Abstract

The rapid growth in the comprehension and insight into genomic and human health data makes it possible to significantly prolong human life expectancy in the next 20 years. The comprehensive implementation of personalized healthcare technology will be the main driver of this change. Personalized health care services will rapidly become more popular as the cost of genome sequencing services rapidly falls within the next three to five years. The vast majority of health and medical providers in the future will begin to provide services based on user genomes and daily health data. The Gene Source Code (GSC) Foundation is committed to building a storage and decentralized application platform for genomic and health data, known as the Gene Source Code System. This system will be built around a human healthcare data ecosystem and create one unique economic mechanism.

The ultimate goal is to encourage users to share genomic and health data within the ecosystem, provide researchers with adequate data support, and provide data users with an easy-to-use interface. The GSC System will create a distributed storage network blockchain system that provides genetic data and individual health data through an integrated interface. Health and medical service providers, including hospitals, chronic disease management agencies, and health consulting service providers, can use the GSC System to build DApps. These applications will provide users with personalized and accurate health services based on their genomic and health data, if they have given prior consent.

The GSC team will firstly create a smart health assistant DApp (which will motivate users to share data and obtain health services)，with which individual users, will receive Gene token rewards when sharing health data, and may also obtain accurate health advice when they have shared enough data. The GSC Foundation believes that it’s every human’s right to live a healthy life and that all people are universally equal even in this era of exponential technological growth. This right should not be affected by a person’s economic condition because each individual is already contributing to the overall health of humankind. The GSC system will be based on secure blockchain technology and data encryption/masking technology so that each individual has ownership of his/her health data and the benefits that come from that ownership.

Table of Contents

I Preface and Background

1.1 Technology and Blockchain's Exponential Growth

The exponential growth of technology has profoundly and rapidly affected everyone's lives in all fields, including the impact of the skyrocketing growth of blockchain technology. It has been nine years since Satoshi Nakamoto invented Bitcoin in 2009. During this period, blockchain technology has rapidly developed and consolidated. Digital cryptocurrency, as represented by Bitcoin, has gradually begun to be understood, accepted and even sought after by the public. Bitcoin, which decentralizes digital virtual assets, is commonly known as digital gold, and Bitcoin users have gradually expanded from crypto geeks to the general public.

At the same time, there are many other exciting developments in the blockchain cryptocurrency industry such as the Ethereum Project, created by Vitalik Buterin and Gavin Wood in 2013, which has been very successful. Due to the ease of use of the Ethereum design concept, more and more decentralized applications, DApps (Distributed Applications), are being built on the technology platform of Ethereum. In addition, there are now many high-profile blockchain projects such as Zcash, Dash, EOS, etc. The rapid development and evolution of blockchain technologies will cause more and more industries to embrace the application of blockchain.

1.2 The Rapid Development of Genomic and Healthcare Big Data

At the same time, in the field of health science, the rapid development of technology has presented both opportunities and challenges. In the foreseeable future, mankind will benefit from sharing genetic big data, and as a result, improved healthcare will greatly extend lifespans and eliminate the vast majority of diseases. As the price of genomic sequencing data continues to drop rapidly, scientists and citizen scientists will have more access to vast amounts of genomic and human health data to study and overcome specific medical and healthcare problems. As the costs of gene sequencing decrease, scientists and citizen scientists will be able to access vast amounts of genetic and physical fitness data for academic or medical purposes. In addition, the research and review of other species' DNA will be beneficial as well. These studies will greatly advance our knowledge of medicine, materials, and energy.

1.3 Using Genetic Data

Genes are substances that transmit genetic information. This information is the source of birth, aging, illness, and death. Gene data contains a large amount of health information and it is closely related to an individual's health. At present, genetic data mainly includes the following aspects:

(1) Assisted clinical diagnosis

Many diseases show a resemblance to other diseases which leads to difficulties in discriminating between them clinically. It also may lead to an incorrect diagnosis. If genetic testing is used to find the cause of the disease at the genetic level, it can help clinicians differentiate between various diseases and even correct clinical diagnosis mistakes.

(2) Screening carriers

One of the most common health screenings is for Down syndrome. Traditional screening for Down syndrome is performed on serum. The detection rate is only 65-75% so false negatives are inevitable. Non-intrusive prenatal genetic testing can accurately detect children with Down syndrome, including 18-trisomy syndrome and 13-trisomy syndrome. In addition, the screening of related disease-causing genes among high-risk groups with a family history of single-gene disorder (especially recessive disorder) can promptly detect the virulence gene carriers in the family and analyze the pathogenic risk among offspring. These 7 screening results provide family members with informative suggestions on how to prevent the defective genes from affecting the next generation.

(3) Guided treatment

The strength of a prescription is decided after a number of extensive tests are performed on groups. The doses decided this way are appropriate for most people, but not for everyone. Some normal dosage prescriptions may cause death because of the genetic differences among patients. Then medicine, which is supposed to help the patient, may in the end, further endanger a patient's health. This phenomenon is called adverse drug reactions (ADR).

One example is the drug warfarin, which is an anticoagulant (a drug that prevents blood clotting). Patients taking this medicine may greatly reduce their risk of thrombosis; however, if the anticoagulant is prescribed at too high a dosage, the blood can fail to coagulate properly and

continuous bleeding may threaten the patient's life. In the human body, there is an enzyme called CYP2C9 which can metabolize this anticoagulant, break it down into small molecules, and destroy its anticoagulant effect. Under normal circumstances, warfarin is metabolized after it completes its therapeutic action, and does no harm to the human body. However, if a person has mutations in CYP2C9 production, his or her metabolic function slows down and becomes a poor metabolizer. As a result, warfarin is metabolized slowly and accumulates in the body, eventually causing continuous bleeding.

The role of genetic testing is crucial in this example. It can, firstly, determine whether a patient's CYP2C9 production has a mutation, and then which metabolic type they belong to. Once this is known, the dosage of the drug can be based on that individual's metabolism type. If their metabolism is strong, then the dose can be increased somewhat; if it is weak, then the dosage should be reduced and the drug should be closely monitored in the blood. This not only guarantees the healing effect of the drug but also causes no ADR.

There are countless instances in which the genetic testing of the patient's metabolism plays an important role in determining the correct dosage. Various genes are involved in the metabolism of drugs for depression, diabetes, asthma, and osteoporosis. Although genetic testing can guarantee the correct dosage, a person's individual genes are seldom considered when prescribing drugs, and as a result, some prescribed medicine causes adverse reactions to certain patients from time to time.

(4) Predict the risk of an individual's illness

Modern medicine believes that disease comes from both genes and external factors. Almost all diseases are related to genes. In some cases, genetic mutations cause dysfunctions in the human body and cause physiological functions to deviate from normal conditions, thereby creating opportunities for diseases. At the same time, negative external factors (an unclean environment, a harsh climate, a poor diet, radiation, lack of exercise, etc.) amplify the problematic consequences of genetic mutations, and this combination of external and internal factors causes disease. Colon cancer is an example.

The metabolism of epithelial cells in the colon is controlled by a gene called APC. If a person's APC gene cannot properly control the normal metabolism of epithelial cells, the uncontrolled epithelial cells may undergo excessive proliferation. This is the genesis of colon cancer. The DCC

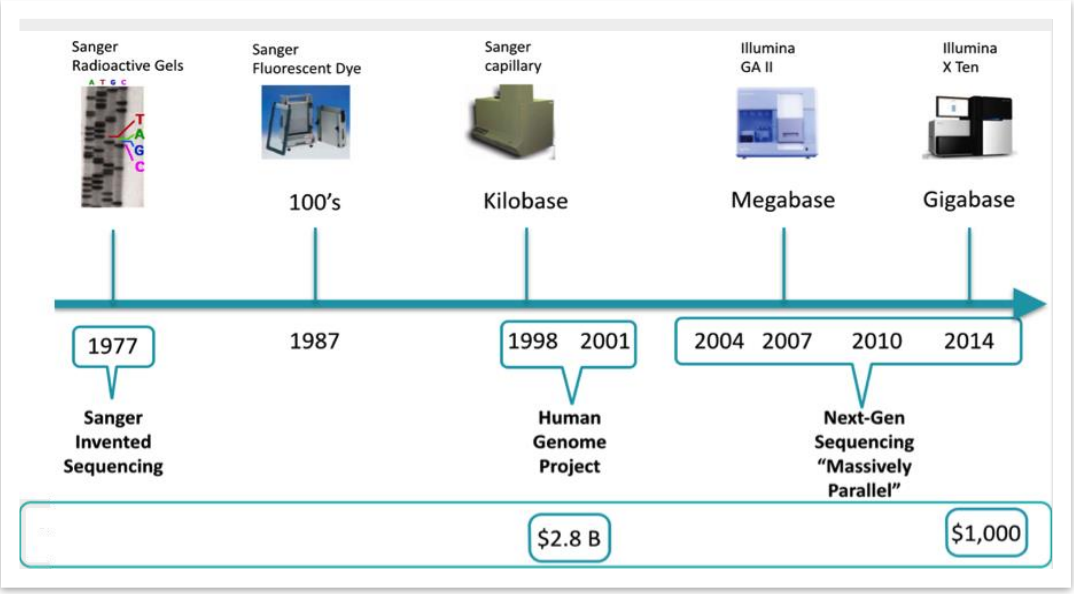
and P53 genes in the human body inhibit these cells from becoming cancerous. If DCC and P53 genes are also mutated and cannot effectively prevent the cancerous transformation of these cells, the body faces the threat of colon cancer. Therefore, we believe that the defects in the APC, DCC and P53 genes are closely related to the occurrence of colon cancer.

People with these defective genes are faced with an imminent risk of colon cancer. Through testing, before the subepithelial neoplasia spreads and evolves into colon cancer, the warning signs of cancer may be discovered. The disease can be prevented beforehand through measures such as improving health habits, improving one's living environment, scientific preventative healthcare, and the avoidance of other adverse factors.

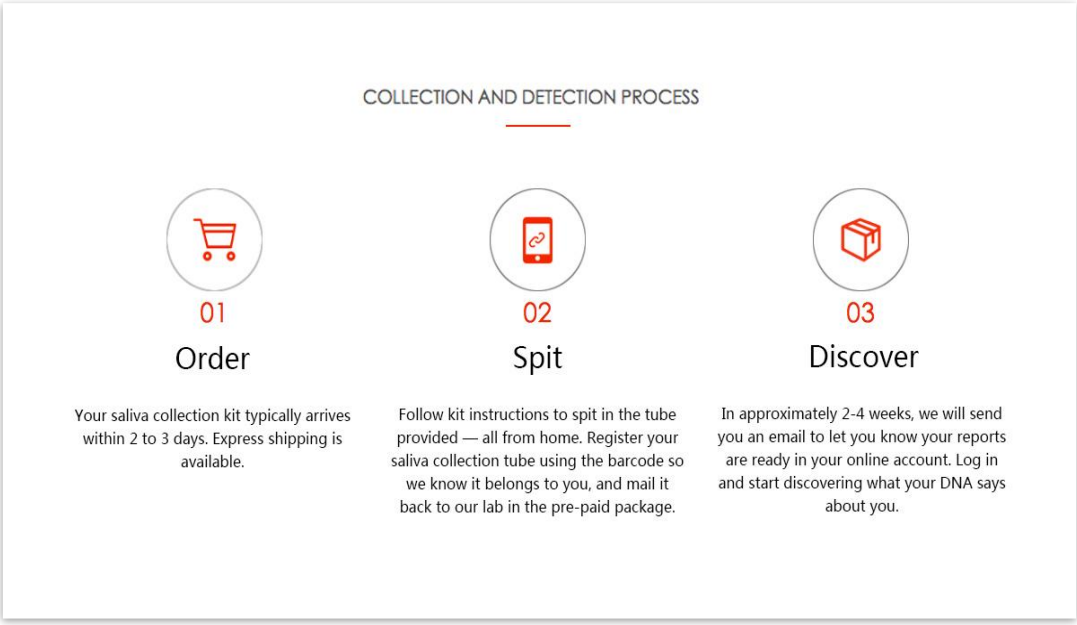
1.4 The Current Status of Genome Sequencing and Health Big Data

DNA sequencing methods have developed rapidly, from manual to capillary electrophoresis. The year 2001 was a milestone year. Through the efforts of many countries, the HGP (human genome project), was completed after 13 years of work and at a cost of 437 million USD.

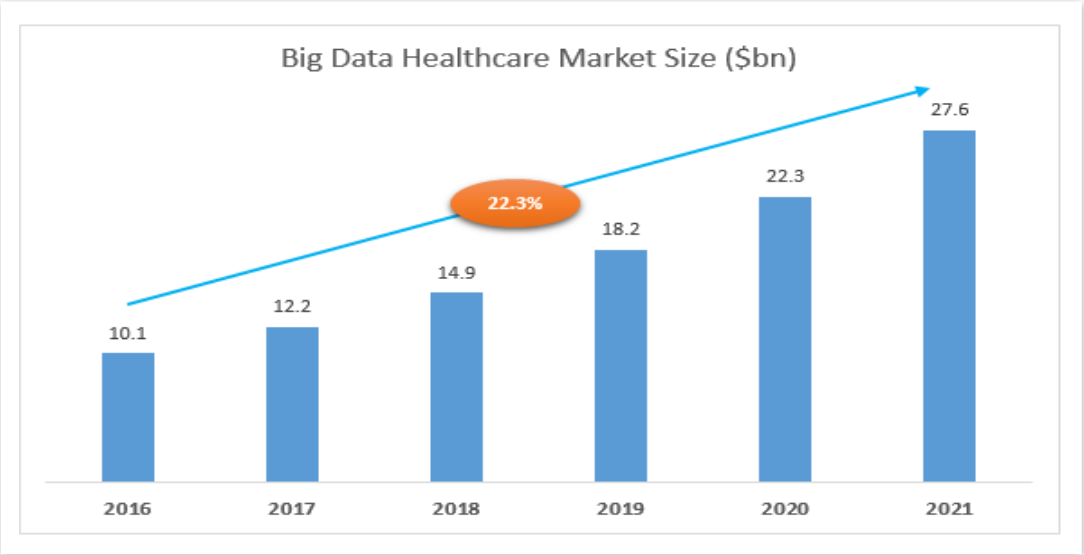
At the beginning of the 21st century, with the advent of high-throughput sequencing technologies, both sequencing time and sequencing costs have been greatly reduced. In 2007, the birth of the first complete human genomic sequence map was accomplished in three months at the cost of only 1.5 million USD. As technology continues to evolve, sequencing fluxes continue to grow. For example, Illumina's sequencer now generates billions of base pair data instead of millions per day. Also, the cost of sequencing genome data for an individual has been reduced from $10 million a decade ago to less than $1,000 dollars.



It's likely, with advances in technology that sequencing costs will continue to decline, and it is expected that the cost of a single genome will fall to $100 in the next 3-5 years. Nowadays, the development of the DTC (Direct-to-Consumer) industry makes the gene sequencing process relatively easy. Most of the genome-level tests can be performed with non-invasive sample collection methods, for example, extracting the DNA completely through oral cells (saliva). And this method can already meet the quality criterion of both first generation and second generation sequencing.



Although, the era of big data has come, the main focus is on internet behavior and media data. As an increasing number of wearable fitness devices enter the market and the Internet of things becomes popular, the focus on human big data analysis is expanding. As the fitness big data era evolves, human health and fitness data will grow exponentially. It is widely accepted that collecting fitness data is more valuable to human well-being than tracking more traditional health data.



http://fusionanalyticsworld.com/big-data-healthcare-market-reach-27-6bn-2021/ 11

II Industry Pain Points

2.1 Limited Ways to Access Genome and Healthcare Big Data

At present, there are still many problems in the field of human genome research. The rapid development of human genome technology requires huge amounts of genome sample data and vast health-related datasets; however, the available methods for integrating this data are very limited. Applications integrating large pools of genomic and health data are still limited to scientific research fields. Moreover, scientists and researchers need to spend a large portion of research funding to obtain data, such data sharing between subjects is quite limited, and the majority of this data is only from patients; there are very limited datasets from healthy people available.

2.2 Scattered Healthcare and Medical Data in Private Companies and Hospitals is Difficult to Review and Utilize

Although it is well known that genomic data can guide people towards adapting a more healthy lifestyle, there is no incentive for the healthy population to take the initiative to obtain their own genomic data. Meanwhile, health-related data such as exercise data, diet data, medical records, and so on, are often passively stored in various private enterprises, and there is almost no sharing of this valuable health data and thus no possibility of related research.

Source:	***Data Type***
Clinical	Electronic records, medical records, etc.
Health Check	Blood glucose, blood pressure, blood lipids, heart rate, height, weight, vision, etc.
Behavioral	Movement, use of electronic equipment, etc.
Genetic	Genome information, family history, etc.
Environmental	Genome information, family history, etc.
Habitual	Smoking, alcohol consumption, poor diet, etc.
Basics	Gender, age, etc.

2.3 The Storage of Genome and Health Big Data Presents Challenges

The existing storage and computation technologies are obviously challenged: an individual's original data volume can usually reach 30G to 100G, while the traditional way of calculating an individual genome pipeline requires about 2-3 days. Despite the current compression algorithms and parallel processing mechanisms for genomic data which can finish within hours, there is still a great need for massive storage and increased computational ability when dealing with a large number of individual genome samples. There is still lack of sample data of both human genome data and multi-dimension healthcare big data, since most research institutes focus only on the combination of genome and tumor data.

2.4 People Have Limited Access to Their Healthcare Data and Receive Little Value from Collecting It

People are constantly generating vast amounts of data, and health-related data is a crucial part of that data. The sharing, aggregation, analysis, and utilization of this data will have enormous added

value. However, the source of this data, and the data contributors, have never been involved in this process nor benefitted from the value of their own data. Therefore, the standardization and systematic storage of human genome and health big data is an urgent problem to be solved. In the future, there will be many applications that can locate and integrate this data to provide personalized health advice to users if the user has given consent. This process is of great importance to everyone's health and longevity and will aid in eliminating disease.

III Blockchain Technology and New Opportunities

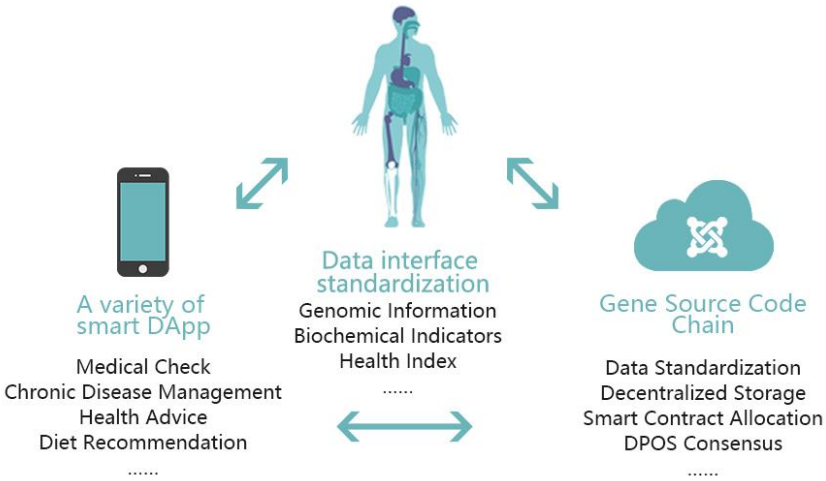
3.1 Blockchain Technology – the Disruptive Innovation

Blockchain technology is based on a decentralized network, combining encryption algorithms, data scheme innovation, and consensus mechanisms to ensure the consistency and continuity of distributed databases. The technology enables instant verification of data, traceability, anonymity, and failure shielding, thus creating an efficient, economical, and secure value-sharing system. Therefore, the equality, security and mass computing power of blockchain technology provides a completely new potential for solving the problem of storing and sharing human genomic data.

3.2 Blockchain Technology Drives Genomic Data Storage and Sharing

Encryption protocol and hashing mechanisms ensure that data encryption and transmission can be shielded from any interference by third parties and also ensure limited access to the user's vital health data. The smart contract mechanism of blockchain technology will provide individual users with the initiative to share their own genome and health data since blockchain technology can provide confidence and protect the fairness of the incentive mechanism. Decentralization of blockchain storage nodes will provide a solid platform for providing researchers with equal access to and sharing of genomic and healthcare data. Decentralized storage nodes with blockchain technology also enable the formation of distributed networks to provide vast amounts of storage for genetic and healthcare data, with the capability to design consensus mechanisms in a more secure manner.

Based on the advantages of blockchain technology above, the Gene Source Code Foundation will build a innovative distributed genetic storage system and set up an application platform based on genomic and healthcare data. We hope this system will center on establishing a comprehensive health data utilization economic ecosystem. The mechanism can encourage users to share genomic and health data within the ecosystem, and users can also get a range of health advice from the service providers. Those activities also provide researchers with adequate data for specific scientific research. The Gene Source Code System will build a distributed storage system of genetic and personal fitness data. Health and medical care service providers, including hospitals, chronic disease management agencies, and health consulting services can develop DApps on the Gene Source Code Chain System. The applications can provide personalized, precise service on the basis of user's permission, given genomic and health data.



IV The Gene Source Code Ecosystem

4.1 Components of the Gene Source Code System

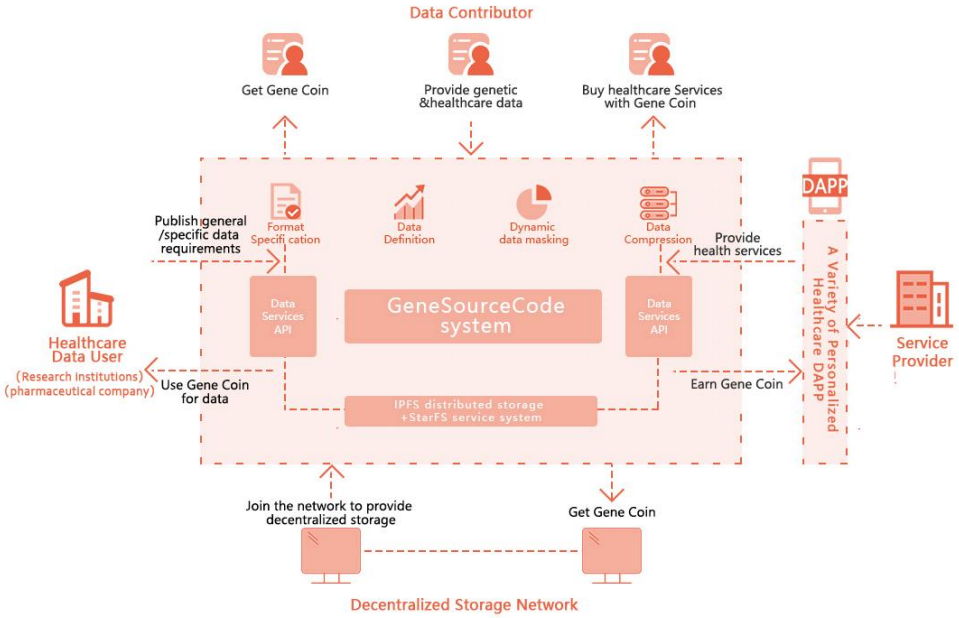
This ecosystem will consist mainly of the following parties:

(1) Data Contributors

(2) Data Users

(3) Storage Provider

(4) Service Providers



Gene Source Code Ecosystem Diagram

4.2 Interpretation of the Ecosystem

(1) Data Contributor:

Data contributors, who consistently contribute genomic and health data to ecosystems were previously limited to individual user contributors. Individuals need to be clear about their willingness to contribute health information continually and sign an electronic agreement. They can then gradually share information on exercise, diet, physical exams, and medical records by downloading the Personal Healthcare Companion app. Smart contracts will ensure automatic and timely data storage, validation and encryption in the distributed storage system, and the deposit of corresponding Gene Currency into the user wallet in a simple and convenient manner. In the initial phase of ecosystem construction, the Gene Source Code system will issue corresponding Gene tokens according to the quantity, quality and degree of confidence of the information provided by the user. Gene token distribution will work on the following principles:

Principle 1: The value of individual data decreases over time

The value of individual health data decreases as time goes by. The earlier contributors begin to contribute data, the more Gene tokens they receive.

Principle 2: The value of the data increases, as data dimensions increase

The value of individual healthcare data will increase as the amount of contributed data increases. There are several dimensions of healthcare data, such as genomic data, exercise data, health check records, medical record data, etc. There are also subdivisions in each dimension. The richness of these data dimensions is of great value for interdisciplinary research. As a result, the more individual healthcare data dimensions available, the more Gene Coins will be granted.

Principle 3: Continuous data with incremental value

It will be more rewarding to consistently provide personal health data. For example, in the case of exercise records, diet, etc., when a certain threshold of continuous contribution is exceeded, additional Gene token rewards will be awarded.

Principle 4: The addition of genome data and health data will result in greater rewards

Sharing of genome data and health check data will garner more Gene Coins. Genome data and health check record data are at the core of the electronic health record system, and the foundation of the Gene Source Code system. Contributions on these will generate additional Gene Coin rewards.

The above principle is implemented by smart contracts within the health assistant DApp of the Gene Source Code system which rewards contributers according to the rule of gene token value assessment. After the construction of the ecosystem, the Gene Source Code system will develop a fully functional SDK and encourage developers to develop numerous personalized health management DApps with gene token rewards as an intrasystem financial stimulus. The SDK provides a unified encryption and desensitization API to standardize the format of the accumulated genomic and fitness data.

At the same time, data contributors, through Gene Coin, can also purchase a variety of genomics and health data-centric health services provided by Service Providerss within the Gene Source Code system such as: Sequencing Services, Health Counseling Services, Genetic Disease Counseling Services, Sports Nutrition Recommendations, and even a wide variety of other types of services based on genomic data that may emerge in the future.

(2) Data Users

Above all, data users, such as university or private research labs will go through rigorous regulatory compliance checks before they can use any data. And they can also publish specific research projects using the Personal Healthcare Companion App to recruit specific data contributors and

provide targeted and enriched individual health data. At the same time the Gene Source Code system will be more open for data users. Some permissions such as calling service providers' API directly and applying bio-information data tools or artificial intelligence assisted mining modules on fitness data will be granted to users. Data users will be divided into business users and research users, the Gene Source Code system will provide discounted prices of Gene token prices to encourage the use of platform data in scientific research. Currently, the demand for health and genomics data from laboratories and R&D facilities at universities and enterprises is growing significantly, especially for the well-defined and continuous data sets.

(3) Storage Providers

The storage provider gets Gene Coin by providing storage space. In the early days, the Gene Coin obtained by the storage provider will be provided by the Gene Source Code system and later by the consumer who stores the data. Service providers, data contributors, data users all may need data storage services. All data will be stored using powerful encryption and standard data masking methods.

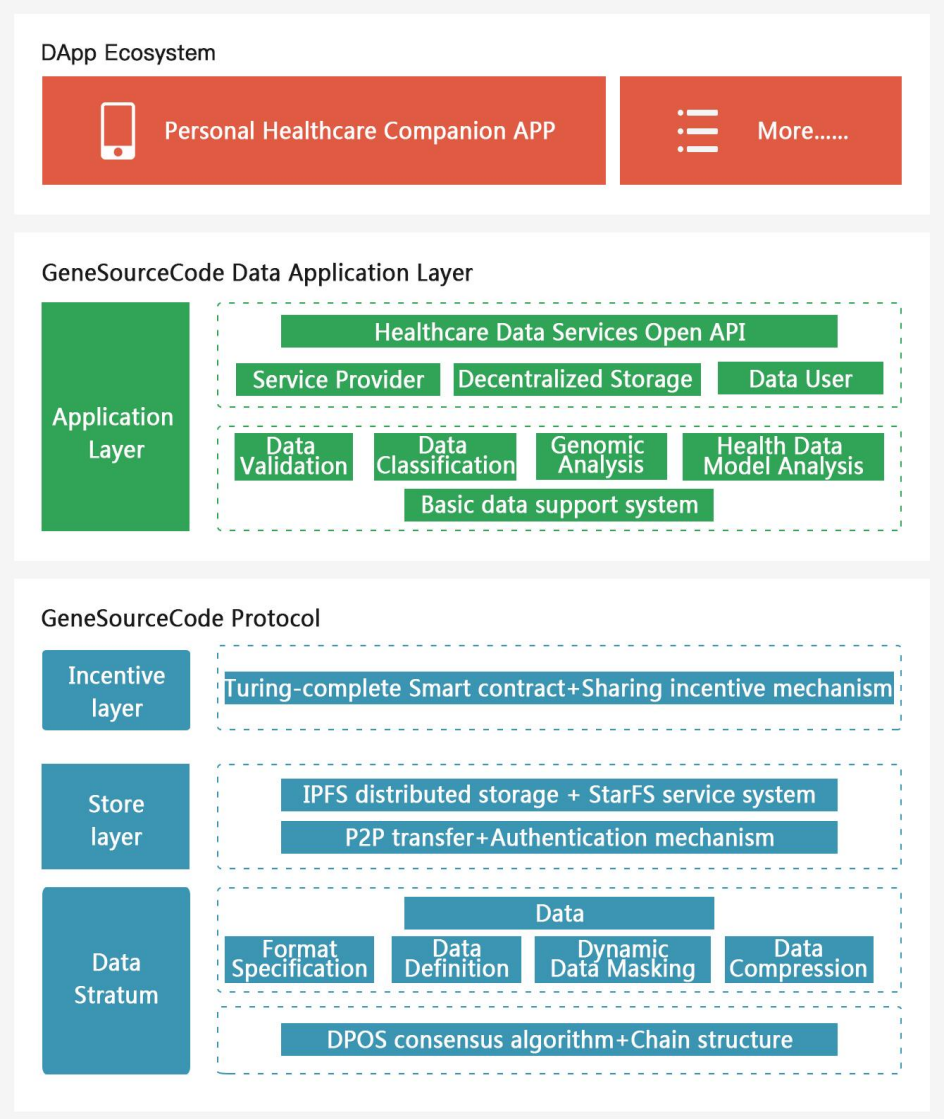
(4) Service Providers

Service providers include sequencing companies, data analysis companies, hospitals, health management companies etc. Opportunities to serve users on the Gene Source Code system will be screened by the Gene Source Code Foundation. The foundation's screening criteria will be announced to the public directly for public scrutiny. Users will be able to use Gene Coin to get a variety of health services from Service Providers.

V Blockchain Solutions

5.1 Gene Source Code Architecture

The Gene Source Code design uses a simplified, four-layer technological model, including: Data layer, Storage layer, Contract layer and Application layer. Each level consists of the following components:



Data layer: consists of standard blockchain chain structure.

Storage layer: mainly based on IPFS and StarFS decentralized storage system to solve storage allocation problems.

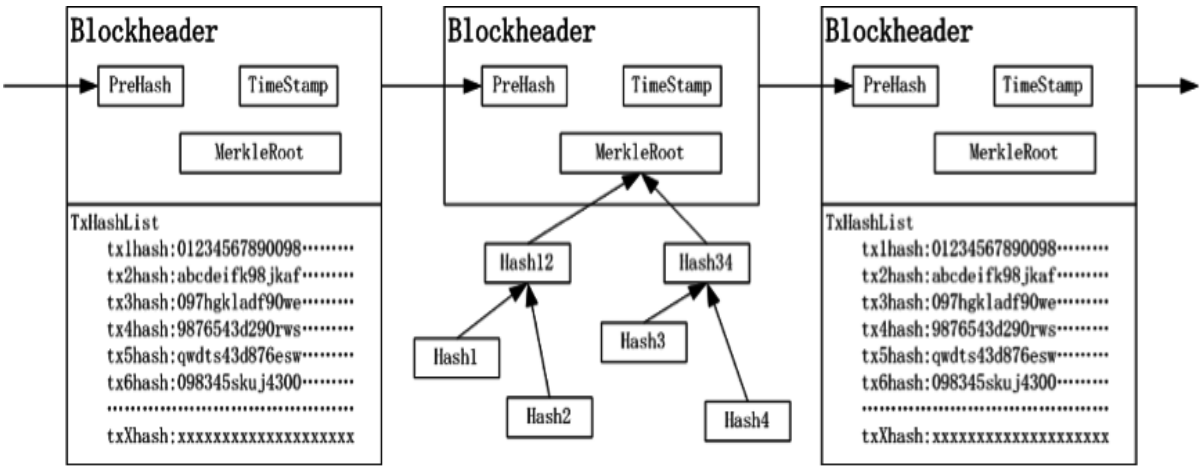
Incentives layer: Consolidation of Gene Coin based on an incentive smart contract.

Application layer: Gene Source Code data support system and data service support system.

5.2 Data Layer Architecture Schema

The data layer is a standard part of blockchain structure. The data layer consists of a sequence of time-stamped blocks, each of which consists of a block header and transaction data. The Data layer

uses a standard chain block chain structure, Merkle tree, hash function, asymmetric encryption, time stamping and other technologies. As shown below:



For the specific block design, we will use each block to save the user transaction records, timestamps and the last block hash abstract code as in the traditional scheme. We will be likely to choose IPFS positioning in the source code digest in which we record the hash values for user genes and health data, and Gene Coin transaction histories across the blockchain system to create a decentralized system for storage and trading genomics and health related big data. This system is currently targeted at gene and health big data collection and exchange. In the future it will be possible to apply it to all kinds of system data. In the design of consensus algorithms, the consensus algorithm of the distributed ledger decides who in the Gene Coin ecosystem will generate a block with its own recorded version of the ledger.

The Gene Source Code system will adopt a self-designed consensus algorithm called Athena to reach the over all consensus. The exact technical description of the Athena implement will be officially published in yellow paper edition 1.0.

For the market version of the DApp, which will be released in December, we will employ mixed consensus DPoS+PBFT used in side chain/Private Chain of EOS. DPoS is famous for being used by EOS whose efficient performance has made a strong impression on users. However, DPoS is far from satisfactory when resolving forks occurring on the block chain. When a malicious node creates several distinct versions of the latest block, several forks occur on the block chain, so that a consensus on which fork is genuine cannot be reached. The coming producers will produce blocks

along the fork which they choose until one fork gets the highest blockchain height. Then, all block producers will come to produce blocks for this longest fork. The duration of absence of consensus of DPoS is relatively long. Yet, DPoS+PBFT can solve this problem perfectly. DPoS+PBFT chooses several nodes as candidate block producers as DPoS does. Then, all stakeholders can vote among these candidates. Every vote is counted with respect to its voter's asset.

To deal with the malicious behaviors of Byzantine nodes which produce several distinct versions of blocks to cause divergence, DPoS+PBFT requires that all producers reach a consensus on the content of every block produced. PBFT requires that all producers propagate the hash of their respective version of the block among producers. The producers agree that the producer taking charge of the current time slot can only produce the version of block which gets over two thirds of votes. In this way, the problem of malicious forks is mitigated.

The GSC blockchain will integrate the most globally common genomic and fitness data format. We will make up missing formats and put all kinds of data and data types in the ecosystem of the Gene Source Code system. Meanwhile, the system design will also integrate a desensitization mechanism for genomic and fitness data and an advanced data compression mechanism.

5.3 Storage Layer Architecture Schema

The storage layer of the Gene Source Code is mainly based on the IPFS distributed storage system. Through a peer-to-peer distributed version of the file system, all connected through the same file system computing device. The use of content-based addresses, which is what users will search for, can make each storage node access faster and more secure.

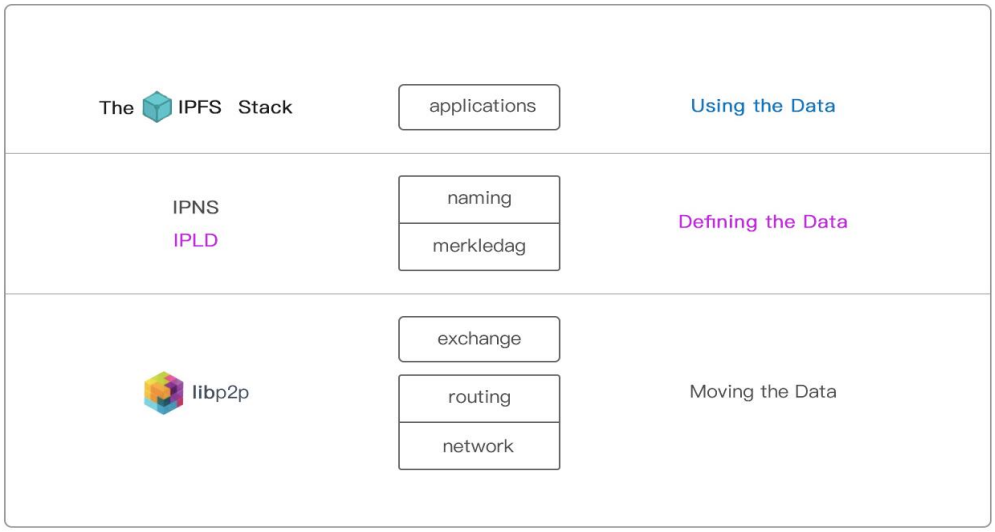
When gene and health data is added to the node of the Gene Source Code, it gets a new name. The name is actually a cryptographic hash, which is calculated from the contents of the file. This hash is guaranteed by encryption to always represent only the contents of the file. This Hash will be completely different even if you modify only one bit of data in the file.

However, as for the genomic data, especially the results of second generation sequencing where the same person's data Hash value could be slightly different, the Gene Source Code storage will be designed to simplify the genome-wide and second generation exome sequencing while adding a

segmented storage format and a personal identification storage mechanism, thus rendering genomic data viable for storage in the blockchain.

The distributed storage solution of the Gene Source Code serves well for both large and small files, and large files it automatically cuts it into smaller pieces so that its nodes can not only download files from one server like HTTP, but also from multiple servers simultaneously. At the same time, a hierarchical storage mechanism will be added for the transfer and storage of mass health data.

The Gene Source Code distributed storage solution can also solve data sharing and transmission problems. Users can quickly find the URL address of a block through the content digest code of a block using the distributed content summary table (DHT), and transmit the block data to the requesting client at a high speed by P2P technology.

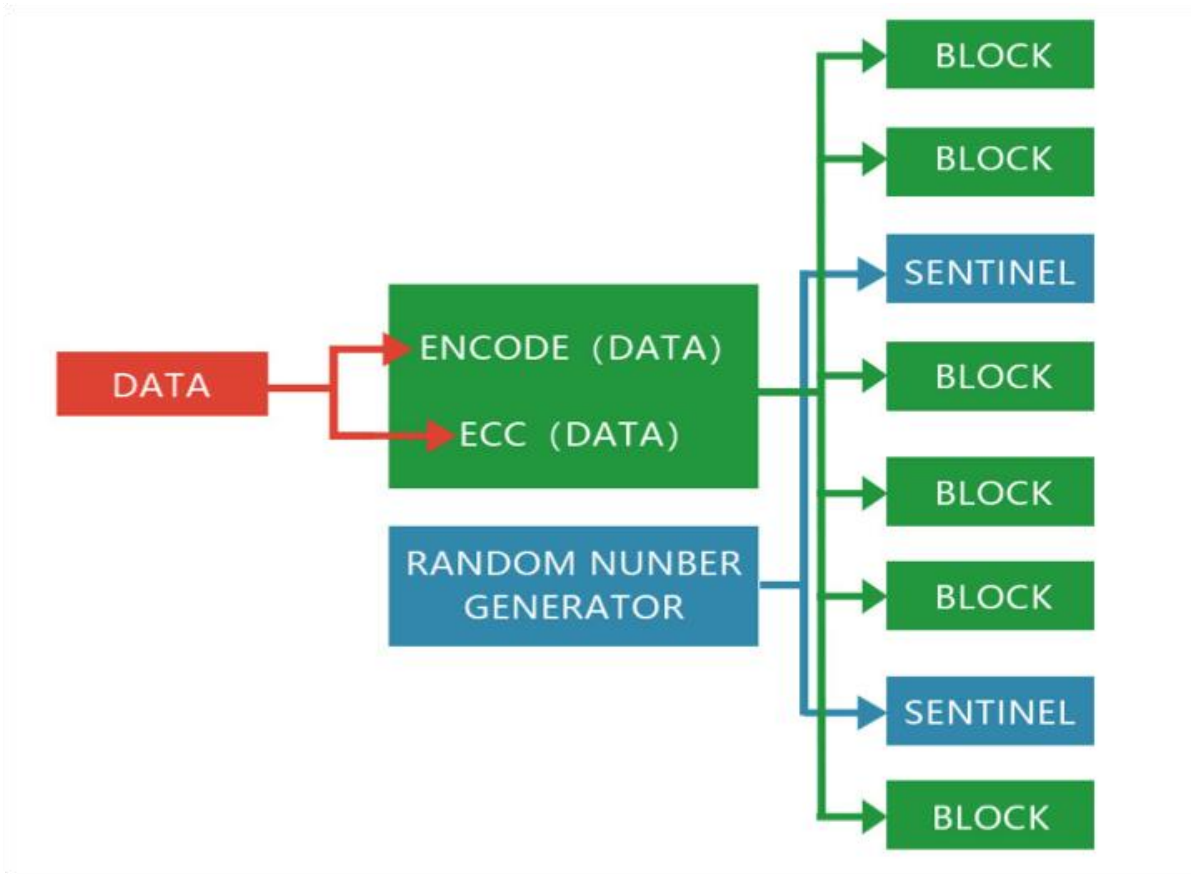


Moreover, we will also develop StarFS service system. This file system allows us to create unlimited expansion of virtual hard disk storage. No matter how big the data to be stored in the future, the network only needs to constantly join the new server running StarFS service to expand storage space. StarFS can also create redundant storage that minimizes data loss in the event of hardware storage failure.

Based on StarFS and IPFS, the Gene Source Code will integrate the world's most versatile genomic and health data formats, defining the missing types to integrate the range of data and definitions into a four-tier Gene Coin system. At the same time the Gene Source Code will also implement a genomics and health data masking mechanism, and the leading data compression mechanism will be integrated into the Gene Source Code data storage solution.

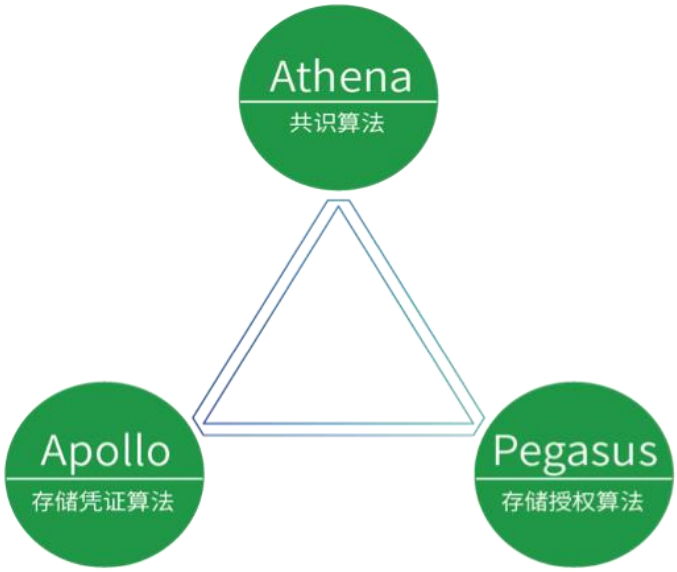
Although the fact that any data on the P2P network is stored by several nodes guarantees the probability of a missing piece of data is trivial, we still have to employ heavy proof schemes on a regular basis, being a provider of health and genomic data sharing service, to ensure the integrity of data and handle data offline problems instantly as they occur. The updates in the field of proof schemes happens continuously.

Storj and Genaro adopts a proof of retrievability algorithm which inserts several sentinel blocks into normal blocks and check out the integrity status of the sentinel blocks regularly to guess the integrity of the whole file. Sia uses a proof of storage algorithm to check out several blocks within a random sampled window every time to determine the condition of the original file. Filecoin relies on proof of replication to detect potential malicious nodes which may degrade the integrity of data with sybil identities, outsourcing and non-dedicated storage space.



Gene Source Code will investigate and weigh the pros and cons of currently popular proof of storage algorithms and independently develop a novel proof of storage algorithm—Apollo. The details of Apollo will be published in technical yellow paper edition 1.0. Sharing health and

genomic data on the Gene Source Code Chain employs a dynamic data access authorization algorithm—Pegasus. It will be built on several dynamic encryption algorithms including a proxy re-encryption framework to cut the overhead of repeated encryptions. Concurrently, the encrypted data will carry identity-related digital watermarking to serve as evidence for prosecuting illegal off-line distributers of the data. The details of Pegasus will be published in yellow paper edition 1.0.



5.4 Incentive Layer Architecture Schema

The incentive layer is composed of a data sharing incentive mechanism based on smart contracts to solve the problem of Gene Coin distribution. The Gene Source Code Foundation will determine the economic balance mechanism through assessment and research to finalize Gene Coin incentives for data contributors and storage providers while ensuring that data users and Service Providers have sufficient indirect incentives, ultimately incorporated into the smart contract system, to achieve the overall incentive and benefit sharing objectives.

5.5 Application Layer Architecture Schema

The Gene Source Code application layer is composed of the Gene Source Code basic data infrastructure system and a Service Providers open interface system. The Gene Source Code basic data support system includes the data validation system(DVS), the data classification system(DCS), the genome analysis system(GAS), and the health data model analysis system(HDMA). These four

subsystems will maintain a solid data application infrastructure. The Gene Source Code will also provide a data open interface system for data Service Providers and data users to get consolidated system data for various services in the future.

5.6 DApp - Personal Healthcare Companion as A Proof of Concept

DApp is short for decentralized application. A traditional web application client-server model is composed of frontend and server. DApp, on the other hand, is a composition of frontend and smart contracts. Smart contracts run on all nodes in a P2P network rather than running on a single server. They exchange data with the block chain but through a centralized database server. All these characteristics defines a DApp. The Gene Source Code chain system will provide a platform running DApps and a toolset for DApp development. Developers can run their own DApp on the Gene Source Code chain system.

A core module of the Gene Source Code system is to provide DApp, a personal healthcare companion for the public. This will be the world's first comprehensive health management DApp based on individual genetic data and multidimensional health data. This App will be available to the public as a personal healthcare companion with gamification functions. The personal healthcare companion will be released in two versions. The first is a promotion version which doesn’t include main chain and peripheral functions. The rest of the functions will be implemented after the main chain is online.

In the meantime, the Gene Source Code foundation will cooperate with health service providers to promote the development and propagation of their DApp. After the completion of the Gene Source Code system and the Personal Healthcare Companion App, the Foundation's next goal is to establish a common distributed system—a common blockchain system for data storage and exchange. Enterprises and some individual users will be able to set up and establish various subdivided data exchange DApps based on this universal distributed system.

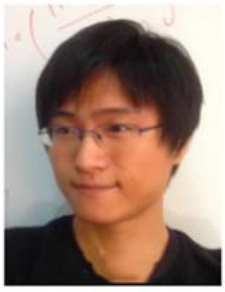
VI Why We Can Do Better

6.1 The Gene Source Code Team and Partners Have Rich BT + IT International Experience

The Gene Source Code team has multidisciplinary experience in block chain technology, bioinformatics, IT architecture, and network security, and can support the research and development of Gene Coin. Asian business partners of GSC's Asian business partner, StarshipGene, has accumulated very substantial genetic testing and health management industry experience. Its major business domain is DTC (Direct to Consumer) which, itself, is a relatively new application breakthrough in the genetic testing industry. The DTC business model, has accumulated a large number of gene testing samples and sequencing and raw data analysis experience.

6.3 The Gene Source Code Project is Endorsed by Forward-Looking Scientists in Well Known Research Institutes

"I'm very excited to hear about GSC's innovation. The cross-border usage of blockchain technology could revolutionize the health industry!"



Xu Luping

CNMM PI in Tsinghua University, Master in Biology and Nanotech, Phd in condensed matter physics in Peking University and ENS, Paris.

“I have heard many innovative ideas, but this one combining blockchain and genomic data storage is ingenious.”

Aard Groen

Dean / Professor of innovative entrepreneurship at University of Groningen and at University of Twente / NIKOS, Netherlands

6.4 Venture Capital Partners

NODE CAPITAL 节点资本

GENESIS 创世资本

StarFund

双花资本

6.5 Eco Partners

华大基因 BGI

星舰基因 StarshipGene

EMTD 怡美通德 Emei Tongde

TIB BEIJING TAIPU-SHUNKANG INSTITUTE FOR LABORATORY MEDICINE

HaploX 海普洛斯

美吉生物 Majorbio

悦微基因 icroread

COLDLAR

阿里云 aliyun.com

京东云 JCloud.com

6.6 Philosophy and Goals of the Gene Source Code Foundation

(1) The Gene Source Code Foundation Philosophy

In the era of exponential development of science and technology, health right must be in the hands of individuals.

(2) Gene Source Code Foundation's Goals

Stages	**Target Names**	**Description**	**Estimated Completion**
First stage	Mars	Personal Healthcare Companion Promotion Edition 0.5 Release	Dec. 2018
Second stage	Solar System	Gene Source Code base system completed	May 2020
Third phase	Milky Way	Universal storage mechanism release	Oct. 2021
Fourth stage	Laniakea	Updated market mechanism for the universal distributed data system	May 2022

VII Gene Token Placement

Total amount of Gene tokens: 2 billion

Private placement	20%
Mining to earn	40%
Incentives for core team and global community ramp up	15%
Foundation development and user community ramp up	15%
Business development	10%

Private placement will target less than 190 investors and aims to raise 20,000 ETH, or the equivalent in BTC

Lock-up Agreement:

Core Gene Source Code team => 4 years, 25% released each year

Business Development => 3 months

VIII Funding Usage

Due to local policy restriction for public funding, we only go through private placement for funding, and haven't got enough fund for prior usage plan , so we adjust the Funding Usage to ensure the DApp development and Mainnet research.

IX Gene Source Code Foundation Governance Mechanism

The Gene Source Code governance mechanism includes a three-tier governance structure: the Holders' Assembly, the Autonomous Committee, and the Executive Committee. The Executive Committee consists of the Technical Committee, the Data Management Committee, the Financial Affairs Committee, the Legal Affairs and Risk Management Committee, and the Marketing and Public Relations Committee. The various committees' responsibilities are divided as follows:

(1) The Technical Committee is responsible for GSC technology management; specific work includes open source code management, code development, code modification, code testing, and code and agenda review.
(2) The Data Management Committee is responsible for data storage standards selection and development, data storage reliability, redundancy principles, and other data management principles.
(3) The Financial Affairs Committee oversees the Ethereum tokens and Gene Tokens including the usage and audit of expenses, routine operations, arrangement of staff expenses, etc.
(4) The Legal and Risk Control Committee is responsible for the registration of domestic and foreign companies, reviewing various agreements and giving professional advice on legal matters.
(5) The Marketing and Public Relations Committee mainly serves the blockchain community and is responsible for technical promotion, product promotion, business cooperation and publicity.

X Risks and Disclaimer

As technology progresses rapidly, the technical solutions described in this white paper may be subject to change. The final implementation will prevail.

Blockchain and genomics are all cutting-edge, emerging technologies. Gene tokens work based on innovative blockchain technology and are the first of its kind in the world in terms of economic incentives and storage mechanisms. Participate with caution, note all technical or model uncertainties, and assess your own risk.

This document is only for informational purposes and does not constitute any representation regarding the sale or purchase of Gene tokens nor the invitation to buy products. This documents does not represent a contractual commitment. The above information or analysis does not constitute an investment decision. This document does not constitute any investment advice or intentions to invest. Once involved, the related parties are agreeing they are aware of and accept the risks associated with the above-mentioned products and are willing to bear any and all consequences. The Gene Source Code Foundation expressly disclaims any losses arising from participation in Gene Source Code projects and secondary market transactions.

XI Customer Service, Contact and Support

Web：http://www.gscchain.org

Twitter：https://twitter.com/gsc_chain

Telegram：https://t.me/genesourcecode

Github：https://github.com/genesourcecodechain

Mediam：https://medium.com/@seq

Email: Future@gscchain.org

Wechat：SSMOM3000