Gene Source Code

Whitepaper 2.4

2018.08

Abstract

The rapid growing comprehension and insight to genomic and human health data make it possible to prolong human life expectancy significantly in the next 20 years. The large scale implementation of personalized healthcare technology will be the main driver of this change. Personalized health care services will rapidly become popular with the dramatic drop in the price of genome sequencing services in 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 Foundation is committed to build the world's first storage and decentralized application platform for genomic and health data, the Gene Source Code System, with the expectation that a human healthcare data ecosystem will be built around this and create one unique Economic Mechanism: Encourage users to share genomic and health data within the ecosystem, provide researchers with adequate data support and provide data users with a friendly interface.

The Gene Source Chain System will create the world's first distributed storage network blockchain system that provides genetic data and individual health data integrated interface. Health and medical service providers, including hospitals, chronic disease management agencies, and health consulting service providers, can use the gene source chain system to build DApps. These applications will provide users with personalized, accurate health services based on genome and health data, subject to user consent.

The Gene Source Chain team will also create a DApp of smart health management assistant, which inspires users to share data and obtain health services, as a prove of concept application. For an individual user, he/she can receive Gene currency rewards while sharing health data, and can also obtain accurate health advice when his/her data is rich enough.

The Gene Source Code Foundation believes that the human right to have a healthy life should be absolutely equal even in an era of exponential technology growth. This right should not be affected by the amount of wealth, simply because each one is already contributing to the health of all humankind. Gene Source Code system will be based on trustful blockchain technology and data encryption/masking technology, so that each individual has ownership of health related data and the benefits within.

Table of Contents

I   Preface and Background

1.1 Technology Exponential Growth and Blockchain Technology Outbreak

The exponential growth of technology has profoundly and rapidly affected everyone's lives in all fields, including the impact of skyrocketing growth in Blockchain technology. It has been nine years since Satoshi Nakamoto invented bitcoin in 2009. During this period, Blockchain technology has been greatly developed and consolidated. Digital cryptocurrency represented by bitcoin began to be gradually understood, accepted and even sought by the people. Bitcoin that decentralizes digital virtual assets is more commonly known as digital gold, and its usage has also gradually expanded from geeks to the general public. At the same time, there are many more exciting developments in the area of Blockchain cryptocurrency industry, such as the Ethereum Project, created by Vitalik Buterin and Gavin Wood in 2013, with great success. Due to the ease of use of the Ethereum design concept, more and more decentralized applications, DApp (Distributed Application), are built on the technology platform of Ethereum. In addition, there are a series of featured Blockchain projects such as Zcash, Dash, EOS and so on. The rapid development and evolution of blockchain technologies will embrace the Blockchain application in various industries.

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 also posed great opportunities and challenges to mankind. In the foreseeable future, mankind will benefit from sharing big data on genes and human healthcare, will greatly extend life spans and eliminate vast majority of diseases.

As the price of genomic sequencing data has dropped rapidly, scientists and citizen scientists will be free to access vast amounts of genomic and human health data to

study and overcome specific medical and healthcare problems. As the price for gene sequencing is descending, scientists and citizen scientists can afford to access vast amounts of genetic and body fitness data for academic or medical purpose. Besides, research and cross study on DNA of other species on earth will be beneficial as well.. These studies will make great progress for mankind in the fields of medicine, materials and energy.

1.3 Application of genetic data

Genes are substances that transmit genetic information. It is the source of birth, aging, illness and death. Gene data contain a large number of health information and are closely related to the health of individuals. At present, genetic data mainly includes the following aspects:

(1)Assisted clinical diagnosis:

Many diseases show resemblance to others which lead to difficulties in discriminating among them in clinical practice and confusions in diagnosis. If genetic testing is used to find the cause of the disease from the genetic level, it can help clinicians differentiate among diseases and even correct mistakes occurred in clinical diagnosis.

(2) Screening carriers:

One of the most common screenings is Down's syndrome screening. Traditional screening for Down's syndrome is performed on serum. The detection rate is only between 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 certain family history of single-gene disorder (especially recessive disorder) can promptly detect the virulence genes carriers in the family and analyze the pathogenic risk among offsprings. The

screening results provide family members with informative suggestions on how to prevent the defective genes from affecting the next generation.

(3) Guided treatment:

The amount of prescription is decided after a number of extensive testing done on groups. The doses decided this way is appropriate for most of people, but not for everyone. Some prescriptions of normal dosage may cause fatal risk because of the genetic difference among people. The medicine which is supposed to cure disease, in turn, put a strain on health. This phenomenon is called adverse drug reactions (ADR).

For example, the drug warfarin is an anticoagulant, a drug that prevents blood clotting. Patients taking this medicine can greatly reduce the risk of thrombosis. However, if the anticoagulant is overdosed, the blood can hardly coagulate and continuous bleeding may threat parent's life. In our body there is an enzyme called CYP2C9, which can metabolize this anticoagulant, break it down into small molecules, and lose its anticoagulant effect. Under normal circumstances, warfarin is metabolized after it completes its therapeutic action, and does no harm to human body. However, if a person has mutations in CYP2C9, his or her metabolic function slows down and become a poor metabolizer, warfarin is metabolized slowly and accumulates in the body, and eventually cause continuous bleeding. The role of genetic testing can be well demonstrated in this case: It can firstly determine whether someone's CYP2C9 has a mutation and determine which metabolic type he belongs to, and then determine the dose of the drug based on the type of metabolism. If it is strongly metabolized, then increase the dose properly; if it is a weak metabolizer, then reduce the dose and pay attention to the drug in the blood. This not only guarantees the therapeutic effect of the drug, but also causes no ADR.

There are so many examples in which the genetic type of metabolism plays an important role in determining the dosage of a drug. Various gene involves in the metabolism of drugs for depression, diabetes, asthma, and osteoporosis. Genetic

testing can guarantees the right dosage of drug. However, personal genetic character is seldom considered in prescription, so that prescribed medicines causes adverse reactions to certain patients from time to time.

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

Modern medicine believes that disease is a joint effect of genes and external factors. Almost all diseases are related to genes. In some cases, genes mutation leads to dysfunction of the human body and deviation of physiological functions from normal conditions, thereby giving opportunities to diseases; at the same time, external factors (environment, Climate, diet, radiation, exercise, etc.) amplify the undesired results caused by mutations in genes, and the combination of external and internal causes disease.

The colon cancer is given as an example. The metabolism of epithelial cells in the colon is controlled by a gene called APC. If a person's APC gene has a problem that it cannot control the normal metabolism of epithelial cells, the uncontrolled epithelial cells may undergo excessive proliferation. It is the initial development of colon cancer. The DCC and P53 genes in the human body inhibit the canceration of these cells. If DCC and P53 genes are also mutated and cannot effectively prevent the cancerous transformation of these cells, the body will have to face the threat of colon cancer. Therefore, we believe that the defects in APC, DCC and P53 gene are closely related to the occurrence of colon cancer. People with such defective genes gets a imminent risk of colon cancer. Through testing, before the proliferation of subepithelial neoplasia has not really evolved into colon cancer, the signs of colon cancer can be discovered. The disease can be prevented beforehand through measures such as improving the living habits, changing the living environment, taking scientific health care and avoiding various adverse effects.

1.4 The Features of Current Genome Sequencing and Health Big Data

DNA sequencing methods have been well developed, from manual to capillary electrophoresis. The year 2001 is a milestone: through the efforts of many countries, the HGP, human genome project has been completed, which took 13 years and cost 437 million US dollars. 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. By 2007, the birth of the first complete human genomic sequence map took only 1.5 million U.S. dollars and three months. As technology continues to evolve, sequencing fluxes continue to grow. For example, Illumina's sequencer generates base pair data from one million per day to billions. And the cost of sequencing genome data reduced to less than $ 1,000 dollars. It can be predicted that with advances in technology, sequencing costs will continue to decline, and we expect that the cost of a single genome will fall to $ 100 in 3-5 years.
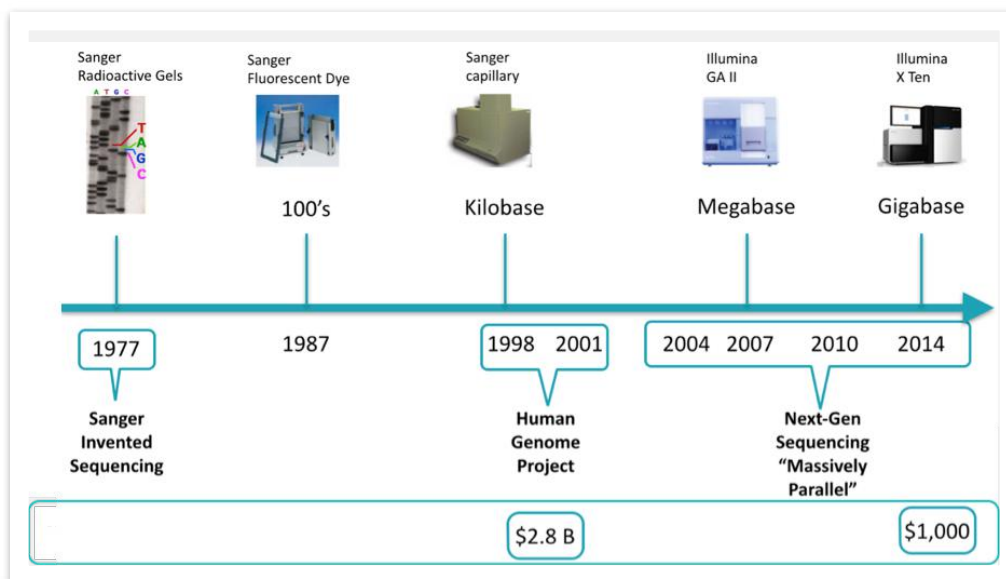
Nowadays, the development of the DTC (Direct to Consumer) industry makes the gene sequencing process relatively easy. Most of the genome-level tests can be performed by noninvasively sample collection methods, for example, extracting the DNA completely through oral cells (saliva). And it can already meet the quality criterion of the first generation and second generation sequencing.

COLLECTION AND DETECTION PROCESS

01 Order
Your saliva collection kit typically arrives within 2 to 3 days. Express shipping is available.

02 Spit
Follow kit instructions to spit in the tube provided — all from home. Register your saliva collection tube using the barcode so we know it belongs to you, and mail it back to our lab in the pre-paid package.

03 Discover
In approximately 2-4 weeks, we will send you an email to let you know your reports are ready in your online account. Log in and start discovering what your DNA says about you.

Although, the era of big data has came, the main focus is put on internet behavior and media data. As an increasing number of wearable fitness devices enter the market and the Internet of things become popular, the focus on human big data analysis is aroused. When the fitness big data era come, the human health and fitness data will grow exponentially. It is widely accepted that collecting fitness data is more valuable to human than traditional concerns can be.

Big Data Healthcare Market Size ($bn)

- 2016: 10.1
- 2017: 12.2
- 2018: 14.9
- 2019: 18.2
- 2020: 22.3
- 2021: 27.6

22.3%

http://fusionanalyticsworld.com/big-data-healthcare-market-reach-27-6bn-2021/
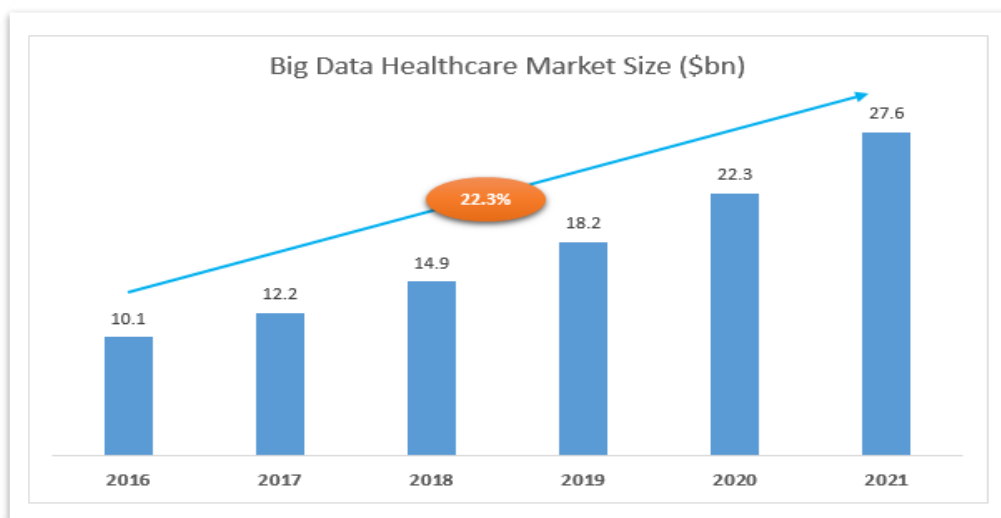
II   Industry Pain Points

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 health related dataset, however, the available methods to integrate those data are very limited. Large scale genomic and health data integrated applications are still limited to scientific research fields. Moreover, scientists and researchers need to pay a big part of research funding to obtain data; these data sharing between subjects is quite limited; the majority of these data are only from the patients, just a few datasets from healthy people.

2.2 Scattered Healthcare and Medical Data in Private Companies and Hospitals is Difficult to Cross-examine and Utilize

Although it is well-known that genomic data can guide people's healthy lifestyles, 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 record data and so on, are often passively stored in various private enterprises, and there is almost no sharing of these valuable health data, and thus no possibility of related research.

Source:	Data Type
Clinical	Electrical record, Med record, etc.
Health Check	Blood glucose, blood pressure, blood lipids, heart rate, height, weight, vision, etc.
Behavioral	Movement steps, use of electronic equipment, etc.
Genetic	Genome information, family history, etc.
Environmental	Type of work, location of residence, passive smoking, etc.
Habitual	Smoking, alcohol consumption, diet ingredients, etc.
Basics	Gender, age, etc.

2.3 The Systematic Storage of Genome and Health Big Data Also Presents Challenges

Due to the enormous volume of personal genome data, the existing storage and computation technologies are obviously challenged: the individual's original data volume can usually reach 30G to 100G, while the traditional way of calculating the personal genome pipeline is about 2-3 days. Despite the current compression algorithms and parallel processing mechanisms for genomic data can accomplish this within hours, there is still a great need for massive storage and computational effort while having large numbers of individual genome samples.

There are still lack of sample data in combination of human genome data and multi-dimension healthcare big data, since the current research institutes focus only on the combination of genome and tumor data.

2.4 People Have Poor Access to Their Healthcare Data and Have Not Been Able to Obtain the Value of Their Own Data Collection

People are constantly generating vast amount of data, of which health related data is a crucial part, and the sharing, aggregation, analysis, utilization and enormous value of these data will have enormous added value, however the source of these data, the data contributors were never involved in this process nor shared 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 call and integrate these data to provide personalized health advice to users on the premise of user authorization. This process is of great importance to the health of all human beings' longevity and the elimination of the disease.

III   Blockchain Technology and New Opportunities

3.1 Blockchain Technology – the Disruptive Innovation

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

3.2 Blockchain Technology Drives Genomic Data Storage and Sharing
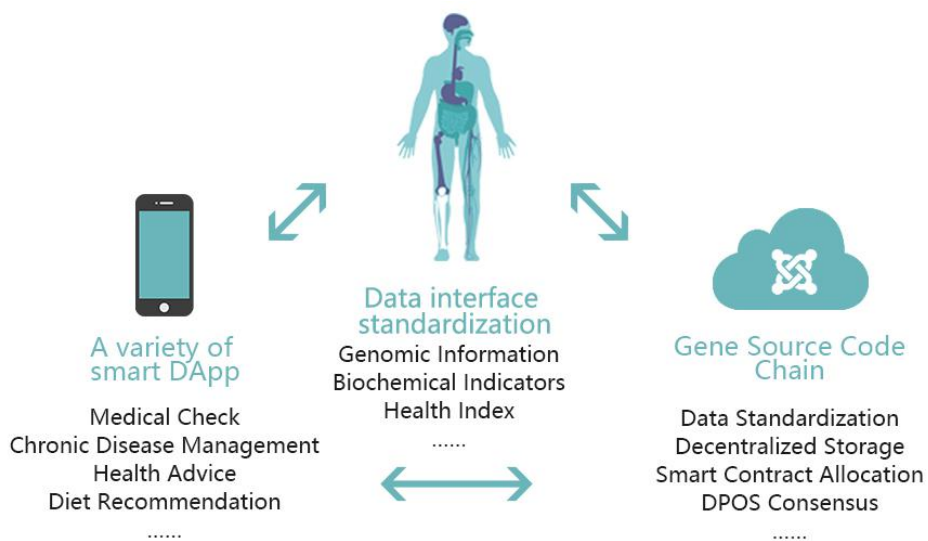
Encryption protocol and hashing mechanism ensure data encryption and transmission can be shield from the interference of third parties in any way, and also ensure limited access to the user's vital health data.The smart contract mechanism of blockchain technology will provide individual users with the 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 researchers to have equal sharing of genomic and healthcare data. Decentralized storage nodes with blockchain technology also enable to form a distributed networks to provide vast amounts of storage for genetic and healthcare data, with the capability to design consensus mechanism in a more secure manner.

Based on the advantages of blockchain technology above, Gene Source Code Foundation will build the world's first 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 large scale health data utilization economic ecosystem. The

mechanism can encourage users to share genomic and health data within the ecosystem, and the users can also get various health advices from the service providers. Those activities also provide researchers with adequate data for specific scientific research.

Gene Source Code System will build the first distributed storage system of gene and personal fitness data. Health and medical care service provider, including hospitals, chronical disease management agency, health consulting services can develop DApp on gene sourcecode chain system. The applications can provide personalized precise service on the premise of user's permission, given genomic and health data.

A variety of smart DApp

Medical Check
Chronic Disease Management
Health Advice
Diet Recommendation
......

Data interface standardization

Genomic Information
Biochemical Indicators
Health Index
......

Gene Source Code Chain

Data Standardization
Decentralized Storage
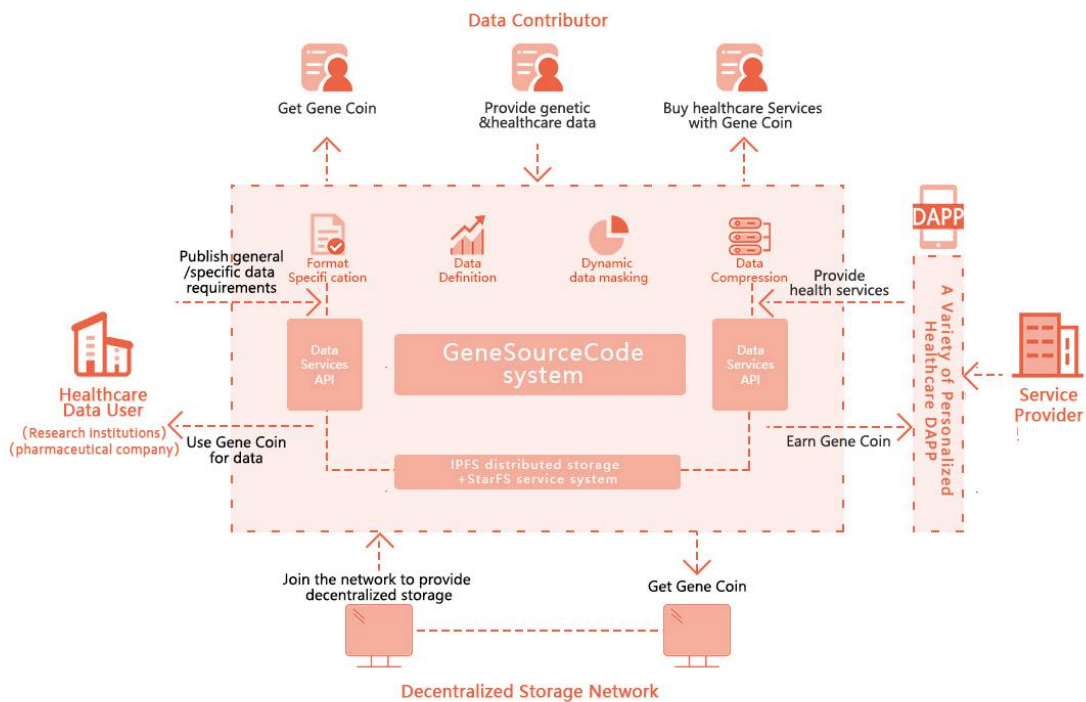Smart Contract Allocation
DPOS Consensus
......

IV The Ecosystem of Gene Source Code System

4.1 The Constituent Units of Gene Source Code System

Within this ecosystem, it will consist mainly of the following abstract parties

(1) Data Contributors

(2) Data User

(3) Storage Provider

(4) Service Provider

Gene Source Code Ecosystem Diagram

4.2 Interpretation of the Ecosystem

(1) Data Contributor

Data contributors, who consistently to 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 examination and medical records by downloading the Personal Healthcare Companion App. Smart contracts will ensure the data store automatically and timely, also validated and encrypted in the distributed storage system, and the 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 the corresponding Gene tokens according to the quantity, quality and degree of confidence of the information provided by the user. Gene token distribution will apply the following principles:

Principle 1: the value of individual data decreases as time goes by
The value of individual health data decreases as time goes by. The earlier contributors continue to contribute data, the more Gene tokens they get.

Principle 2: the value of the data increases, as data dimension increases
The value of individual healthcare data will increase as the dimension of contribution data increases. There are several dimensions of healthcare data, such as genomic data, exercise data, health check record, 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 are available, the more Gene coin will be granted.

Principle 3: Continuous data with incremental value
It will be more rewarding to consistently providing personal health data. For example, exercise records, diet, etc., when a certain threshold is exceeded, such as continuous provision will apply additional Gene token rewards.

Principle 4: genome data, health check data apply more rewards.
Sharing of genome data and health check data will get more Gene coin. Genome data, 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 contributer according to the rule of gene token value assessment..

After the construction of the ecosystem, gene source code system will develop a fully functional SDK and encourage developers to develop numerous personalized health management DApp with gene token rewards as a intrasystem financial stimulus. The SDK provides an 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 providers 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/Corporate research lab will go through rigorous regulatory compliance check 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 Gene Source Code system will be more open for data users. Some permissions such as calling service provider's API directly and applying bioinformation data tool or artificial intelligent assistanted mining modules on fitness data will be granted to users.

Data users will be divided into business users and research users, Gene Source Code system will provide discounted prices of Gene coin to encourage the use platform data in scientific research. Currently, the demand for health and genomics data from laboratories and R & D facilities at universities and enterprises grow significantly, especially for the well-defined and continuous data sets.

(3) Storage Provider

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

(4) Service Provider

Service provider includes sequencing companies, data analysis companies, hospitals, health management companies etc. Opportunities to serve users on the Gene Source Code system will be screened by 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 Schema

5.1 Gene Source Code Architecture Design

The Gene Source Code uses a simplified four-layer technological model, including: Data layer, Storage layer, Contract layer and Application layer.

Each level consists of the following components:
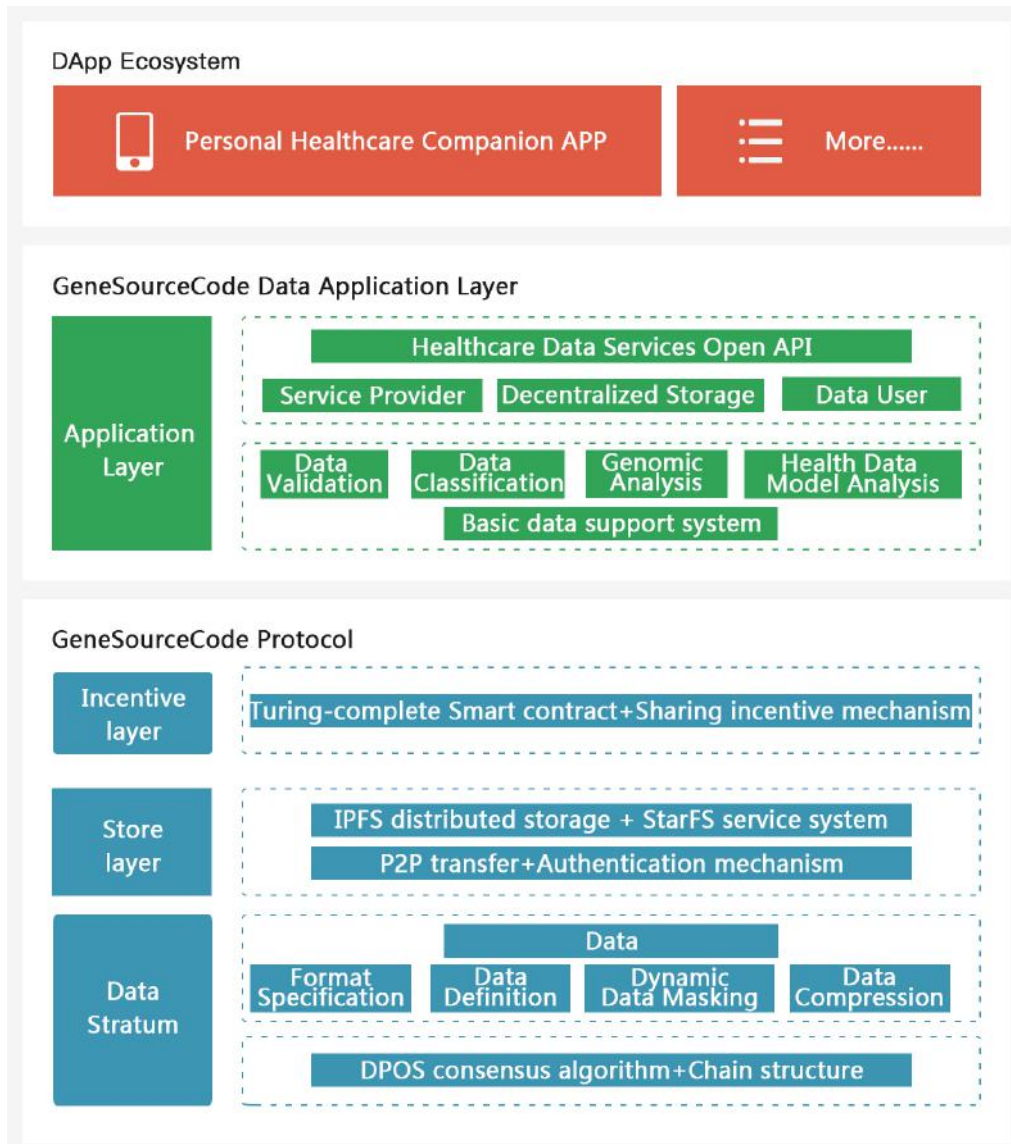
Data layer: consists of standard blockchain chain structure.

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

Incentives layer: Consolidation of Gene coin based on 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. Data layer using a standard chain block chain structure, Merkle tree, hash function, asymmetric encryption, time stamping and other technologies.
As shown below:
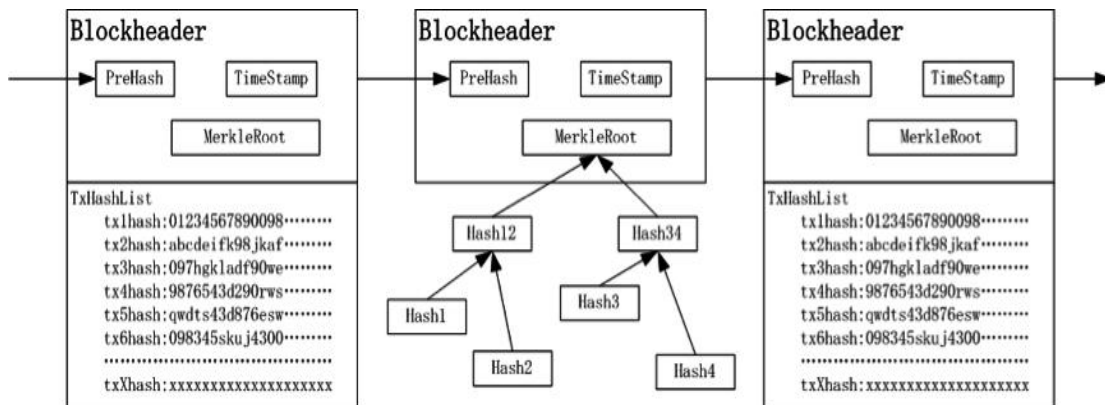
The specific block design, we will use each block to save the user transaction records, timestamps and the last block hash abstract code as the traditional scheme, we will be likely to choose IPFS positioning the source code digest, in which we record the hash values for user genes and health data, and Gene coin transaction histories across the blockchain system to create a decentralized system for storage and trading genomics and healthy big data. This system is currently targeted at gene and health big data collection and exchange, the future will be able to apply to all kinds of system data.

In the design of consensus algorithms, the consensus algorithm of distributed ledger decides who in the Gene coin ecosystem will generate a block with its own recorded version of the ledger.

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

For market version of Dapp which will be released in Decemenber, we will employ mixed consensus DPoS+PBFT used in side chain of EOS. DPoS is famous for being used by EOS whose efficient performance makes a strong impression on users. However, DPoS is far from satisfactory when resolving forks occurring on block chain. When a malicious node creates several distinct versions of the latest block, several forks occur on 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. While, DPoS+PBFT can solve this problem perfectly.

DPoS+PBFT chooses several nodes as candidate block producers as DPoS does. Then, all stake holders can vote among these candidates. Every vote is counted with respect of its voter's asset.

To deal with the malicious behaviors of Byzantine nodes who produce several distinct versions of blocks to cause divergence, DPoS+PBFT requires all producers must reach a consensus on the content of every block produced. PBFT regulates that all producers propagate the hash of their respective version of block among producers. The producers agree on 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 are mitigated.

The blockchain of gene source code 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 gene source code system. Meanwhile, the system design will also integrate the desensitization mechanism for genomic and fitness data and 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 peer-to-peer distributed version of the file system, all with the same file system computing device connected together. The use of content-based address, 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. 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, so that 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 exome of the second generation of sequencing and adding segmented storage format and personal identification storage mechanism, so that the genomic data can be viable to store in the blockchain.

The distributed storage solution of Gene Source Code serves well for both large and small file, and for 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.

Gene Source Code distributed storage solution, can also solve the data sharing and transmission problems. Users can quickly find the URL address of a block through the content digest code of a block through the distributed content summary table (DHT), and transmit the block data to the requesting client at a high speed by P2P technology.
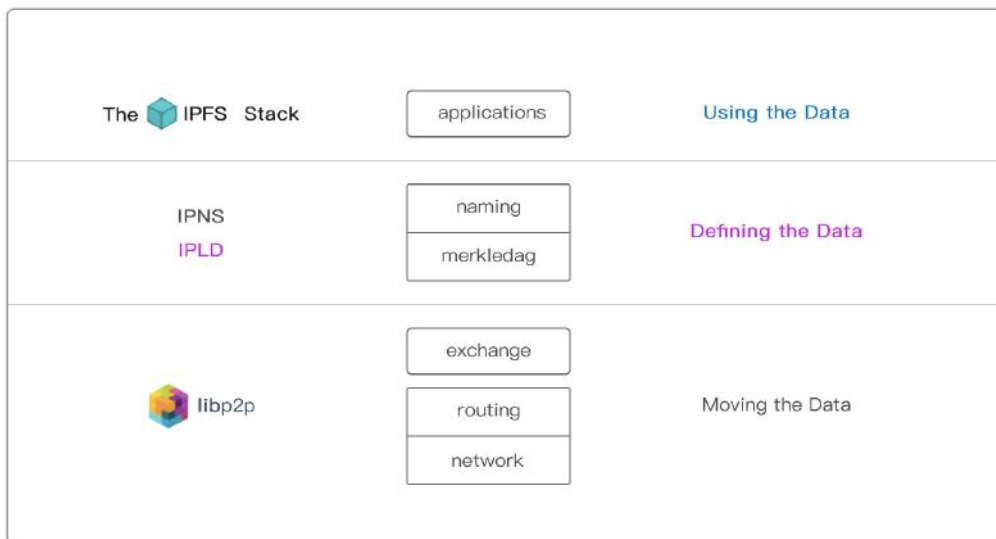
Moreover, we will also develop StarFS service system, this file system allows us to create unlimited expansion of virtual hard disk storage. No matter how big the data to be stored in the future, the network only needs to constantly join the new server running StarFS service to expand storage space. 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 and defines the missing types to integrate types of data and definitions into a four-tier Gene Coin system. At the same time Gene Source Code will also implement genomics and health data masking mechanism, and the leading data compression mechanism integrated into the Gene Source Code data storage solution.

Although the truth that any data on 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 make sure the integrity of data and handle the data offline problem at the moment right after its occurrence. The updates in the field of proof schemes happens continuesly. Storj and Genaro adopts proof of retrievability algorithm, which inserts several sentinel blocks into normal blocks and check out the integrity status of the sentinel blocks later regularly to guess the integrity of the whole file. Sia uses the proof of storage algorithm to check out several blocks within a random sampled window everytime to make sure the condition of the original file. Filecoin relies on the proof of replication to detect potential malicious nodes which may deteriorate the integrity of data with sybil identities, outsourcing and non-dedicated storage space.
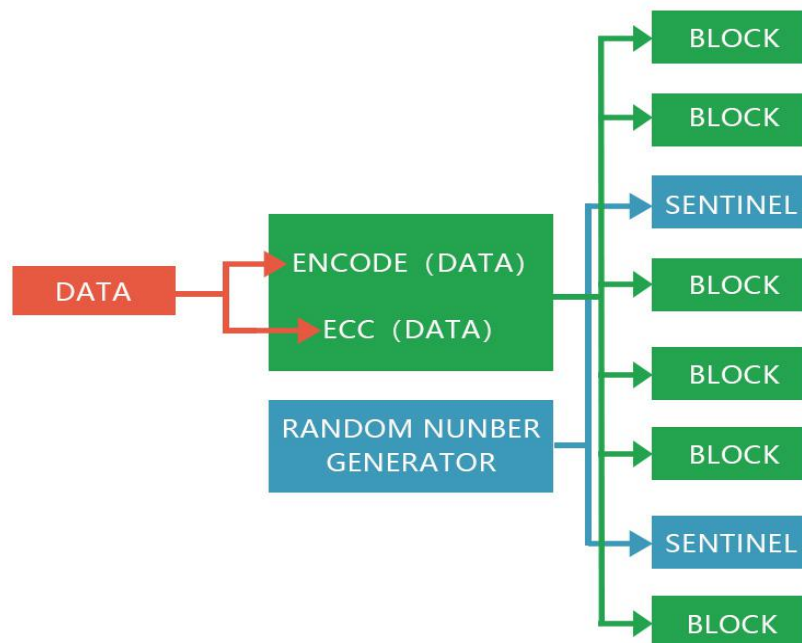
Gene source code will investigate and weigh the pro and cons of currently popular proof of storage algorithms and independently develop a novel proof of storage algorithm, Apollo. The detail of Apollo will be published in technical yellow paper edition 1.0.

Sharing health and genomic data on Gene Source Code Chain employs dynamic data access authorization algorithm, Pegasus. It will be built on several dynamic encryption algorithms including proxy re-encryption framework to cut the overhead of repeated encryptions. In the mean while, the encrypted data will carry identity-related digital watermarking to serve as an evidence for prosecuting illegal off-line distributers of the data. The detail of Pegasus will be published in yellow book edition 1.0.
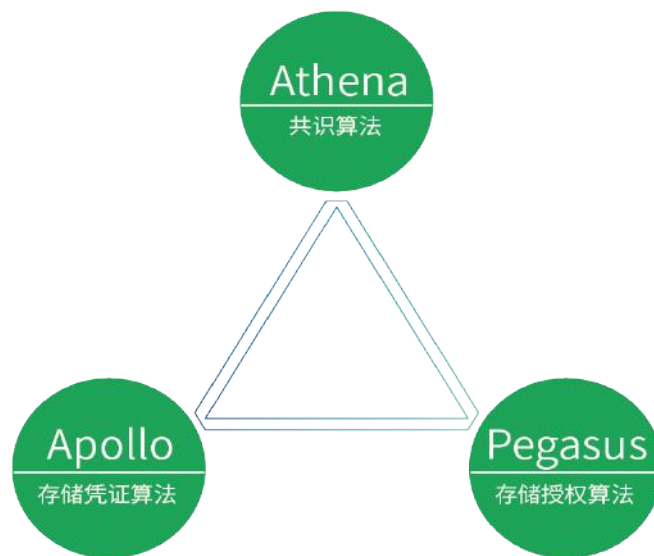
5.4 Incentive Layer Architecture Schema

The incentive layer is composed of 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, and ultimately in the smart contract system, to achieve the overall incentive and benefit sharing.

5.5 Application Layer Architecture Schema

Gene Source Code application layer, is composed of Gene Source Code basic data infrastructure system and service provider open interface system. Gene Source Code basic data support system includes data validation system(DVS), data classification system(DCS), genome analysis system(GAS), health data model analysis system(HDMA). These four subsystems will maintain a solid data application infrastructure.

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 appilcation. Traditional web application of client-server model is composed of frontend and server. While DApp 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 block chain, but with a centralized database server. All these characteristics defines a DApp. Gene source code chain system will provide a platform running DApps and a toolset for DApp development. Developers can run their own DApp on 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. The personal healthcare companion will be released in two versions. The first version is promotion version which doesn't include backbone chain and peripheral functions. The rest of the functions will be implemented after the backbone chain is online.

In the mean time, gene source code foundation will cooperate with health service providers and promote the development and propagation of their DApp.

After the completion of 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 DApp based on this universal distributed system.

VI   Why We Can Do Better

6.1 Team Introduction

William Xie
C++ Hacker
Beijing Institute of Technology Computer Science, Ph.D.

Puzzor Shi
Security Hacker. Chinese Academy of Sciences Information Technology, Ph.D. Microsoft Global Internet Safety Vulnerability Expert. Top 18th Pentester in the world

Peter Tian
Blockchain hacker, Go and Ruby hacker, Crypotocurrency Exchange Development Expert, EOS Development Contributor.

Andy Zhuang
Master's in Engineering from TUDelft
Successfully founded several tech start-ups and an early participant in the Bitcoin movement.
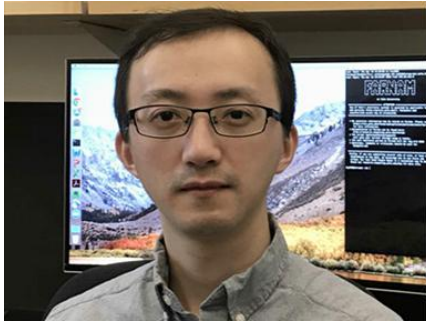
Thomas Yang
Bioinformatics Hacker
Postdoctoral Associate, Yale University
PhD. in Bioinformatics, Tsinghua University
Master in Statistics, UCLA

Charissa Trahms
Charissa worked for eight years in the United States before spending over a decade in China working with some of China's biggest companies such as Lenovo. She holds a degree in Business Management.

6.2 Gene Source Code Team and Partners Have Rich BT + IT Cross-border Experience

Gene Source Code team has multidisciplinary experience of block chain technology, bioinformatics, IT architecture, network security, and can support the research and development of Gene coin.

Asian business partners of Gene Source Code, is StarshipGene, which has accumulated a large number of genetic testing and health management industry experience. And its major business domain is DTC (Direct to Consumer), which itself is a relatively new application breakthrough in genetic testing industry. DTC business model, could accumulated a large number of gene testing samples, and sequencing and raw data analysis experience.

Starship Gene products have been sold in JD.com, WeChat store and many other platforms for sale, and it is also in the Visa Consumer Product Partner Program in year 2017, the company has received millions of angel investment.

6.3 The Gene Source Code Project is Endorsed by Many Far-sighted Scientists in Famous Research Institute

"Gene coin will be a very valuable and innovative project combining blockchain and gene big data."

Lu Zhi

Bioinformatics and Genomics PI in Tsinghua University, Phd and Postdoc in Rochester and Yale University, research focused on RNA Regulation and Precision Medicine.

"I'm very excited to hear about the innovation of Gene Source Code, and the cross-border use of blockchain technology could create a miracle in health!"

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, this one combined blockchain and genomic data storage is of ingenuity."

Aard Groen

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

6.4 Backing by Venture Capitals

6.5 Eco partners

6.6 Philosophy and Goals of Gene Source Code Foundation

(1) Gene Source Code Foundation Philosophy

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

(2) Gene Source Code Foundation Goals

Stages	Target Names	Descriptions	Estimated Finishing time
The first stage	Mars	Release of Personal Healthcare Companion Promotion Edition 0.5	Dec. 2018
The second stage	Solar System	Gene Source Code base system accomplished	May. 2019
The third phase	Milky Way	Release of universal storage mechanism	Oct. 2019
The fourth stage	Laniakea	Market mechanism of universal distributed data system enhanced	May. 2020

VII   Gene Token Placement

Total Amount of Gene token: 2 Billion

Private placement	20%
Mining to earn	40%
Core team incentive and Global development community Build-up	15%
Foundation development and user community build-up	15%
Business Development	10%

Private placement targets less than 190 investors and aims to raise 20000 ETH, or equal BTC

Lock-up Agreement:

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

Business Development part => 3 month

VIII   The Use of Funding

Usage	Description	Percetage
Expand staff	1. Shenzhen Data Center and local operation team 2. St.Jose Data Center and Research Lab	45%
Hardware procurement	The establishment of two data center and partner station in Universities	15%
Operating expenses	Basic operating expenses of Team and Data Centers	20%
Marketing	Build brand and marketing awareness in academic and public domain worldwide	20%

IX　　　Gene Source Code Foundation Governance Mechanism

The Gene Source Code governance mechanism includes the three-tier governance structure of 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, the Marketing and Public Relations Committee. The various departments are responsible for the affairs as follows:

　(1) Technical Committee is responsible for Gene Source Code technology management, specific work including open source code management, code development, code modification, code testing, and code and agenda review.

　(2) Data Management Committee is responsible for data storage standards selection and development, data storage reliability and determination of redundancy principles and other data management principles.

　(3) Financial Affairs Committee arrange the Ethereum token and Gene Token including the usage and audit of expense, routine operation, arrangement of staffing expenses, etc.

　(4) Legal and Risk Control Committee: responsible for the registration of domestic and foreign companies, reviewing various agreements and giving professional advice on legal matters.

　(5) Marketing and Public Relations Commission: The Market and Public Relations Commission mainly serves the blockchain community and is responsible for technical promotion, product promotion, business cooperation and publicity.

X Risk Warning and Disclaimer

As the 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 token innovative blockchain technology is the first of its kind in the world for economic incentives and storage mechanisms. Participate with caution, regardless of the 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 token nor the invitation to buy products nor any contractual or commitment in any way. The above information or analysis does not constitute an investment decision. This document does not constitute any investment advice or investment intention. Once involved, the relevant parties are aware of and accept the risks associated with the products and are willing to bear the 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