

# GGtools 2011: leaner software for genetics of gene expression

VJ Carey (stvjc at channing.harvard.edu)

August 25, 2011

## Contents

<b>1</b>	<b>Introduction; major changes</b>	<b>2</b>
<b>2</b>	<b>A simple exploration of eQTL for a selected gene</b>	<b>2</b>
2.1	Filtering an smlSet for a chromosome-wide test . . . . .	2
2.2	Executing tests and interrogating the results . . . . .	3
2.3	Visualization . . . . .	4
2.4	Checking coincidence with other genomic features . . . . .	8
<b>3</b>	<b>Supporting comprehensive testing</b>	<b>10</b>
3.1	The default behavior of <code>eqt1Tests</code> . . . . .	10
3.2	Reducing the memory footprint for comprehensive ‘same chromosome’ tests	11
3.3	Acquiring test results within specified intervals . . . . .	13
<b>4</b>	<b>Working with multiple populations</b>	<b>16</b>
<b>5</b>	<b>Session information</b>	<b>17</b>

# 1 Introduction; major changes

Since its introduction in 2006, GGtools has provided a number of data structures and tools for exploratory data analysis and hypothesis testing in expression genetics. Since 2006, Bioconductor’s facilities for representing genomes and for exploiting advanced ideas in computing and statistical modeling have evolved considerably, and many components of GGtools/GGBase need to be discarded to promote use of new facilities.

The following major changes have been made.

- **smlSet** instances should not be used for genotyping panels of more than one million loci. A packaging discipline has been introduced. An expression genetics experiment should be managed in a package in which expression data are held in an **ExpressionSet** instance in the **data** folder, and *snpStats* **SnpMatrix** instances are stored in **inst/parts**. After installation, the **getSS** function constructs an **smlSet** instance on the fly – typically with modest memory footprint because only a fraction of available loci are held in memory. **externalize** demonstrates construction of a compliant package for expression genetics data by converting an existing whole- or partial-genome **smlSet** instance into a package.
- **gwSnpTests** has been retained for focused in-memory test computation, but it will be phased out.
- **eqt1Tests** is the primary function for inference in expression genetics. This function takes genotype and phenotype information from an **smlSet**, passes it through testing facilities of *snpStats*, and creates managed out-of-memory storage for voluminous testing results.
- Facilities for metadata handling, e.g., determining SNP locations, gene locations, vocabulary translations, are being phased out. Users must construct location and feature name metadata using up-to-date Bioconductor resources or their own resources.

The remainder of this document illustrates basic activities now supported by the package.

## 2 A simple exploration of eQTL for a selected gene

### 2.1 Filtering an **smlSet** for a chromosome-wide test

We retain a two-chromosome example with expression data derived from the Wellcome Trust GENEVAR project archives, and genotype data from HapMap phase II:

```
> library(GGtools)
> data(hmceuB36.2021)
> hmceuB36.2021
```

```

SnpMatrix-based genotype set:
number of samples: 90
number of chromosomes present: 2
annotation: illuminaHumanv1.db
Expression data dims: 47293 x 90
Phenodata: An object of class "AnnotatedDataFrame"
  sampleNames: NA06985 NA06991 ... NA12892 (90 total)
  varLabels: famid persid ... male (7 total)
  varMetadata: labelDescription

```

A commonly found gene with eQTL is CPNE1 on chromosome 20. We can determine the probe name in use for CPNE1 as follows:

```

> library("illuminaHumanv1.db")
> cpn = get("CPNE1", revmap(illuminaHumanv1SYMBOL))
> use = intersect(cpn, featureNames(hmceuB36.2021))
> if (length(use) == 0) stop("probe not on array")
> use = use[1]

```

We will now trim down the `smlSet` instance to this gene and SNPs on chr20:

```

> hmlit = hmceuB36.2021[probeId(use), ]
> hmlit = hmlit[chrnum("20"), ]
> hmlit

```

```

SnpMatrix-based genotype set:
number of samples: 90
number of chromosomes present: 1
annotation: illuminaHumanv1.db
Expression data dims: 1 x 90
Phenodata: An object of class "AnnotatedDataFrame"
  sampleNames: NA06985 NA06991 ... NA12892 (90 total)
  varLabels: famid persid ... male (7 total)
  varMetadata: labelDescription

```

## 2.2 Executing tests and interrogating the results

The preferred testing procedure is:

```

> tname = function() gsub(".*/.*\\""", "", tempfile())
> e1 = eqtlTests(hmlit, ~male, targdir = tname(), runname = tname())
> e1

```

```
eqtlTools results manager, computed Thu Aug 25 10:50:30 2011
gene annotation: illuminaHumanv1.db
There are 1 chromosomes analyzed.
some genes (out of 1): GI_23397697-A
some snps (out of 119921): rs4814683 rs6076506 ... rs6062370 rs6090120
```

```
> class(e1)

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

We support focused queries into the manager of results. For this application we are computing  $\chi^2(1)$  statistics for each gene-SNP combination, measuring the fit of an additive genetic model with, in this case, adjustment for gender.

```
> e1[rsid("rs6060535"), probeId("GI_23397697-A")]

$`20`
      GI_23397697-A
rs6060535      53.62
attr(,"vmode")
[1] "short"
```

We can acquire the most predictive SNP using `topFeats`.

```
> topFeats(probeId(cpn), mgr = e1, ffind = 1)

rs17093026 rs1118233 rs12480408 rs6060535 rs11696527 rs6058303 rs2425078
      56.21      54.10      53.62      53.62      53.62      53.62      53.62
rs6060578 rs1970357 rs7273815
      53.62      53.62      53.53
```

Here the `ffind` parameter must be set to pick among files that might be managed by the manager bound to `mgr`. See the technical appendix for more details.

## 2.3 Visualization

To visualize the results, using hg18 locations for the SNP, we must acquire location metadata. Note that we are using the 2009 SNPlocs metadata from Bioconductor; for later builds, an `as.GRanges` parameter can be set to help avoid some of the conversion tasks.

```

> library(SNPlocs.Hsapiens.dbSNP.20090506)
> c20 = getSNPlocs("chr20")
> c20 = SNPlocs.Hsapiens.dbSNP.20100427:::SNPlocsAsGranges(c20,
+   "chr20")
> rs20 = paste("rs", elementMetadata(c20)$RefSNP_id, sep = "")
> names(c20) = rs20
> length(c20)

```

```
[1] 323041
```

```
> c20[1:3, ]
```

GRanges with 3 ranges and 2 elementMetadata values

	seqnames	ranges	strand	RefSNP_id	alleles_as_ambig
	<Rle>	<IRanges>	<Rle>	<character>	<character>
rs28753379	chr20	[8572, 8572]	*	28753379	Y
rs28579812	chr20	[8646, 8646]	*	28579812	Y
rs6078030	chr20	[9098, 9098]	*	6078030	Y

```
seqlengths
```

```
chr20
```

```
NA
```

The names of SNP for which tests were computed are

```

> tn = rownames(e1@fflist[[1]])
> tn[1:10]

```

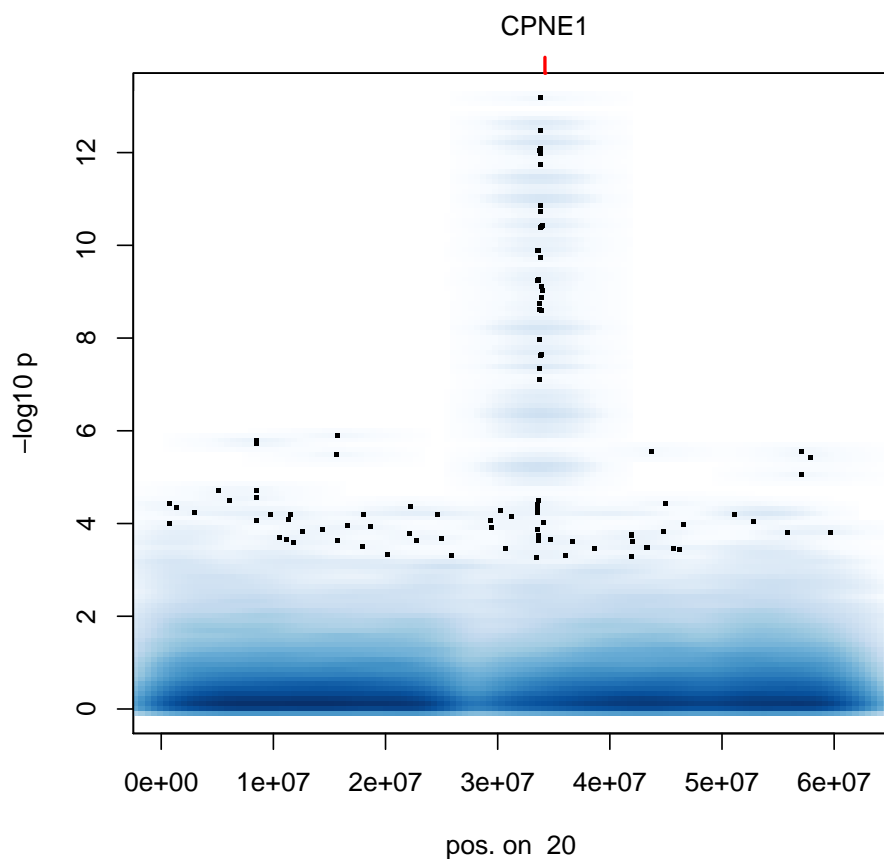
```

[1] "rs4814683" "rs6076506" "rs6139074" "rs1418258" "rs7274499"
[6] "rs6116610" "rs6084963" "rs13043000" "rs6054257" "rs6140075"

```

We can now obtain the manhattan plot:

```
> manhPlot(cpn, mgr = e1, ffind = 1, locGRanges = c20, cex = 3)
```



In general we will prefer to have this information in a browser track. To accomplish this, add the test results to the GRanges instance for locations as ‘score’ and export a wig file.

We have to align the SNPs with locations with those for which we have tests. First, restrict the locations.

```
> okn = intersect(tn, names(c20))
> c20ok = c20[okn]
```

Now obtain the -log10 p values:

```
> chis1 = e1[, probeId(cpn)][[1]][okn, ]
> ml10p = -log10(1 - pchisq(chis1, 1))
```

Bind to the location data:

```
> elementMetadata(c20ok)$score = ml10p
> c20ok[1:5, ]
```

GRanges with 5 ranges and 3 elementMetadata values

	seqnames	ranges	strand	RefSNP_id	alleles_as_ambig
	<Rle>	<IRanges>	<Rle>	<character>	<character>
rs4814683	chr20	[ 9795, 9795]	*	4814683	K
rs6076506	chr20	[11231, 11231]	*	6076506	K
rs6139074	chr20	[11244, 11244]	*	6139074	M
rs1418258	chr20	[11799, 11799]	*	1418258	Y
rs7274499	chr20	[12150, 12150]	*	7274499	M

	score
	<numeric>
rs4814683	0.1016386
rs6076506	0.0000000
rs6139074	0.3876541
rs1418258	0.1016386
rs7274499	0.0000000

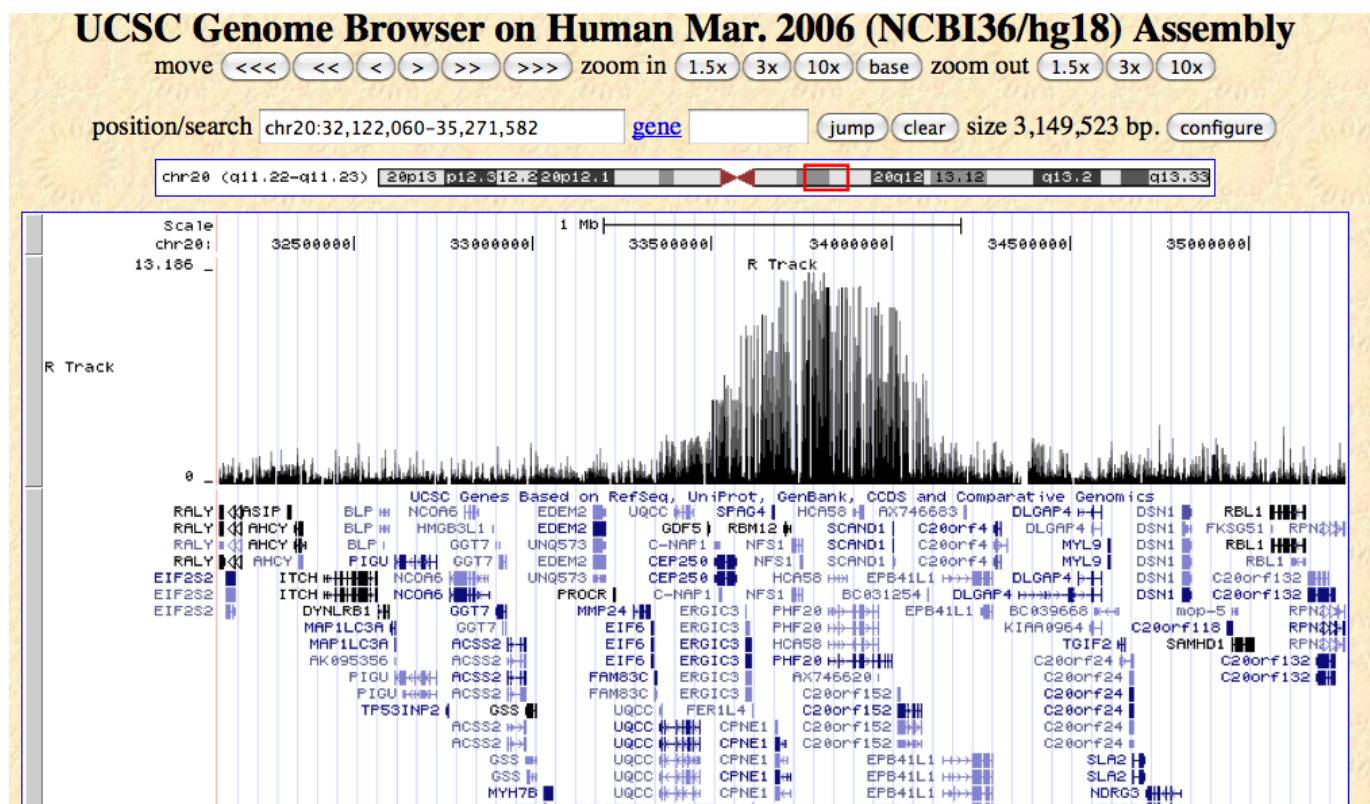
seqlengths

chr20  
NA

and export:

```
> library(rtracklayer)
> export(c20ok, "cpne1.wig")
```

This is imported as a custom track and viewed as:



## 2.4 Checking coincidence with other genomic features

When we store the statistical results in a `GRanges` instance, we have capacity to quickly assess genomic context of high-scoring SNP. The top 20 SNP and their locations are

```
> top200 = c20ok[order(elementMetadata(c20ok)$score, decreasing = TRUE)[1:200]]
> top200[196:200]
```

`GRanges` with 5 ranges and 3 `elementMetadata` values

	seqnames	ranges	strand	RefSNP_id	alleles_as_ambig
	<Rle>	<IRanges>	<Rle>	<character>	<character>
rs11697672	chr20	[34088036, 34088036]	*	11697672	W
rs10439606	chr20	[33531574, 33531574]	*	10439606	R
rs4911507	chr20	[33539412, 33539412]	*	4911507	Y
rs4911509	chr20	[33550723, 33550723]	*	4911509	S
rs6079979	chr20	[15756805, 15756805]	*	6079979	R
	score				
	<numeric>				



```
rs11697672 6.007190
rs10439606 5.921447
rs4911507 5.921447
rs4911509 5.921447
rs6079979 5.907904
```

```
seqlengths
chr20
NA
```

We see that we have SNP with p-values at most  $10^{-5}$ . We will now check whether any of these lie in regulatory regions defined by ORegAnno, confined to 10MB near CPNE1 on chr20.

```
> library(rtracklayer)
> ss = browserSession()
> genome(ss) = "hg18"
> GG = GRanges(seqnames = "chr20", IRanges(3e+07, 4e+07))
> oq = ucscTableQuery(ss, "oreganno", range = GG)
> toq = track(oq)
> ov = subsetByOverlaps(toq, top200)
> ov
```

```
UCSC track 'oreganno'
```

```
UCSCData with 3 rows and 3 value columns across 1 space
```

	space	ranges		name	score	strand
	<factor>	<IRanges>		<character>	<numeric>	<character>
1	chr20	[33650951, 33651421]		OREG0004040	0	-
2	chr20	[33706235, 33706695]		OREG0004046	0	-
3	chr20	[33935478, 33936028]		OREG0025894	0	+

For further discovery of the regulatory feature characteristics, consider

```
> tableName(oq) = "oregannoAttr"
> featdat = getTable(oq)
> featdat[featdat$id %in% ov$name, ]
```

Retrieving the table can take a while, so we suppress it in this vignette, but on 8 Mar 2011 the result was:

```
> featdat[ featdat$id %in% ov$name, ]
```

	id	attribute	attrVal
5809	OREG0004040	type	REGULATORY REGION
5810	OREG0004040	Gene	FER1L4

5811	OREG0004040	Dataset	Stanford ENCODE Dataset
5812	OREG0004040	EvidenceSubtype	Transient transfection luciferase assay
5833	OREG0004046	type	REGULATORY REGION
5834	OREG0004046	Gene	RBM12
5835	OREG0004046	Dataset	Stanford ENCODE Dataset
5836	OREG0004046	EvidenceSubtype	Transient transfection luciferase assay
86472	OREG0025894	type	REGULATORY REGION
86473	OREG0025894	TFbs	CTCF
86474	OREG0025894	Dataset	CTCF ChIP-chip sites (Ren lab)
86475	OREG0025894	EvidenceSubtype	ChIP-on-chip (ChIP-chip)

enabling us to learn more about the annotations of regions in which high-scoring SNP are found.

Clearly other UCSC-mediated resources can be used similarly to help interpret SNP scores.

## 3 Supporting comprehensive testing

### 3.1 The default behavior of `eqt1Tests`

`eqt1Tests` will, by default, compute scores for all gene-SNP combinations in the `smlSet` instance to which it is applied. With a multicore system, some concurrency can be achieved. To illustrate the activity, we take only 20 genes sampled from the illumina platform and compute all eQTL tests for SNP on chromosomes 20 and 21.

```
> NG = length(featureNames(hmceuB36.2021))
> hmct = hmceuB36.2021[sample(1:NG, size = 20, replace = FALSE),
+   ]
> applier = lapply
> chkmult = function() {
+   length(grep("^multicore$", installed.packages()[, 1])) >
+   0
+ }
> if (chkmult()) {
+   library(multicore)
+   options(cores = min(c(multicore:::volatile$detectedCores,
+   6)))
+   applier = mclapply
+ }
> unix.time(e2 <- eqt1Tests(hmct, ~male, geneApply = applier, targdir = tempdir(),
+   runname = tname()))

    user  system elapsed
20.687   1.863  13.658
```

We are now managing 3.4 million tests. These are stored out of memory using the *ff* system for memory-mapped disk representations of R objects.

```
> e2@fflist[[1]]

ff (closed) short length=2398420 (2398420) dim=c(119921,20) dimorder=c(1,2)

> sapply(e2@fflist, dim)

      20      21
[1,] 119921 50165
[2,]      20      20
```

It is unpleasant to use the at-sign *fflist* notation but until a richer set of use cases is identified, this primitive idiom will be used.

In general we will break up a comprehensive run by gene sets – typically one run for all genes on a chromosome. In this case, we will generate 22 *eqtlTestManager* instances; these can be unified with a *cisTransDirector* or *multiCisDirector* object. We will discuss that later.

### 3.2 Reducing the memory footprint for comprehensive ‘same chromosome’ tests

In the example given above, the data for genotypes on chr20 and chr21 are simultaneously in memory. Up to now, we have often stored all genotype data in memory when using *smlSets*, leading to multiple Gb of RAM consumption, causing conflicts and loss of speed. We now describe an approach to building and using package-based experiment representations, which allow reduced and more flexible use of resources.

We can take a unified *smlSet* instance and turn it into a package using the *externalize* function. We also install it:

```
> edname = paste("exdem", tname(), sep = "")
> if (interactive() & file.exists(edname)) stop(paste("will not remove existing '",
+   edname, "' folder"))
> externalize(hmceuB36.2021, edname)

now install exdemfile237fc982
NULL

> tar(tna <- paste(edname, ".tar.gz", sep = ""), edname, compression = "gzip")
> install.packages(tna, repos = NULL, type = "source")
> try(system(paste("rm -rf", tna)))
> try(system(paste("rm -rf", edname)))
```

Now we can use `getSS` to create tailored `smlSet` instances on the fly.

```
> gc()
```

```

      used  (Mb) gc trigger  (Mb) max used  (Mb)
Ncells 3472942 92.8    5241317 140.0  5241317 140.0
Vcells 13829237 105.6   25379777 193.7  24039357 183.5

```

```
> s1 = getSS(edname, "20")
```

To get a tailored `smlSet`, use `getSS("exdemfile237fc982", [chrvec])`  
available chromosomes are named 20 21

```
> gc()
```

```

      used  (Mb) gc trigger  (Mb) max used  (Mb)
Ncells 3474086 92.8    5241317  140  5241317 140.0
Vcells 19567293 149.3   26728765  204  24039357 183.5

```

```
> rm(s1)
```

```
> gc()
```

```

      used  (Mb) gc trigger  (Mb) max used  (Mb)
Ncells 3473657 92.8    5241317  140  5241317 140.0
Vcells 13829659 105.6   26728765  204  24039357 183.5

```

This technology is used in the `makeDiagDirector` function. This function requires a list that maps genes to chromosomes. In this example, we take 20 genes from each of chromosome 20 and 21 to define the map. All same-chromosome tests for these 40 genes will be computed. No concurrency is used.

```

> mymap = lapply(c("20", "21"), function(x) get(x, revmap(illuminaHumanv1CHR))[1:20])
> mymap = lapply(mymap, function(x) intersect(x, featureNames(hmceuB36.2021)))
> names(mymap) = c("20", "21")
> unix.time(md <- makeDiagDirector(edname, mymap, rhs = ~male,
+   sleeplen = 2, runname = tname(), targdir = tname()))

```

```

      user  system elapsed
32.245    1.320   38.644

```

```
> md
```

multiCisDirector instance with 2 eqtlTestsManagers.  
elements: 20 21

It would be typical to save the director object (here named `md`) to disk, so that the ff files that have been computed can be reused conveniently.

### 3.3 Acquiring test results within specified intervals

Here we will discuss how to use `cisProxScores` with the director computed above. The calling sequence is

```
cisProxScores(smlSet, fmla, dradset, direc = NULL, folder, runname,  
             geneApply = mclapply, saveDirector = TRUE, snpGRL=NULL,  
             geneGRL=NULL, snpannopack="SNPlocs.Hsapiens.dbSNP.20090506", ffind=NULL, ...)
```

so in fact one can pass an `smlSet` instance as input. However, if a `multiCisDirector` instance is available, that can be used, avoiding the computation of the tests. If the director is used, we need only specify

- **dradset**: the collection of nested paired intervals around genes within which scores will be collected
- **geneGRL**: a list of `GRanges`, with one element per chromosome, providing coordinates of gene start and stop locations
- **snpGRL**: a list of `GRanges`, with one element per chromosome, providing coordinates of SNPs with names
- **ffind**: an index value into the manager-specific lists for the scores we wish to retrieve; in most cases this will be 1

The names of probesets scored in `md` can be obtained:

```
> ps = lapply(md@mgrs, function(x) colnames(x@fflist[[1]]))  
> names(ps) = names(md@mgrs)  
> ps  
  
$`20`  
[1] "GI_10518498-S" "GI_10800147-S" "GI_11037054-S" "GI_11184232-S"  
[5] "GI_11496985-S" "GI_11641410-A" "GI_12408676-S" "GI_12669910-S"  
[9] "GI_12707574-S" "GI_13027601-S" "GI_13027806-S" "GI_13129013-S"  
[13] "GI_13129143-S" "GI_13259532-S" "GI_13376346-S" "GI_13376446-S"  
[17] "GI_13435378-S" "GI_13518027-S" "GI_13540544-S" "GI_13699829-I"  
  
$`21`  
[1] "GI_10092616-S" "GI_10835182-S" "GI_11037060-A" "GI_11037060-I"  
[5] "GI_11038653-S" "GI_11038670-S" "GI_11140814-S" "GI_11342663-S"  
[9] "GI_11345479-S" "GI_13376728-S" "GI_13562087-S" "GI_14043059-S"  
[13] "GI_14249201-S" "GI_14249687-S" "GI_14602454-I" "GI_14602456-A"  
[17] "GI_14602456-I" "GI_14602458-S" "GI_15011912-S" "GI_15451880-A"
```

The `GRanges` instance is then obtained as

```

> getgr = function(x, map) abs(sapply(mget(x, map), "[", 1))
> st = lapply(ps, getgr, illuminaHumanv1CHRLOC)
> en = lapply(ps, getgr, illuminaHumanv1CHRLOCEND)
> sp = names(ps)
> grl = list()
> for (i in 1:length(ps)) {
+   grl[[i]] = GRanges(seqnames = sp[i], IRanges(st[[i]], en[[i]]))
+   names(grl[[i]]) = ps[[i]]
+ }
> names(grl) = names(ps)

```

SNP locations from hg18 can be obtained as follows (we did chr20 above):

```

> seqlevels(c20) <- sub("chr", "", seqlevels(c20))
> c21 = getSNPlocs("chr21")
> c21 = SNPlocs.Hsapiens.dbSNP.20100427:::SNPlocsAsGranges(c21,
+   "chr21")
> rs21 = paste("rs", elementMetadata(c21)$RefSNP_id, sep = "")
> names(c21) = rs21
> seqlevels(c21) <- sub("chr", "", seqlevels(c21))
> srl = list(`20` = c20, `21` = c21)

```

Now compute scores in intervals to 50kb around gene, then from 50kb to 100kb, then from 100kb to 500kb:

```

> csc = cisProxScores(direc = md, dradset = c(0, 50000, 1e+05,
+   5e+05), snpGRL = srl, geneGRL = grl, geneApply = lapply,
+   ffind = 1)

```

```
2021202120212021
```

This yields a deeply nested structure; a hint of best scores per gene within intervals comes from

```

> lapply(csc, lapply, sapply, function(x) max(unlist(x)))
$`FL0e+00.0e+00`
$`FL0e+00.0e+00`$`20`
GI_10518498-S GI_10800147-S GI_11037054-S GI_11184232-S GI_11496985-S
      5.32      8.27      1.29      7.32      3.33
GI_11641410-A GI_12408676-S GI_12669910-S GI_12707574-S GI_13027601-S
      2.53      2.54      3.56      1.78      2.08
GI_13027806-S GI_13129013-S GI_13129143-S GI_13259532-S GI_13376346-S
      5.40      6.55      6.73      1.49      13.62
GI_13376446-S GI_13435378-S GI_13518027-S GI_13540544-S GI_13699829-I

```

3.00	3.46	4.71	4.12	1.71
------	------	------	------	------

\$`FL0e+00.0e+00`\$`21`

GI_10835182-S	GI_11037060-A	GI_11037060-I	GI_11038653-S	GI_11038670-S
2.92	3.25	4.42	0.41	6.68
GI_11342663-S	GI_11345479-S	GI_13376728-S	GI_13562087-S	GI_14043059-S
8.52	5.30	0.06	0.99	2.67
GI_14249201-S	GI_14602454-I	GI_14602456-A	GI_14602456-I	GI_14602458-S
0.40	2.22	0.81	3.29	4.74
GI_15451880-A				
9.88				

\$`FL0e+00.5e+04`

\$`FL0e+00.5e+04`\$`20`

GI_10518498-S	GI_10800147-S	GI_11037054-S	GI_11184232-S	GI_11496985-S
8.23	7.61	6.34	8.57	7.94
GI_11641410-A	GI_12408676-S	GI_12669910-S	GI_12707574-S	GI_13027601-S
6.14	10.23	5.50	5.18	6.10
GI_13027806-S	GI_13129013-S	GI_13129143-S	GI_13259532-S	GI_13376346-S
7.72	12.26	5.66	4.35	5.16
GI_13376446-S	GI_13435378-S	GI_13518027-S	GI_13540544-S	GI_13699829-I
6.63	4.69	6.58	6.93	7.57

\$`FL0e+00.5e+04`\$`21`

GI_10835182-S	GI_11037060-A	GI_11037060-I	GI_11038653-S	GI_11038670-S
4.09	2.79	2.42	7.39	3.05
GI_11342663-S	GI_11345479-S	GI_13376728-S	GI_13562087-S	GI_14043059-S
4.98	4.42	3.76	2.04	6.39
GI_14249201-S	GI_14249687-S	GI_14602454-I	GI_14602456-A	GI_14602456-I
10.35	3.46	2.53	6.73	4.01
GI_14602458-S	GI_15451880-A			
5.63	4.12			

\$`FL5e+04.1e+05`

\$`FL5e+04.1e+05`\$`20`

GI_10518498-S	GI_10800147-S	GI_11037054-S	GI_11184232-S	GI_11496985-S
4.62	3.87	4.71	7.51	6.62
GI_11641410-A	GI_12408676-S	GI_12669910-S	GI_12707574-S	GI_13027601-S
3.69	7.64	3.56	6.30	5.15
GI_13027806-S	GI_13129013-S	GI_13129143-S	GI_13259532-S	GI_13376346-S

9.90	10.48	8.78	8.52	4.61
GI_13376446-S	GI_13435378-S	GI_13518027-S	GI_13540544-S	GI_13699829-I
6.56	2.60	3.12	5.80	8.37

\$`FL5e+04.1e+05`\$`21`

GI_10835182-S	GI_11037060-A	GI_11037060-I	GI_11038653-S	GI_11038670-S
4.15	3.13	4.54	6.07	8.43
GI_11342663-S	GI_11345479-S	GI_13376728-S	GI_13562087-S	GI_14043059-S
7.83	5.14	3.67	4.59	3.59
GI_14249201-S	GI_14249687-S	GI_14602454-I	GI_14602456-A	GI_14602456-I
10.57	6.13	4.04	7.39	5.63
GI_14602458-S	GI_15451880-A			
6.84	8.05			

\$`FL1e+05.5e+05`

\$`FL1e+05.5e+05`\$`20`

GI_10518498-S	GI_10800147-S	GI_11037054-S	GI_11184232-S	GI_11496985-S
9.41	9.50	10.61	7.56	9.76
GI_11641410-A	GI_12408676-S	GI_12669910-S	GI_12707574-S	GI_13027601-S
7.70	15.79	7.26	7.57	12.89
GI_13027806-S	GI_13129013-S	GI_13129143-S	GI_13259532-S	GI_13376346-S
10.64	10.94	9.73	6.95	9.71
GI_13376446-S	GI_13435378-S	GI_13518027-S	GI_13540544-S	GI_13699829-I
11.25	8.21	9.11	14.71	9.68

\$`FL1e+05.5e+05`\$`21`

GI_10092616-S	GI_10835182-S	GI_11037060-A	GI_11037060-I	GI_11038653-S
9.70	8.50	8.99	12.33	11.62
GI_11038670-S	GI_11342663-S	GI_11345479-S	GI_13376728-S	GI_13562087-S
8.60	6.93	10.37	9.00	8.09
GI_14043059-S	GI_14249201-S	GI_14249687-S	GI_14602454-I	GI_14602456-A
7.72	7.27	22.05	9.73	8.05
GI_14602456-I	GI_14602458-S	GI_15011912-S	GI_15451880-A	
12.24	10.32	6.55	14.13	

## 4 Working with multiple populations

Here we will discuss how to take multiple population-specific directors and obtain a director representing the sums across populations.



## 5 Session information

```
> sessionInfo()
```

```
R version 2.13.1 (2011-07-08)
```

```
Platform: i386-apple-darwin9.8.0/i386 (32-bit)
```

```
locale:
```

```
[1] C
```

```
attached base packages:
```

```
[1] tools      splines    stats      graphics  grDevices  utils      datasets  
[8] methods    base
```

```
other attached packages:
```

```
[1] exdemfile237fc982_0.0.0  
[2] multicore_0.1-5  
[3] SNPlocs.Hsapiens.dbSNP.20090506_0.99.1  
[4] illuminaHumanv1.db_1.10.0  
[5] GGtools_3.10.2  
[6] ff_2.2-3  
[7] bit_1.1-7  
[8] GenomicRanges_1.4.8  
[9] org.Hs.eg.db_2.5.0  
[10] rtracklayer_1.12.4  
[11] RCurl_1.6-9  
[12] bitops_1.0-4.1  
[13] IRanges_1.10.6  
[14] annotate_1.30.0  
[15] AnnotationDbi_1.14.1  
[16] GGBase_3.12.0  
[17] RSQLite_0.9-4  
[18] DBI_0.2-5  
[19] snpStats_1.2.1  
[20] Matrix_0.9996875-3  
[21] lattice_0.19-33  
[22] survival_2.36-9  
[23] Biobase_2.12.2
```

```
loaded via a namespace (and not attached):
```

```
[1] BSgenome_1.20.0  
[2] Biostrings_2.20.3  
[3] KernSmooth_2.23-6
```

- [4] SNPlocs.Hsapiens.dbSNP.20100427\_0.99.2
- [5] XML\_3.4-2
- [6] grid\_2.13.1
- [7] xtable\_1.5-6