DNA Methylation
Practical 3 - Annotation

Allan McRae
a.mcrae@uq.edu.au

Sonia Shah
sonia.shah@uq.edu.au

Overview

This practical will cover

- Annotating methylation probes

- Investigating cross-binding issues

Read in Data

We will use the results from the EWAS performed in the previous practical session

```
load("ewas.ret.Robj")


ls()
names(ewas.ret)
head(ewas.ret$p.value)
```

Extract and Sort p-values

We will work with the results from the SVA analysis

```
pval = sort(ewas.ret$p.value[,"sva"])
head(pval)

probe = names(pval)
```

Remove Sex Chromosome SNPs

We saw that many of the significant differences in methylation between the sexes were on the X and Y chromosome. These are “not interesting”, so lets filter the results to only include tosomal chromosomes.

```
library(meffil)

autosomal = meffil.get.autosomal.sites("450k")

ap = is.element(probe, autosomal)
head(ap, n=1000)

probe = probe[ap]
pval = pval[ap]
```

Create a QQ plot for the autosomal probes

How Many Significant Probes

Calculate a Bonferroni significance threshold and count the number of autosomal probes with smaller p-values

```
n = length(pval)

thresh = 0.05/n

length(which(pval < thresh))
```

Get Probe Annotation

There are many probe annotation files available for the Illumina HumanMethylation450k array. We will use the annotation from a package required for meffil

```
library(IlluminaHumanMethylation450kanno.ilmn12.hg19)
data(IlluminaHumanMethylation450kanno.ilmn12.hg19)

IlluminaHumanMethylation450kanno.ilmn12.hg19

anno = getAnnotation(
          IlluminaHumanMethylation450kanno.ilmn12.hg19)
head(anno)
```

Extracting Annotation

Probe “cg00804338” is the most significant autosomal probe. Lets extract some information about that

```
anno["cg00804338",]
```

To extract a particular piece of information, use

```
anno["cg00804338",]$SourceSeq
```

What Sort of Questions To Ask?

We could ask a variety of questions here...

e.g.

* Where are the significant probes in relation to CpG islands?
* Does any of this differ between cases and controls?

BLAST

One issue with DNA methylation arrays is that probes can potentially cross bind with other regions of the genome.

When testing for sex differences, cross-binding to the sex chromosomes can create false positive results.

Randomly select a few probes from the top 100 most significant results and extract their probe sequence. We will test these for non-specific binding.

BLAST

Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi

- Select “Nucleotide BLAST”

- Paste the probe sequence

- For database, choose “Reference geomic sequences”
- Type “human” for organism

- Select “blastn” in the Program Selection

Check what chromosomes the probe binds to. A commonly used threshold is binding to 80% of the probe length (40bp) with a high identity and no gaps in the alignment.

Group Activity

Select a random significant probes

Extract the probe sequence

Do a BLAST search on the probe to determine if it cross-binds to another chromosome

Record your answer on the whiteboard

```
probe[sample(1:6558, 1)]
```