

**HGBC** Beta 版 0.8

Human Genome Block Chain

基因价值生态链白皮书



摘要.....	1
一、项目背景.....	1
1.1 人类对基因的探索 .....	1
1.2 基因数据的应用 .....	3
1.3 现实存在的问题 .....	5
1.3.1 普通用户 .....	5
1.3.2 研究机构 .....	6
1.3.3 服务方 .....	7
二、我们的愿景.....	8
三、生态体系.....	9
3.1 HGBC 团队 .....	10
3.2 用户 .....	10
3.3 测序服务商.....	10
3.4 项目方 .....	11
3.5 服务方 .....	11
四、产品方案.....	11
4.1 基因宝藏 APP .....	11
4.1.1 任务盒子 .....	12
4.1.2 研究所 .....	12
4.1.3 宝藏商店 .....	13
4.2 数据分析云平台 GAIC .....	17
4.3 第三方开发平台 .....	19
五、激励计划.....	20
5.1 奖励池 .....	20
5.2 测序早鸟奖励 .....	21
5.3 数据挖矿 .....	21
六、商业前景.....	23
6.1 基因应用交易 .....	23
6.2 服务交易分成 .....	24
6.3 项目方服务费 .....	24
七、技术说明.....	24
7.1 平台架构 .....	24
7.2 基因组测序与分析 .....	25



---

7.3 存储与确权.....	26
7.4 数据安全保护 .....	28
八、HGBC Token .....	29
8.1 应用场景.....	29
8.2 分配方案.....	30
九、发展路线.....	31
十、HGBC 核心团队.....	32
十一、战略顾问团队 .....	35
十二、生态合作伙伴 .....	38
12.1 测序服务商 .....	38
12.2 解读服务商 .....	38
12.3 健康管理服务商 .....	38
十三、风险提示.....	39
十四、联系我们.....	39



# 摘要

基因是生命遗传信息的载体，基因数据作为生命的基础数据是每个人重要的数据资产，在营养饮食、运动健身、美容护肤、健康管理、药物研发、亲权确认、科学研究等方面具有重要应用。全基因组测序得到的是一种通用的基因数据，可用于各种基因检测服务，而这些数据被基因检测企业所把控，用户不能掌控自己的数据，在不同企业之间共享获得服务。另一方面，以生物制药企业、生物公司、科研机构等为代表的项目方对基因数据以及多维度的医疗健康数据有强烈的需求，却没有可信的数据源。在此情况下，HGBC（Human Genome Block Chain）项目通过引入区块链技术，把基因组数据完整确权给个人，让每个人切实拥有自己数据的所有权，同时，利用分布式存储和加密技术解决数据的隐私和安全问题。通过解构行业分工，形成以用户为核心，测序商、服务提供方、项目方参与的新生态。用户数据作为生态中的数字资产，在医疗健康、科学研究等领域流动，让个人获得数字资产收益的同时，加快生命健康的探索，造福人类社会。

## 一、项目背景

### 1.1 人类对基因的探索

基因，是遗传物质的基本单位，也是一切生物组成的基础。人类基因组是指人所携带的全部遗传信息，由 23 对染色体组成，包含 30 多亿个碱基对，拥有超过 21000 个基因。基因组不仅可以通过复制将遗传信息传递给下一代，还可以使遗传信息得到表达，不同人种之间头发、肤色、眼睛、鼻子等不同就是由基因差异所致。

人类对基因的探索不足百年，却取得了瞩目的成就。在 2000 年，人类第一份全基因组草图绘制完成，该项目被称为“人类基因组计划”，由美国、英国、法国、德国、日本和中国科学家组成，历时 11 年，共计花费 30 亿美元。随后，测序技术飞速发展，2014

年，Illumina 公司发布 HiSeq X Ten 高通量基因测序仪，使全基因组测序成本下降到 1000 美元以内。测序成本的下降趋势远超摩尔定律<sup>1</sup>。如图 1 所示。

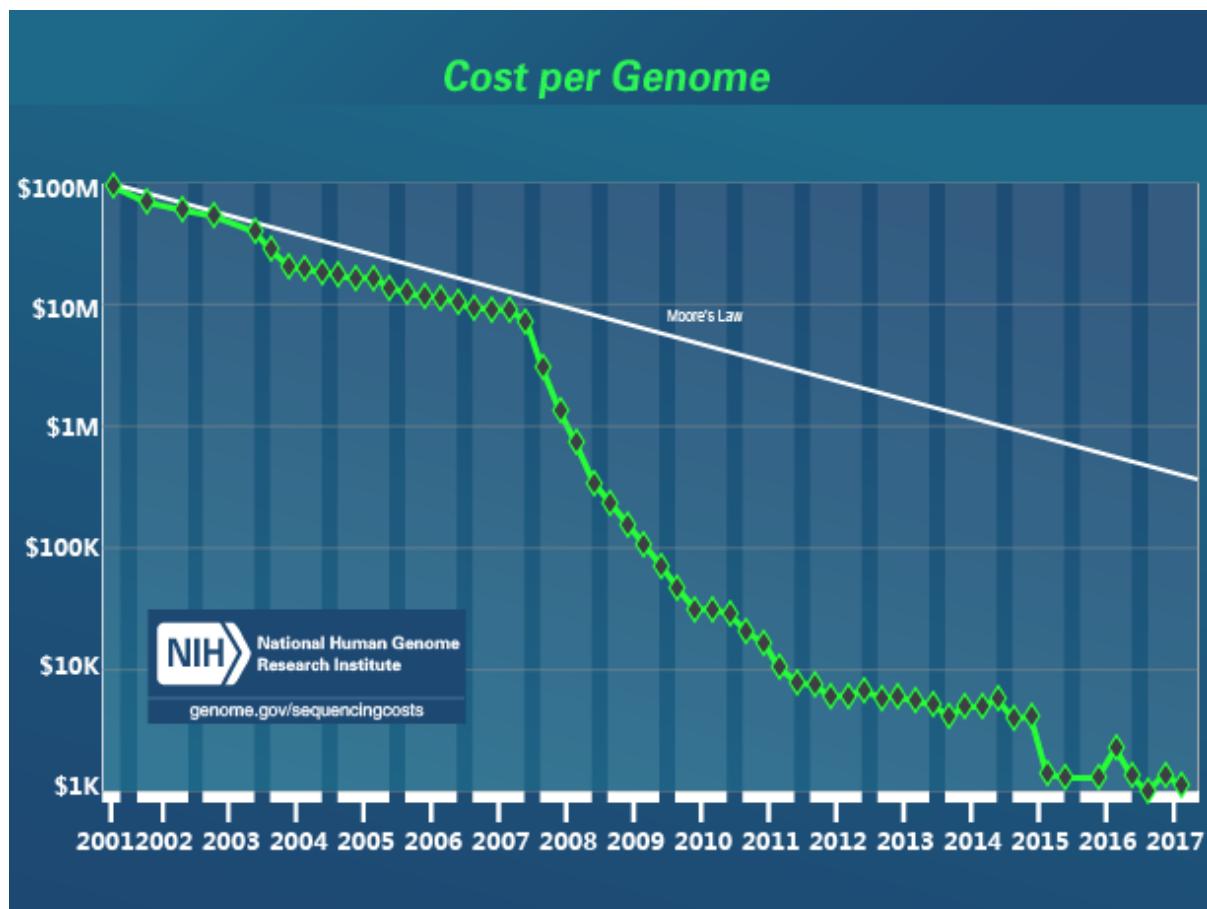


图 1 测序成本下降趋势

测序成本的降低使大规模人群基因组测序成为可能，而这正是精准医疗的基础。2015 年，美国率先发起了“精准医疗”计划，一期投入 2.15 亿美元，希望招募 100 万名志愿者进行基因组测序<sup>2</sup>。中国紧随其后，由中国科学院北京基因组研究所领衔发起“中国十万人基因组计划”<sup>3</sup>，中国江苏启动“百万人群基因组测序计划”<sup>4</sup>，通过队列研究和百万人基因组 DNA 测序，建立中国人群特有的遗传信息数据库。

<sup>1</sup> <https://www.genome.gov/27541954/dna-sequencing-costs-data/>

<sup>2</sup> <https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>

<sup>3</sup> <http://tv.people.com.cn/n1/2017/1229/c61600-29736110.html>

<sup>4</sup> <http://leaders.people.com.cn/n1/2017/1029/c58278-29614857.html>



## 1.2 基因数据的应用

测序成本的下降，不仅推动了各国家政府的“精准医疗计划”，也推动了个人基因检测市场的快速发展。面向个人的基因检测产品覆盖了人的全生命周期，包含孕前基因筛查遗传病、孕中胎儿染色体非整倍体疾病筛查、新生儿相关疾病检测、生长发育指导、疾病风险评估、健康管理、用药指导、康复监测、饮食健康等多种应用场景。

**孕前基因检测：**隐性遗传病是指致疾病的基因在两条染色体上都具备时，携带该基因型的个体才会发病；而只有一个致病基因型时，该个体只是携带者，并不发病。当父母都是隐性遗传病基因的携带者时，只有 1/4 概率生下不含致病基因的健康宝宝。通过孕前基因检测筛查，可以有效知晓准父母是否携带有相关隐性遗传病的致病基因，从而为生育健康宝宝做好充足准备。如果发现父母均携带有隐性遗传病基因，可以通过目前成熟的辅助生殖技术获得不再携带致病基因型的后代，阻断相关疾病在家族内的传递。

**孕期基因检测：**无创 DNA 产前检测通过对母体静脉血游离 DNA（包括胎儿游离 DNA）进行基因测序，可以准确检测胎儿是否患有唐氏综合征（T21）、爱德华氏综合征（T18）、帕陶氏综合征（T13）三大染色体非整倍体疾病<sup>5</sup>。

**新生儿耳聋基因检测：**中国共有听力障碍患者 2780 多万，占全国 8269 万残疾人的 33%，数量居全国各类残疾之首，每年新增 3 万多聋儿<sup>6</sup>。耳聋与遗传存在很大关系，可以分为因遗传基因引发的迟发型渐进性耳聋和药物性耳聋。对大前庭导水管综合征新生儿，给予生活指导，降低耳聋发生；对药物性耳聋携带新生儿，给予用药指导，预防因药致聋；对耳聋基因突变携带新生儿，可指导其未来婚配生育，预防后代发生耳聋。

**健康风险评估：**是指对个人或群体的健康状况及未来患病风险性的量化评估。个人进行健康风险基因检测，可以更好了解自己的真实健康风险，指导自己改变或修正不健康的行为。2013 年，美国演员安吉丽娜·茱莉发表文章《我的医疗选择》<sup>7</sup>：“...我是 BRCA1 基因携带者，这种基因存在着缺陷，会导致罹患乳腺癌和卵巢癌的风险大幅提高。尽管不同女性存在着个体差异，但据医生估计，我患乳腺癌的几率高达 87%，患卵巢

<sup>5</sup> <https://www.sciencedirect.com/science/article/pii/S0140673697021740>

<sup>6</sup> <http://finance.sina.com.cn/roll/2016-11-08/doc-ifxxnffr7130870.shtml>

<sup>7</sup> <https://www.nytimes.com/2013/05/14/opinion/my-medical-choice.html>



瘤的几率也达到 50%。只有很少一部分乳腺癌病例是由遗传性的基因突变引起的。对于 BRCA1 基因缺陷携带者来说，平均有 65% 的几率患上乳腺癌”。基因检测在癌症风险评估中发挥了重要作用，通过个人风险评估的结果，有针对性的采取预防方案。

**精准用药**：根据个体的差异选择正确的药物，正确的剂量给药。华法林是临床应用最广泛的口服抗凝药物，人们发现编码华法林代谢和药效相关酶的基因存在个体差异和种族差异。以遗传药理为基础的药物剂量调整得到各方面的重视，FDA 修改说明书建议华法林在用药前应做 CYP2C9、VKORC1、CYP4F2 等相关基因型检测，确定个体用药差异<sup>8</sup>。

**身份鉴定**：人体细胞有总数约为 30 亿个碱基对的 DNA，每个人的 DNA 都不完全相同，人与人之间不同的碱基对数目达几百万之多，因此通过分子生物学方法显示的 DNA 图谱也因人而异，由此可以识别不同的人。通过对 DNA 鉴定还可以判断两个人之间的亲缘关系。

**健康生活**：基因上的差异也影响着每个人生活的方方面面。对运动健身来说，ACTN3 基因与爆发力有关<sup>9</sup>；PAPSS2 基因与运动积极性有关<sup>10</sup>；BDKRB2、LPL 基因与有氧运动减重效果有关<sup>11</sup>等。对于日常饮食，FTO 基因异常容易导致肥胖<sup>12</sup>；APOA2 基因突变的人不适合高脂肪食物；ALDH2 和 ADH1B 基因影响着酒精代谢能力<sup>13</sup>；APOE E4 基因型的人摄入鱼肝油会有血脂升高的风险<sup>14</sup>；TMEM18，MC4R 基因上的变异与抑郁-暴饮暴食相关<sup>15</sup>。只有从基因层面认识自己，才能更好更健康的生活。

<sup>8</sup> <https://www.pharmgkb.org/label/PA166104776>

<sup>9</sup> 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.

<sup>10</sup> 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.

<sup>11</sup> 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.

<sup>12</sup> Wang L, Yu Q, Xiong Y, et al. Variant rs1421085 in the FTO gene contribute childhood obesity in Chinese children aged 3~C6years[J]. Obesity research & clinical practice, 2013, 7(1): e14-e22.

<sup>13</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3860439/>

<sup>14</sup> 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.

<sup>15</sup> 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.



## 1.3 现实存在的问题

快速发展的基因测序技术和大幅下降的测序成本，极大的刺激了个人基因检测市场，推动了生命科学的研究。在基因行业从业十多年的我们，不仅看到了市场的快速发展，也看到了行业中存在的各种问题。

### 1.3.1 普通用户

对普通用户来说，受限于基因检测的专业知识，存在以下问题：

#### 1、基因数据确权机制缺乏，用户权益无法得到保障。

基因数据属于医疗健康大数据的重要组成部分。医疗大数据能够获得新知识，实现新决策，创造新价值，因而医疗大数据显然具有财产属性，能够成为权利的客体。医疗大数据所产生的权利属于信息权，在民法权利分类上属于物权之一种，具备物权的四项权能：占有权、使用权、收益权、处分权。具体来说，医疗大数据的所有者享有对大数据信息的排他的占有权；享有自行使用或许可他人使用的使用权；享有因自行使用或许可他人而产生的收益权；享有将数据转让给他人或抛弃的处分权。

用户对基因数据的所有权没有广泛的机制保障，基因检测企业常把用户的基因数据保存在自己的服务器上，以期获取更大的价值；同时，由于基因数据本身应用所需要技能专业性强，普通用户即使能够下载到自己的基因数据也无法直接分析使用。用户在基因数据的价值变现上处于无奈的境地，不能够享受自己基因数据的收益。2015年1月6日，生物制药巨头 Genentech 公司购买了基因检测公司 23andMe 3000 名帕金森患者的信息，通过采集全基因组数据，试图鉴定该疾病的靶点，并为此支付 6000 万美元，作为中间商的 23andMe 获得了所有收益<sup>16</sup>。

基因数据作为每个人生来拥有的一项财产，由于数据采集和处理能力的限制，被中心化的机构所占据，数据的名义所有权和实质支配权相分离。使用区块链技术将基因数据确权给个人，本质上是将数据的所有权和支配权归还给个人。

<sup>16</sup> <https://www.fiercebiotech.com/r-d/23andme-strikes-60m-genentech-deal-to-continue-pivot-to-database-driven-r-d>

## 2、基因检测过程繁琐，周期较长。

用户购买一份全基因检测产品到最终拿到检测结果，会经历如下过程：

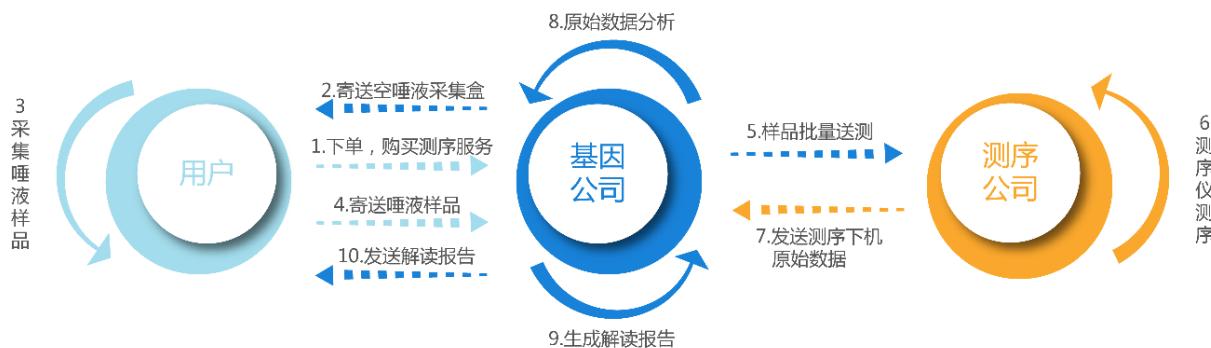


图2 基因检测流程图

整个过程大约 30 天左右，过长的时间周期严重影响了用户的产品体验，同时，测序本身费用占据了产品很大一部分成本，而全基因组测序得到的基因数据是一种通用的标准数据，可以应用到不同的基因检测产品中。目前缺乏可信的数据共享机制和平台，用户购买不同公司的基因检测产品时，都有重复的进行采样、测序、分析等过程，不仅增加了用户数据泄露的风险，同时也造成了用户不必要的经济支出。

以用户为核心，以技术为保障，建立可信的生态平台，用户只需一次测序，就能享受不同服务方提供的各种基因检测服务，提高数据的复用性，为用户带来切实利益。

### 1.3.2 研究机构

对于科学的研究和药物研发机构来说，存在以下问题：

#### 1、科研样本数据获取难度大，周期长。

对于以人群基因组数据为基础的科研项目，第一步便是招募志愿者，如美国“精准医疗计划”，从 2015 年开始，累计投入 2.15 亿美元，用于招募 100 万志愿者，建立全国研究队列。志愿者的招募筛选是一切研究的起始和基础，在花费大量人力物力和时间成本，成功招募到志愿者后，需要对志愿者进行 DNA 采样、基因组测序、数据分析等标准流程，之后才进入项目的核心研究阶段。

如果有一个聚集大量用户和用户数据的生态平台，研究机构只需从用户手里购买标准化的数据就可以直接用于研究，大大节约了成本，提高了效率。



## 2、样本量小，数据维度少，表型数据不完善不规范。

精准医学、精准健康管理等不断发展，对基因组数据、医疗健康数据的价值挖掘和应用提出了更高的要求。人类 21000 多个基因，还有不少功能是未知的，很多疾病、性状、行为、心理等相关的遗传因素还没有找到，很大程度上是因为数据的样本量不够大，数据的维度不够丰富，数据的时间线不够长。一般研究项目的样本量从几十个到几万个不等，研究数据的维度往往在基本的表型数据与基因变异，基因表达或甲基化中一个维度或两个维度之间的分析，由于样本获取困难，样本的数据基本都是少数的时间点的断点数据。

## 3、用户参与度低。

在基因技术发展和基因数据应用的过往历程中，基本是以科学的研究机构、药物研发机构为中心的运行模式，个人往往被动参与，甚至自己的样品被用于科学研究而不知，用户不能够直接从数据中受益，进而导致用户的参与度低，项目研发阻力大，效率低。

要实现海量样本数据汇集、全维度数据采集、长时间数据汇总，就必须对数据采集和应用的传统中心化机构模式进行革新，变成以个人为中心的运行模式：

- 让个人用户获得奖励后积极参与进来，从而获得海量样本；
- 让个人用户能够在贡献更多维度数据后获得更多奖励；
- 让个人用户能够长时间提供医疗健康数据后获得更多奖励；

### 1.3.3 服务方

对于基因产品和服务提供方来说，存在以下问题：

#### 1、样本数据量小，缺乏表型数据，产品研发成本高。

对于大多数面向大众基因产品和服务提供方来说，并没有雄厚的资金实力招募志愿者进行产品研发，而基因产品的研发又离不开大规模人群的样本量和多维度的表型数据。基因检测公司目前普遍做法是先将小数据量下开发的不成熟产品推向市场，再根据不断沉淀的数据迭代产品。

#### 2、大众对基因认知程度低，用户群体小。



虽然基因是生命的基础组成，但普通大众对基因的认知程度比较低，基因检测产品不被大众所接受，只有一小部分的人购买过基因检测产品。对于基因检测产品提供方来说，不仅需要专注于研发产品，还需要承担起教育大众，培育市场的职责。

### 3、产品周期长，体验差。

考虑到用户数据的隐私和安全性，加之中心化企业驱动力不足，用户的基因数据在各个基因检测公司之间不能流通，形成了数据孤岛。基因检测公司每获得一个新用户，都需要经历复杂的过程和漫长的周期才能获得用户的基因数据，进而提供相应的服务。这种模式下，不仅损坏了产品的体验，也限制了产品的快速迭代。

## 二、我们的愿景

随着基因组测序技术的飞速发展，个人基因组测序的成本快速降低，人人可及的基因组数据将会成为每个人重要的数字资产。基因组数据在个人拥有所有权并可以自由、便捷支配后，其经济价值和社会价值将以个人为中心实现，变革以往以企业、科研机构为中心的模式，实现平等、民主、高效的运营模式。

HGBC 通过分布式存储、加密技术、区块链方案实现个人基因组数据及相关医疗健康数据的存储和确权，解决用户对数据的安全、隐私之忧，从而为个人积极参与到基因组测序、基因组数据应用、医疗健康表型数据的采集扫清技术障碍，为个人、机构和社会赋能，加快人类基因组数据的宝藏挖掘，实现让 1 亿人健康活到 121 岁的愿景。

对于个人，HGBC 将帮助每个人切实拥有自己的基因组数据和医疗健康数据，能够自主决定参与到多种科学项目中去，既可以获得经济利益的回报，还能够享受研究成果的科学红利。基于基因组数据的“GPS”导航匹配功能，个人将能够直接以最低成本、最快时间获取最适合自己的医疗健康服务方案，从而实现健康、长寿，优化人生的目标。HGBC 将提供开发环境和配套激励机制，促进基于基因组数据的基因应用开发，目标是未来 10 年平台拥有 10000 种基因应用。

对于机构，HGBC 将帮助药物研发、科学研究、保险经营、健康管理等机构直接连接上个人用户，经过用户允许可以快速获得最全面、最完善的基因数据及医疗健康数据，为

其自身业务的研究、产品研发提供大数据的支持，提高研发效率，降低成本，创新业务模式。

对于社会，HGBC 确权基因数据给个人后，赋能个人用户平等参与到科学技术的发展创新中，并能够获得经济收益和技术进步收益，能够享受到最匹配其自身基因信息的医疗健康服务方案，降低经济成本和时间成本的支出。社会整体在科学的研究、医疗保险、社会保险的支出将随着 HGBC 生态建设的完善逐步降低，运营效率提升，使得基因组数据成为未来健康生活的基础设施，提供公共服务的各国政府也将因此受益。

### 三、生态体系

HGBC 生态是由 HGBC 团队主导，以用户为核心，测序服务商、项目方、服务方共同参与建设的基因价值生态平台。

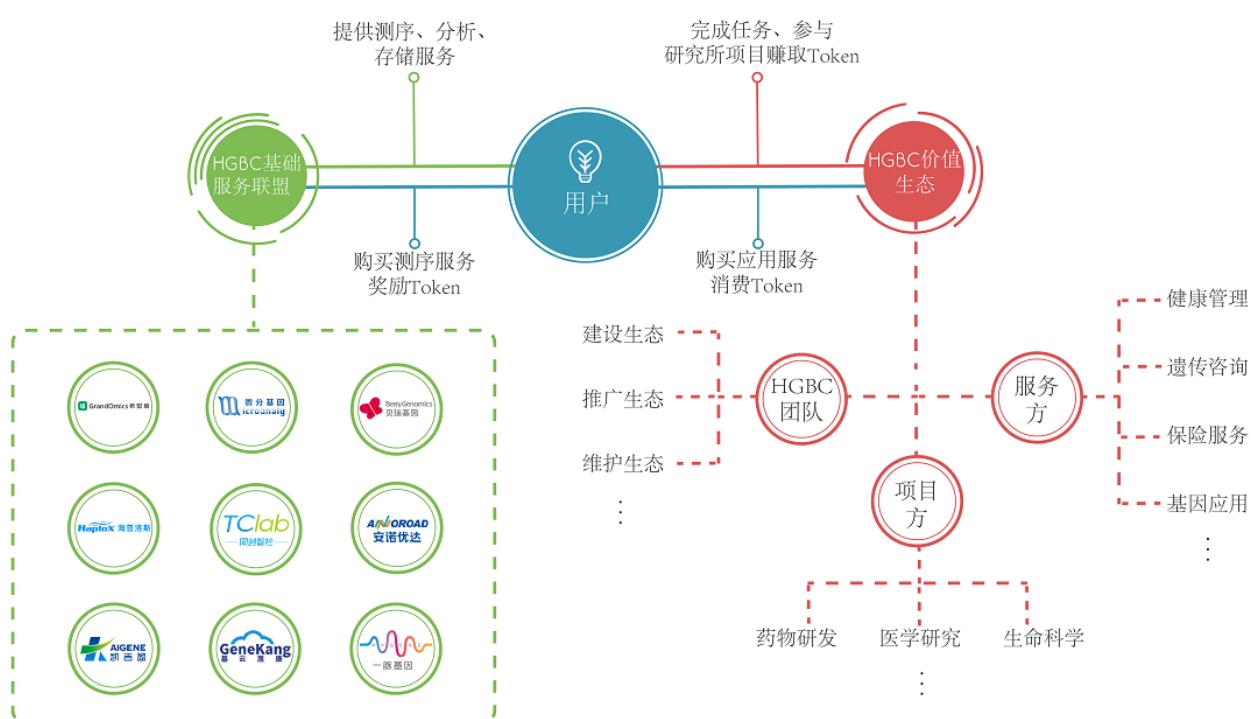


图 3 HGBC 生态体系



### 3. 1 HGBC 团队

HGBC 团队是项目发起方，同时也是 HGBC 生态建设的开发者和维护者。为了维护 HGBC 生态繁荣，团队一方面积极探索区块链技术变革基因行业生产关系，一方面积极与行业上下游企业建立联盟，推动生态建设。除此之外，团队充分发挥在基因行业中的技术优势，积极扶持第三方企业以基因数据为基础，开发丰富的基因应用。

### 3. 2 用户

用户是 HGBC 生态的核心。在 HGBC 生态中共有两类用户，一类是拥有自己基因组数据的 X 用户，一类是还未拥有自己基因组数据的 W 用户。

X 用户可以参与到项目方发起的研究项目中，获得 Token 收益，也可以使用 Token 购买服务方提供的各种基因应用和服务。用户除了会获得生态平台提供的一次性测序赠送 Token，还会获得更高的 Token 奖励。

W 用户通过购买测序服务商提供的测序服务，获得自己的基因组数据成为 X 用户，获得 Token 奖励，也可以通过完成任务盒子中的任务，获得 Token 奖励。

### 3. 3 测序服务商

测序服务商是拥有基因组测序能力的厂商，使用全球先进的测序仪提供测序服务。测序服务商经 HGBC 基金会认证审核，只有具有合格资质的企业才能加入到生态系统，为用户提供测序服务。HGBC 基金会会定期从测序周期、测序数据质量、服务质量等多方面对测序服务商进行审核，对于评分较低的测序服务商，基金会有权暂停或取消其服务资格，保障用户权益。



### 3. 4 项目方

项目方是指对基因和医疗健康数据有需求的企业，可以是药物研发企业、生物科技企业、大学研究机构等。项目方需要具有合法的资质并且信息对生态公开透明。项目方在 HGBC 生态中发布的研究项目，需经过 HGBC 基金会审核通过。

### 3. 5 服务方

服务提供方，是指基于用户数据为用户提供各项服务的机构或个人。这些服务包括但不限于疾病风险评估、健康管理、药物风险检测、遗传筛查、个体特质检测、遗传咨询、基因应用、保险保障等。服务提供方和项目方之间的身份可互换，项目方可将自己的研究成果发布到 HGBC 生态中，成为服务提供方，为用户提供服务。

## 四、产品方案

HGBC，是基于区块链构建的基因与医疗数据价值生态网络。一种重塑产业链，以用户为中心的全新的协作模式。以区块链技术保障用户唯一掌控自己的基因和医疗健康数据，并自主决策数据在生态中的流转。配合完整的 Token 经济模型，最大程度的提升数据价值，打造共赢生态。

HGBC 生态产品由基因宝藏 APP、数据分析云平台、第三方开发平台组成。

### 4. 1 基因宝藏 APP

用户可以通过"基因宝藏 APP"领取数据挖矿收益、参与研究项目、购买应用与服务等，参与到 HGBC 的生态建设。

HGBC 将为注册用户分配唯一的数字身份 ID，并写入到区块链账本中。

基因宝藏 APP 主要由任务盒子、研究所、宝藏商店三部分组成。



#### 4.1.1 任务盒子

为了让每个用户更好的享受到 HGBC 生态价值，我们提出了数据挖矿的概念。数据挖矿是根据用户完成任务盒子中的任务，获得算力，算力越高在相应时间内获得的 Token 奖励越高。

#### 4.1.2 研究所

用户可以在研究所自主选择参与科研项目。用户和研究项目之间是一种双相选择的关系，是数据使用权的一次交易。主要环节包括：

- ① 项目方在“项目发布服务平台”上发布自己的项目，项目内容要包括项目介绍、数据要求、报酬、参与人数等信息。
- ② 触发智能合约，分配项目 ID，锁定项目方承诺的报酬（Token），设置项目结束条件等。
- ③ 配置完成的项目由平台推送到研究所中。
- ④ 用户在研究所浏览项目，自主决定是否参加。
- ⑤ 用户确认参加后，系统会将根据用户数据自动生成的数据标签推送给项目方。
- ⑥ 项目方根据数据标签筛选合适的用户，发送数据购买请求。
- ⑦ 数据购买请求由研究所以短信和 App 消息的形式推送给用户。
- ⑧ 用户收到请求后，确认是否加入项目。
- ⑨ 用户同意参与项目，用户数据中项目所需部分被取出，并使用项目方公钥对数据进行加密。
- ⑩ 平台将加密后的数据推送到项目方的项目管理系统中。项目方收到数据后，使用自己的私钥解密数据。
- ⑪ 项目方成功获得数据后，触发智能合约，将项目方锁定的报酬（Token）分发给用户，整个交易完成。

采用智能合约和非对称加密数据传输的方式，彻底保障数据交易过程中的传输安全。  
整体流程图下图所示。

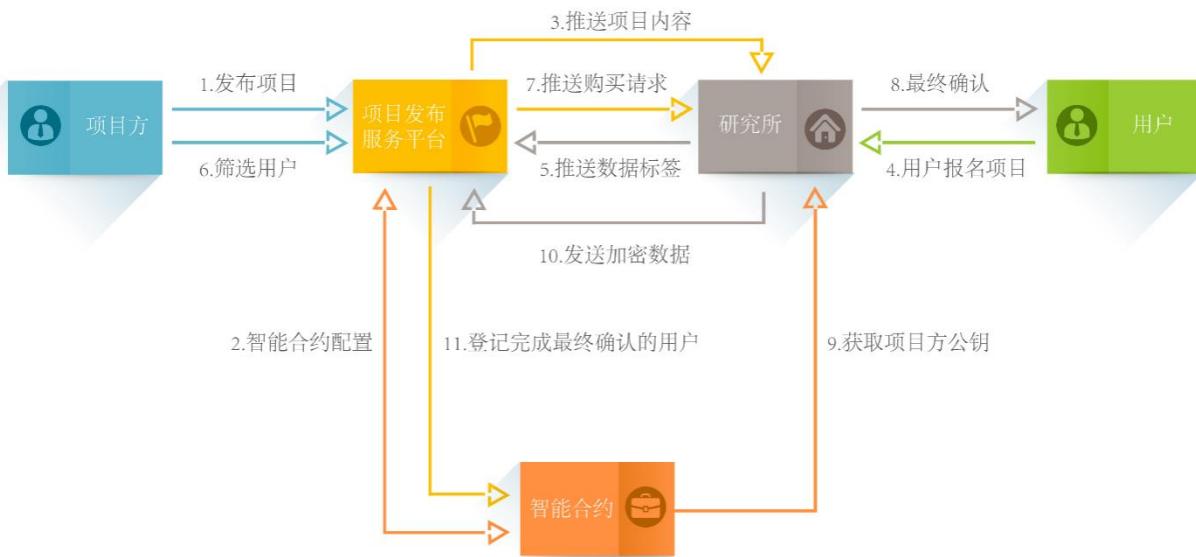


图 4 项目参与流程图

注：1. 数据标签是根据用户的基因数据、医疗健康数据、表征数据自动计算得到的数据特征。2. 服务发布交易平台是项目方发布项目，管理项目，获取数据的地方。

#### 4.1.3 宝藏商店

应用与服务是 HGBC 链的价值呈现方式，HGBC 团队开发了宝藏商店，作为应用与服务的载体。宝藏商店主要向用户提供基于基因数据的各类健康及娱乐应用，以及专业的、一对一的个性化服务。应用和服务由 HGBC 团队和第三方团队共同开发。

1 ) 应用：基因应用为用户解读基因密码提供多样化的选择。比如基因信息的祖源、营养、运动等方面的个性化解读。

2 ) 服务：由专业人员提供的一对一服务，为用户提供个性化基因密码解读和个性化服务定制，包括遗传咨询服务和健康管理服务等。

团队在 HGBC 项目发起之前，已有丰富的基因解读及应用开发经验，拥有全基因组范围内的解读开发能力，成功开发了“酒量 PK 赛”、“基因亲密度”、“乳腺癌/卵巢癌风险评估”等应用。

## 案例一：全基因组检测报告

个人全基因组健康评估套餐，利用领先的第二代测序技术，对人基因组中 30 亿个碱基对进行测序和分析，并结合最新的科研进展，根据个人特有的遗传信息，提供个性化的健康建议、生育指导、用药提示、饮食建议、营养指导等全生命周期健康管理方案，让用户一次检测，终生受益。



图 5 全基因组检测报告图

## 案例二：酒量 pk 赛

通过基因数据判断个体对酒精的代谢能力，从而判断其“酒量”大小，并引入“PK”对比机制，兼具娱乐与健康生活两大用途。在不饮酒的情况下，即可实现“酒量”大小的比拼。根据具体基因型给出饮酒建议，实现健康饮酒，减少醉酒带来身体损伤。

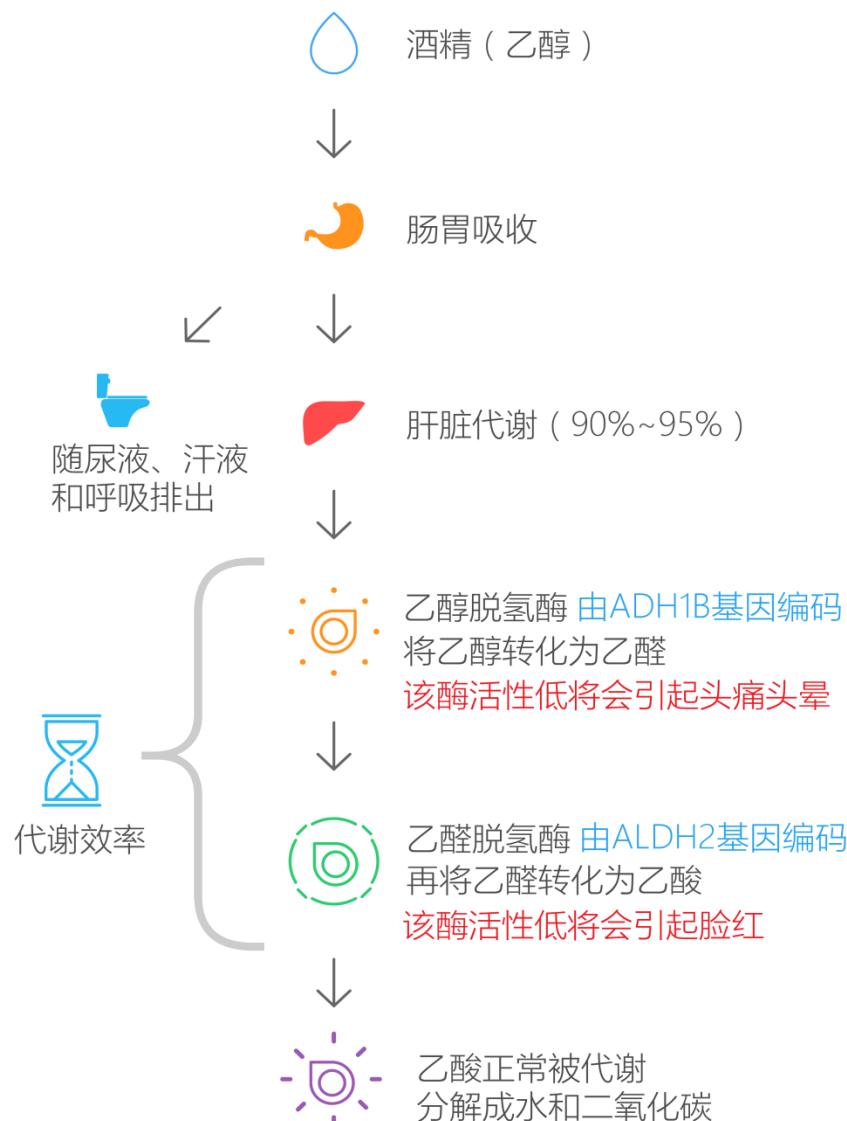


图 6 酒精代谢的生化途径

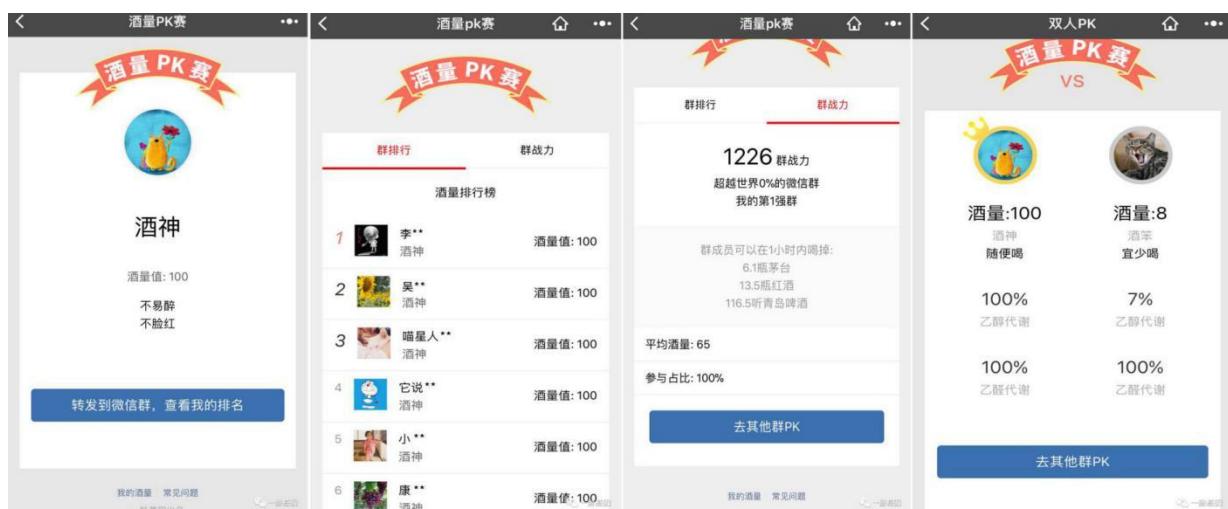


图 7 酒量 PK 小程序

### 案例三：基因亲密度

通过对比双方的全基因组数据，得出两者的基因层面的亲密度，量化人类群体中任意个体之间的亲缘关系。



图 8 基因亲密度图

### 案例四：乳腺癌/卵巢癌风险评估

大约 5% ~ 10% 的乳腺癌是由遗传引起的，这意味着基因缺陷（基因突变）从患者父母传递给了下一代。目前，我们根据 BRCA1、BRCA2 两个基因几十个位点及多种风险因子开发出了的乳腺癌/卵巢癌患病风险评估方案。风险因子指的是影响个体患病几率的因素。



图 9 乳腺癌/卵巢癌风险评估图

## 4. 2 数据分析云平台 GAIC

基因组数据从测序仪器上产出后，无法直接进行分析和解读，需要对原始数据进行预处理，包括检查数据质量、清洗、比对、提取突变信息等步骤。

具体过程如图 10 所示。

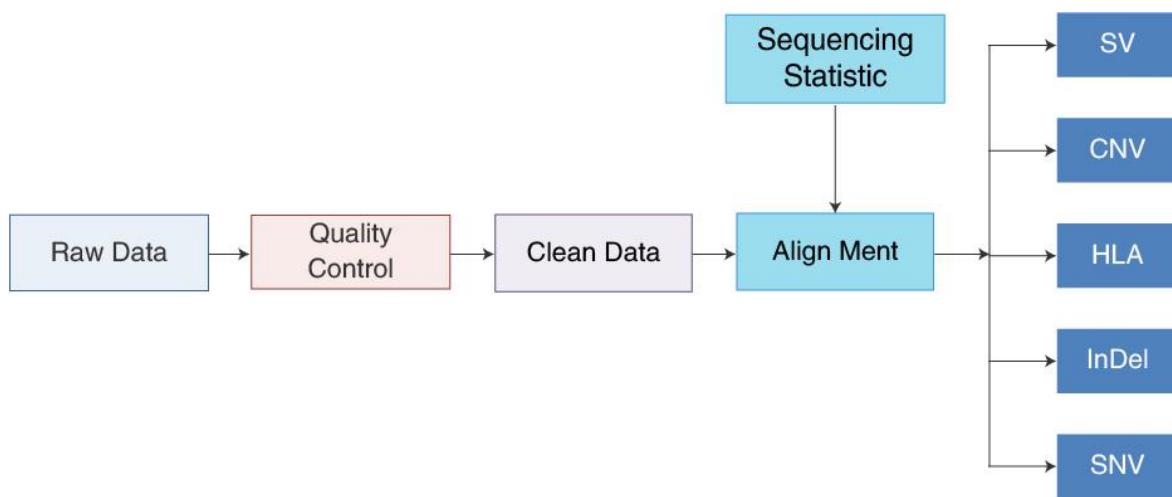


图 10 基因组数据处理标准流程

首先对原始测序数据进行质量值过滤以获得高质量序列，将过滤后的序列（ reads ）比对到人的参考基因组上。统计测序的效果，包含有效数据量、测序覆盖度、测序深度等指标，用来评估数据是否合格。对于合格的数据，根据比对结果，可进行个体水平、群体层面的数据分析和挖掘。

基因组数据分析需要很高的专业知识和庞大的计算资源，HGBC 团队已经开发出专业的数据分析云平台 GAIC ( Genome Analysis and Interpretation in Cloud )，为生态建设提供基础的数据分析能力。

GAIC 基于云平台高效的计算能力和海量的存储能力，提供基因大数据的基础分析和安全存储服务，自动完成从原始数据到比对、基因注释的每一个环节。

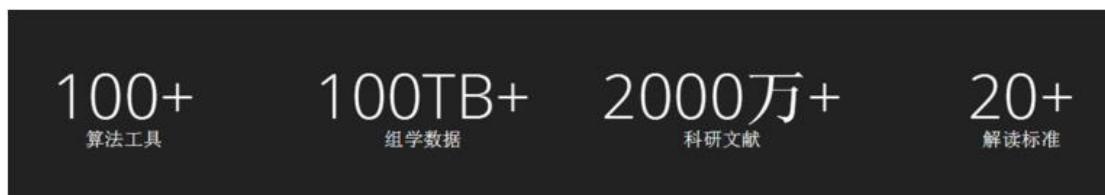


图 11 GAIC 平台图

### 4.3 第三方开发平台

基于基因和医疗健康数据的应用与服务的开发，是 HGBC 重要组成部分。HGBC 团队将提供丰富完善的开发者工具，帮助开发者更好，更快的构建应用与服务。这些工具包括多种高级语言的 SDK、API 等。开发平台的宗旨是帮助开发者专注于业务层面，降低开发难度和成本。

通过与第三方开发者紧密协作的方式，释放个人开发者和小团队开发者的能力，让 HGBC 生态上的应用与服务百花齐放，也让用户拥有更多的选择和更大的价值。

## 五、激励计划

HGBC 生态设立有专门的奖励池来奖励参与生态建设的用户，并采用生态原生 Token ( HGBC ) 作为唯一奖励。HGBC 生态采用 GSS ( Gradient Subsidy of sequencing , 测序梯度补贴 ) 和 PODS ( Proof of Data Stake , 数据权益证明 ) 两种算法机制来保障奖励的公平合理。

### 5.1 奖励池

奖励池 (Reward Pool) 是一个根据时间和基因组测序参与人数动态调整的 HGBC Token 池。奖励池中的 Token 有两方面用途，一方面用于对早期参与基因组测序用户的 一次性奖励；另一方面用于用户的数据挖矿奖励。奖励池中的 token 占总发行量的 50%，共 15 亿个。其中 3 亿用于测序参与奖励，12 亿用于数据挖矿奖励。

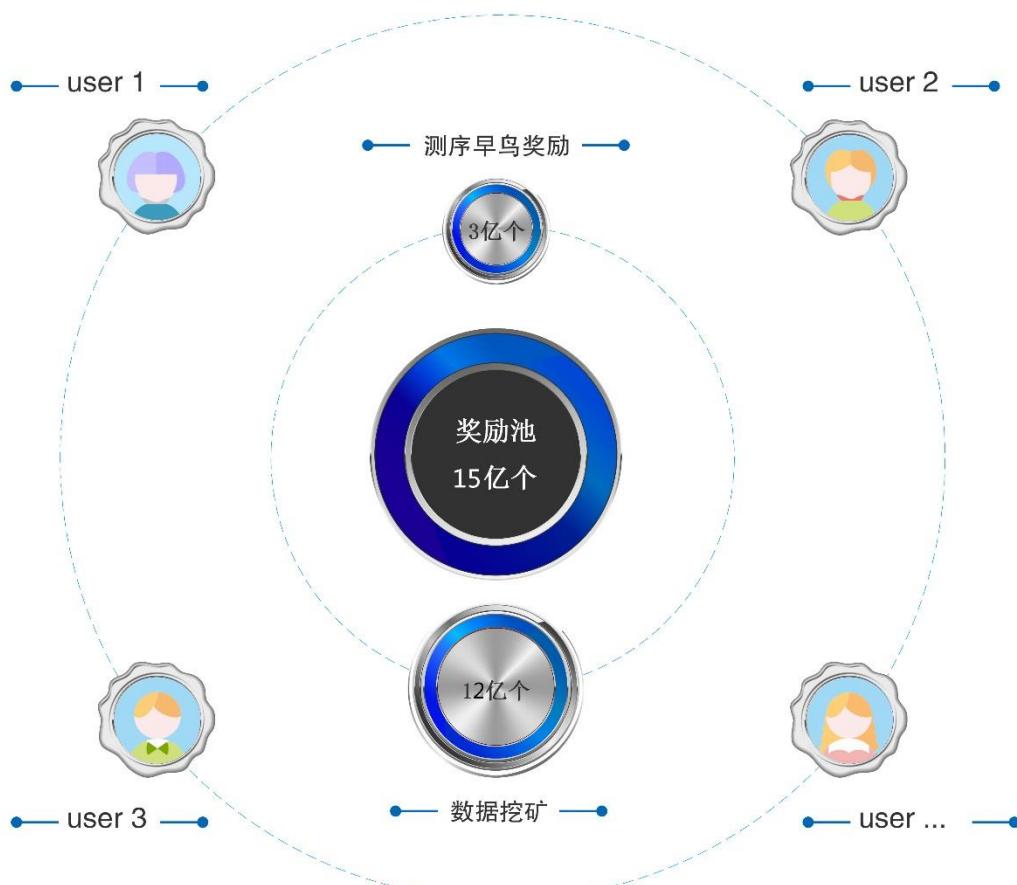


图 12 HGBC Token 奖励池



## 5. 2 测序早鸟奖励

对于早期参与基因测序的用户，HGBC 会拿出一定比例（占总发行 Token 的 10%）的 HGBC Token 作为一次性奖励。这样做，一方面是为了补贴用户基因组测序的成本，另一方面是激励更多的人参与到生态建设中，享受基因数据的价值。

考虑到随着测序技术的发展，测序成本的快速下降。同时，随着生态的逐步完善，HGBC Token 的价值也会越来越高。因此，在基因测序用户的人数增加的同时，采用梯度下降的方式降低测序补贴 Token 数量。HGBC 生态完成百万人基因组计划之后，我们预估测序成本已经下降到非常低的价格，人人都能接受。因此，百万人计划完成之后不再发放测序奖励，但用户还是可以通过数据挖矿获得 Token 奖励。

HGBC Token 采用 GSS ( Gradient Subsidy of sequencing , 测序梯度补贴 ) 算法进行奖励。算法规则如下表：

计划名称：	百人计划	千人计划	万人计划	十万人计划	百万人计划
限定人数：	100	1,000	10,000	100,000	1,000,000
人均奖励：	100,000	30,000	6,000	1,000	100
总量：	一千万	三千万	六千万	一亿	一亿

## 5. 3 数据挖矿

加入到 HGBC 生态中的每个用户都可以根据数据权益证明 ( Proof of Data Stake , PODS ) 机制持续获得数字资产 ( HGBC Token ) 奖励，称之为“数据挖矿”。

由于用户享有自身基因和医疗健康数据的所有权和收益权，并且为 HGBC 生态建设做出了贡献。因此，“数据挖矿”的本质是对这种权益与贡献的一种数字资产奖励。算力的本质是为量化这种权益与贡献而创造出的便于计算的单位。用户单位时间内所能挖到的数字资产与挖矿算力呈正相关关系。

挖矿算力值是根据用户参与的各种挖矿任务累加得到。挖矿任务可以到基因宝藏 app 上任务盒子中领取，一些任务举例如下：



- ① 绑定多维度的基因数据。
- ② 参与表征问卷调查。
- ③ 绑定体检机构、健康管理机构账户。
- ④ 绑定可穿戴设备。

数据权益证明 ( Proof of Data Stake , PODS ) 是一种结合用户算力和参与度设计出的 Token 奖励机制。

#### 名词术语 :

- ① 奖励周期  $t$  : 两次相邻奖励发放之间的时间称之为奖励周期。系统每天 01:00 统一结算并向用户发放。
- ② 用户算力  $U_n$  : 当前奖励周期下用户的有效算力。如果用户 120 小时内未领取奖励 Token , 用户的参与度变成 0 , 用户的算力将被冻结 , 不计入全网算力 , 也不再发放奖励。用户重新领取奖励后 , 算力重新激活 , 变为有效算力。
- ③ 全网算力 : 当前奖励周期下所有有效用户的算力之和。有效用户指的是指算力未被冻结的用户。
- ④ 全网 Token 产量  $R_m$  : HGBC 会按照固定比例产生一定时间内恒定量的 Token。
- ⑤ 奖励 Token 数量  $T_n$  : 每一个奖励周期 , 用户可以获得一定的奖励 Token。

#### 参数值设定 :

- ① 用于数据挖矿的 Token 总量  $a$  : 12 亿个。
- ② 奖励周期  $t$  : 24h。
- ③ 衰减周期  $d$  : 每次产量减少的周期 , 计  $d = 365$  天 ( ~1 年 ) 。
- ④ 衰减因子  $\gamma$  : 每次产量减少的比例 , 计  $\gamma = 30\%$  。
- ⑤ 初始 HGBC Token 产量  $c$  : 开始奖励第一个周期年 , 每个奖励周期产生的奖励 Token 约为 1013698 个 , 第一年共奖励 3 亿 7 千万个。第  $n$  年奖励数量如下公式 :

$$C_1 = 370,000,000$$

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

#### Token 奖励规则 :

单个奖励周期内 , 用户的算力  $U_n$  通过用户参与任务获得的算力累加得到。计算公式如下。



$$U_n = \sum_i^k a_i S_i$$

其中  $S_i$  是用户第  $i$  个任务的得分， $a_i$  是第  $i$  个任务的权重值，默认权重为 1。 $S_i$  乘以  $a_i$  得到第  $i$  个任务的算力值。

每天 01:00，HGBC 网络会通过 PODS 机制发放奖励 Token。这些 Token 极少量用于生态基础设施支出，绝大部分直接发放给用户。用户实际得到的 Token 数量按以下公式计算。

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

其中， $R_m$  是当前奖励周期内发放的总 Token 量； $\sigma$  为奖励 Token 里用于生态基础设施（服务器、存储等）支出的比例，具体数值随着生态实际支出而变化， $\sigma$  最大不超过 0.02，在千人基因组计划完成前  $\sigma = 0$ ； $t$  是当奖励周期内，生态内有效用户的总数。

## 六、商业前景

HGBC 项目的愿景是通过区块链技术解决基因组数据确权和流通问题，以用户为中心，变革以往以企业、科研机构为中心的模式，为个人、机构和社会赋能，加快人类基因组数据的宝藏挖掘。随着用户的规模越来越大，生态的参与方越来越多，HGBC 项目的商业价值和社会价值将逐步释放，并最终实现指数级增长。

HGBC 的商业收入主要来自以基因应用的交易收入、服务交易的分成和项目方的服务费。

### 6. 1 基因应用交易

宝藏商店中的应用服务有来自第三方提供的，也有来自 HGBC 团队开发的。团队有丰富的基因产品开发经验，在 HGBC 项目之前，已完成多款基因应用的开发。

HGBC 团队开发的基因应用交易收入作为项目的收入来源之一。



## 6. 2 服务交易分成

除团队开发的基因应用和服务之外，第三方提供的应用服务平台将收取一定的费用。在 HGBC 生态建设早期，为鼓励更多的服务方加入到生态中，平台会减免手续费甚至补贴服务方。

## 6. 3 项目方服务费

项目方在生态中发布的项目包含商业项目和公益项目，平台对于商业项目会收取一定的服务费。对公益项目，平台不会收取任何费用。

# 七、技术说明

## 7. 1 平台架构

HGBC 生态平台技术架构由基础层、核心层和客户端三层组成，其中基础层分为存储服务和区块链服务；核心层由基因宝藏 App 和开发平台组成。如图 13 所示。

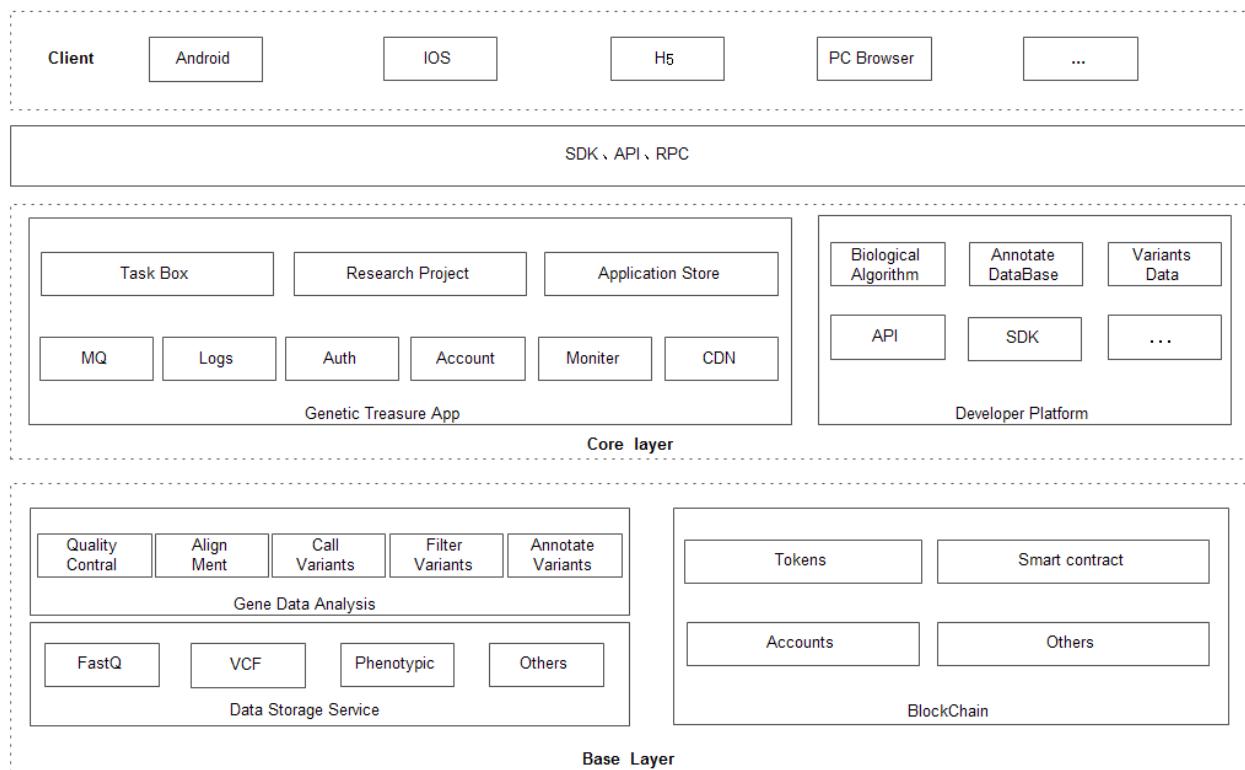


图 13 HGBC 平台架构图

## 7.2 基因组测序与分析

为了发挥基因组数据的最大价值，HGBC 生态主要以人全基因组数据为主（后期根据需要会增加其它类型测序数据）。人全基因组测序采用高通量基因测序仪，测序层数不低于 30 层，产出数据量 90Gb 左右。

高通量测序流程分为以下步骤：

- 采集样本（HGBC 平台采用唾液样本的方式）
- 提取 DNA
- 构建 DNA 测序文库
- 测序仪测序
- 下机数据处理

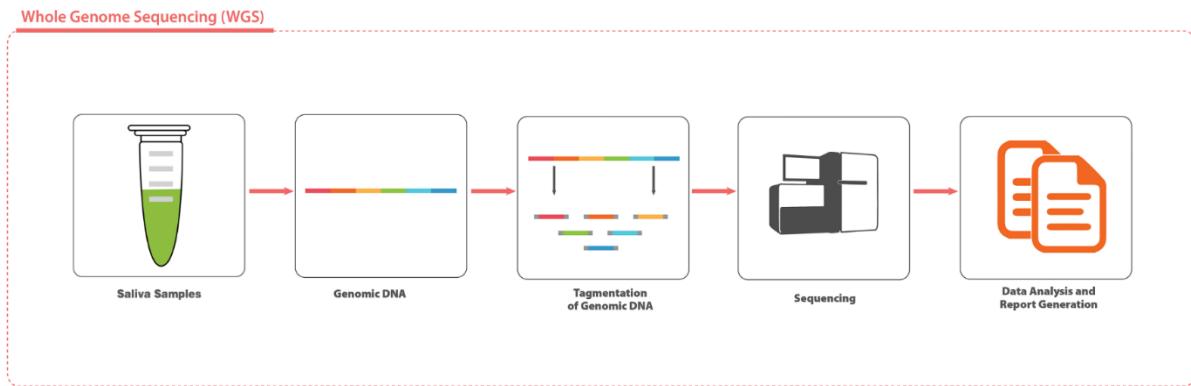


图 14 高通量测序流程

### 7.3 存储与确权

目前通用的人全基因组重测序采用 30 倍的测序深度，产出 FASTQ 格式的基因组原始文件大约 200G（压缩后 60G 左右）。在目前的区块链架构中，并没有很好的解决大数据量的存储问题，如以太坊的 Storage 持久化存储，采用 key-value 形式，其中 key 和 value 的长度都是 256bit，区块链存储成本非常昂贵，并且不支持大数据量直接存储。对于大数据量上链问题，通用的解决方案是将文件的 Hash（哈希）值和所属用户的账户信息写入区块链中，大数据文件本身存储于中心化的存储服务中。

HGBC 平台需要为用户提供一种安全、可靠、低成本、去中心化的数据存储方案，让用户切实掌控自己的数据，上述通用解决方案并不是基因大数据在区块链上存储的终极方案。因此，我们构建一种分布式存储架构—HGBCStorage，底层基于玩客云存储服务，采用领先的加密计算，所有数据通过加密通道传输，无需担心隐私泄露。定制的 Paragon 文件保护系统，用户的每一份数据都受到极致保护，保障硬盘数据安全无忧，真正成为用户的贴心数据管家。

基因组数据测序分析完成后，对产出的数据按照固定大小进行切分，将切分后的数据块进行加密，加密后的数据块分散的存储于 HGBCStorage 不同节点中，同时对每个数据块求 Hash 值，将不同块的 Hash 值以“默克尔树”的数据结构进行组织，使用树的根节点（也称为该数据指纹）能唯一还原此基因组数据本身。

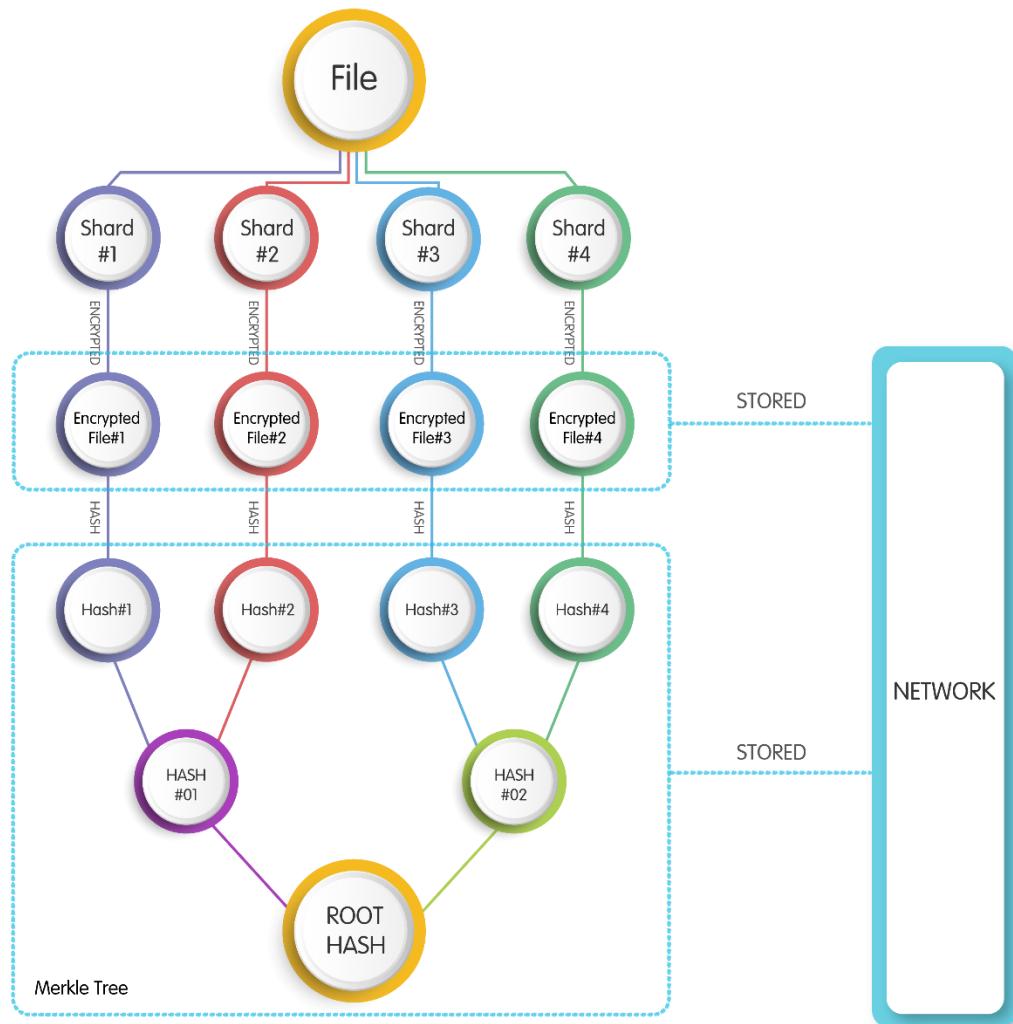


图 15 基因组数据存储流程

HGBCStorage 存储服务将基因组的数据指纹使用用户区块链账号的公钥进行加密，并记录在区块链上，全网广播。加密后的数据指纹只有用户通过自己的私钥才能解密，进而还原数据。

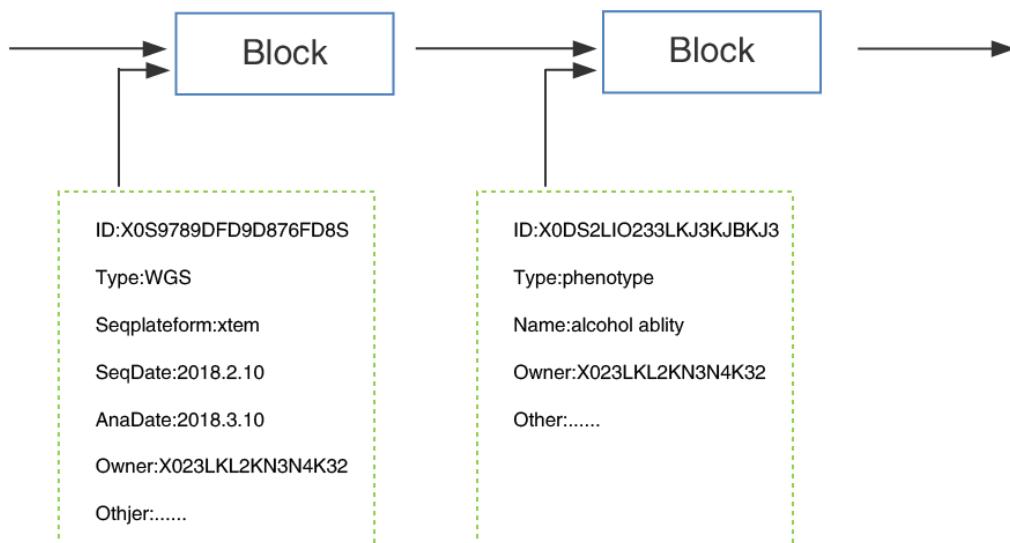


图 16 基因组数据上链确权

HGBCStorage 最终希望用户自主存储和管理自己的数据。由用户将基因组数据接入到自己的玩客云设备中，通过共享和授权，让 HGBC 生态机构访问数据，并将数据使用记录记录在区块链中，用户拥有自己数据的完全掌控权。

## 7.4 数据安全保护

HGBC 生态中包含用户个人、项目方、服务提供方、平台方等多种角色，用户个人数据在多种角色中流转，用户隐私保护尤为重要。HGBC 生态平台用户的隐私保护包括以下 4 个方面：

### 1. 伪匿名隐私保护

用户加入 HGBC 生态链后，会有一个全网唯一的昵称，用户在生态中的所有行为都以昵称为用户标识进行操作，进而将用户的真实身份与生态中的链上身份进行了有效隔离。如用户参与科研项目，项目方可见的是用户的昵称。

### 2. 数据加密保护

用户的基因组和表型数据以加密的形式存储在用户私人玩客云存储设备中，用户对自己的数据实现完全控制。

### 3. 单向匿名机制



HGBC 生态中，测序商、项目方、服务提供方的身份是经过基金会认证审核通过的，对用户是完全透明的，而用户的身份信息是伪匿名的，并且所有交易信息都会被永久的存储在区块链中。

#### 4. 用户授权机制

HGBC 中涉及到用户数据的交易服务，都需要得到用户本人的授权确认。如项目方获取用户数据时，生态平台会向用户发送站内信或短信，用户授权同意后此交易智能合约才会广播全网。

## 八、HGBC Token

HGBC 生态基于 ERC20 发行一种在生态中流通 Token—HGBC。HGBC Token 共发行 30 亿枚，并且永不增发。

### 8. 1 应用场景

HGBC Token 是 HGBC 生态中的一种“通证”，在生态中具有以下使用场景：

1. **测序奖励**：对购买基因组测序服务的用户进行 Token 奖励；
2. **激励**：对完成任务盒子中任务的用户给予激励；
3. **发布项目**：项目方需要使用 HGBC Token 发布科学项目；
4. **交易结算**：用户参与项目方的科研项目后，使用 HGBC Token 向用户结算费用；
5. **购买服务**：用户使用 HGBC Token 购买服务方提供的服务；
6. **平台燃烧**：收取项目方、服务方一定的手续费；

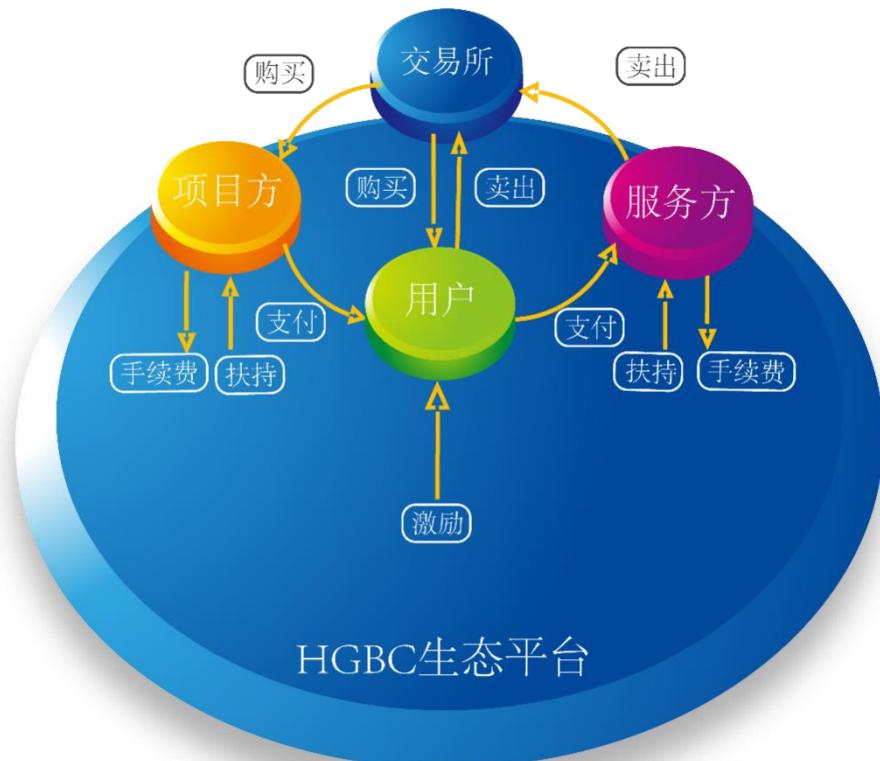


图 17 HGBC Token 流转模型

HGBC 生态是一个快速发展平台，随着生态的不断完善，作为生态中唯一流通的 Token，HGBC 的价值也会不断提升。

## 8.2 分配方案

总计 30 亿 HGBC Token 永不增发，分配方案如下图所示。

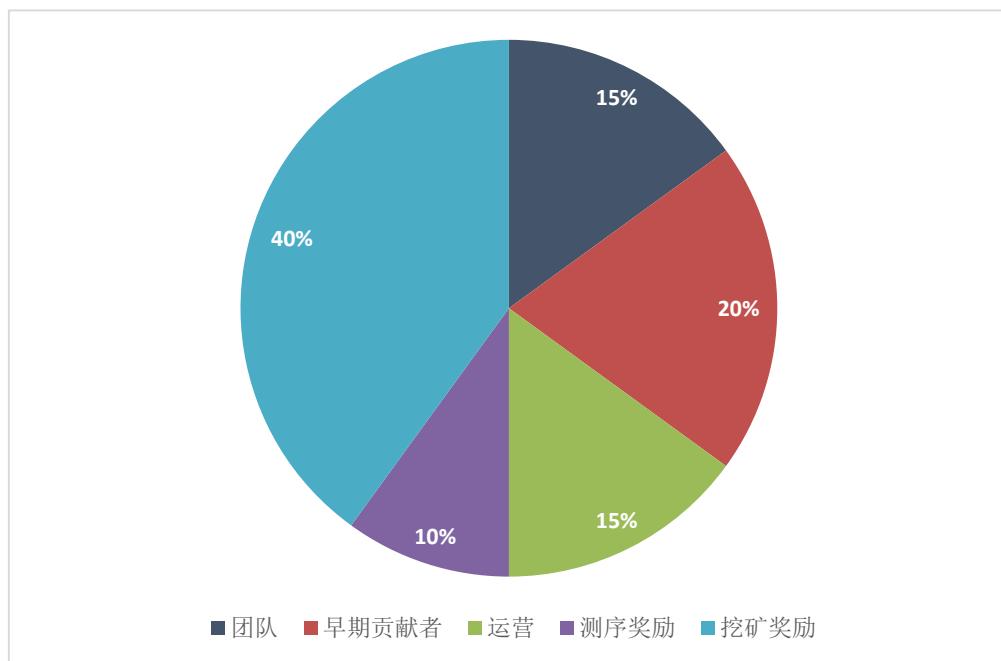


图 18 HGBC Token 分配方案

- 团队 15% , 首年释放全部总量的 5% , 剩余部分在四年锁定期内分批释放。
- 早期贡献者 20% , 包括基石投资者和私募。
- 运营 15% , 用于生态建设和宣传推广。
- 测序奖励 10% , 用于奖励前 100 万拥有基因组的用户。
- 挖矿奖励 40% , 用于奖励参与生态建设的用户。

## 九、发展路线

2015 年 Q2 , 百人基因组计划完成

2015 年 Q4 , 全基因组解读知识库 V1.0 构建完成

2016 年 Q3 , 参与中国 “精准医学计划”

2017 年 Q3 , HGBC 项目萌芽与概念论证

2018 年 Q1 , 区块链技术调研与选型

2018 年 4 月 , 参与组建 “中国基因检测联盟”

2018 年 5 月 , 启动千人基因组计划

2018 年 Q3 , 基因宝藏 APP 上线

2018 年 Q4 , 宝藏商店与研究所功能上线

2019 年 Q1 , 开放第三方开发平台

2019 年 Q2 , 百万人基因组计划启动

...

## 十、HGBC 核心团队



图 19 创始人 : 郝向稳

郝向稳是中国科学院北京基因组研究所基因组学硕士。在基因行业有 15 年的工作经验 , 曾在华大基因、怡美通德、 Illumina 等业内机构任职。 2014 年创办北京基云惠康科技有限公司 , 专注人基因组数据分析、解读和基因应用开发 , 释放基因组数据的价值 , 应用到健康生活场景中。



图 20 联合创始人：林晶

林晶，计算机硕士，主攻大数据处理方向。2015 年带领团队完成基于 Hadoop 的人全基因组数据分布式处理，实现了 5 倍的速度提升。2016 年研究区块链技术，完成多个区块链项目。



图 21 CTO：张旭

张旭，互联网行业从业 11 年，先后在百度、奇虎 360 等任职，其中手机卫士用户量达到 4 亿，从 2013 年开始接触比特币，目前主要从事区块链方向的研究，带领团队制作过多个区块链项目，如以太狗。



图 22 HGBC 核心团队

HGBC 团队其他成员由遗传学、生物学、计算机、生物信息以及区块链背景的资深人士组成，在全基因组数据分析、解读和基因应用的开发方面具有丰富的经验。

## 十一、战略顾问团队



图 23 陈磊

陈磊，迅雷 CEO、网心科技 CEO。清华大学计算机科学与技术系学士学位，德州大学奥斯汀分校计算机系硕士学位。先后在谷歌和微软任职，并曾担任腾讯云总经理和腾讯开放平台副总经理，2017 年担任迅雷 CEO。2014 年获年度中国云计算最具影响力人物，2015 年被评为互联网行业年度杰出人物。



图 24 郑杰

郑杰，树兰医疗管理集团总裁，上海大学树兰医工研究院院长，浙江数字医疗卫生技术研究院常务副院长，浙江加州国际纳米技术研究院精准医学中心副主任，开放医疗与健康联盟 OMAHA 发起人，中国卫生信息学会健康医疗大数据产业发展与信息安全专业委员会常委，中华医学会数字医学分会国际数字医学学会副主席，中国非公立医疗机构协会

医院管理分会副会长，浙江省医学会数字医学分会常务委员，中国物联网医疗专委会委员。《颠覆医疗》推荐序作者，《未来医疗》、《数字医疗》译者。



图 25 余国良 博士

余国良博士，百华协会的理事，美国华人生物医药科技协会创始会长，美国浙江商会会长和中国 PDX 产学研联盟常务主席。美国加州大学伯克莱分校博士，哈佛大学医学院博士后。余国良博士联合创建了近 10 家生物技术公司并担任董事长，Epitomics 公司以 1.7 亿美元被并购。2013 年他加入中美冠科担任董事长，把公司扭亏为盈成功上市。余国良在重大新药创制有丰富的实战经验，先后在中美创建了 5 家新药研发公司。



图 26 王学刚

王学刚，贝壳社联合创始人，金准基因董事长。人类基因组计划参与科学家之一，中国转化医学联盟副主席，中国生物医药技术协会转化医学分会委员，中国电子学会生命电子学分会第七届委员会常务委员，浙江省“千人计划”评审专家，百华协会成员。



图 27 Han Cao, Ph.D.

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



## 十二、生态合作伙伴

### 12.1 测序服务商



### 12.2 解读服务商



### 12.3 健康管理服务商





## 十三、风险提示

随着项目和技术的发展，本白皮书中描述的内容方案会有变动的可能性，我们会更新白皮书，但不承诺更新及发布时间。

本白皮书是一份“概念性”文件，仅用于向公众传递 HGBC 项目方案，不作为任何个人或机构用于投资 HGBC 项目的法律依据。□

HGBC 官方未授权任何第三方机构或个人代理募资，请勿和任何非官方机构或个人发生直接投资行为。如发生损失，HGBC 官方不承担任何责任。

## 十四、联系我们

官方网站：<http://www.hgbc.io>

电子邮箱：[service@hgbc.io](mailto:service@hgbc.io)

微信群秘：



Telegram：[https://t.me/hgbc\\_cn](https://t.me/hgbc_cn)