Combining expression and CGH data on breast cancer cell lines: Neve2006

VJ Carey, package maintainer, <stvjc@channing.harvard.edu>

October 13, 2026

Contents

1 Introduction **1**

2 CGH feature data **2**

3 Gene expression data in relation to copy number **3**

1 Introduction

The basic source information is obtainable by interrogating objects in the package.

```
> library(Neve2006)
> data(neveRMAmatch)
> neveRMAmatch

ExpressionSet (storageMode: lockedEnvironment)
assayData: 22283 features, 50 samples
  element names: exprs
protocolData: none
phenoData
  rowNames: 600MPE AU565 ... ZR75B (50 total)
  varLabels: ind cellLine ... reductMamm (15 total)
  varMetadata: labelDescription
featureData: none
experimentData: use 'experimentData(object)'
  pubMedIds: 17157791
Annotation: hgu133a

> experimentData(neveRMAmatch)
```

```
Experiment data
  Experimenter name: Neve RM
  Laboratory: Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley,
  Contact information:
  Title: A collection of breast cancer cell lines for the study of functionally distinc
  URL:
  PMIDs: 17157791

  Abstract: A 116 word abstract is available. Use 'abstract' method.
```

```
> data(neveCGHmatch)
> neveCGHmatch
```

```
cghSet (storageMode: lockedEnvironment)
assayData: 2696 features, 50 samples
  element names: exprs
protocolData: none
phenoData
  rowNames: 600MPE AU565 ... ZR75B (50 total)
  varLabels: ind cellLine ... reductMamm (15 total)
  varMetadata: labelDescription
featureData
  featureNames: RP11-82D16 RP11-62M23 ... RP11-247J14 (2696 total)
  fvarLabels: Clone Chrom kb kbGenome
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
  pubMedIds: 17157791
Annotation: Neve2006.caArrayDB
```

```
> all.equal(sampleNames(neveRMAmatch), sampleNames(neveCGHmatch))
```

```
[1] TRUE
```

We can see that there are 50 arrays from the U133A platform and 50 CGH samples, and the sample names match. The PMID for the primary manuscript is provided; additional information on URLs where the base data can be found should be added.

2 CGH feature data

Let's look at metadata provided on the first five clones in the CGH platform:

```
> featureData(neveCGHmatch)[1:5,]
```

```
An object of class 'AnnotatedDataFrame'
  rowNames: RP11-82D16 RP11-62M23 ... RP11-51B4 (5 total)
  varLabels: Clone Chrom kb kbGenome
  varMetadata: labelDescription
```

The genomic location of clone sequence is tied completely to the build of the genome used to compute the location. This information is not provided in the data sources. We know from the paper that scanning and OncoBAC arrays were used, but no information about the details of annotation are immediately available.

Figure 5 of the main paper associates clone RP11-265K5 with a number of genes, including PROSC and BRF2. Let's see where this clone lives:

```
> pData(featureData(neveCGHmatch))[grep("RP11-265K5",
+   featureNames(neveCGHmatch)),]

                Clone Chrom       kb kbGenome
RP11-265K5 RP11-265K5     8 38125.49  1429440
```

If we use Map Viewer at NCBI for Homo sapiens build 36.2, we find this clone located at about 37180K, in the vicinity of genes WHSC1L1, LETM2, FGFR1, PPADPC1B, DDHD2, BAG4, LSM1, STAR, ASH2L. Some of these are noted in Figure 5. The cytoband is 8p12.

3 Gene expression data in relation to copy number

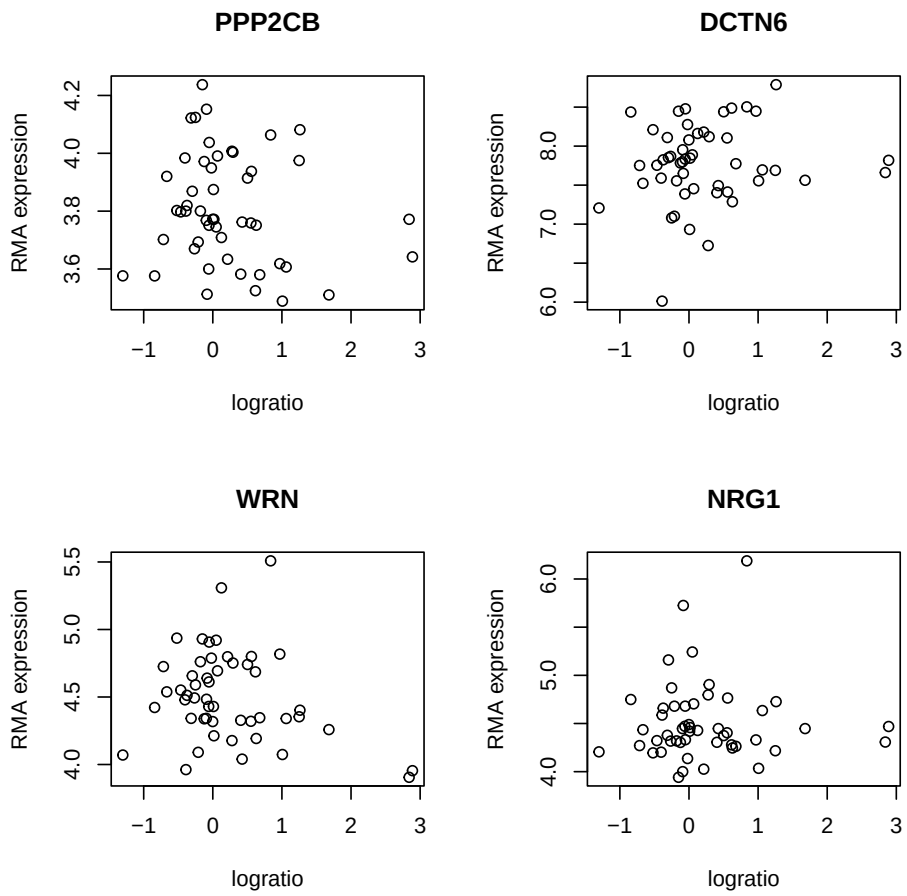
Let's identify probe sets annotated to 8p12.

```
> library(hgu133a.db)
> library(annotate)
> cb = as.list(hgu133aMAP)
> G8p12ind = grep("8p12", unlist(cb))
> ps8p12 = names(unlist(cb)[G8p12ind])
> nevex = exprs(neveRMAmatch)[ps8p12,]
> syms = as.character(unlist(lookUp(rownames(nevex), "hgu133a", "SYMBOL")))
```

The logratios that measure copy number will be used 'raw'.

```
> nevlr = as.numeric(logRatios(neveCGHmatch)["RP11-265K5",])

> par(mfrow=c(2,2))
> for (i in c(2,8,11,20))
+  plot( nevlr, nevex[i,], xlab="logratio", ylab="RMA expression",
+    main=syms[i])
> par(mfrow=c(1,1))
```



Also of interest is the expression distribution for SFRP1.

Connection of these associations to phenotype is straightforward with proper use of phenoData components of the ExpressionSet and cghSet objects. A unified representation of the two assays that includes support for linking clone and probe set is in development.