

Overview of *GGtools* for genetical genomics

VJ Carey <stvjc@channing.harvard.edu>

October 13, 2006

1 Introduction

We use the term *genetical genomics* to refer to data analysis activities that link genotypic information such as SNP configurations to gene expression phenotype. The *GGtools* package includes various demonstration resources and analysis tools for these activities. We will attach the library and have a look at a basic demonstration resource.

```
> library(GGtools)
```

```
KernSmooth 2.22 installed  
Copyright M. P. Wand 1997
```

```
> data(chr20GGdem)  
> class(chr20GGdem)
```

```
[1] "racExSet"  
attr(,"package")  
[1] "GGtools"
```

```
> chr20GGdem
```

```
racExSet instance (SNP rare allele count + expression)  
rare allele count assayData:  
  Storage mode: environment  
  featureNames: rs4814683, rs6076506, ..., rs6062370, rs6090120 (117417 total)  
  Dimensions:  
           racs  
Features 117417  
Samples    58  
  
expression assayData  
  Storage mode: environment
```

```

featureNames: 1007_s_at, 1053_at, ..., AFFX-r2-P1-cre-3_at, AFFX-r2-P1-cre-5_at (8793)
Dimensions:
  exprs
Features    8793
Samples      58

```

phenoData

```

rowNames: NA06985, NA06993, ..., NA12892 (58 total)
varLabels and varMetadata:
  sample: hapmap id

```

Experiment data

```

Experimenter name: Cheung VG
Laboratory: Department of Pediatrics, University of Pennsylvania, Philadelphia, Penns
Contact information:
Title: Mapping determinants of human gene expression by regional and genome-wide asso
URL:
PMIDs: 16251966

```

Abstract: A 180 word abstract is available. Use 'abstract' method.

Annotation [1] "hgfocus"

The `racExSet` class is an extension of the `eSet` class. It represents expression data from the hgfocus chip on 48 individuals in the CEU CEPH group, and SNP data obtained from their HapMap genotyping results.

The data are organized into an 8793 by 58 matrix of expression values accessible with the `exprs` method, and an 117417 by 58 of rare allele counts:

```
> dim(exprs(chr20GGdem))
```

```
[1] 8793    58
```

```
> dim(snps(chr20GGdem))
```

```
[1] 117417    58
```

```
> snps(chr20GGdem)[1:5, 1:5]
```

	NA06985	NA06993	NA06994	NA07000	NA07022
rs4814683	2	0	0	2	1
rs6076506	0	0	0	0	NA
rs6139074	2	0	0	2	1
rs1418258	2	0	0	2	1
rs7274499	0	0	0	0	NA

We need some genetic metadata about SNPs; these are culled from SNP genotyping panels released on a chromosome-by-chromosome basis for CEPH participants by HapMap project:

```
> data(chr20meta)
> chr20meta[1:4, ]
```

	pos	strand
rs4814683	9795	+
rs6076506	11231	+
rs6139074	11244	+
rs1418258	11799	+

A basic task is to compute a screen (over the genome, or, more practically, over a chromosome) of genotypic determination of expression. The `snpScreen` method helps with this; we illustrate an example related to results in Cheung and Spielman 2005:

```
> chr20GGdem = exclMono(chr20GGdem)
> S100 = snpScreen(chr20GGdem, chr20meta, genesym("CPNE1"), ~.,
+   lm, gran = 30)
> S100
```

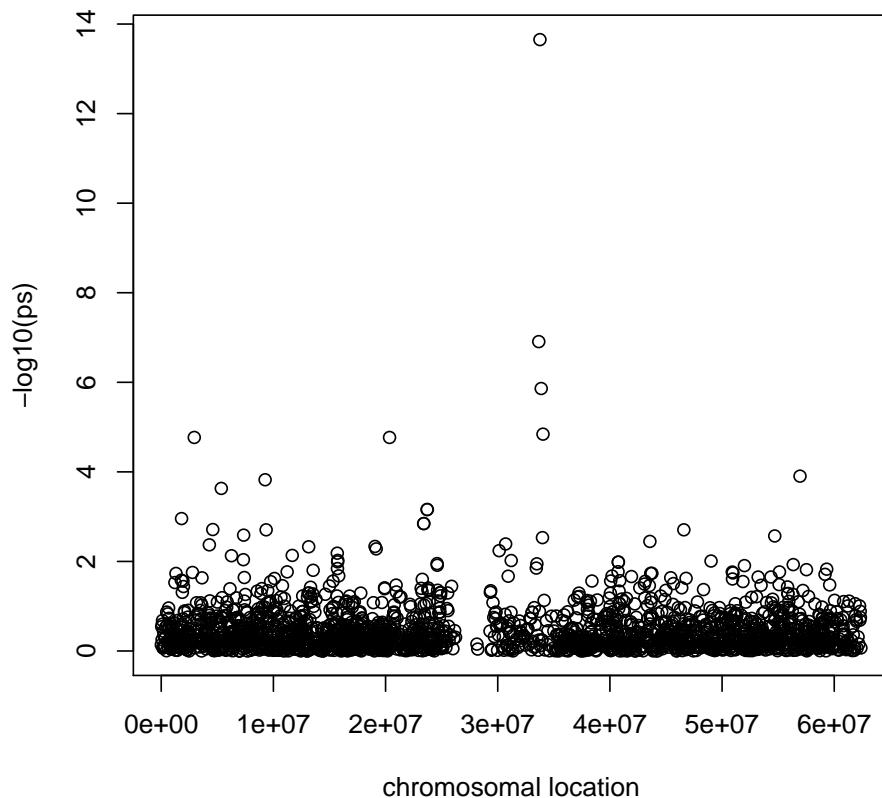
GGtools snpScreenResult for call:

```
snpScreen(racExSet = chr20GGdem, snpMeta = chr20meta, gene = genesym("CPNE1"),
  formTemplate = ~., fitter = lm, gran = 30)
```

There were 2125 attempted fits,
and 2125 were successful.

A primitive display is obtained as follows. We know that `lm` was used, so the relevant p-values are in the coefficient component of the summarized fit objects.

```
> ps = as.numeric(sapply(S100, function(x) try(summary(x)$coef[2,
+   4])))
> plot(S100@locs, -log10(ps), xlab = "chromosomal location")
```



2 Performance-oriented specialization

The `snpScreen` method illustrated above is very general (can accommodate and retain results of any R modeling function) but fairly slow. We have added an R function `fastAGM` for fast fitting of an additive genetic model (equivalent to but much faster than using `lm`).

```
> ut = unix.time(sCPNE1 <- snpScreen(chr20GGdem, chr20meta, genesym("CPNE1"),
+   ~., fastAGM, 75))
> ut

[1] 0.2 0.0 0.2 NA NA

> sCPNE1

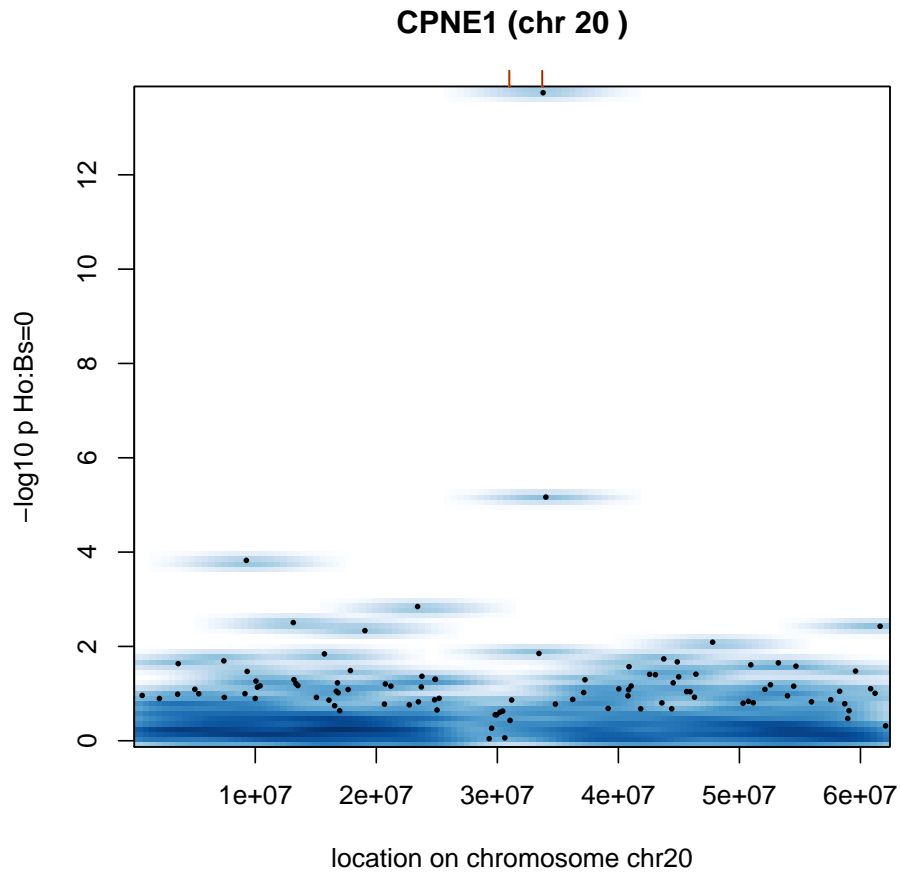
GGtools snpScreenResult for call:
snpScreen(racExSet = chr20GGdem, snpMeta = chr20meta, gene = genesym("CPNE1"),
```

```
formTemplate = ~., fitter = fastAGM, gran = 75)
There were 581 attempted fits,
and 581 were successful.
```

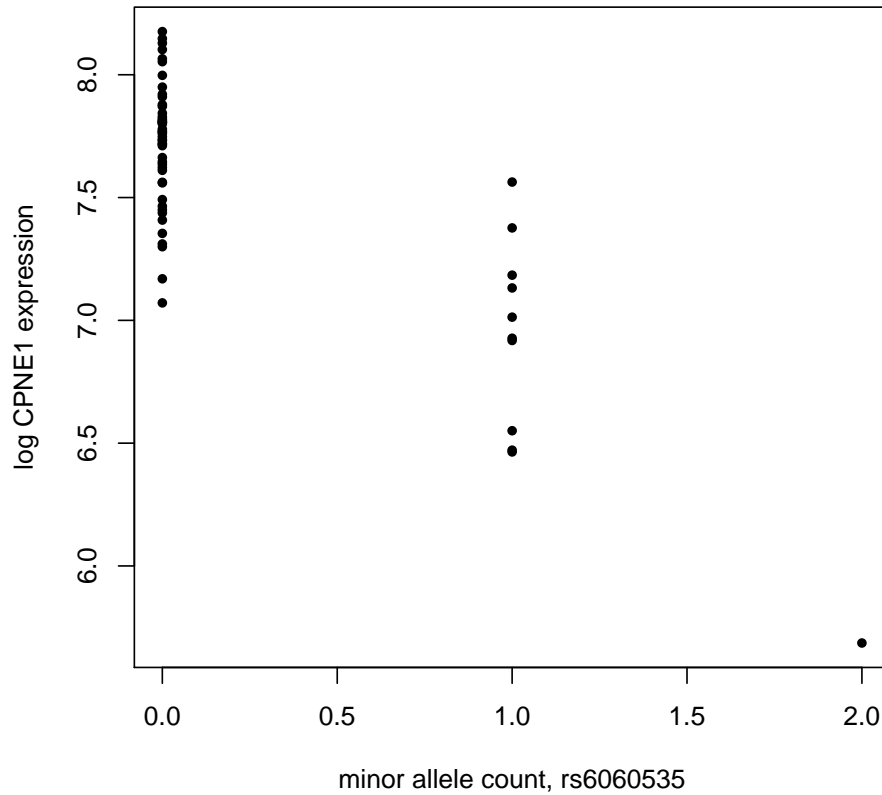
```
> wm = which.min(pp <- extract_p(sCPNE1))
> pp[wm]
```

```
rs2425109
1.817641e-14
```

```
> plot_mlp(sCPNE1, chr20meta)
```



```
> plot_EvG(chr20GGdem, genesym("CPNE1"), "rs6060535")
```



3 Appendix: Package documentation for *GGtools*

Information on package 'GGtools'

Description:

```

Package:      GGtools
Title:        software and data for genetical genomics (c) 2006 VJ
               Carey
Version:      1.2.0
Author:       stvjc <stvjc@channing.harvard.edu>
Description:  dealing with hapmap SNP reports, GWAS, etc.
Depends:      R (>= 2.2.0), methods, Biobase (>= 1.11.26), hgfocus,
               geneplotter(>= 1.11.8)
LazyData:    yes

```

```

Maintainer:    stvjc <stvjc@channing.harvard.edu>
License:       Artistic (see COPYING)
Packaged:      Thu Aug 10 11:02:04 2006; stvjc
Collate:       snpMeta.R AllClasses.R AllGenerics.R HapMapUtils.R
               exclMono.R countRare.R fastAGM.R genoString.R oneFit.R
               racExSet-methods.R snpScreenResult-methods.R snps3Pto.R
               updateObject.R zzz.R
Built:         R 2.4.0; i386-pc-mingw32; 2006-10-13 16:01:10; windows

```

Index:

HM2rac	compute rare allele count from a hapmap file
HMworkflow	function to bind together HapMap genotyping results and expression data
geneLocs	gene metadata from NCBI
genoStrings	create a character vector of genotype value strings
make_racExSet	create a racExSet from simpler constituents
plot_EvG	plot expression vs genotype
racExSet-class	Class "racExSet" for combining RareAlleleCount representations of SNPs, gene expression data, and other phenotype data
snpMeta-class	Class "snpMeta" -- HapMap-based metadata structures for SNPs
snpScreen	compute model fits over a sequence of SNPs
snps	accessor for genotype data in a ggExprSet

Further information is available in the following vignettes in directory 'D:/biocbld/1.9d/tmpdir/Rinst526859155/GGtools/doc':

GGoverview: GGtools overview (source)

Session information for this vignette build:

```
> sessionInfo()
```

```

R version 2.4.0 (2006-10-03)
i386-pc-mingw32

```

```
locale:
LC_COLLATE=English_United States.1252;LC_CTYPE=English_United States.1252;LC_MONETARY=E
```

```
attached base packages:
```

```
[1] "tools"      "methods"    "stats"      "graphics"   "grDevices"  "utils"
[7] "datasets"   "base"
```

```
other attached packages:
```

```
GGtools geneplotter    annotate    hgfocus    Biobase
"1.2.0"    "1.12.0"    "1.12.0"   "1.14.0"   "1.12.1"
```