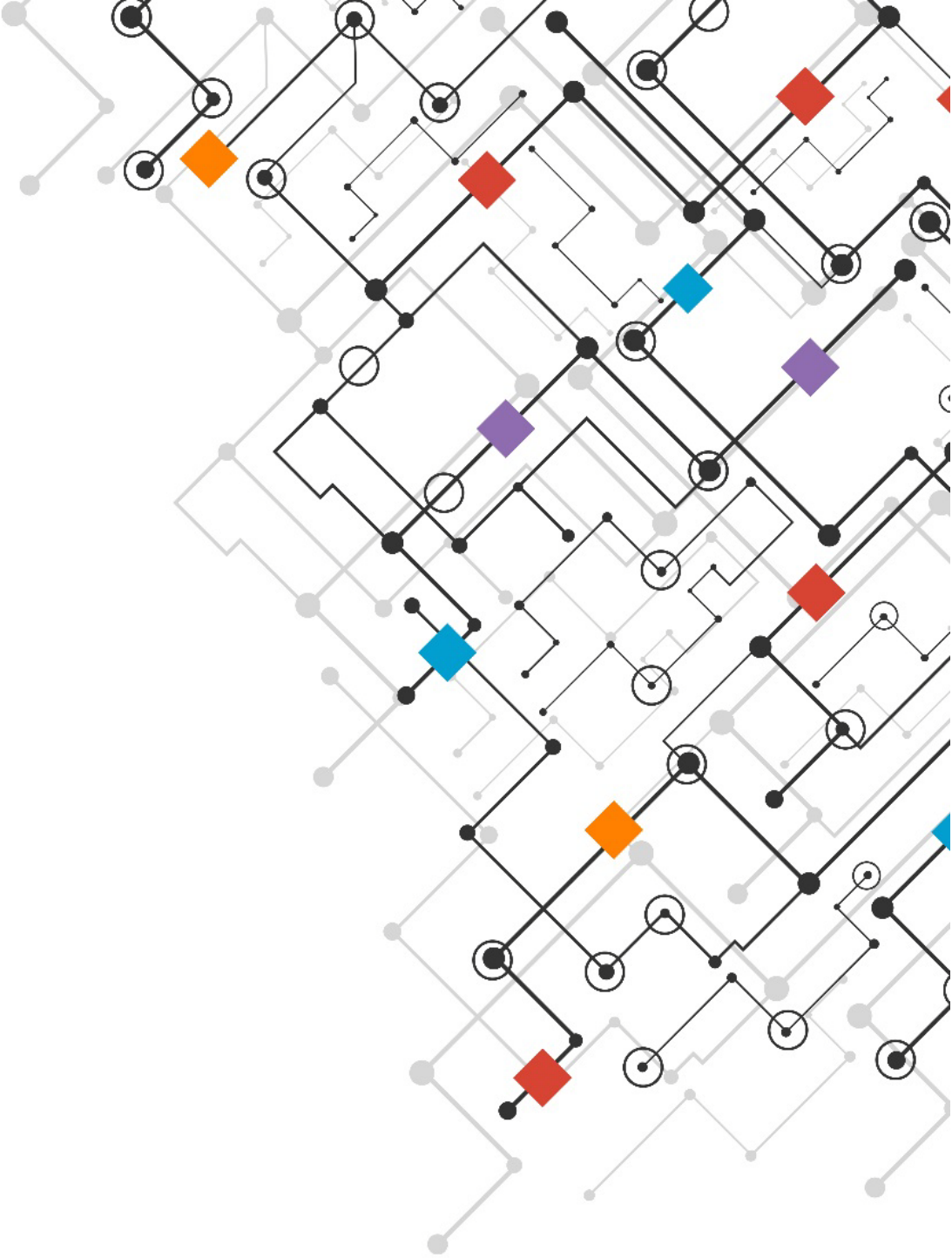

HGBC

Human Genome Block Chain

基因价值生态链白皮书

CONTENTS

ABSTRACTS

Genes are the carrier of genetic information, and genetic data as the basic data of life is an important data asset for everyone. Genetic data can be used to develop many applications in human nutrition diet, exercise and fitness, skin care, health management, drug development, parental confirmation, and scientific research. Whole genome sequencing produces a common genetic data that can be used for a wide variety of genetic testing services. The first problem facing the industry is that the data is isolated and distributed in a single genetic testing company (only for a single-purpose genetic application), users can not control their own data, can not use their own genetic data to share between enterprises to get a variety of other genetic testing services. The second problem facing the industry is that biopharmaceutical companies, biotech companies, and research institutions have strong demand for genetic data and multi-dimensional medical health data, but there is no credible and reliable data source. In this case, HGBC (Human Genome Block Chain) introduces the blockchain technology to completely identify the genomic data to individuals, so that each person can actually have the ownership of their own data, and at the same time, use distributed storage and encryption technology to solve the data privacy and security issues. By refactoring the division of labor in the industry, we will establish a new ecosystem with users as the core, sequencing companies, service providers and project parties. HGBC and other third-party genetic testing companies can develop a variety of applications on the platform for nutritional diet, exercise, skin care, health management, drug development, parental recognition, and scientific research, allowing individuals to obtain the benefits of digital assets while accelerating the exploration of life and health for the benefit of human society.

1.BACKGROUND

1.1 Exploration of genes

Genes are the basic unit of genetic material and the basis of all biological components. The human genome refers to all the genetic information carried by humans. It consists of 23 pairs of chromosomes, contains more than 3 billion base pairs, and has more than 21,000 genes. The genome can not only transmit genetic information to the next generation through replication, but also enable the expression of genetic information. The differences in hair, skin color, eyes, nose, etc. between different races are caused by genetic differences.

Humans have been exploring genes for less than a hundred years, but they have achieved remarkable results. In 2000, the first genome-wide sketch of human beings was completed. The project was called the Human Genome Project and consisted of scientists from the United States, Britain, France, Germany, Japan, and China. It lasted for 11 years and cost a total of $3 billion. Subsequently, sequencing technology developed rapidly. In 2014, Illumina released the HiSeq X Ten high-throughput gene sequencer, which reduced the cost of whole-genome sequencing to less than $1,000. The downward trend in sequencing costs far exceeds Moore's Law. [1]

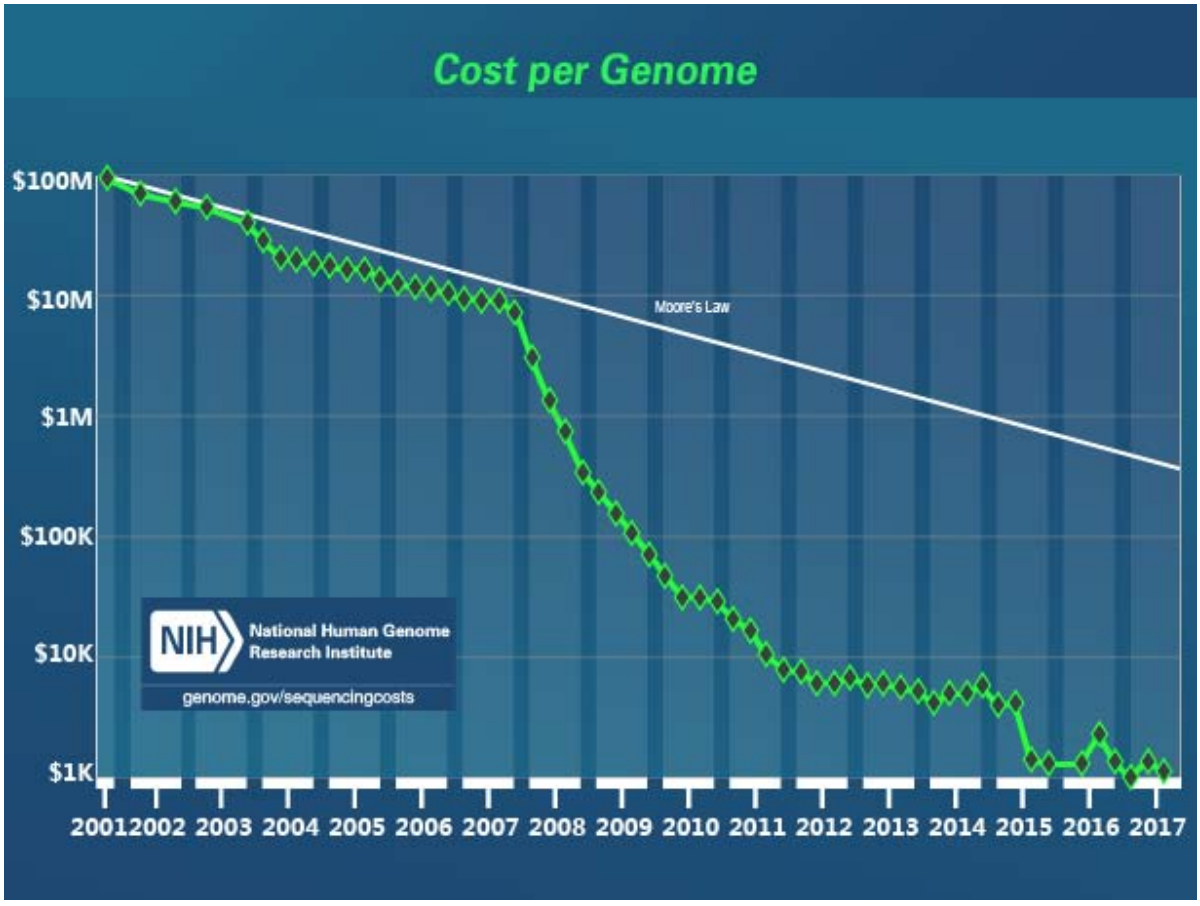

Figure 1: Sequencing cost reduction trend

[1] https://www.genome.gov/27541954/dna-sequencing-costs-data/

The reduction in sequencing costs makes it possible to sequence large-scale populations of genomes, which is the foundation of precision medicine. In 2015, the United States took the lead in launching the “Precision Medical” program, with an investment of US$215 million in the first phase, hoping to recruit 1 million volunteers for genome sequencing. In China, the “100,000 Genome Project of China” was initiated by the Beijing Institute of Genomics of the Chinese Academy of Sciences, and the “Millions of People's Genome Sequencing Project” was launched in Jiangsu, China, with the aim of establishing a genetic information database for Chinese populations through the sequencing of millions of human genomes.

1.2 Application of genetic data

The decline in sequencing costs has not only promoted the “precise medical plan” of various governments, but also promoted the rapid development of the personal genetic testing market. Personal-oriented genetic testing products cover the entire life cycle of humans, including pre-pregnancy genetic screening for genetic diseases, fetal chromosome aneuploidy disease screening, neonatal-related disease testing, growth and development guidance, disease risk assessment and health management, medication guidance, rehabilitation monitoring, diet health and other application scenarios.

Pre-pregnancy genetic testing: Recessive genetic disease is a common genetic-related disease. When the recessive gene of the disease is present on both chromosomes, the individual carrying the genotype will have symptoms of the disease; When a chromosome carries a disease-causing genotype (the other chromosome does not carry it), the individual is only a carrier and does not have symptoms of the disease. When parents are both carriers of recessive gene, only a quarter of the probability of giving birth to a healthy baby without a disease-causing gene. Through pre-pregnancy genetic testing screening, it is possible to effectively know whether a couple carries a causative gene associated with a recessive genetic disease, thereby preparing for a healthy baby. If parents are found to carry recessive genetic disease genes, they can obtain offspring that no longer carry disease-causing genotypes through the current mature assisted reproductive technology, and block the transmission of related diseases within the family.

Genetic testing during pregnancy: Non-invasive DNA prenatal testing By genetic sequencing of maternal venous blood free DNA (including fetal free DNA), it is possible to accurately detect whether the fetus has three major chromosome aneuploidy diseases: Down syndrome (T21), Edwards syndrome (T18), Pada's syndrome (T13). [2]

Neonatal deafness genetic testing: China has more than 27.8 million hearing-impaired patients, accounting for 33% of the nation's 82.69 million disabled people. The number is the highest in all kinds of disability in the country, with an annual increase of more than 30,000 deaf children. Deafness has a great relationship with heredity and can be divided into delayed-type progressive deafness and drug-induced deafness caused by genetics. For newborns with large vestibular aqueduct syndrome, give life guidance and reduce the incidence of deafness; For drug-induced

[2] https://www.sciencedirect.com/science/article/pii/S0140673697021740

deafness carrying newborns, give medication guidance to prevent drug-induced deafness; For deafness gene mutations carrying newborns, give advice on family planning to prevent deafness in offspring.

Health Risk Assessment: A quantitative assessment of the health status and future risk of illness.

Through health risk genetic testing, you can better understand your true health risks and guide yourself to change or correct unhealthy behaviors. In 2013, American actor Angelina Jolie published an article entitled "My Medical Choice"[3]: "...I am a carrier of the BRCA1 gene. This gene has defects that can lead to a significant increase in the risk of breast and ovarian cancer. Although there are individual differences among different women, according to doctors, my chance of breast cancer is as high as 87%, and the risk of ovarian cancer is also 50%. Only a small number of breast cancer cases are caused by genetic mutations. For BRCA1 gene-deficient carriers, there is an average of 65% chance of developing breast cancer." Genetic testing plays an important role in cancer risk assessment, and targeted prevention measures are adopted through the results of individual risk assessments.

Precise medication: Choose the right medicine and the right dose according to the individual's differences. Warfarin is the most widely used oral anticoagulant in the clinic. It has been found that there are individual and ethnic differences in the genes encoding warfarin metabolism and drug-related enzymes. The dose adjustment of pharmacological pharmacology-based drugs has received attention in all aspects. The FDA revised the instructions to recommend that warfarin should be tested for genotypes such as CYP2C9, VKORC1, and CYP4F2 before using the drug to determine the difference in drug use. [4]

Identification: Human cells have a total of about 3 billion base pairs of DNA. The DNA of any two people is not exactly the same. The number of different base pairs between humans is several million. Therefore, the DNA profile displayed by molecular biological methods varies from person to person, thereby identifying different people. It is also possible to judge the genetic relationship between two people by identifying the DNA.

Healthy living: Genetic differences also affect every aspect of everyone's life. For exercise and fitness, ACTN3 gene is related to explosive power[5]; PAPSS2 gene is related to exercise enthusiasm[6]; BDKRB2 and LPL genes are related to aerobic exercise weight loss effect[7]. For daily diet, FTO gene abnormality is easy to cause obesity[8]; APOA2 gene mutation is not suitable for high-fat foods;

---

[3] https://www.nytimes.com/2013/05/14/opinion/my-medical-choice.html

[4] https://www.pharmgkb.org/label/PA166104776

[5] Yang N, MacArthur D G, Gulbin J P, et al. ACTN3 genotype is associated with human elite athletic performance[J]. The American Journal of Human Genetics, 2003, 73(3): 627-631.

[6] De Moor M H, Liu Y, Boomsma D I, et al. Genome-Wide Association Study of Exercise Behavior in Dutch and American Adults[J]. Medicine and Science in Sports and Exercise, 2009, 41(10): 1887-1895.

[7] Garenc C, Pérusse L, Bergeron J, et al. Evidence of LPL gene-exercise interaction for body fat and LPL activity: the HERITAGE Family Study[J]. Journal of Applied Physiology, 2001, 91(3): 1334-1340.

[8] Wang L, Yu Q, Xiong Y, et al. Variant rs1421085 in the FTO gene contribute childhood obesity in Chinese children

ALDH2 and ADH1B genes affect alcohol metabolism[9]; APOE E4 genotypes have risk of elevated blood lipids when ingesting cod liver oil[10]; Mutations in the TMEM18 and MC4R gene are associated with depression-overeating[11]. Only by knowing yourself at the genetic level can you live a better and healthier life.

1.3 various problems in the industry

The rapid development of gene sequencing technology and the drastic reduction of sequencing costs have greatly stimulated the personal genetic testing market and promoted life science research. We have been in the genetic industry for more than ten years, not only seeing the rapid development of the market, but also seeing various problems in the industry.

1.3.1 Users

1.3.1.1 Lack of genetic data ownership protection mechanism, user rights can not be guaranteed.

Gene data is an important part of medical health big data. Using medical big data can gain new knowledge, make new decisions, and create new value, so medical big data obviously has property attributes and can be the object of rights. The rights generated by medical big data belong to the right of information, and belong to the property rights in the classification of civil law rights. They have four powers of property rights: possession, use rights, income rights, and disposal rights. Specifically, the owner of medical big data enjoys exclusive possession of big data information; enjoys the right to use or license others; and enjoys the right to income arising from the use or licensing of others; enjoys the right to dispose of data or to abandon it.

There is no mechanism guarantee for the ownership of genetic data by users. Genetic testing companies often store their genetic data on their own servers in order to obtain greater value. At the same time, because of the expertise required for genetic data applications, even the genetic data can be downloaded, ordinary users can not directly analyze the use it. Users are in a situation of helplessness in the realization of the value of genetic data, and cannot enjoy the benefits of their own genetic data. On January 6, 2015, biopharmaceutical giant Genentech purchased information(genetic data from 3000 Parkinson's patients) from the genetic testing company 23andMe. By collecting whole genome data, it tried to identify the target of the disease and paid 60 million US dollars for it. The 23andMe got all the benefits. [12]

---

aged 3¨C6years[J]. Obesity research & clinical practice, 2013, 7(1): e14-e22.

[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3860439/

[10] Daiello LA, Gongvatana A, et al. Association of fish oil supplement use with preservation of brain volume and cognitive function. [J]. Alzheimers Dement. 2015 Feb;11(2):226-35.

[11] Hotta K, Nakamura M, Nakamura T, et al. Association between obesity and polymorphisms in SEC16B,TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population[J]. Journal of human genetics, 2009,54(12): 727-731.

[12] https://www.fiercebiotech.com/r-d/23andme-strikes-60m-genentech-deal-to-continue-pivot-to-database-driven

As a property owned by each human being, genetic data is occupied by centralized institutions due to the limitation of data collection and processing capabilities, and the nominal ownership and substantive dominance of data are separated. The use of blockchain technology to identify genetic data to individuals is essentially the return of ownership and control of data to individuals.

1.3.1.2 The gene detection process is cumbersome and has a long cycle.

When a user purchases a whole genetic testing product and finally gets the test result, he will go through the following process:

1. Users place orders and purchase gene sequencing services from gene companies.
2. Gene company sends users a genetic data collection box.
3. The user collects the sample himself (the cells in the saliva).
4. The user sends the sample (the cells contained in the saliva) back to the gene company.
5. Gene company sends samples to sequencing companies in batches.
6. Sequencing companies use gene sequencing instruments for gene sequencing.
7. The sequencing company sends the sequencing raw data to the gene company.
8. Gene company analyzes raw data for sequencing.
9. Gene company generates an interpretation report of sequencing raw data.
10. Gene company sends the interpretation report to users.

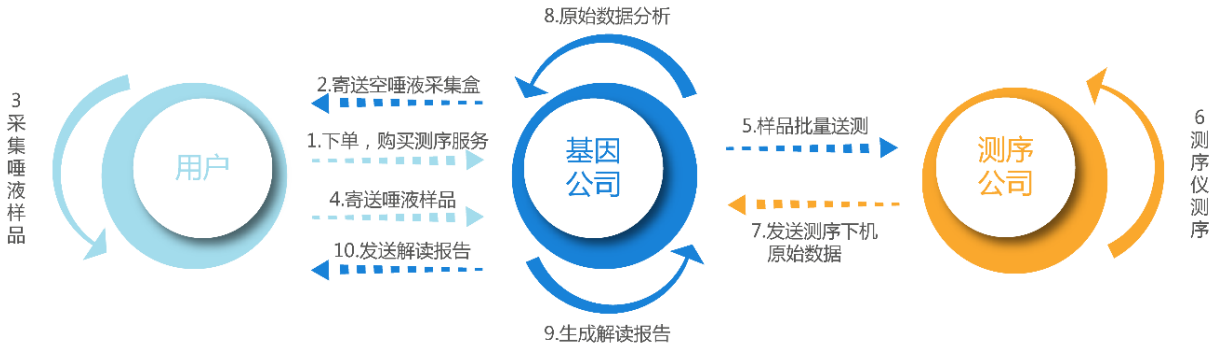

Figure 2 The genetic testing process

The whole process is about 1-2 months, and the long time seriously affects the user's product experience. At the same time, the cost of sequencing itself accounts for a large part of the cost of the product. Gene data from whole genome sequencing is a common standard data that can be applied to different genetic testing products. At present, there is no reliable data sharing mechanism and platform. When users purchase genetic testing products of different companies, they have to repeatedly perform sampling, sequencing, analysis and other processes, which not only increases the risk of user data leakage, but also causes unnecessary Economic expenditure.

With users as the core and technology as the guarantee, a credible ecological platform can be established. Users can enjoy various genetic testing services provided by different service providers only by one sequencing, improve data reusability, and bring tangible benefits to users. .

1.3.2 Scientific research / Drug research and development institutions

1.3.2.1 The data of scientific research samples is difficult to obtain.

For the research projects based on population genomic data, the first step is to recruit volunteers, such as the US “Precision Medical Plan”. Starting in 2015, a total investment of US$215 million will be used to recruit 1 million volunteers and establish a national study. queue. First, it takes a lot of manpower, material and time to recruit volunteers, DNA sampling, genome sequencing, data analysis, etc., and finally enter the core research stage of the project.
If there is an ecological platform that aggregates a large number of users and user data, the research organization can directly use the standardized data from the user's hands and can be directly used for research, which greatly saves costs and improves efficiency.

2. The sample size and the data dimension is small, and the phenotypic data is not standardized.

Precision medicine and precise health management put forward higher requirements for the mining and application of genomic data and medical health data. Many functions of human 21000+ genes are unknown. Many genetic factors related to diseases, traits, behaviors, and psychology have not been found, largely because the sample size of the data is not large enough, and the data is not rich in dimensions, and the timeline of the data is not long enough. The sample size of general research projects ranges from tens to tens of thousands, Scientists often analyzing only basic phenotypic data and gene mutations, gene expression or methylation. Due to sample acquisition difficulties, the sample data is basically breakpoint data.

3. User participation is low.

In the past history of genetic technology development and application of genetic data, it is basically a mode of operation centered on scientific research institutions and drug research and development institutions. Individuals often participate passively, and even do not know that their samples are used for scientific research. Users cannot directly benefit from the data, resulting in low user participation and low project development efficiency.
To achieve massive sample data collection, full-scale data collection, and long-term data aggregation, it is necessary to innovate the traditional centralized organization model of data collection and application, and become an individual-centric operation mode:
Set enough rewards to motivate individual users to actively participate, so that a large sample can be obtained on the platform; Allow individual users to earn more rewards after contributing more dimensional data; Allow individual users to receive more rewards after providing medical health data for a long time.

1.3.3 Gene product/service providers

1.3.3.1 The sample data volume is small, lack of phenotypic data, and the product development cost is high.

For most gene products/service providers, there is no strong financial strength to recruit volunteers for product development, but the research and development of genetic products cannot be separated from the sample size and multi-dimensional phenotypic data of large-scale population. It is common practice for genetic testing companies to bring immature products developed under small data volumes to the market and then iterate products based on constantly evolving data.

1.3.3.2 The user group is small, and the public has low awareness of genetic products/services.

Although genes are the basic component of life, the general public has a low level of knowledge about genes. Genetic testing products are not known to ordinary people. Only a small number of people have purchased genetic testing products. Gene products/service providers cannot focus solely on R&D products, but also take on the responsibility of educating the public and nurturing the market.

1.3.3.3 The product cycle is long and the experience is poor.

For the privacy and security of user data, and the enterprises with centralized architecture do not have enough motivation to share data, the user's genetic data cannot be circulated among the various genetic testing companies, forming many isolated data islands. Every time a genetic testing company acquires a new user, it needs to go through a complicated process and a long cycle to obtain the user's genetic data, and then provide the corresponding genetic product/service. This mode not only damages the product experience, but also limits the rapid iteration of the product.

2. OUR VISION

The HGBC plans to build a multi-dimensional gene medical health database with a scale of 100 million user groups. With the rapid development of genome sequencing technology, the cost of personal genome sequencing is rapidly decreasing, and personal genome data will become an important digital asset for everyone. After an individual has ownership of personal genomic data and is free and easy to control personal genomic data, economic and social value will be realized for individuals and businesses. In the past, the model centered on enterprises and scientific research institutions was inefficient, and the new model is an equal, democratic, and efficient operation model.

The HGBC realizes the storage and ownership confirmation of personal genomic data and related medical health data through distributed storage/encryption/blockchain solution, solves users' concerns about data security and privacy, encourages the public to actively participate in genome sequencing and genomic data application, motivates gene product/service providers to develop countless interesting and useful applications, accelerates the development of treasures of human genome data, and improves the health and longevity of 1 billion people.

2.1 For individuals

The HGBC will help everyone to have ownership of his/her own genomic data and medical health data, and be able to participate in a variety of scientific research projects at his/her own discretion, in return for economic benefits, and to enjoy the benefits of research results. Based on the “GPS” navigation matching function of genomic data, individuals will be able to directly obtain the most suitable medical health service plan at the lowest cost and the fastest time, thus achieving the goal of health, longevity and life optimization.

2.2 For research institutions / gene product providers

The HGBC will help drug development organizations, scientific research institutions, insurance operators, and health management institutions to directly connect to individual users. With the permission of users, they can quickly obtain the most comprehensive and complete genetic data and medical health data, and provide big data for their product development. In the new model, R&D efficiency has increased and R&D costs have decreased. The HGBC will provide a development environment and supporting incentives to facilitate the development of genetic applications based on genomic data. The goal is to have more than 10,000 genetic applications in the next 10 years.

2.3 For society

After HGBC confirms the genetic data to individuals, it empowers individual users to participate in the development and innovation of science and technology, and can obtain economic benefits and technological progress, and can enjoy the medical health service plan that best matches its own

genetic information, reducing the economy costs and time costs. The social expenditures in scientific research, medical insurance, and social insurance will gradually decrease with the improvement of HGBC ecological construction, and operational efficiency will increase, making genomic data the infrastructure for healthy living in the future, and government agencies providing public services will also benefit.

3. ECOSYSTEM

HGBC Ecology is user-centered and is a genetic value ecosystem built and participated by HGBC team, sequencing service provider, gene product provider and gene service provider.

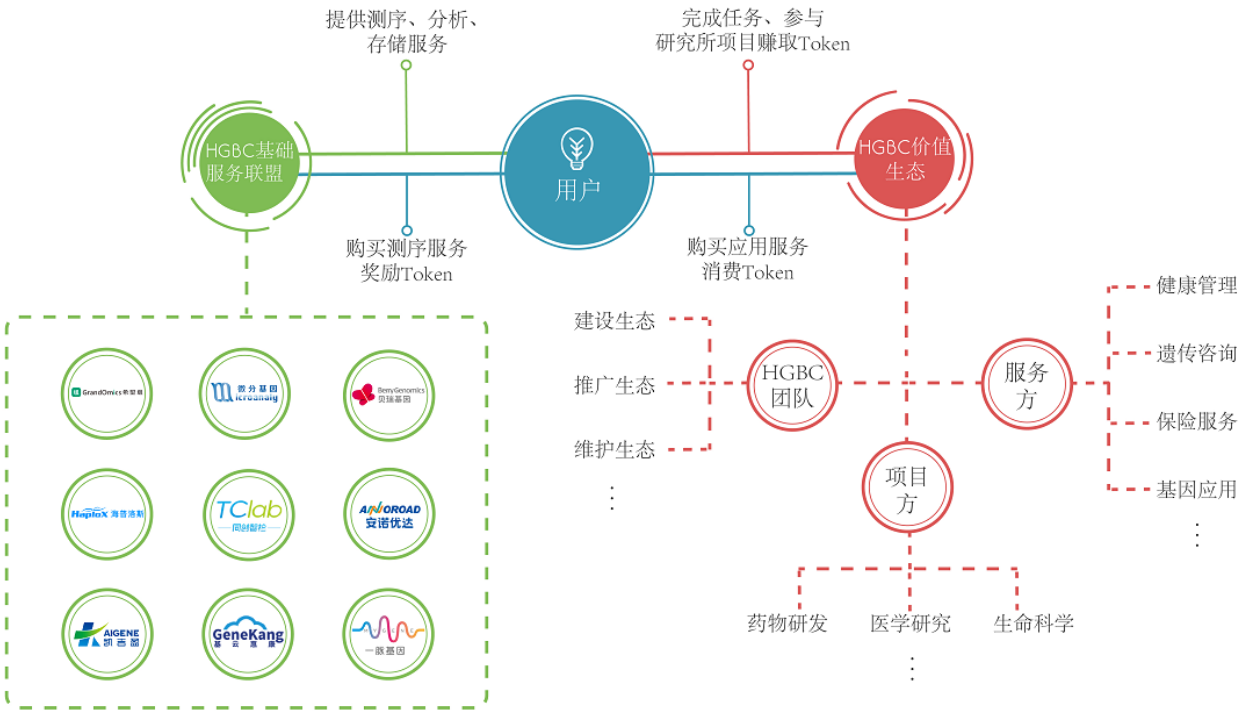

Figure 3: HGBC ecosystem

3.1 HGBC team

The HGBC team is the project sponsor and the developer and maintainer of the HGBC ecosystem. In order to maintain the ecological prosperity of HGBC, the team actively explored the blockchain technology to change the production relationship of the genetic industry, and actively established alliances with upstream and downstream enterprises to promote ecological construction. In addition, the team develops some genetic products/services on the platform alone, and more importantly, attracts third-party companies to develop more genetic applications on the platform.

3.2 Users

Users are at the heart of the HGBC ecosystem. There are two types of users in the HGBC ecosystem. One is the X user who has completed his own genome data after sequencing, and the other is the W user who has not yet performed gene sequencing and has no genomic data.

X users can participate in research projects initiated by the project side to obtain Token benefits,

and can also use Token to purchase various genetic applications and services provided by the service providers. When the user completes the gene sequencing, the user will receive the gift Token provided by the platform, and will receive the daily Token reward when the genomic data remains on the platform.
By purchasing a sequencing service provided by a sequencing service provider, W users can obtain their own genomic data to become X users, obtain Token rewards, and obtain Token rewards by completing tasks in the task box.

3.3 Sequencing Service Provider

The sequencing service provider is a manufacturer with genome sequencing capabilities and uses a qualified sequencer to provide sequencing services. The sequencing service provider has been audited by HGBC, and only qualified companies can join the ecosystem to provide sequencing services to users. HGBC will regularly review the sequencing service providers from the sequencing cycle, sequencing data quality, and service quality. For the lower-ranking sequencing service providers, HGBC has the right to suspend or cancel its service qualifications to protect user rights.

3.4 Project side

The project side refers to enterprises that have demand for genetic and medical health data, and may be drug research and development enterprises, biotechnology companies, and university research institutions. The project party must have legal qualifications and the information is transparent to the ecology. The research project released by the project party in the HGBC ecology shall be approved by HGBC.

3.5 Gene Products/Service Providers

A gene product/service provider is an organization or individual that provides products and services to users based on user data. These products and services include, but are not limited to, disease risk assessment, health management, drug risk testing, genetic screening, individual trait testing, genetic counseling, genetic applications, insurance coverage, and more. The identity between the service provider and the project side is interchangeable, and the project side can publish its research results to the HGBC ecosystem and become a product/service provider to provide users with gene products/services.

4.SOLUTIONS

HGBC is an ecological network of genetic and medical data based on blockchain construction. It is a new user-centered collaboration model that reshapes the genetic industry chain. The platform uses blockchain technology to ensure that users control their genetic and medical health data, maximize the value of data and create a win-win ecosystem under the Token incentive system. *The HGBC solution consists of: (1)the Gene Treasure APP (Darwin Planet APP), (2)the Data Analysis Cloud Platform, (3)and a third-party development platform.*

4.1 Gene Treasure APP (Darwin Planet APP)

Users can participate in the HGBC ecosystem through the "Gene Treasure APP" (ie Darwin Planet APP) to collect data mining revenue, participate in research projects, purchase applications and services.
The HGBC will assign a unique numeric identity ID to the registered user and write it to the blockchain ledger.
Gene Treasure APP (Darwin Planet APP) consists of a Task Box, a Research Institute, and a Treasure Shop.

4.1.1 Task Box

In order to make each user better enjoy the ecological value of HGBC, we put forward the concept of “data mining”. “Data mining” means that the user can complete the “computing power” when completing the task in the task box. The higher the “computing power”, the higher the Token reward obtained in the corresponding time.

4.1.2 Research Institute

Users can choose to participate in research projects at the Institute. The relationship between the user and the research project is a two-phase selection relationship, which is a transaction of data usage rights. The Institute adopts smart contracts and asymmetric encrypted data transmission to completely guarantee the transmission security in the data transaction process. The overall flow chart is shown below.
(Note: 1. The data tag is a data feature that is automatically calculated based on the user's genetic data, medical health data, and characterization data. 2. The service release trading platform is where the project party releases the project, manages the project, and obtains the data.)

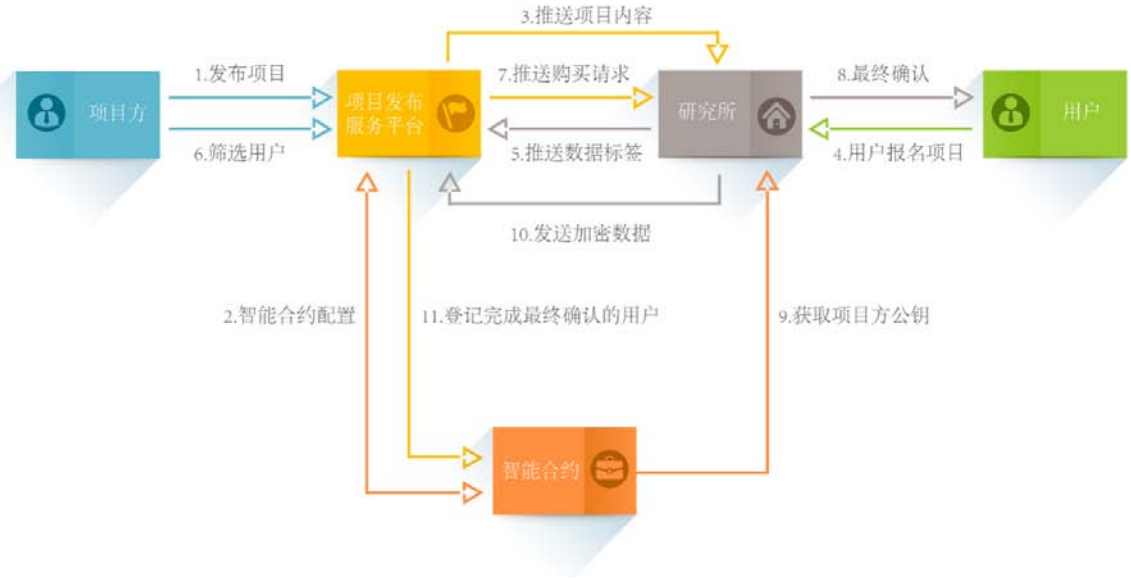

Figure 4: project participation flow chart

1. The project party publishes its own project on the “Project Release Service Platform”. The project content should include information such as project introduction, data requirements, compensation, and number of participants.
2. Trigger the smart contract, assign the project ID, lock the promised token (Token), set the project end condition, and so on.
3. The initial configuration of the project is pushed to the Institute by the platform.
4. The user browses the project at the institute and decides whether or not to participate.
5. After the user confirms the participation, the system will push the data label automatically generated based on the user data to the project party.
6. The project side filters the appropriate users based on the data tags and sends a data purchase request.
7. The data purchase request is sent to the user in the form of a text message and an App message.
8. After receiving the request, the user confirms whether to join the project.
9. If the user agrees to participate in the project, the required portion of the project data is taken out and the data is encrypted using the public key of the project side.
10. The platform pushes the encrypted data to the project management system of the project side. After the project side receives the data, it decrypts the data using its own private key.
11. After the project side successfully obtains the data, the smart contract is triggered, and the compensation (Token) locked by the project side is distributed to the user, and the entire transaction is completed.

4.1.3 Treasure store

Applications and services are the value presentation of the HGBC chain, and the HGBC team developed the Treasure Store as a vehicle for applications and services. The Treasure Store offers

users a wide range of health and entertainment applications based on genetic data, as well as professional, one-to-one personalized services. Applications and services are jointly developed by the HGBC team and third-party teams.

Applications: Gene applications offer a diverse range of options for users to interpret genetic codes, such as the individualized interpretation of the origin, nutrition, and sport of genetic information.

Services: One-to-one services provided by professionals to provide users with personalized genetic code interpretation and personalized service customization, including genetic counseling services and health management services.

Before the launch of the HGBC project, the team had extensive experience in gene interpretation and application development, and had a genome-wide interpretation and development capability. It successfully developed the “PK Race”, “Gene Intimacy” and “Breast Cancer/Ovarian Cancer Risk Evaluation" and other applications.

Case 1: Genome-wide test report

The Personal Whole Genome Health Assessment Package uses the leading second-generation sequencing technology to sequence and analyze 3 billion base pairs in the human genome, combined with the latest scientific advances, to provide users with a full lifecycle health management program based on their unique genetic information: personalized health advice, birth guidance, medication tips, dietary advice, nutrition guidance, etc., allowing users to test once and benefit for the rest of their lives.

Figure 5: genome-wide test report

Case 2: Drinking PK Competition

Through the genetic data to determine the individual's ability to metabolize alcohol, to determine its "drinking ability" size, and introduce a "PK" comparison mechanism, which has both entertainment and healthy life. Give drinking advice based on specific genotypes, achieve healthy drinking, and reduce body damage caused by drunkenness.

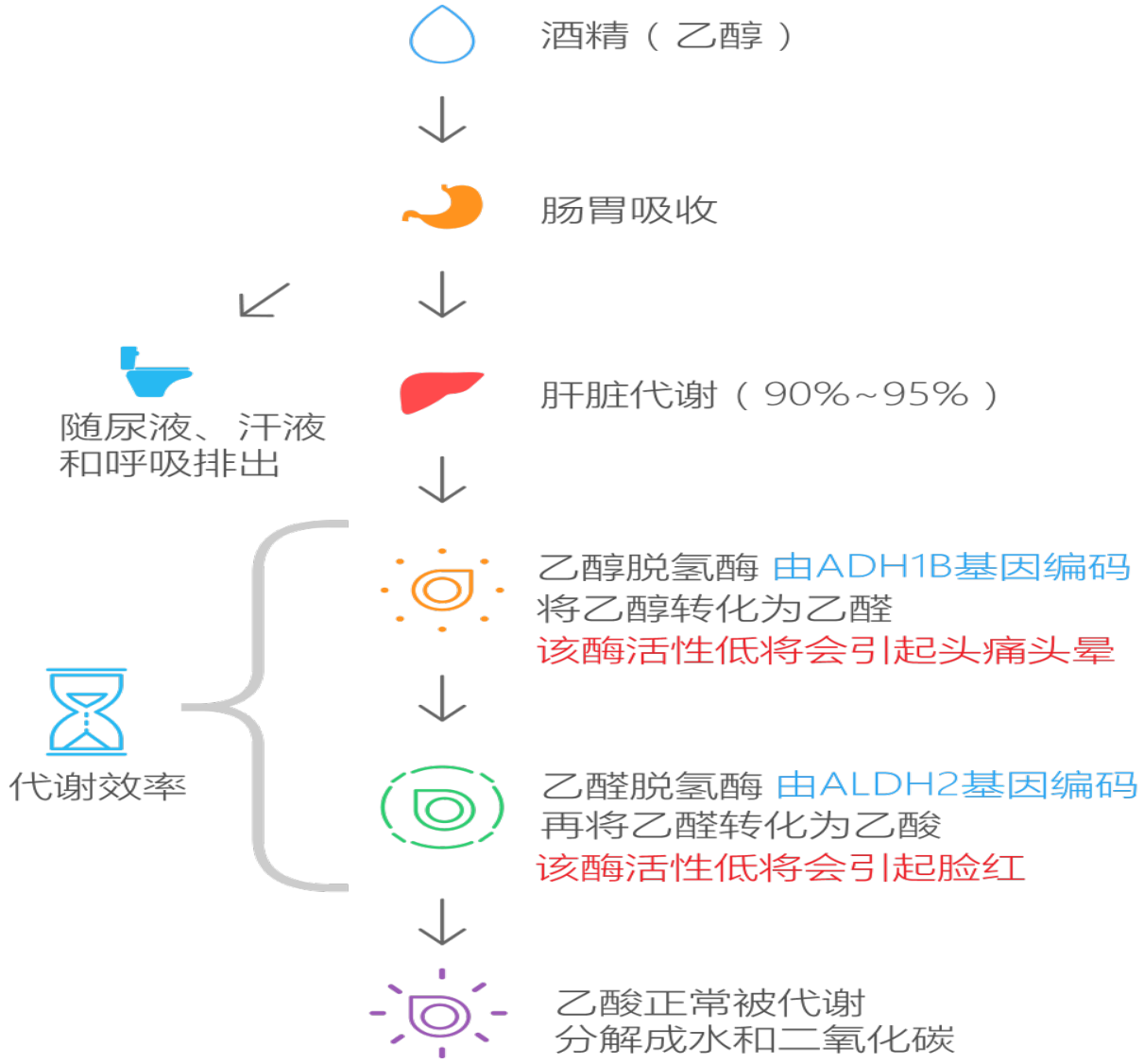

Figure 6: Biochemical pathways for alcohol metabolism

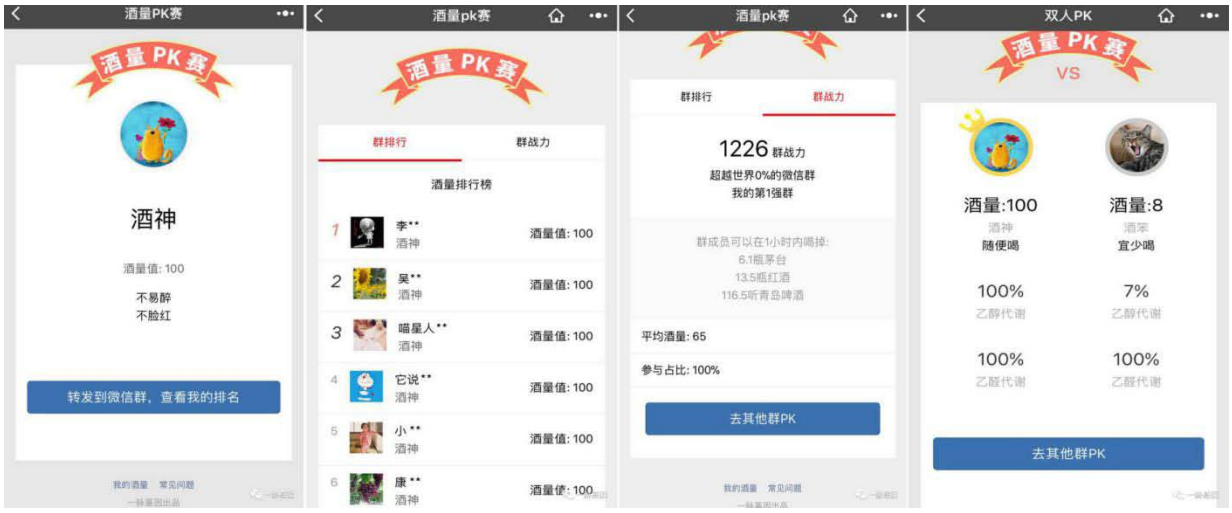

Figure 7: Alcohol PK applet

Case 3: Gene intimacy

By comparing the genome-wide data of both sides, the genetic level of the two is obtained, and the genetic relationship between any individual in the human population is quantified.

Figure 8: gene intimacy applet

Case 4: Breast Cancer / Ovarian Cancer Risk Assessment

About 5% to 10% of breast cancers are caused by heredity, which means that genetic defects (gene mutations) are passed from the patient's parents to the next generation. At present, we have developed a breast cancer/ovarian cancer risk assessment program based on dozens of sites of BRCA1 and BRCA2 genes and various risk factors. The risk factor refers to the factors that affect the probability of an individual's illness.

Figure 9: Breast cancer / ovarian cancer risk assessment applet

4.2 Data Analysis Cloud Platform GAIC

When the genomic data is produced from the sequencing instrument, it cannot be directly analyzed and interpreted. The raw data needs to be preprocessed, including checking the data quality, cleaning, comparing, and extracting the mutation information.
The specific process is shown in Figure 10.

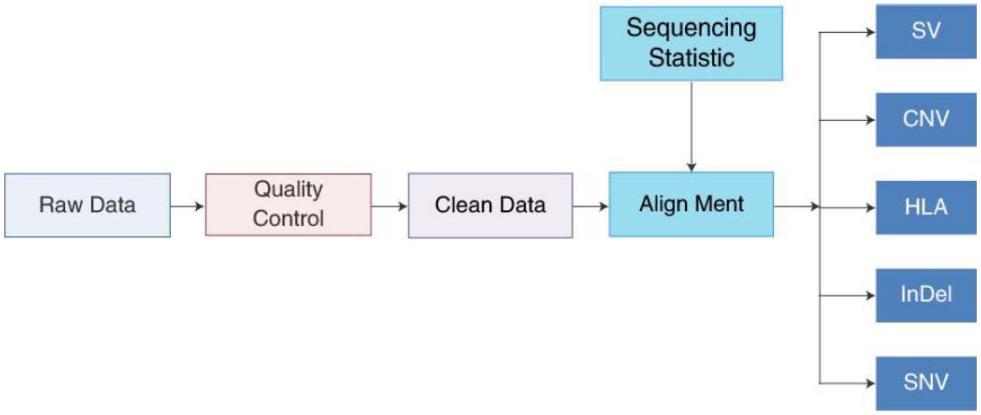

Figure 10: Genome data processing standard process

The raw sequencing data is first subjected to mass value filtering to obtain a high quality sequence, and the filtered reads are aligned to the human reference genome. The effect of statistical sequencing, including effective data volume, sequencing coverage, sequencing depth and other indicators, is used to assess whether the data is qualified. For qualified data, based on the comparison results, data analysis and mining at the individual level and group level can be performed.

Genomic data analysis requires high expertise and huge computing resources. The HGBC team has developed the professional data analysis cloud platform GAIC (Genome Analysis and Interpretation in Cloud) to provide basic data analysis capabilities for ecological construction.

Based on the efficient computing power and massive storage capacity of the cloud platform, GAIC provides basic analysis and secure storage services for gene big data, and automatically completes every step from raw data to comparison and gene annotation.

Figure 11: GAIC platform

4.3 Third Party Development Platform

The development of applications and services based on genetic and medical health data is an important part of HGBC. The HGBC team will provide a rich set of developer tools to help developers build applications and services faster and faster. These tools include SDKs, APIs, and more in a variety of high-level languages. The purpose of the development platform is to help developers focus on the business level and reduce development difficulty and cost.

By working closely with third-party developers, the ability of individual developers and small team developers is released, allowing HGBC's ecological applications and services to flourish, and allowing users to have more choices and greater value.

5. TOKEN INCENTIVE PLAN

HGBC Ecology has set up a special reward pool to reward users involved in ecological construction, using the blockchain-based pass Token as the only reward. HGBC Ecology uses GSS (Gradient Subsidy of sequencing) and PODS (Proof of Data Stake) to ensure fair and reasonable rewards.
A Token is distributed in the ecosystem based on the ERC20 contract on Ethereum. HGBC Token has issued a total of 3 billion Tokens and will never issue any new tokens.

5.1 Application Scenario

HGBC Token is a kind of “pass-through” in the HGBC ecosystem. It has the following usage scenarios in the ecology:

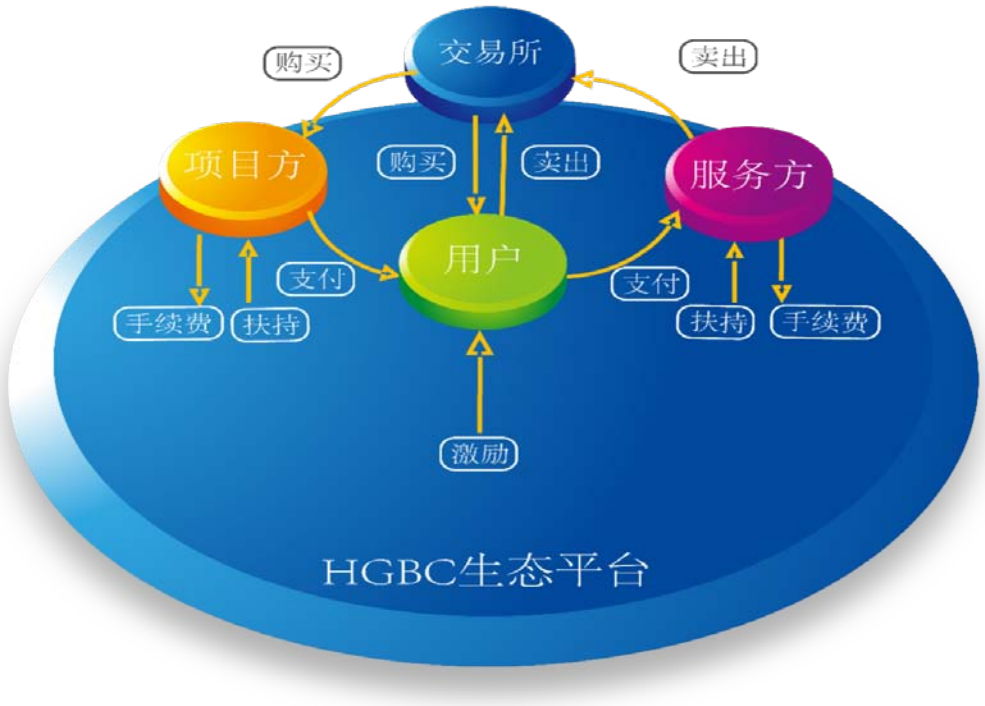

Figure 17: HGBC Token flow model

1. Sequencing rewards: Token rewards for users who have completed genome sequencing and provided genomic data to the platform;
2. Incentive: Give incentives to users who complete tasks in the task box;
3. Release project: The project party needs to lock a part of HGBC Token to release scientific research projects;
4. Transaction settlement: After the user participates in the research project of the project side, the project party uses the HGBC Token to pay the user;
5. Purchase service: The user uses HGBC Token to purchase the services provided by the gene service provider;

6. Purchase of products: Users use HGBC Token to purchase products provided by the gene product provider;
7. Platform burning: The platform collects a certain amount of HGBC token from the project party and the service party as the processing fee.
HGBC Ecology is a rapid development platform. With the continuous improvement of ecology, as the only token in the ecology, the value of HGBC will continue to increase.

5.2 Distribution plan

A total of 3 billion HGBC Tokens are allocated as follows (never issue any new tokens).

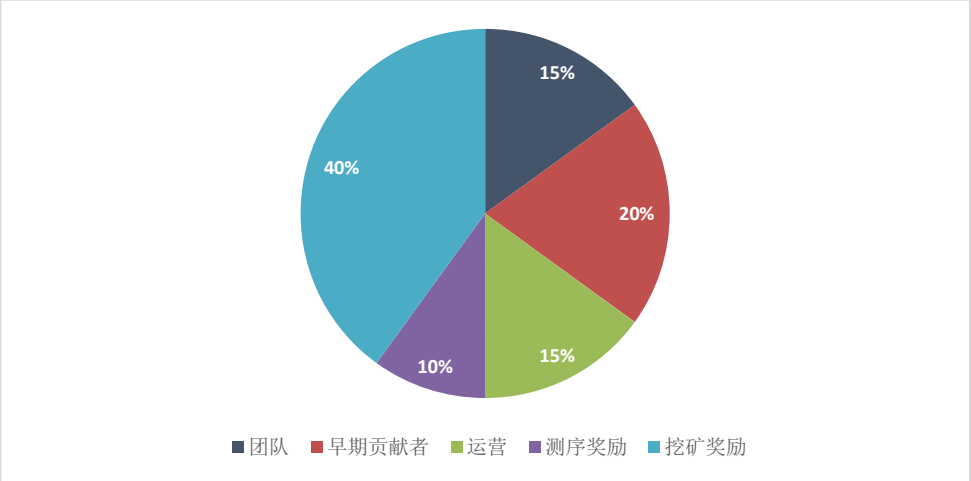

Figure 18: HGBC Token Distribution

(1) 15% to the HGBC team, released in batches during the five-year lock-up period.
(2) 20% to early contributors (funds or other resources).
(3) 15% is used for operating expenses for ecological construction and promotion.
(4) 10% for sequencing rewards, used to reward the first 1 million users who have completed gene sequencing and submitted genomic data.
(5) 40% is used for mining rewards to reward users involved in ecological construction.
The last two are collectively referred to Reward pool(50% in total).

5.3 Sequencing Early Bird Rewards

For users who are involved in early gene sequencing, HGBC will use a certain percentage (10% of the total issued Token) HGBC Token as a one-time reward, subsidize the cost of genome sequencing, encourage more people to participate in ecological construction, enjoy the value of the data.

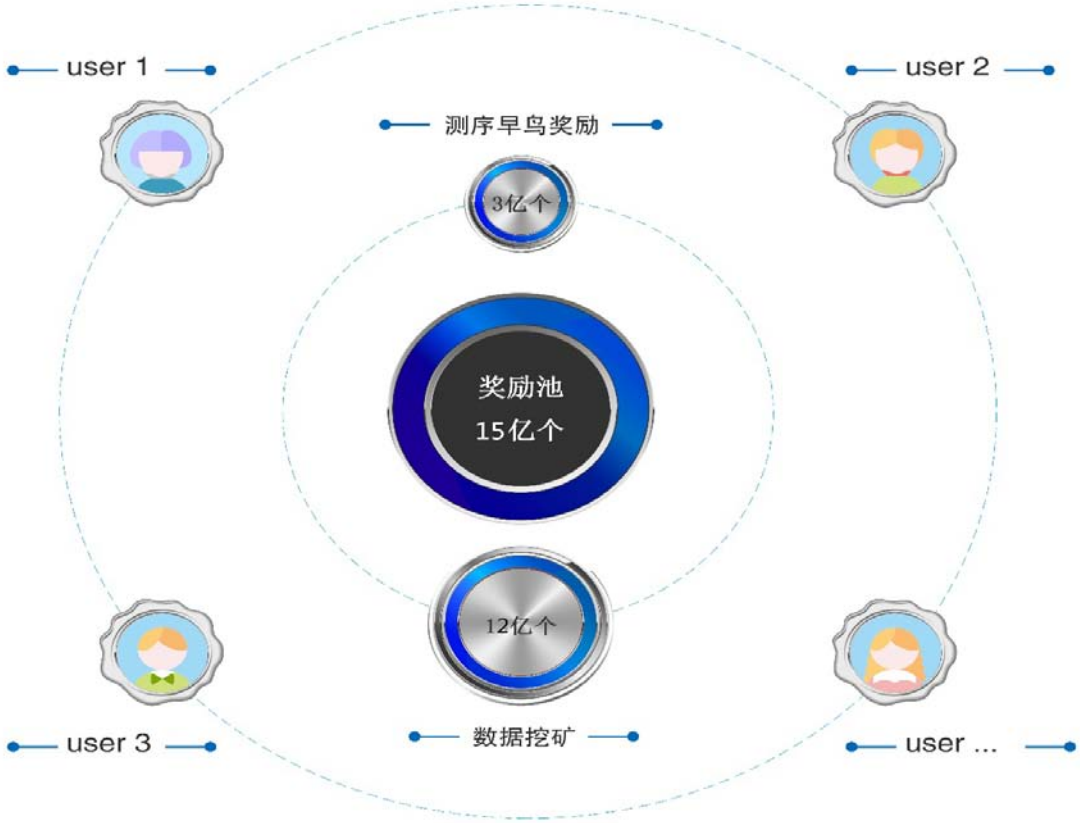

Figure 19: HGBC Token Reward Pool

Considering the development of sequencing technology, the cost of sequencing has dropped rapidly. At the same time, with the gradual improvement of the ecology, the value of HGBC Token will become higher and higher. Therefore, while the number of users of gene sequencing increases, the number of sequencing subsidies Token is reduced by using a gradient decrease. After HGBC Ecology completed the Million Human Genome Project, we estimate that sequencing costs have dropped to very low prices and are acceptable to everyone. Therefore, after the completion of the Million Human Genome Project, HGBC will no longer issue sequencing rewards, but users can still obtain Token rewards through data mining.

The HGBC Token is rewarded using the GSS (Gradient Subsidy of sequencing) algorithm. The algorithm rules are as follows:

NAME	100 Human Genome Project	1,000 Human Genome Project	10,000 Human Genome Project	100,000 Human Genome Project	Million Human Genome Project
Number of people	100	1,000	10,000	100,000	1,000,000
Per capita reward	100,000	30,000	6,000	1,000	100
Total amount	10,000,000	30,000,000	60,000,000	100,000,000	100,000,000

5.4 Data mining

Users who join the HGBC ecosystem can continue to receive HGBC Token rewards under the Proof of Data Stake (PODS) mechanism, called "data mining."
Since users have ownership and income rights to their own genetic and medical health data, the essence of “data mining” is a digital asset reward for this kind of equity and contribution.
Here, the concept of “computing power” is introduced. “Computing power” is a unit created to quantify this kind of rights and contributions. The digital assets obtained by “mining” in unit time are strictly positively correlated with “computing power”.
The mining "computing power" is obtained by accumulating various tasks involved by the user. The mining task can be obtained from the task box on the Gene Treasure app. Some examples of tasks are as follows:

- Bind multi-dimensional genetic data.
- Participate in the characterization questionnaire.
- Bind the medical examination institution and health management institution account.
- Bind the wearable device.

Proof of Data Stake (PODS) is a Token reward mechanism designed in combination with user “computing power” and participation.

Terminology

- **Reward period t**: The time between two adjacent awards is called the reward period. The system settles at 01:00 every day and issues tokens to users.
- **User computing power Un**: The effective computing power of the user under the current reward cycle. If the user does not receive the reward token within 120 hours, the user's computing power will be frozen, not counted in the whole network computing power, and no longer reward this user. After this user re-logs in to the APP to receive the reward, the computing power of this user is reactivated and becomes effective computing power.
- **Total network computing power**: the sum of the computing power of all valid users under the current reward cycle. A valid user is a user who is not frozen.
- **Total network Token production Rm**: HGBC will generate a constant amount of Token for a certain period of time according to a fixed ratio.
- **Reward Token Quantity Tn**: Each reward period, the user can get a certain reward token.

Parameter value setting

- **Total Token for data mining a**: 1.2 billion.
- **Reward period t**: 24h.
- **Decay period d**: The period of each production reduction, d = 365 days (~1 year).
- **Attenuation factor γ**: the ratio of each production reduction, γ = 30%.
- **HGBC Token Production c**: In the first cycle year, the reward Token generated in each reward cycle is approximately 1013698, with a total of 370 million rewards. The number of awards for

the nth year is as follows:

$$C_1 = 370{,}000{,}000$$

$$C_n = C_{n-1} - C_{n-1} * \gamma \ (n \geq 2, \gamma = 30\%)$$

Token Reward Rules

In a single reward cycle, the user's computing power Un is the sum of the computing power obtained by the user in completing each task. Calculated as follows.(Where Si is the score of the user's i-th task, ai is the weight of the i-th task, and the default weight is 1. Multiply Si by ai to get the power of the i-th task.)

$$U_n = \sum_{i}^{k} a_i S_i$$

Every day at 01:00, the HGBC network will issue reward tokens through the PODS mechanism. The number of Tokens actually obtained by the user is calculated according to the following formula.

$$T_n = \frac{U_n}{\sum_{s=1}^{t} U_s}(1-\sigma)R_m(\sigma < 0.02)$$

Among them, Rm is the total amount of Tokens issued during the current reward period; σ is the proportion of the rewards used in the ecological infrastructure (server, storage, etc.), and the specific value varies with the actual ecological expenditure, and σ is not more than 0.02. σ = 0 before the completion of the Thousand Human Genome Project; t is the total number of valid users within the ecosystem during the current reward cycle.

6. BUSINESS PROSPECTS

The vision of the HGBC project is to solve the problem of ownership and circulation of genomic data through blockchain technology, to change the business model centered on enterprises and scientific research institutions, to empower individuals, institutions and society, and to accelerate the mining of human genome data. As the scale of users grows larger and the number of ecological participants increases, the commercial value and social value of the HGBC project will gradually be released, and eventually achieve exponential growth.
HGBC's business income mainly comes from: directly providing users with genetic products/services developed by the HGBC team, sharing of genetic products/services developed by third-party developers, and service fees from research institutions and enterprises.

6.1 Sales of genetic products/services developed by HGBC team.

The genetic products/services in the APP are from third parties and from the HGBC team. The HGBC team has extensive experience in genetic product/service development. Prior to the HGBC project, the HGBC team had developed several gene products/services.
The HGBC team directly provided users with the genetic products/services developed by the HGBC team as one of the project's revenue sources.

6.2 Revenue share from Third-party developers providing genetic products/services.

In addition to the sales from genetic products/services developed by the HGBC team, the platform will charge a fee for sales revenue generated by third-party genetic products/services provided to users. In the early stage of HGBC ecological construction, in order to encourage more third-party developers to join the ecology, the platform will reduce the fee or even subsidize third-party developers.

6.3 Service fees from research institutions and enterprises.

The projects released by scientific research institutions and enterprises in the HGBC ecosystem include commercial projects and public welfare projects, and the platform will charge a certain service fee for commercial projects. For public welfare projects, the platform will not charge any fees.

7. TECHNOLOGY

7.1 Platform Architecture

The HGBC eco-platform technology architecture consists of three layers: the base layer, the core layer and the client layer. The base layer is divided into storage services and blockchain services; the core layer consists of mobile app and development platform.

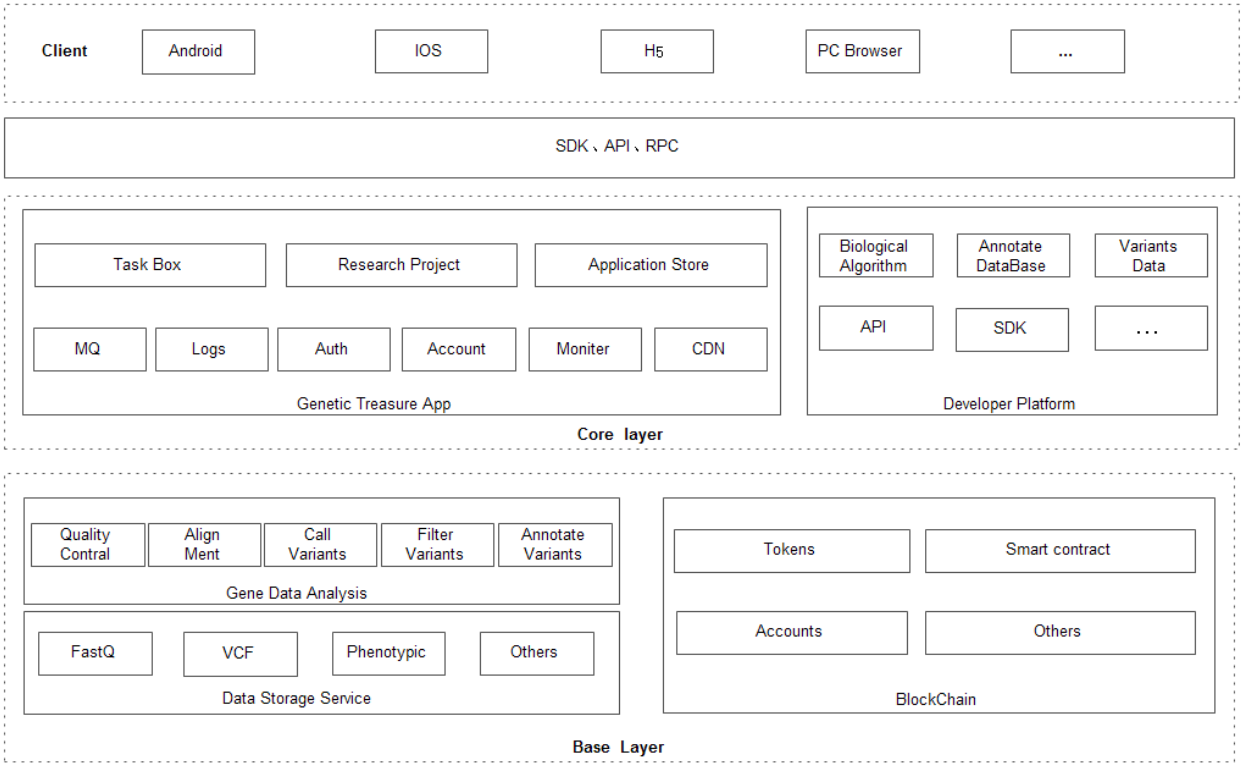

Figure 20: HGBC platform architecture

7.2 Genomic sequencing and analysis

In order to maximize the value of genomic data, the data is based on human genome-wide data (other types of sequencing data will be added later as needed). Human genome sequencing uses a high-throughput gene sequencer with a sequencing layer of no less than 30 layers and a yield of 90 Gb.

The high-throughput sequencing process is divided into the following steps:

- Sample collection (HGBC platform uses saliva samples)
- Extract DNA
- Construction of DNA sequencing libraries
- Sequencer sequencing
- Data processing

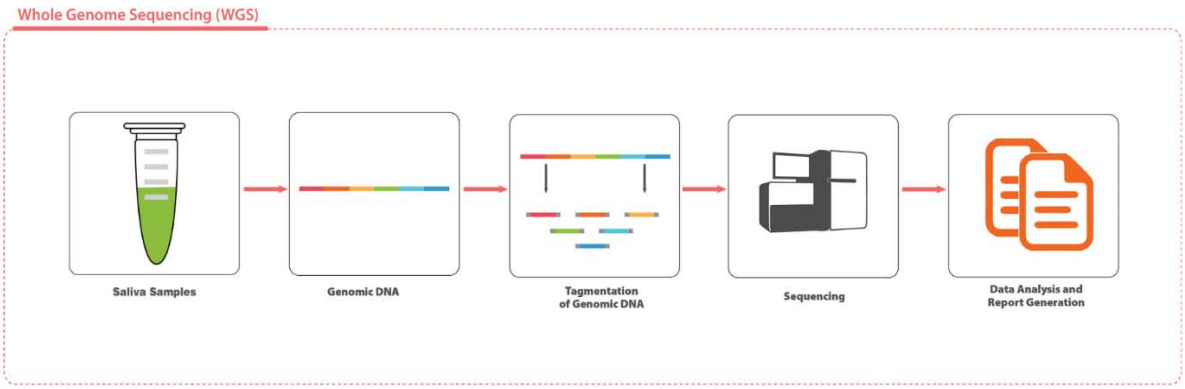

Figure 21: High-throughput sequencing process

In order to ensure the quality of genetic data, the basic operating mechanism of the gene sequencing process is as follows:

- The user selects the sample collector service provider independently through HGBC;
- The sample collector service provider obtains the order and random code with the sample (HGBC smart contract release), and the service provider pastes the sample random code to the sample collector and delivers it to the user;
- The user uses the mobile phone to scan the QR code to bind the sample and the collector, and the HGBC smart contract issues the Token to the sample collector service provider, and the user expresses the sample to the DNA extraction and library construction service organization(smart contract randomly selected);
- After the DNA extraction and library construction is completed, the smart contract randomly selects three DNA sequencing organizations, the sequencing service provider completes the sequencing and uploads, and the HGBC smart contract cross-validation verifies the authenticity and encrypts the data, and the genetic sequencing service provider obtains the Token reward ;

Note: If the sequencing service provider data is fraudulent, the smart contract (HGBC) will blacklist it and permanently discard it and publish the record on the whole network, returning the token to the user.

7.3 Storage and Ownership Confirmation

At present, human genome-wide resequencing uses a sequencing depth of 30 times, and the original genome file of FASTQ format is about 200G (about 60G after compression). In the current blockchain architecture, there is no good solution to the problem of large data storage, such as Ethereum's Storage persistent storage, in the form of key-value, where the length of the key and value are 256bit, blockchain Storage costs are very expensive and do not support direct storage of large amounts of data. For large data volume issues, the general solution is to write the Hash value of the file and the account information of the user into the blockchain, and the big data file itself is stored in the centralized storage service.

The HGBC platform needs to provide users with a safe, reliable, low-cost, decentralized data storage solution that allows users to control their own data. The centralized storage solution is not good for gene big data in the blockchain.

Therefore, we build a distributed storage architecture – HGBC Storage, based on blockchain storage services, the bottom layer uses encryption computing, all data is transmitted through the encrypted channel, no need to worry about privacy leaks. The customized Paragon file protection system protects every piece of data of the user, ensuring the security of the hard disk data and truly becoming the user's intimate data steward. After the sequencing analysis of the genomic data is completed, the output data is segmented according to a fixed size, and the segmented data block is encrypted, and the encrypted data blocks are distributed and stored in different nodes of the HGBC Storage, and each data block is obtained at the same time. Hash value of different blocks is organized by the data structure of "Merkel tree", the root node of the tree (also called the data fingerprint) can be used to uniquely restore the genomic data itself. HGBC Storage encrypts the data fingerprint of the genome using the public key of the user blockchain account and records it on the blockchain for broadcast on the entire network. The encrypted data fingerprint can only be decrypted by the user through its own private key, thereby restoring the data.

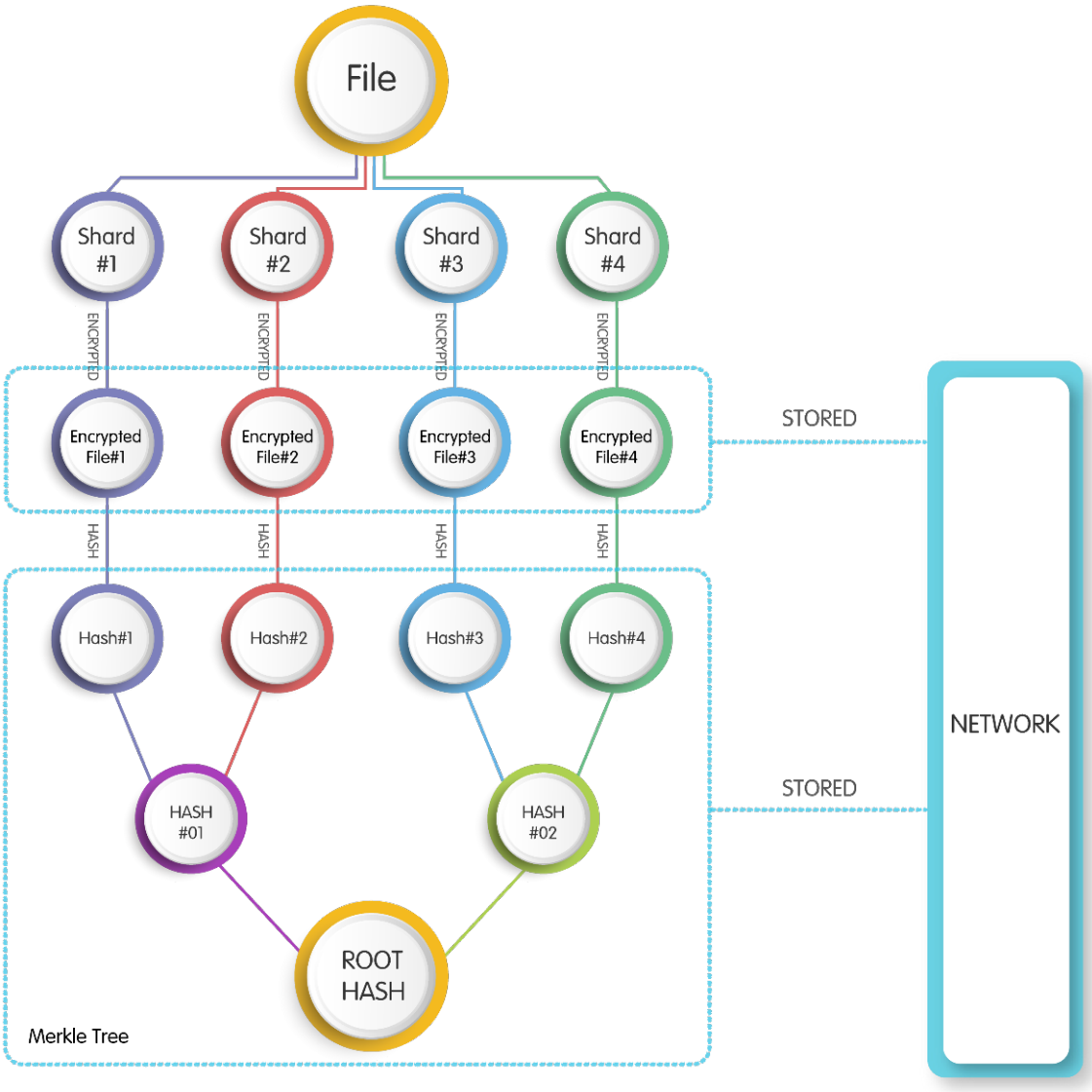

Figure 22: Genome data storage process

HGBC Storage ultimately wants users to store and manage their own data. The user accesses the genomic data to his own device. Through sharing and authorization, the HGBC ecosystem accesses the data and records the data usage record in the blockchain. The user has full control over his own data.

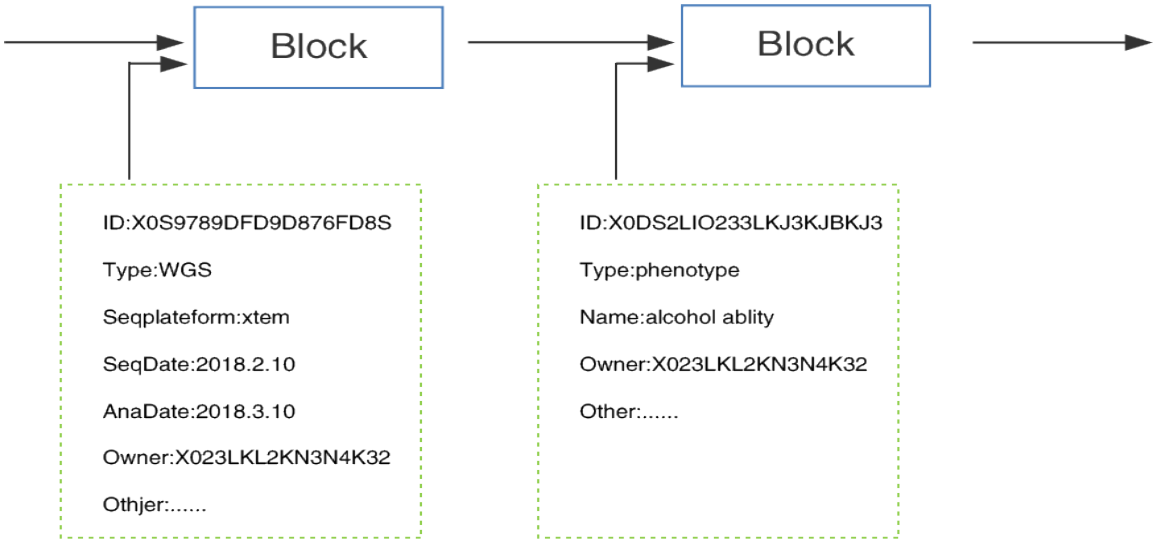

Figure 23: Genomic data chaining confirmation

7.4 Data Security Protection

The HGBC ecosystem includes multiple roles such as individual users, project parties, service providers, and platform parties. User personal data flows through multiple roles, and user privacy protection is especially important. The privacy protection of HGBC eco platform users includes the following four aspects:

7.4.1 Pseudo-anonymous privacy protection

After the user joins the HGBC eco-chain, there will be a unique nickname for the whole network. All the actions of the user in the ecology are operated by the nickname of the user identifier, and the user's real identity is effectively isolated from the chain identity in the ecology. If the user participates in the research project, the project party can see the user's nickname.

7.4.2 Data encryption protection

The user's genomic and phenotypic data is stored in encrypted form, and the user has complete control over his/her data.

7.4.3 One-way anonymity mechanism

In the HGBC ecosystem, the identity of the sequencer, project party, and service provider is verified

by the HGBC Foundation certification, and is completely transparent to the user, and the user's identity information is pseudo-anonymous, and all transaction information will be permanently Stored in the blockchain.

7.4.4 User authorization mechanism

The transaction service involving user data in HGBC needs to be authorized by the user. If the project party obtains user data, the platform will send a station letter or text message to the user, and the user can authorize it to proceed.

8. STRUCTURE: Decentralized Autonomous Organization (DAO)

8.1 Organizational Structure

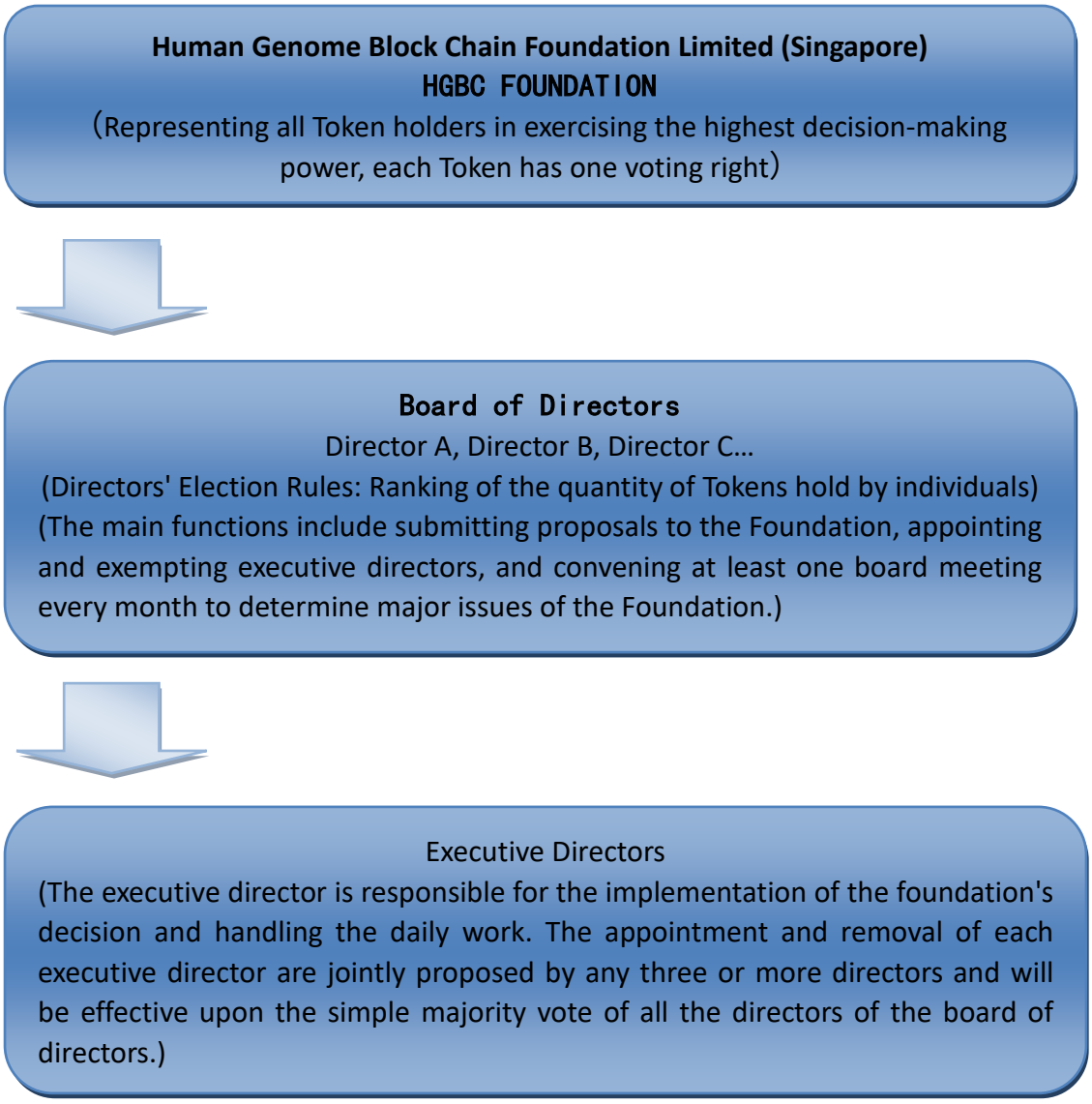

Figure 24: Organizational structure

8.1.1 HGBC Foundation

The HUMAN GENOME BLOCK CHAIN FOUNDATION LIMITED registered in Singapore represents the decision-making power of all token holders.

8.1.2 Board of directors

The permanent establishment of the Foundation is a board of directors composed of several directors. (Directors' Election Rules: Ranking of the quantity of Tokens hold by individuals).

8.1.3 Executive directors

The board of directors has a number of executive directors (the appointment and removal of executive directors is jointly nominated by any three or more directors, and is passed by a simple majority vote of all directors). The executive director is responsible for the Foundation and daily work of the project.

8.1.4 Monthly meeting

A board meeting is held at least once a month, with an executive director presiding over the meeting and reporting to all directors on the progress and financial status of the Foundation.

8.1.5 Auditing

After the first phase of financing is in place, the Foundation will hire an accounting firm with credibility to issue audit reports on a regular basis.

8.2 Operational Mechanism:

"Proposal-Voting-Execution-Acceptance" (the whole process is visualized in the APP)

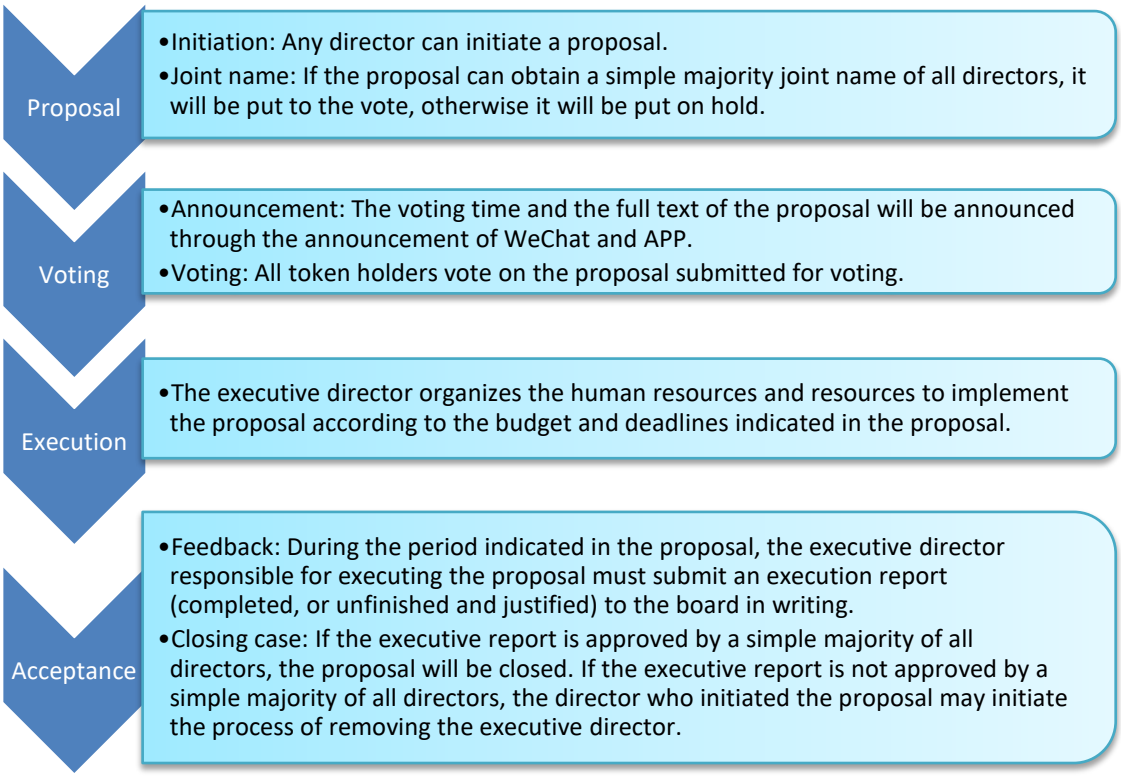

Figure 25: Decision Mechanism

8.2.1 "Proposal"

Any director of the board of directors may initiate a proposal according to the general template (including the content of the event, the budget ceiling, the deadline for completion, and the executive director who assumes responsibility). If it is able to obtain a simple majority joint name of all directors, it shall be submitted for voting, otherwise it shall be put on hold.

8.2.2 "Voting"

After the WeChat and APP announcements, all token holders vote on the proposal submitted by the board of directors within a given time, each token has one voting right, if the number of voting rights in favor is greater than the number of voting rights against , the proposal was approved and submitted for execution, otherwise it was put on hold.

8.2.3 “Execution”

The executive director organizes the Foundation's human and financial resources to implement the proposal according to the budget ceiling and completion deadline indicated by the proposal.

8.2.4 “Acceptance”

During the agreed completion period, the executive director responsible for executing the proposal shall submit an execution report (completed, or unfinished and justified) to the board of directors in writing. If the executive report is approved by a simple majority of all directors, the proposal will be closed. If the executive report is not approved by a simple majority of all directors, the director who initiated the proposal may initiate the process of removing the executive director.

9. PROGRESS AND WORK ARRANGEMENTS (2018.06)

9.1 Summary of previous work

After a year of small-scale operation, HGBC has completed the whole process closed loop from product, platform, technology and operation, and achieved the following results:

- The client APP is online and put into use, with tens of thousands of active users and thousands of sequencing users (users who have completed genome-wide sequencing);
- The back-end technology platform completes the overall framework construction;
- Development and commercialization of custom sequencing products based on blockchain;
- Initial establishment of the sales system;
- Established an operations team to meet basic needs;
- The exploration of the token application case.

9.2 Next stage work plan

9.2.1 Technical improvement

- Phase II development of the HGBC platform;
- Client 2.0 development;
- Development of distribution agent CRM management system;
- Customized development of sequencing products;
- Third-party developer API interface and development platform;
- Institutional version of APP development;
- Expansion and optimization of the sales system.

9.2.2 Team expansion

- Expand the HGBC operations team;
- Expand the product distribution agency team;

9.2.3 Business Development

- Closed loop of the third-party developer system of gene products/services, through 3-5 years of efforts, to achieve more than 10,000 gene applications on the platform;
- To become a common platform for the relevant industries in the genetics industry, and through 3-5 years of efforts, the number of companies in the relevant industries related to the access platform will exceed 3,000;

9.2.4 Market financing

(1) HGBC adheres to the principle of “doing things first, refinancing”. Under the premise that the team has already completed the first stage of work, the HGBC will start financing in the private equity market to expand capital investment and start the next stage of work.
(2) Project management and project operation should be aligned with the mainstream blockchain project. Tokens should enter the mainstream trading market of the blockchain, and the token value will be gradually released.

10. TEAM

Figure 26: Hao Xiangwen (founder)

Mr. Hao Xiangwen, Master of Genomics, Beijing Genomics Institute, Chinese Academy of Sciences, CEO of Beijing Ji Yun Hui Kang Technology Co., Ltd., is one of the earliest pioneers in the genetics industry in China. He has 16 years of experience in the genetics industry. He was in Huada Gene and Yi Mei Tong De, Illumina and other industry organizations. Focusing on human genomic data analysis, interpretation and gene application development to release the value of genomic data and apply it to healthy living scenarios.

Figure 27: Ren Wenqi

Dr. Ren Wenqi, PhD in Economics from Peking University, Master of Law from Peking University, and Executive Director of Beijing Baohai Baoshan Data Technology Co., Ltd.

Figure 28: Li Jia

Mr. Li Jia, former senior product manager of Alibaba Group and founder of Zhong Liu Ji Shui Culture Communication, has many years of experience in the Internet industry and is well versed in the design and operation of Internet product rules.

The other members of the HGBC team are comprised of seniors in genetics, biology, computing, bioinformatics, and blockchain backgrounds, and have extensive experience in genome-wide data analysis, interpretation, and gene application development.

Strategic advisory team

Figure 29: Chen Lei

Mr. Chen Lei, CEO of Thunder, CEO of Network Technology. He holds a bachelor's degree in computer science and technology from Tsinghua University and a master's degree in computer science from the University of Texas at Austin. He has served in Google and Microsoft, and served as general manager of Tencent Cloud and deputy general manager of Tencent Open Platform. In 2017, he served as CEO of Thunder. In 2014, he was the most influential person in China's cloud computing in the year. In 2015, he was named an outstanding figure in the Internet industry.

Figure 30: Zheng Jie

Mr. Zheng Jie, President of Shulan Medical Management Group, Dean of Shulan Medical Research Institute of Shanghai University, Executive Vice President of Zhejiang Digital Medical and Health Technology Research Institute, Deputy Director of Precision Medical Center of California International Nanotechnology Research Institute, Open Medical and Health Founder of OMAHA of the Alliance, Member of the Standing Committee of the Health Information Big Data Industry Development and Information Security Committee of the China Health Information Society, Vice Chairman of the International Digital Medicine Society of the Digital Medicine Branch of the Chinese Medical Association, Vice President of the Hospital Management Branch of the China Non-Public Medical Institutions Association, Zhejiang Province Member of the Standing Committee of the Digital Medical Branch of the Medical Association and member of the China Internet of Things Medical Committee. "Subversion of the medical" recommended author, "Future Medical", "Digital Medical" translator.

Figure 31: Yu Guoliang

Dr. Yu Guoliang, Director of the Baihua Association, Founding President of the American Chinese Biomedical Science and Technology Association, President of the Zhejiang Chamber of Commerce in the United States and Executive Chairman of the China PDX Industry, Research and Research Alliance. Ph.D., University of California, Berkeley, postdoctoral fellow at Harvard Medical School. Dr. Yu Guoliang co-founded nearly 10 biotechnology companies and served as chairman, and Epitomics

was acquired for $170 million. In 2013, he joined the China-US Crown Section as chairman. Yu Guoliang has rich practical experience in the creation of major new drugs, and has established five new drug research and development companies in China and the United States.

Figure 32: Wang Xuegang

Mr. Wang Xuegang, co-founder of Shell Society, chairman of Jin Zhun Gene. One of the scientists involved in the Human Genome Project, Vice Chairman of the China Translational Medicine Alliance, Member of the Translational Medicine Branch of the China Biomedical Technology Association, Standing Committee Member of the 7th Committee of the Life Electronics Branch of the Chinese Institute of Electronics, and the “Thousand Talents Program” Evaluation Expert of Zhejiang Province Member of the China Association.

Figure 33: Han Cao, Ph.D.

Mr. Han Cao, Founder and Chief Scientific Officer, BioNano Genomics, Inc. He is the expert in the single molecule analysis, complex genome assembly and biomedical relevant structural variation discovery field, such as the currently best assembled individual human/Asian genome.

11. ECOLOGICAL PARTNERS

11.1 Sequencing Service Provider

11.2 Interpretation service providers

11.3 Health Management Service Provider

全科Q-Care

诏曰

YENKUNG+
言康

12. DESCRIPTION

As projects and technologies evolve, the content scenarios described in this white paper are subject to change. We will update the white papers but do not commit to updates and release times.
This white paper is only intended to communicate the HGBC program to the public and is not used as a legal basis for any individual or institution to invest in an HGBC project.
HGBC officially does not authorize any third party agency or individual to raise funds for HGBC.

13. CONTACT US

Website: http://www.hgbc.io

APP download: https://fir.im/hgbc

Email: service@hgbc.io

Telegram: https://t.me/hgbc_cn

Wechat: