Overview over the DKFZ kidney data package

Wolfgang Huber

October 22, 2014

```
> library(kidpack)
```

The package contains five data objects: two for the processed data, including sample information (phenoData) and probe (genes) information, and three for the raw data, including spotting information and array processing information.

The data was measured at the German Cancer Research Centre in 2002 by Holger Sültmann [1]. He hybridized labeled cDNA from around 85 renal cell cancer biopsies that had been obtained at the University of Göttingen to cDNA arrays that he had produced himself. The cDNA arrays use the two-color Stanford-type spotted cDNA technology, with 4224 different clones spotted in duplicate. About half of the clones were selected for being expressed in kidney according to a previous study on whole genome arrays, and the other half are from Bernd Korn's (RZPD) 'onco collection'. Each sample was hybridized twice. 175 chips were scanned and digitized. After quality control, we selected one representative (good) chip for each sample, resulting in a set of 74. These are presented in the *exprSet* named `eset`.

1 What is it good for?

There were three different subtypes of renal cell cancer (RCC): clear cell (cc), papillary (p), and chromophobe (ch). These pheno-variables may be used for classification or differential expression. The gene expression is quite strongly associated with the subtype.

Other interesting phenovariables are the survival variables `(progress, rf.survival)` and `(died, survival.time)`. Obviously, the two are highly correlated. The binary variable `m` indicates whether metastases were present (and known) at the time of surgery. The association of the gene expression data with these variables is more subtle. Perhaps only wishful thinking.

The manuscript has been submitted. As soon as it is accepted, final, and public, the preprint will be made availabe in the doc directory of the package. Until then, please contact me (WH) directly and I can send you the most current version by email.

2 Processed data

```
> data(eset)
> data(cloneanno)
```

For later use, we define some plot colors for the `type` variable:

```
> unique(pData(eset)$type)

[1] "ccRCC" "pRCC"  "chRCC"

> cols <- c("red", "blue", "darkgreen")
> names(cols) <- c("ccRCC", "pRCC", "chRCC")
```

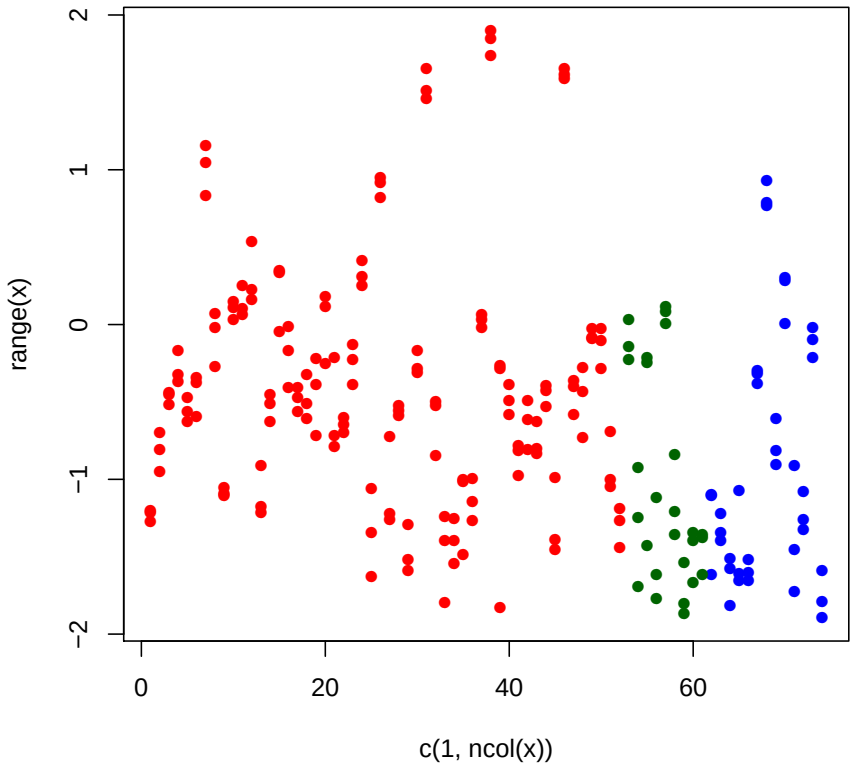
The chips contained three different clones that all probed for Fibronection 1:

```
> sel <- grep("fibronectin 1", cloneanno$description)
> cloneanno[sel, ]

     plate SrcRow SrcCol imageid       AccNumber spot1 spot2
2119     6     11      3  136798 r36450,"r36451"  1612  5964
2626     7     14     15  324061 w46461,"w46530"  1730  6082
2627     7     14     19  324997 w48576,"w48721"  1731  6083
                                                          description
2119  human mrna for fibronectin (fn precursor),"fibronectin 1 : fn1","fn1"
2626  human mrna for fibronectin (fn precursor),"fibronectin 1 : fn1","fn1"
2627  human mrna for fibronectin (fn precursor),"fibronectin 1 : fn1","fn1"
       vendor
2119 IMAGp950
2626 IMAGp950
2627 IMAGp950
```

Let's plot the expression values:

```
> eo <- eset[sel, order(pData(eset)$type)]
> x  <- exprs(eo)
> plot(c(1, ncol(x)), range(x), type="n")
> for(i in 1:nrow(x))
+   points(x[i, ], col=cols[pData(eo)$type], pch=16)
```



3 Raw data

Let's have a look at the raw data

```
> data(qua)
> data(hybanno)
> data(spotanno)
> s1 <- cloneanno$spot1[sel]
> s2 <- cloneanno$spot2[sel]
> s1

[1] 1612 1730 1731

> qua[s1, "fg.green", 1:3]
```

```
           1       2       3
1612 23.2745  7.4989 12.4166
1730 18.4937 10.3098 16.4911
1731 17.2570  8.1597 12.4474
```

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

```
             filename patientid slideid
1 00-P09206_E44-1.txt        87   E44-1
2 00-P09206_E63-3.txt        87   E63-3
3 00-U00363_E34-1.txt        86   E34-1
```

The columns `cloneanno$spot1`, `cloneanno$spot2` are of class `numeric`, with values from 1 to 8704. They refer to the rows of `spotannoanno`.

The column `spotanno$probe` is of class `numeric`, with values from 1 to 4224, referring to the rows of `cloneanno`.

References

[1] Gene expression in kidney cancer is associated with novel tumor subtypes, cytogenetic abnormalities and metastasis formation. Holger Sueltmann, Anja von Heydebreck, Wolfgang Huber, Ruprecht Kuner, Andreas Buness, Markus Vogt, Bastian Gunawan, Martin Vingron, Laszlo Fuezesi, and Annemarie Poustka (Division of Molecular Genome Analysis, German Cancer Research Center, Heidelberg). *Submitted 2004.*