceed with the questionnaire in case they did not want to provide three opportunities or challenges (e.g., some participants were only able to name one opportunity but multiple challenges, or vice versa). At the end of the brainstorming questionnaire we asked participants to provide basic demographic information and an email address so that we could invite them for the subsequent Delphi phases.

Overall, participants provided us with 33 entries for potential opportunities and 39 entries for potential challenges. We consolidated the set of responses by manually aggregating duplicate responses, unifying terminology, and grouping similar opportunities/ challenges [27, 29]. In order to augment the lists of opportunities and challenges, we further conducted a review of related literature and white papers. To identify scientific publications addressing the use of DLT in genomics, we searched pertinent scientific databases that cover a wide range of journals and conferences: ACM Digital Library, EBSCOhost, IEEE Xplore, ProQuest, ScienceDirect, and PubMed. We searched each database with the following string in title, abstracts and keywords: *("distributed ledger technolog*" OR "blockchain*" OR "DLT") AND ("genom*" OR "genetic*" OR "DNA" OR "proteonomic*")* and limited our search to peer-reviewed articles to ensure high quality of articles. Our search yielded 573 articles, of which we deemed only five relevant for this research. We also included eight available white papers of relevant businesses. Subsequently, two researchers coded potential opportunities and challenges in these articles and white papers independently and compared their results.

The data consolidation process and literature review resulted in a list of 17 opportunities and 15 challenges for the application of DLT in genomics. In order to ensure that participants' thoughts were adequately captured and represented in the consolidated list, we asked participants to approve the list of opportunities/ challenges as suggested by Paré, et al. [27]. Participants' feedback led to minor adjustments, (e.g., rephrasing certain aspects), but did not result in adding new or removing existing opportunities/challenges.

3.2.2. Narrowing down and ranking phases. Extant literature suggests 20 or less items for the ranking phase [29], since a large number of items reduces participants ability to properly distinguish the ranks of individual items [30]. The brainstorming phase yielded less than 20 opportunities as well as less than 20 challenges. Thus, and in an attempt to reduce panel attrition, we decided to skip the narrowing down phase and directly proceed to the ranking phase, although the narrowing down phase might in some cases provide additional insights.

In the ranking phase, we presented the 17 opportunities and 15 challenges in random order and asked participants to rank them according to the following rationales: For the opportunities, we asked participants to rank them with regard to their value for the genomics community from highest to lowest. For the challenges, we asked participants to rank them with respect to which challenges deserve the genomics community's attention and resources from most to least. Participants were also asked to justify their rankings, although this was not mandatory and could be skipped. We used Kendall's coefficient of concordance (W) to measure the degree of consensus between the experts' rankings (Paré et al. 2013; Schmidt 1997; Singh et al. 2009). It attains values between 0 and 1, whereby consensus is considered strong for $W \geq 0.7$, moderate for $0.7 > W \geq 0.5$, and weak for $0.5 > W \geq 0.3$ (Schmidt 1997). For lower values no consensus is presumed. Moreover, a Friedman test was used to calculate the mean rank for each opportunity/challenge (Friedman

1937). Nine participants signaled their willingness to participate in additional Delphi rounds after the brainstorming phase, with only 3 actually completing the ranking phase by the time of writing this paper.

4. Results

Based on the inputs of experts during the Delphi study and our literature review, we were able to identify 17 opportunities (see Tab. 3) and 15 challenges (see Tab. 4) for the application of DLT in genomics. In order to structure identified opportunities and challenges, we drew on the TOE-framework [31] and categorized opportunities and challenges into technology (i.e., opportunities/challenges related to the technology itself), organization (i.e., opportunities and challenges related to organizational structures of the genomics community), and environment (opportunities and challenges related to factors that are external to the genomics community).

4.1. Opportunities for DLT in genomics

4.1.1. Technology. Most identified opportunities (9 out of 17) belong to the technology category. Owing to the inherent characteristics of genome data and their high sensitivity, protection of genetic privacy is a major objective in genomics. Consequently, we identified several technological opportunities for DLT in genomics that support the protection of data owners' genetic privacy. DLT, for example, warrants high degrees of confidentiality by ensuring that genome data is only disclosed to third parties with consent of the data owner. Likewise, and completing the so-called CIA triad, DLT also warrants high degrees of integrity by ensuring that genome data is not altered without data owners' prior agreement, as well as high degrees of availability of DLT-based platforms and services (and thus genome data) due to being inherently decentralized and removing single points of failure. On the one hand DLT further enables data owners to maintain their genome data anonymously, meaning that the tracing of data owners' real identities requires a prohibitive amount of effort, while on the other hand also affording high levels of transparency by ensuring the traceability of all actions performed on the data. From the perspective of data users (e.g., researchers, medical professionals, insurers) the durability as well as the accuracy of genome data records are important factors for an effective use of such data. Towards this end DLT affords the establishment of a permanent genome data record (opportunity: durability), as well as verifying the correctness of genome data (opportunity: data accuracy). Adding to this, DLT-based systems reduce storage requirements for those in charge of governing the genome data by enabling their distributed storage. Finally, many contemporary DLT concepts rely on compute intensive tasks as part of their consensus protocols (e.g., proof-of-work). DLT therefore might allow for the solving of compute intensive tasks in genomics by enabling the distribution of compute tasks to nodes of a DLT network, for example, as part of the proof-of-work principle.

4.1.2. Organization. We categorized 7 out of the 17 opportunities into the organization category. Based on its distributed and decentralized nature, DLT eliminates the need for central authorities. Thus, it supports decentralization within the genomics community by, for example, helping to break up extant data silos or affording the removal of (costly) intermediaries. While the current genomics ecosystem is relatively closed with several large closed-access databases and limited possibilities for individuals to actively participate, the openness of DLT (i.e., many DLT-based systems are open to everyone) promotes the active participation of all stakeholders (e.g., researchers, data owners) in the generation, sharing, and processing of genome data. Another frequent problem within genomics pertains to heterogeneity of different information systems, which often hinders effective exchange of genome data between different entities. DLT supports interoperability by facilitating the standardization of technologies and policies. Similarly, many processes in genomics (especially the granting and revocation of consents) are complex and tedious. DLT supports the formalization and subsequently automatic execution of processes in genomics through the utilization of smart contracts. An often-raised question about genome data relates to the ownership of the data. To this end, DLT not only allows to enforce property rights over personal genome data by enabling the verification and administration of said property rights, but also gives data owners direct control over their data, by enabling data owners to grant and revoke access rights to their genome data flexibly and on-demand, based on their changing data sharing and access preferences. Closely related to the question of ownership is the question of how to reimburse entities for the disclosure or processing of genome data. Thereby, application of DLT in genomics allows for the creation of a token economy, which enables the exchanging of value in the form of tokens (e.g., monetary value, access to services) for the disclosure and/or processing of genome data.

Table 3. Opportunities for DLT in genomics.

Cat.	Opportunity	Src
Technology	**Anonymity.** DLT enables data owners to maintain (e.g., store, share) their (genome) data without their real identities being traceable since it is either not possible or requires unreasonable efforts.	●
	Availability. DLT increases the probability to which a DLT-based (genome) data platform or service can be reached in a fully functioning condition, due to it being a decentralized system.	●
	Confidentiality. DLT is capable of ensuring that (genome) data is only disclosed to third parties with consent of the data owner, where the data owner defines the granularity and form of the (genome) data.	●
	Data accuracy. DLT allows to verify the correctness of (genome) data.	●
	Data storage. DLT reduces data storage requirements for individual data owners by enabling the distributed storage of (genome) data.	◑
	Distributed computing. DLT allows the solving of compute-intensive problems by enabling the distribution of compute tasks to nodes of a DLT network (e.g., as part of proof-of-work).	◑
	Durability. DLT affords the establishment of a permanent (genome) data record.	●
	Integrity. DLT is capable of ensuring that (genome) data is not altered without prior agreement of the data owners.	●
	Transparency. DLT affords the traceability of the handling of (genome) data (e.g., data access, data processing, data search) by means of logging all performed actions.	●
Organization	**Decentralization.** DLT eliminates the need for central authorities, thus helping to break up data silos and affording the removal of intermediaries.	●
	Flexibility. DLT enables the on-demand granting and revocation of access rights to (genome) data based on data owners' changing data sharing and access preferences.	●
	Interoperability. DLT supports the exchange of (genome) data among different information systems by means of facilitating standardization of technologies and policies.	◑
	Openness. DLT promotes the active participation of all stakeholders (e.g., researchers, data owners) in the generation, sharing, and processing of (genome) data.	●
	Ownership. DLT allows for the verification and administration of property rights on personal (genomic) data.	◑
	Process automation. DLT affords the formalization of processes in genomics through smart contracts for the purpose of automating these processes.	●
	Token economy. DLT enables the exchanging of value via of tokens (e.g., monetary value, access to services) for the disclosure and/or processing of (genome) data.	●
Env.	**Interdependent privacy.** DLT allows to enforce that relatives approve of a data owner sharing their genome data by means of smart contracts.	◐

Column Source: ◐ = *opportunity originated from brainstorming phase only;* ◑ = *opportunity originated from literature review only;* ● = *opportunity originated from brainstorming phase and review*

4.1.3. Environment. Finally, one opportunity belongs to the environment category. Kinship is a central characteristic of genome data, meaning that blood relatives share certain genetic traits with each other [15]. This could potentially also impede the privacy of blood relatives when sharing one's genome data. DLT supports the preservation of such interdependent privacy by requiring that affected relatives approve of one sharing their genome data prior to the actual sharing.

4.2. Challenges for DLT in genomics

4.2.1. Technology. Similar to the opportunities, the majority of challenges (7 out of 15) belong to the technology category. Although DLT receives tremendous attention from research and practice it is still a relatively immature technology with many unresolved questions and issues. Especially in terms of security and the sensitivity of genome data, technology maturity or more precisely the lack thereof remains an issue. Adding to this, data handling (i.e., the retrieval and management of genome data) and data storage are two challenges deeply rooted in DLT characteristics. Most DLT-concepts were designed for small transactional data and are unable to store genomics-sized data sets on ledger [4, 6]. Furthermore, current implementations of DLT lack the capabilities to efficiently query genome data (e.g., accessing arbitrary parts of a dataset or streaming data). Towards this end, many extant proposals for DLT-based platforms in genomics are also based on Blockchain, whose primary consensus protocols are based on proof-of-work, which relies on compute-intensive tasks. Although other less resource-consuming consensus mechanisms (e.g., proof-of-stake) have been proposed, the dominance of inefficient consensus protocols wastes precious computing resources that could be used for solving genomics-related problems instead. As pointed out before, ensuring integrity of data is an often-cited opportunity of DLT. However, at the same time DLT's rigidness could also pose a serious challenge in terms of integrity, namely if those rightfully interested in changing data are being hindered to do so (e.g., due to prohibitively expensive consensus mechanisms or simply a lost encryption key).

Similar to integrity, durability was not only identified as an opportunity but also as a challenge. Due to rigidness of DLT-based systems and data replication,

Table 4. Challenges for DLT in genomics.

Cat.	Challenge	Src
Technology	**Data handling.** DLT-based systems are not designed to retrieve or manage high volumes of (genome) data efficiently.	●
	Data storage. DLT-based systems are not designed to store genomics-size data sets on-ledger.	●
	Durability. Disclosing (genome) data via DLT is a permanent, difficult-to-reverse decision.	●
	Efficiency. Consensus mechanisms in DLT-based systems such as proof-of-work waste computing resources.	◑
	Integrity. Altering (genome) data that is stored, shared, or managed via DLT-based systems might require much effort or be impossible.	●
	Technology maturity. DLT is an immature technology with potential security issues.	●
	Transparency. Transparency of distributed ledger systems might impede individuals' privacy.	●
Organization	**Conflicting interests.** Different stakeholders might have conflicting interests, which are impossible to resolve.	●
	Ease of use. Getting, maintaining, and managing access to (genome) data via DLT requires much effort.	●
	Interoperability. Effective use of DLT for genomics requires interoperability between a multitude of diverse information systems.	●
	Novelty. DLT is largely unknown to the genomics community and poorly understood.	●
	Token economy. Rewarding data owners with tokens for sharing their (genome) data can incentivize mindless data sharing.	◐
Env.	**Emergency access.** Situations that require access to health-related data by third parties without data owners' prior approval (e.g., by medical professionals in an emergency, when data owners are unconscious), might violate foundational principles of DLT.	◑
	Openness. Openness of DLT-based systems can result in new attack vectors and misuse of the system and stored (genome) data.	●
	Uncertain regulation. Regulations around DLT and genetic privacy are constantly changing and evolving differently in different parts of the world.	●

Column Source: ◐ = *challenge originated from brainstorming phase only;* ◑ = *challenge originated from literature review only;*
● = *challenge originated from brainstorming phase and review*

completely removing information that has once been stored from the system is difficult and, in many cases, impractical. Another often cited opportunity for the application of DLT in genomics is transparency. Yet, such uncompromising transparency could also impede individuals' privacy by providing information or enabling the deduction of information that one does not wish to be available to others (e.g., logs of an individual sharing their genome data with a cancer research center could indicate that this person suffers from cancer).

4.2.2. Organization. We categorized five challenges as organizational challenges. While we identified technology maturity as a technology-related challenge, novelty of DLT also poses a challenge to the genomics community itself as its benefits and risks are largely unknown or at least poorly understood. Not only might this lead to ineffective implementations but also to outright rejection of this technology. Likewise, getting, maintaining, and managing access to DLT-based systems still requires comparably much effort. Overall, most DLT-based systems still suffer from poor ease of use, which could also hinder adoption within the genomics community. The effective application of DLT within genomics requires interoperability between a multitude of diverse information systems. Although DLT could facilitate standardization of technologies and policies within genomics, the current situation is far from this ideal, essentially creating a chicken-and-egg problem in terms of interoperability. Adding to this, effective application of DLT also requires the formalization of certain aspects (e.g., access requirements). Here, different stakeholders might have conflicting interests, which could be difficult or impossible to resolve. From an ethical point of view, rewarding data owners with tokens for sharing their genome data might result in mindless data sharing. Due to the unique characteristics of genome data, future uses and potential avenues for privacy infringements are difficult to foresee, even for professionals. Introducing monetary incentives into this ecosystem could encourage data sharing without being able to adequately assess the potential issues.

4.2.3. Environment. Lastly, the environment category comprises three challenges. A frequent requirement within health information technology is a so-called emergency access, where in cases of emergency medical professionals are granted access to patient data without their prior consent. Such access might violate fundamental principles of DLT and could be impossible or at least be highly difficult to realize. Next, openness refers to the system promoting the active participation of involved stakeholders. However, at the same time such an open system design could also invite malicious users into the system and result in new attack vectors or the misuse of the system and especially stored genome data. The third and last challenge in this category refers to uncertain regulation. Since DLT is a relatively young technology its true benefits as well as its problems for societies around the globe are largely unknown. In an attempt to keep up with the rapid technological advances, regulation in different parts of the world is constantly changing and adapting to the new technological realities. Consequently, changes in

Table 5. First round rankings.

Opportunity	Mean rank	Overall rank
Confidentiality	4.00	1
Ownership	4.33	2
Flexibility	5.67	3
Decentralization	6.33	4
Openness	6.67	5
Interdependent privacy	7.00	6
Anonymity	7.67	7
Transparency	9.00	8
Durability	9.67	9
Interoperability	10.00	10
Availability	10.33	11
Integrity	10.67	12
Data accuracy	11.33	13
Distributed computing	11.67	14
Token economy	12.33	15
Process automation	12.67	16
Data storage	13.67	17
		Kendall's W = 0.346
Challenge	**Mean rank**	**Overall rank**
Uncertain regulation	1.33	1
Durability	5.67	2
Ease of use	6.00	3
Emergency access	6.00	3
Novelty	6.00	3
Technology maturity	6.00	3
Openness	6.33	7
Efficiency	8.67	8
Integrity	9.33	9
Interoperability	9.67	10
Token economy	10.00	11
Data storage	10.33	12
Transparency	10.33	12
Conflicting interests	11.00	14
Data handling	13.33	15
		Kendall's W = 0.450

regulation around the world could render certain applications of DLT illegal or impossible.

4.3. Ranking of opportunities and challenges

Based on the inputs we received from our panel so far, the top three opportunities are (1) confidentiality, (2) ownership, and (3) flexibility, whereas the top three challenges are (1) uncertain regulation, (2) durability, and (3) ease of use. Although Kendall's W must be considered low for both rankings ($W_{Opportunities} = 0.346$, $W_{Challenges} = 0.450$), overall consensus seems to be higher for the challenges ranking. Table 5 shows the mean ranks as well as the resulting overall ranks of the first round for all elicited opportunities and challenges.

5. Discussion

5.1. Principal findings

The objective of this research was to identify salient opportunities and challenges for the application of DLT in genomics. To this end, we conducted a ranking-type Delphi study with an international panel of experts and augmented the results of the brainstorming phase with a comprehensive literature review, which altogether yielded 17 diverse opportunities as well as 15 diverse challenges for the application of DLT in genomics. Consistent with current efforts on DLT in genomics, our Delphi study and literature review yielded token economy as one potential opportunity. However, contrary to the currently predominant focus on the use of DLT to establish genome data markets, results of our Delphi study also yielded several opportunities deemed more important than the creation of a token economy (e.g., confidentiality, ownership, or flexibility). Furthermore, uncertain regulation was by far deemed as the most important challenge, which is also consistent current debates around the application of DLT in critical contexts such as health care or finance. Since DLT has received tremendous attention from researchers and practitioners in the health care sector [23], an obvious question is whether the opportunities and challenges presented here are unique to genomics. While many of the opportunities and challenges certainly also apply to health care in general, we also see some opportunities (e.g., interdependent privacy) and challenges (e.g., data handling and data storage) that are unique or at least more relevant to the genomics context. Moreover, the relative rankings of individual opportunities and challenges might be different for the genomics context than for a general health care context.

Interestingly, many of the identified opportunities were at the same time also named as a potential challenge. A prime example for this is transparency, which overall might be an opportunity for the genomics community but could at the same time also infringe individuals' privacy. We think that there are two potential reasons for this that can also be found in the presented challenges. First, DLT is a novel and still poorly understood phenomenon in the genomics community. With time and increasing knowledge about DLT, we might be able to address several challenges, transforming them into pure opportunities. Second, many of the conflicting interests between the diverse stakeholders involved in genomics might be unresolvable, creating a reality in which the very same DLT characteristic is an opportunity for some stakeholders, while being a challenge to others.

Looking at the sources of identified opportunities and challenges, we see that most opportunities (12 out of 17) as well as most challenges (12 out of 15) were actually named in both, the brainstorming phase of our Delphi study and in extant literature. Although this surprising finding seemingly contradicts our initial statement that current discussions about DLT in genomics are predominantly positive, it must be noted that potential challenges were often only named implicitly. Moreover, sections discussing potential challenges were rather short and often downplayed named challenges.

5.2. Implications

Our work has several implications for research and practice. For practitioners, including those researchers enticed by the utilization of DLT for their work, our research highlights several benefits brought forward by the application of DLT. It especially shows that there are many benefits of DLT beyond the creation of mere data markets. However, our results also show the presence of a diverse set of challenges that need to be overcome in order to realize the full potential of DLT in genomics. Those interested in the utilization of DLT in genomics should therefore develop strategies to address the most pressing challenges.

For research, we are among the very first to systematically elicit opportunities and especially challenges for the nascent phenomenon that is DLT in genomics, as well as their relative importance. We thereby add to research a better and contextualized understanding of DLT. In particular, we strengthen the importance of genomics as a research context for the application of DLT that, although similar to the application of DLT in general health care, possesses some unique features like interdependent privacy issues. We also highlight the very existence of the dual roles of some opportunities that at the same time are regarded as challenges (or vice versa). Starting from here, researchers enticed by the application of DLT in genomics can use our lists of opportunities and challenges to further investigate specific opportunities or challenges. The dual role of some opportunities/challenges also warrants further, in-depth investigations.

5.3. Limitations and future research

Despite this study being a first step towards a more nuanced contemplation of the application of DLT in genomics, our research is not without limitations. First, the results presented in this study are to a large degree based on the inputs of a limited number of experts. This is mainly due to the fact that there is a shortage of professionals with expertise in DLT and genomics and that many experts, especially those working for businesses in the DLT-in-genomics space, did not respond to our invitations. Although the number of experts who participated in our study is within established thresholds, we tried to account for this limitation by also conducting a comprehensive review of related white papers and research articles. We are thus confident that our results provide a comprehensive picture of the opportunities and challenges for DLT in genomics. Another limitation of our work pertains to the fact that so far only 3 experts completed the first ranking phase, which is why our rankings must be considered preliminary at best. Although such high panel attrition is not uncommon in Delphi studies, we will continue to invite the remaining experts to complete the ranking and also investigate other means (e.g., a survey) to establish a more robust ranking of the presented opportunities and challenges. Towards this end, and drawing on the conflicting interests challenge, future research should also investigate the potential existence of different relative rankings of the presented opportunities and challenges for different stakeholders in genomics.

6. Conclusion

Although DLT has recently emerged as a hot topic within the genomics community, little is known about the actual opportunities and challenges for the application of DLT in genomics. Within this research we systematically elicited 17 opportunities and 15 challenges for DLT in genomics as well as their relative importance. Thereby, we make important contributions to practice and research. For practice, we highlight application opportunities for DLT in genomics beyond genome data markets as well as important potential pitfalls that need to be addressed. For research, we underline genomics as a promising application area for DLT and support a more nuanced, less hyped view of this phenomenon. We also lay the foundations for researchers interested in differences between the application of DLT in genomics and in health care in general.

7. References

[1] Kannengießer, N., Lins, S., Dehling, T. and Sunyaev, A., Mind the Gap: Trade-Offs between Distributed Ledger Technology Characteristics. Preprint available at https://arxiv.org/abs/1906.00861.

[2] Lavazova, O., Dehling, T. and Sunyaev, A., From Hype to Reality: A Taxonomy of Blockchain Applications. In Proceedings of the 52nd HICSS (Maui, HI, USA, 2019).

[3] Weber, I., Gramoli, V., Ponomarev, A., Staples, M., Holz, R., Tran, A.B. and Rimba, P., On availability for blockchain-based systems. In Proceedings of the 2017 IEEE 36th SRDS (Hong Kong, China, 2017).

[4] Shabani, M. Blockchain-based platforms for genomic data sharing: a de-centralized approach in response to the governance problems? Journal of the American Medical Informatics Association, 26, 1 76-80.

[5] Lee, S.-J., Cho, G.-Y., Ikeno, F. and Lee, T.-R. BAQALC: Blockchain Applied Lossless Efficient Transmission of DNA Sequencing Data for Next Generation Medical Informatics. Applied Sciences, 8, 9 1471.

[6] Ozercan, H.I., Ileri, A.M., Ayday, E. and Alkan, C.J.G.r. Realizing the potential of blockchain technologies in genomics. Genome Research, 28, 9 1255-1263.

[7] Trujillo, J.L., Fromhart, S. and Srinivas, V. Evolution of blockchain technology: Insights from the GitHub platform. https://www2.deloitte.com/insights/us/en/industry/financial-services/evolution-of-blockchain-github-platform.html (retrieved 2019/05/28).

[8] Lamport, L., Shostak, R. and Pease, M. The Byzantine generals problem. ACM Transactions on Programming Languages and Systems, 4, 3 382-401.

[9] Zhang, K. and Jacobsen, H.-A., Towards dependable, scalable, and pervasive distributed ledgers with blockchains. In Proceedings of the 2018 IEEE 38th ICDCS (Vienna, Austria, 2018).

[10] Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf (retrieved 2019/04/14).

[11] Wohrer, M. and Zdun, U., Smart contracts: security patterns in the ethereum ecosystem and solidity. In Proceedings of the 2018 IWBOSE (Campobasso, Italy, 2018).

[12] World Health Organization Genomics and world health: Report of the Advisory Committee on Health Research. https://apps.who.int/iris/bitstream/handle/10665/42453/a74580.pdf (retrieved 2019/05/30).

[13] Aronson, S.J. and Rehm, H.L. Building the foundation for genomics in precision medicine. Nature, 526, 7573 336.

[14] Wetterstrand, K.A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) https://www.genome.gov/sequencingcostsdata (retrieved 2019/06/11).

[15] Naveed, M., Ayday, E., Clayton, E.W., Fellay, J., Gunter, C.A., Hubaux, J.-P., Malin, B.A. and Wang, X. Privacy in the genomic era. ACM Computing Surveys, 48, 1 6:1-6:44.

[16] Thiebes, S., Lyytinen, K. and Sunyaev, A., Sharing is About Caring? Motivating and Discouraging Factors in Sharing Individual Genomic Data. In Proceedings of the 38th ICIS (Seoul, South Korea, 2017).

[17] Clayton, E.W. Ethical, legal, and social implications of genomic medicine. New England Journal of Medicine, 349, 6 562-569.

[18] Weidman, J., Aurite, W. and Grossklags, J. On Sharing Intentions, and Personal and Interdependent Privacy Considerations for Genetic Data: A Vignette Study. IEEE/ACM Transactions on Computational Biology and Bioinformatics.

[19] Aktan-Collan, K., Haukkala, A., Mecklin, J.-P., Uutela, A. and Kääriäinen, H. Comprehension of cancer risk one and 12 months after predictive genetic testing for hereditary non-polyposis colorectal cancer. Journal of Medical Genetics, 38, 11 787-792.

[20] Terry, S.F. The tension between policy and practice in returning research results and incidental findings in genomic biobank research. Minnesota Journal of Law, Science & Technology, 13 691.

[21] Nebula Genomics Blockchain-enabled genomic data sharing and analysis platform. https://icorating.com/upload/whitepaper/dhVxJUvg3HBgFiwBCx6fgHoIB8IN8SF1ktevH8Os.pdf (retrieved 2019/06/10).

[22] Mackey, T.K., Kuo, T.-T., Gummadi, B., Clauson, K.A., Church, G., Grishin, D., Obbad, K., Barkovich, R. and Palombini, M. 'Fit-for-purpose?'–challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Medicine, 17, 1 68.

[23] Roman-Belmonte, J.M., De la Corte-Rodriguez, H. and Rodriguez-Merchan, E.C. How blockchain technology can change medicine. Postgraduate Medicine, 130, 4 420-427.

[24] Talukder, A.K., Chaitanya, M., Arnold, D. and Sakurai, K., Proof of Disease: A Blockchain Consensus Protocol for Accurate Medical Decisions and Reducing the Disease Burden. In IEEE SmartWorld (Guangzhou, China, 2018).

[25] Thiebes, S., Schlesner, M., Brors, B. and Sunyaev, A. Distributed Ledger Technology in genomics: A call for Europe. European Journal of Human Genetics, 2019, 1-2.

[26] Okoli, C. and Pawlowski, S.D. The Delphi method as a research tool: an example, design considerations and applications. Information & Management, 42, 1 15-29.

[27] Paré, G., Cameron, A.-F., Poba-Nzaou, P. and Templier, M. A systematic assessment of rigor in information systems ranking-type Delphi studies. Information & Management, 50, 5 207-217.

[28] Miles, M.B. and Huberman, A.M. Qualitative data analysis: An expanded sourcebook. Sage Publications Inc., Newbury Park, CA, 1994.

[29] Schmidt, R.C. Managing Delphi surveys using nonparametric statistical techniques. Decision Sciences, 28, 3 763-774.

[30] Schmidt, R., Lyytinen, K., Keil, M. and Cule, P. Identifying software project risks: An international Delphi study. Journal of Management Information Systems, 17, 4 5-36.

[31] Bosch-Rekveldt, M., Jongkind, Y., Mooi, H., Bakker, H. and Verbraeck, A. Grasping project complexity in large engineering projects: The TOE (Technical, Organizational and Environmental) framework. International Journal of Project Management, 29, 6 728-739.

8. Appendix

Table A1. Delphi Panel Characteristics.

Characteristics	Panel profile (n=12)
Sex	Female: 25,00% Male: 75,00%
Age (years)	Avg.: 43 Min.: 29; Max.: 66
Nationality	U.S.: 25,00% German: 16,67% Other: 8,33%
Experience in genomics (years)	Avg.: 7,25 Min.: 1; Max.: 25
Experience in DLT (years)	Avg.: 2,79 Min.: 0; Max.: 12