From 生信菜鸟团

千人基因组计划的重要性我也不想多说了，由于时间跨度比较长，最终的数据不只是一千人，**最新版共有NA编号开头的1182个人，HG开头的1768个人！**它的官方网站是：有一个ppt讲得很清楚如何通过官网做的data portal来下载数据：<https://www.genome.gov/pages/research/der/ichg-1000genomestutorial/how_to_access_the_data.pdf> 我不喜欢可视化的界面，我比较喜欢直接进入ftp自己翻需要的数据，千人基因组计划不仅仅有自己的ftp站点，而且在NCBI，EBI和sanger研究所里面也有数据源可以下载， 是非常丰富的生信入门资源！

ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/

ftp://ftp.sanger.ac.uk/pub/1000genomes/

ftp://ftp.ebi.ac.uk/pub/databases/1000genomes/

 ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp

**千人基因组计划测了5个大的人种，25个亚人种**，具体介绍如下：

09/08/2014 12:00AM 1,663 [20131219.populations.tsv](file:///C:\1000genomes\ftp\20131219.populations.tsv)

09/09/2014 12:00AM 97 [20131219.superpopulations.tsv](file:///C:\1000genomes\ftp\20131219.superpopulations.tsv)

其实对大部分人来说，除非你想下载千人基因组计划的原始数据来学习生物信息学分析流程，不然用不着这个ftp站点的，它自己在EBI里面的有一个非常好用的可视化界面来浏览千人基因组计划的variation结果

**千人基因组计划 — 基因组浏览器：** <http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>

<http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=rs35761398>  chr1:24201919:24201920

<http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=2501432>  chr1:24201920

<http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=2502992>  chr1:24201919

在千人基因组计划里面看一个rs就能看到各种人群信息：

<http://browser.1000genomes.org/Homo_sapiens/Variation/Population?r=1:24201420-24202420;v=rs2501432;vdb=variation;vf=1849472>

这些人群信息，可以画一个网路图！ 只需要变化rs ID号即可，当然并不是所有的rs ID号都在千人基因组计划里面有显示的。

还有一个java软件-可视化检测千人基因组数据

<http://bioinformatics.oxfordjournals.org/content/early/2016/03/17/bioinformatics.btw147.short?rss=1>

<http://limousophie35.github.io/Ferret/>

但是好像不是很好用！

**在千人基因组计划的ftp主站点里面可以下载所有数据。**

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/

ftp://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/

直接看最新版的数据，共有NA编号开头的1182个人，HG开头的1768个人！

ftp://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/phase3/data/

也可以按照人种来查看这些数据：ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\_collections/1000\_genomes\_project/data/

每个人的目录下面都有 四个数据文件夹

Oct 01 2014 00:00    Directory alignment

Oct 01 2014 00:00    Directory exome\_alignment

Oct 01 2014 00:00    Directory high\_coverage\_alignment

Oct 01 2014 00:00    Directory sequence\_read

这些数据实在是太丰富了！

也可以直接看最新版的vcf文件，记录了这两千多人的所有变异位点信息！

可以直接看到所有的位点，具体到每个人在该位点是否变异！

ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/

不过它的基因型信息是通过MVNcall+SHAPEIT这个程序call出来的，具体原理见：<http://www.ncbi.nlm.nih.gov/pubmed/23093610>

而且网站还提供一些教程：ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\_collections/1000\_genomes\_project/working/

我们肯定可以在千人基因计划的官网下载测序数据，主要是vcf格式的突变！

* Coriell Catalog website: [1000 Genomes Project](https://catalog.coriell.org/1/NHGRI/Collections/1000-Genomes-Collections/1000-Genomes-Project)
* 1000 Genomes website: [browser.1000genomes.org/index.html](http://browser.1000genomes.org/index.html) (by SNP ID)
* 1000 Genomes website: [www.1000genomes.org/data](http://www.1000genomes.org/data) (bulk data)

**但是关于它的表达数据，就不是那么简单了！**

The most important available existing expression datasets involving 1000g individuals are probably the following:

**RNAseq (mRNA & miRNA) on 465 individuals (CEU, TSI, GBR, FIN, YRI)**

Pre-publication RNA-sequencing data from the Geuvadis project is available through [http://www.geuvadis.org](http://www.geuvadis.org/web/geuvadis/home)

<http://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-1/samples.html>  
<http://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-2/samples.html>

**RNAseq on 60 CEU individual [1]**

<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-197>

**Expression arrays on about 800 HapMap 3 individuals with a lot of overlap with 1000g data [1,2]**

[http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-198](http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-264)  
<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-264>

**RNAseq for 69 YRI individuals [3]**

<http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-19480>