




2016 杭州·云栖大会
THE COMPUTING CONFERENCE

云栖社区
yq.aliyun.com

服务：数据驱动的基因组分析与解读

2016
The Computing Conference

主办单位： 杭州

 Alibaba Group
阿里巴巴集团

战略合作伙伴：

陈钢
CTO@WeGene



扫码观看大会视频

目录 content

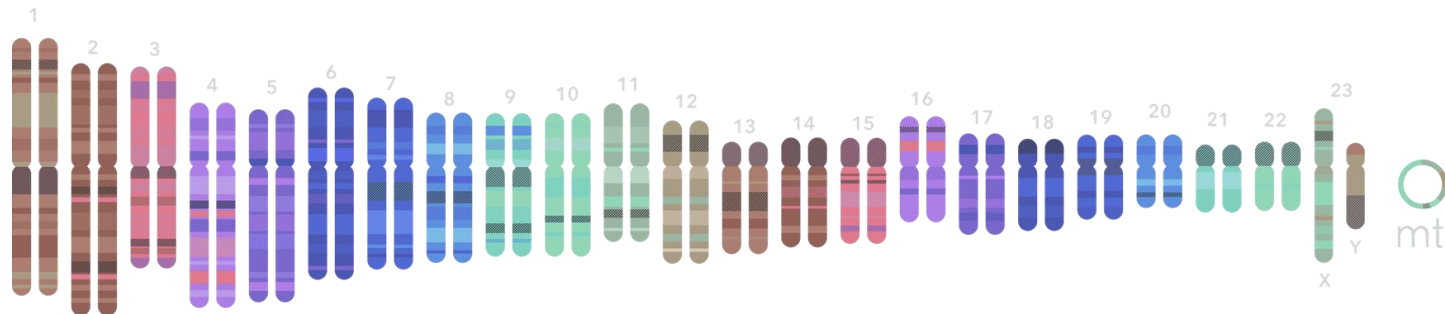
个人基因组为用户做什么？
我们在阿里云生态中的实践
未来的基因组数据平台



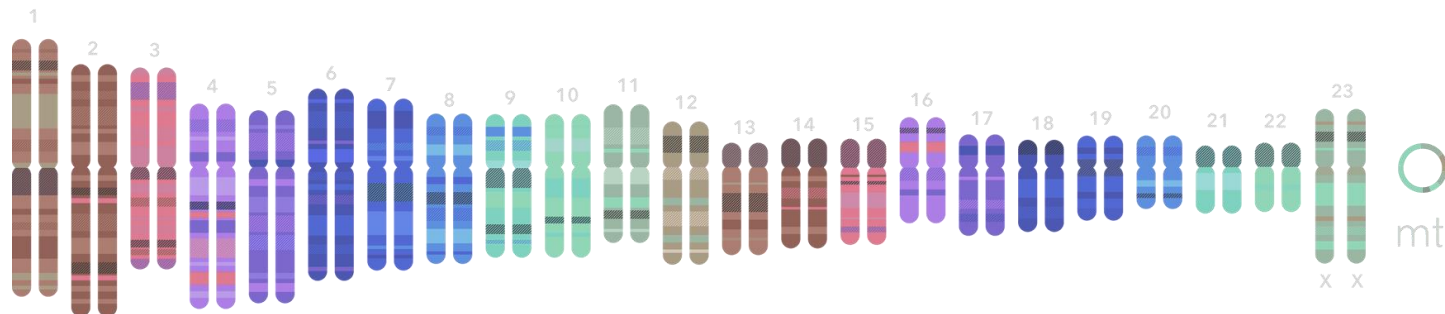


一、个人基因组为用户做什么？





$22 \times 2 + XX$ OR $22 \times 2 + XY$





云南元谋人
170万年前



周口店北京人
60万年前







成吉思汗

遗传后裔1600万人



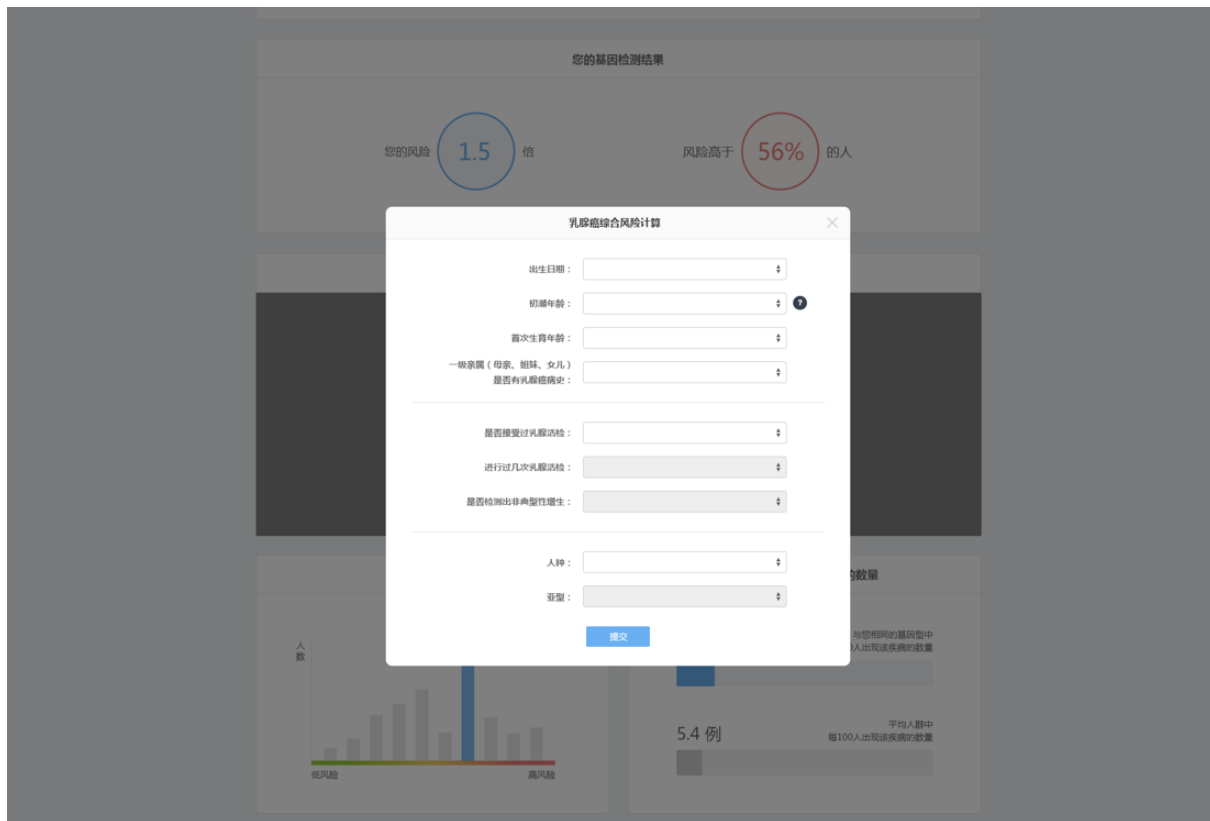
营养代谢 10项 +
健康风险 100项 +
遗传疾病 30项 +
药物反应 40项 +
遗传特征 40项 +
原始数据 60万位点

基因检测的局限性

- 该检测无法覆盖所有与该项目相关的位点
- 基因只是影响性状的一部分因素，其他可能的因素包括生活环境和生活方式等
- 该结果不能用于临床诊断
- 该结果会随着参与人数的增多不断优化







您的乳腺癌综合风险 ?

年龄 : 30

初潮年龄 : 12-13

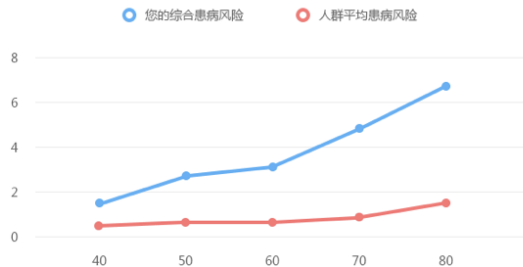
首次生育年龄 : 未育

一级亲属 (母亲、姐妹、女儿) 是否有乳腺癌病史 : 未知

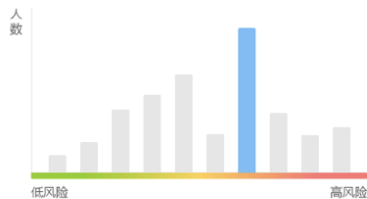
乳腺活检结果 : 未知

人种及亚型 : 黄棕人种 - 中国

更新数据重新计算



您在人群中的位置



该疾病每100人中出现的数量

8.1 例

与您相同的基因型中
每100人出现该疾病的数量

5.4 例

平均人群中
每100人出现该疾病的数量



扫码观看大会视频

微解读

用户自己创建解读项目



立刻创建自己的解读项目, 并与 WeGene 的众多用户分享

*微解读内容均为用户创建, 微基因不对解读内容负责。?

除了微基因官方报告的200多项解读之外, WeGene 原始数据库的60万个位点等着被你探索。

如果你有以下经历:

- 1、“生信”狂热者, 喜欢自己解读原始数据。
- 2、自己关心的项目微基因官方报告里面暂时没有。
- 3、关注最新基因研究成果。

立刻创建

* 目前只支持单位点影响研究。

* 创建项目的位点须在[[微基因原始数据库](#)]中。



扫码观看大会视频

微解读

用户自己创建解读项目

60+



我要自然白!

OCA2基因编码的P蛋白被认为是一种内在膜蛋白,涉及到小分子的传递,特别是作为黑色素前体的酪氨酸。该基因的某些突变会导致色素沉着的变化,包...

发起人: xraywu

[查看详情](#)



善解人意

催产素受体基因 (OXTR) 编码的蛋白OXTR, 遍布于神经和生殖系统中, 是催产素分子的受体。研究显示, 当编码该蛋白基因的DNA序列中某特定位置的A...

发起人: cedro

[查看详情](#)



食指长还是无名指长?

食指与无名指长度之比 (2D: 4D) 通常被用来反映个体出生前的雄激素水平。然而, 在单纯的遗传层面, 也有一些因素影响着两者长度之比。一项基于不同...

发起人: xraywu

[查看详情](#)



工作太累不是老板的错?

有部分人群在工作中更有可能进入“精疲力竭”的状态。这种状态也是所谓的“倦怠综合征”的一种表现。容易在工作中精疲力尽的人群在抑郁、睡眠障碍及...

发起人: xraywu

[查看详情](#)



我只是骨架大而已

腰臀比是用来测量身体脂肪分布情况及代谢率的重要指标。一项基于大规模人群的全基因组关联分析发现了多个与腰臀比有关的遗传位点。其中, rs6905288...

发起人: penguin650

[查看详情](#)



长得捉急怪我咯!

一项在欧洲人中的研究表明, 你看上去是不是“长得捉急”也与基因有关。在该项研究中, 研究者发现在 MC1R 基因上的多个位点决定了一个人看上去是不是比...

发起人: xraywu

[查看详情](#)



你是勇士还是贤者?

位于 COMT 基因上的位点 rs4680 (Val158Met) 是一个被广大研究者们关注的位点。COMT 基因与个体的多巴胺水平密切相关, 因此会影响到个人的性格...

发起人: xraywu

[查看详情](#)



对新事物的好奇心

研究表明, 位于 DRD4 基因上的位点 rs1800955 可能与人们对新事物的探索精神有关。拥有突变型位点 (C 521T) 的人群显示出对新事物更强烈的探索精神。...

发起人: xraywu

[查看详情](#)



扫码观看大会视频



WeGene 研究所

一个特正经的人类研究中心



抽烟那些事

发起人: 济宁医学院

大多数抽烟的人, 会在睡醒后有来一根的习惯, 美其名曰:
醒脑, 然而一根烟往往是不足的.....



今天睡的好不好?

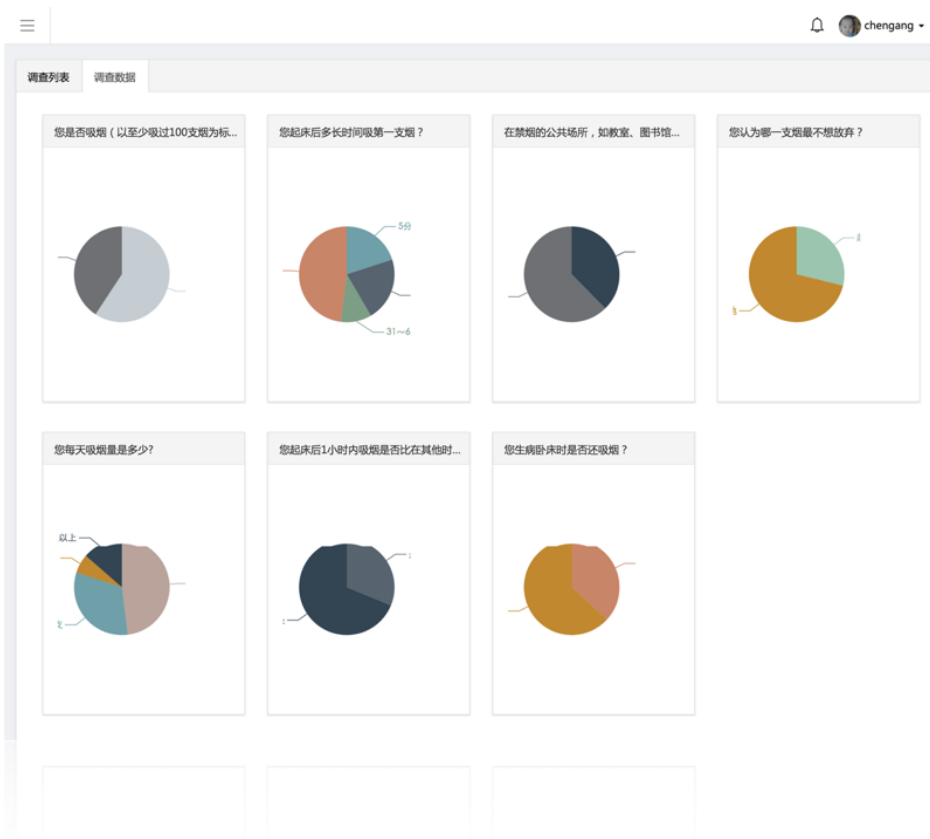
发起人: WeGene

走过了很多路, 听过了很多道理, 但却总是过不好这三分之
一生?



扫码观看大会视频

济宁医学院
尼古丁重度成瘾 400+



扫码观看大会视频

WeGene：基因组数据

Fitbit：运动数据

若饭：定量饮食

Quantified Self !



扫码观看大会视频

WEGENE 开放平台
 申请 文档

WeGene 开放平台

立即申请

```

6      "RS111033199": "GG",
7      "RS1805137": "TT",
8      "RS28936415": "GG",
9      "RS4565946": "CC"
10   },
11   {
12       "I3002517": "TT",
13       "I5012679": "CC",
14       "I5012680": "GG",
15       "RS1051730": "CT",
16       "RS411033199": "GG"
    
```

互联互通的基因组数据

每一个 WeGene 的用户都拥有自己的基因组数据。通过 WeGene 的 RESTful API，您可以在自己的各种应用中使用和解读这些基因组数据。

36.4%
24.4%
16.2%
10.9%
7.2%
4.9%

不懂基因也能开发基因应用

WeGene 开放平台不仅提供用户基因分型结果，还为开发者提供祖源、疾病、运动等分析结果。您不必成为基因专家就能开发基因应用。

开始编写 WeGene 应用吧

我们已经为您准备好了 PHP、Go 等编程语言的实例，Java、JavaScript 等更多编程语言的实例正在路上。拿回去改改就能开张了！



扫码观看大会视频

Relationship GO!

联系我们 Dropdown

亲缘关系鉴定

此功能只针对WeGene用户的基因检测数据，授权您的数据与他人进行对比即可。

WeGene的亲缘关系只是数据解读，不具法律效力，不承担相关法律责任。

开始测试

相关网站

[WeGene](#)

一键分享



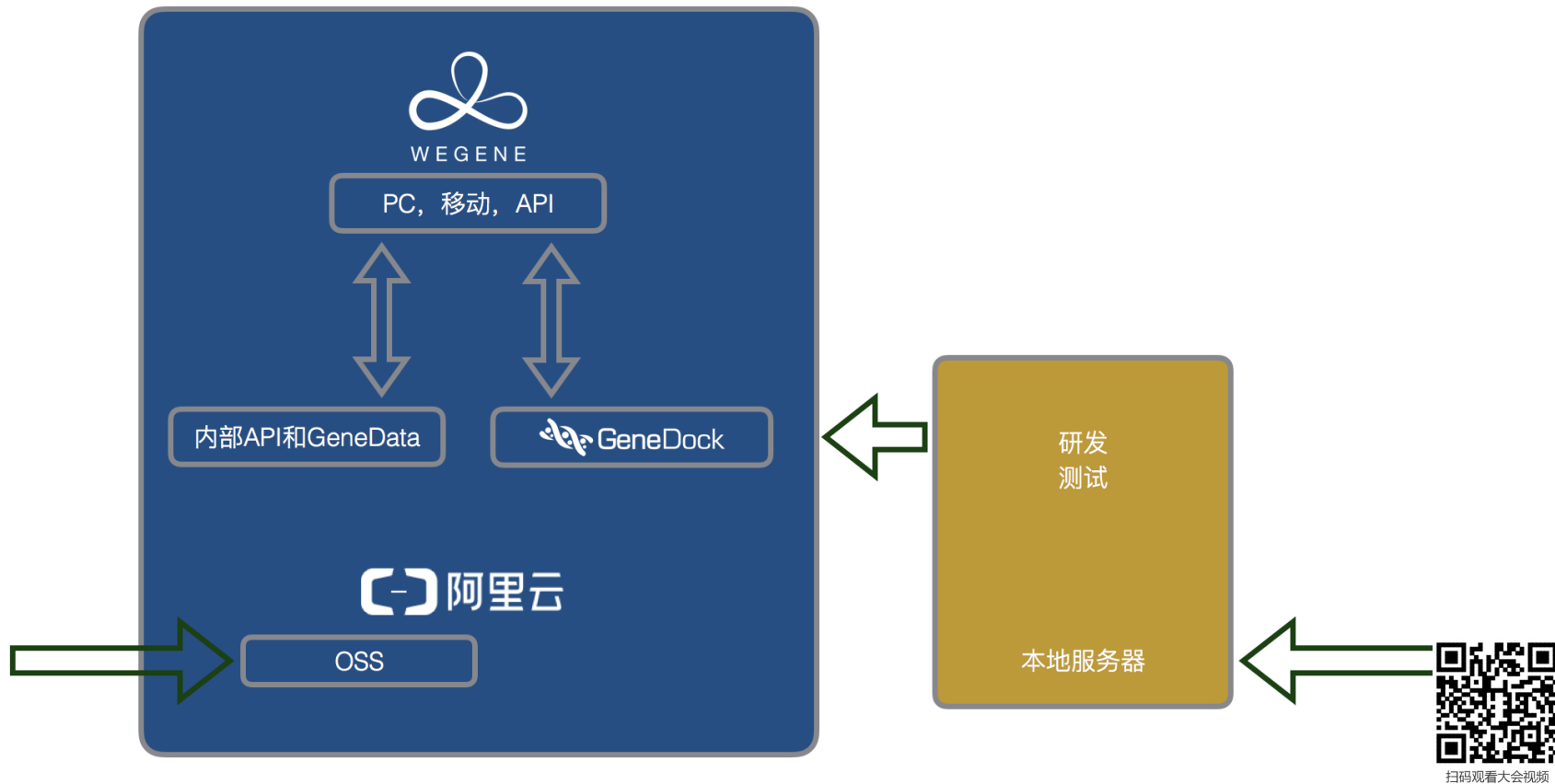
扫码观看大会视频



二、我们在阿里云中的实践







数据存储

RDS：预处理后的基因组数据，表型数据，网站数据

OSS：基因组原始数据，大型的计算结果：

全基因组数据：FASTQ，BAM，约每用户200GB，约50用户

芯片数据：CEL，约每用户200MB，约7000用户

GeneDock：正在处理的全基因组数据和芯片数据

本地存储：所有数据的备份



数据传输

- 芯片数据：实验室用OSS客户端上传OSS
- 全基因组数据：通过GeneDock的专线上传GeneDock系统
- 23andme等数据：通过API直接存储RDS



计算

- 本地服务器：
 - 主要用于研发，或者是处于alpha测试阶段的产品，Celery调度
 - 基于Spark的芯片数据质控
- ECS：
 - 实时的小规模计算，例如简单的遗传风险评估；
 - 常规Web服务
- GeneDock：
 - 全基因组数据分析





三、未来



基因组数据的二次分析

统计分析，机器学习，数据可视化



全基因组数据存储、分析和可视化

Genomics API，上层应用，标准化



扫码观看大会视频

亚洲人群的Genome Graph

参考数据集，自学习系统，开放服务



扫码观看大会视频

2016 The
Computing
Conference
THANKS

