how to make an smlSet from hapmap data

VJ Carey

April 14, 2011

1. **Raw data acquisition:** Obtain the hapmap files from the bulk data download. A typical filename is

   `genotypes_chrY_YRI_r23_nr.b36_fwd.txt.gz`

2. use snpStats read.HapMap.data to obtain the associated SnpMatrix and support data frame. We do this for the 24 main chromosome files. We save the SnpMatrix for chromosome n to C[nn].rda. Be careful with the ordering of filenames – should match desired ordering of chromosomes.

3. **Create a list of SnpMatrix of genotype data:**

```
> ofi = dir(patt = "C.*rda")
> allsm = list()
> cn = rep(NA, 24)
> for (i in 1:24) {
+     cat(i)
+     load(ofi[i])
+     fn = gsub(".rda", "", ofi[i])
+     allsm[[i]] = get(fn)[[1]]
+     cn[i] = as.character(get(fn)[[2]][1, "Chromosome"])
+     print(fn)
+     rm(fn)
+     gc()
+ }
```

   Don't forget to give names 1:22, X, Y to the list elements.

4. Create an environment and assign the list created above to symbol `smList` in that environment. This environment is a valid value for the smlEnv slot of a smlSet instance.

5. The chromInds slot gives numerical indices indicating which chromosomes are included; see hmceuB36.2021 in GGtools for an example.

6. remaining slots are as in ExpressionSet