ceu1kg: resources for exploring the 1000 genomes data on individuals of central European ancestry in Bioconductor

VJ Carey

December 4, 2010

1 Introduction

Using results of next generation sequencing experiments, a consortium of geneticists produced calls for SNP at approximately 8 million loci of the genomes of individuals of central European ancestry.

NOTE: All code in this vignette is unevaluated for Bioconductor release 2.7, because of last-minute changes to the snpMatrix package by H-T Leung. You can build this vignette when using Bioconductor release 2.8, by changing all eval=FALSE to TRUE.

Genotype calls are held in a list of snp.matrix instances:

```
> library(ceu1kg)
> if (!exists("ceu1kg.sml")) data(ceu1kg.sml)
> names(ceu1kg.sml)
> ceu1kg.sml[[1]]
```

Metadata about the loci are provided in GRanges instances:

```
> if (!exists("ceu1kgMeta_20")) data(ceu1kgMeta_20)
> ceu1kgMeta_20[1:3]
```

The Bioconductor *GGdata* package includes HapMap phase II genotypes on 90 CEU individuals in 30 trios, coupled with expression data as distributed at the Sanger GENEVAR project (`ftp://ftp.sanger.ac.uk/pub/genevar/`). The 1KG genotypes are available for 43 of these 90 and the associated genotype plus expression data for these 43 are stored in an 'smlSet' container:

```
> if (.Platform$OS.type != "windows") {
+     data(ceu1kg)
+     ceu1kg
+ }
```

2   Session information

```
> sessionInfo()
```