Vignette for *Fletcher2013a*: gene expression data from breast cancer cells after FGFR2 signalling.

Mauro AA Castro[*], Michael NC Fletcher[*], Xin Wang, Ines de Santiago,
Martin O'Reilly, Suet-Feung Chin, Oscar M Rueda, Carlos Caldas,
Bruce AJ Ponder, Florian Markowetz and Kerstin B Meyer [†]
florian.markowetz@cancer.org.uk
kerstin.meyer@cancer.org.uk

April 27, 2023

Contents

[*]joint first authors

[†]Cancer Research UK - Cambridge Research Institute, Robinson Way Cambridge, CB2 0RE, UK.

1 Description

The package *Fletcher2013a* contains time-course gene expression data from MCF-7 cells treated under different experimental systems in order to perturb FGFR2 signalling (further details in the documentation of each dataset). The data comes from Fletcher et al. [1] (Cancer Research UK, Cambridge Research Institute, University of Cambridge, UK) where further details about the background and the experimental design of the study can be found. The R scripts provided in this vignette help to reproduce the differential expression analysis, including some initial checks assessing the consistency of the expression patterns across the sample groups. The second part of this study is available in the package *Fletcher2013b*, which has been separated for ease of data distribution. The package *Fletcher2013b* reproduces the network analysis as described in Fletcher et al. [1] based on publicly available data and the dataset presented in this package.

2 Differential expression analysis

The log2 gene expression data used in this vignette is presented in the form of 3 `ExpressionSet` objects called `Exp1`, `Exp2` and `Exp3`, all containing detailed phenotype information needed to run the differential expression analysis.

```
>    library(Fletcher2013a)
>    data(Exp1)
>    data(Exp2)
>    data(Exp3)
```

The differentially expressed (DE) genes are assessed by the linear modelling framework in the *limma* package [2]. Here the complete limma analysis is executed in a 4-step pipeline that (***i***) extracts the gene expression data and all relevant information from the `ExpressionSet` objects, (***ii***) prepares the design and fits the linear model, (***iii***) sets the contrasts according to the experimental groups and (***iv***) runs the eBayes correction and decides on the significance of each contrast for a p-value < 1e-2 and global adjustment method = 'BH'.

```
>    Fletcher2013pipeline.limma(Exp1)
>    Fletcher2013pipeline.limma(Exp2)
>    Fletcher2013pipeline.limma(Exp3)
```

Having fitted the linear model with default options, the pipeline should save all results in the current working directory in the form of 3 data files called `Exp1limma.rda`, `Exp2limma.rda` and `Exp3limma.rda`, including a set of figures showing the number of DE genes inferred in each analysis (see Figures 1a, 2a and 3a).

3 E2 and FGFR2 signatures

A pre-processed differential expression dataset is also available for the function Fletcher2013pipeline.deg, which can be used to retrieve the DE genes inferred in the limma analysis:

```
>    deExp1 <- Fletcher2013pipeline.deg(what="Exp1")
>    deExp2 <- Fletcher2013pipeline.deg(what="Exp2")
>    deExp3 <- Fletcher2013pipeline.deg(what="Exp3")
```

This wrapper function extracts consolidated gene lists organized according to the limma contrasts, including random gene lists that can be used in subsequent analyses. **DT**: DE genes in vehicle time-course contrasts; **E2**: DE genes in E2 vs. vehicle contrasts; **E2FGF10**: DE genes in E2+FGF10 vs. E2 contrasts; **E2AP20187**: DE genes in E2+AP20187 vs. E2 contrasts; **Tet** DE genes in Tet vs. vehicle contrasts (i.e. PlusTet.VEH vs. MinusTet.VEH); **TetDT**: DE genes in Tet time-course contrasts; **TetE2**: DE genes in Tet+E2 vs. Tet contrasts; **TetE2FGF10**: DE genes in Tet+E2+FGF10 vs. Tet+E2 contrasts; **random**: random gene lists (for a diagram representing all limma contrasts, please see Supplementary Figures 1, 2 and 3 in [1]).

4   Principal components analysis

In order to assess the experimental variation and the consistency among the sample groups, the pipeline performs a principal components analysis (PCA) on the gene expression matrices taken from the set of differentially expressed genes derived from the limma analysis. The PCA pipeline uses the R function *prcomp* with default options [3].

```
>    Fletcher2013pipeline.pca(Exp1)
>    Fletcher2013pipeline.pca(Exp2)
>    Fletcher2013pipeline.pca(Exp3)
```

The PCA pipeline should save all results in the current working directory in the form of 3 pdf files showing the two first principal components of each experimental system (see Figures 1b, 2b and 3b).

5   Follow-up on differentially expressed genes

Next we plot the overlap among three of these lists summarizing the number of DE genes called in each of the experiments (Figure 4a).

```
>    library(VennDiagram)
>    Fletcher2013pipeline.supp()
```

Additionally, the microarray data were confirmed in independent biological replicates by performing quantitative RT-PCR for a number of selected genes. IL8 is one of the most strongly induced genes and the previous pipeline also reproduces the main results from the follow-up on this gene. The increased IL8 mRNA expression is detected similarly by two microarray probes (Figure 4b) and by RT-PCR (Figure 4c). Furthermore, in line with increased expression we find that IL-8 secretion increased after FGF10 stimulation (Figure 4d). This secretion was blocked by PD173074 confirming that the effect is FGFR specific. The weighted Venn can be reproduced using the R-package **Vennerable**) (see *package installation* section):

```
>    #note: in order to run this option, the source code for the R-package Vennerable
>    #should be installed from the R-Forge repository (http://R-Forge.R-project.org)
>    library(Vennerable)
>    vv <- list(Exp1=deExp1$E2FGF10, Exp2=deExp2$E2AP20187, Exp3=deExp3$TetE2FGF10)
>    plotVenn(Venn(vv), doWeights=TRUE)
```
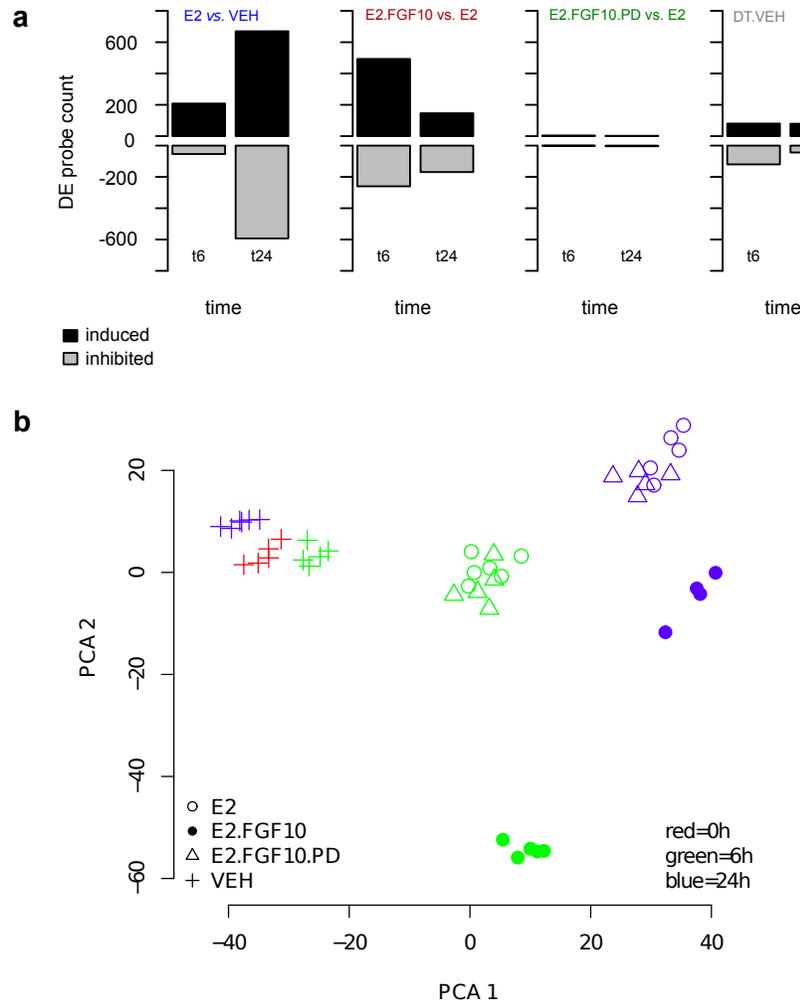
Figure 1: **Endogenous FGFRs perturbation experiments.** Summary of the differential expression analysis showing significant gene counts for P<0.01. (a) Bar charts depicting the number of probes significantly deregulated at each time point after stimulation. (b) PCA analysis of variation observed for the differentially expressed genes (n=2141) in the microarray analysis.
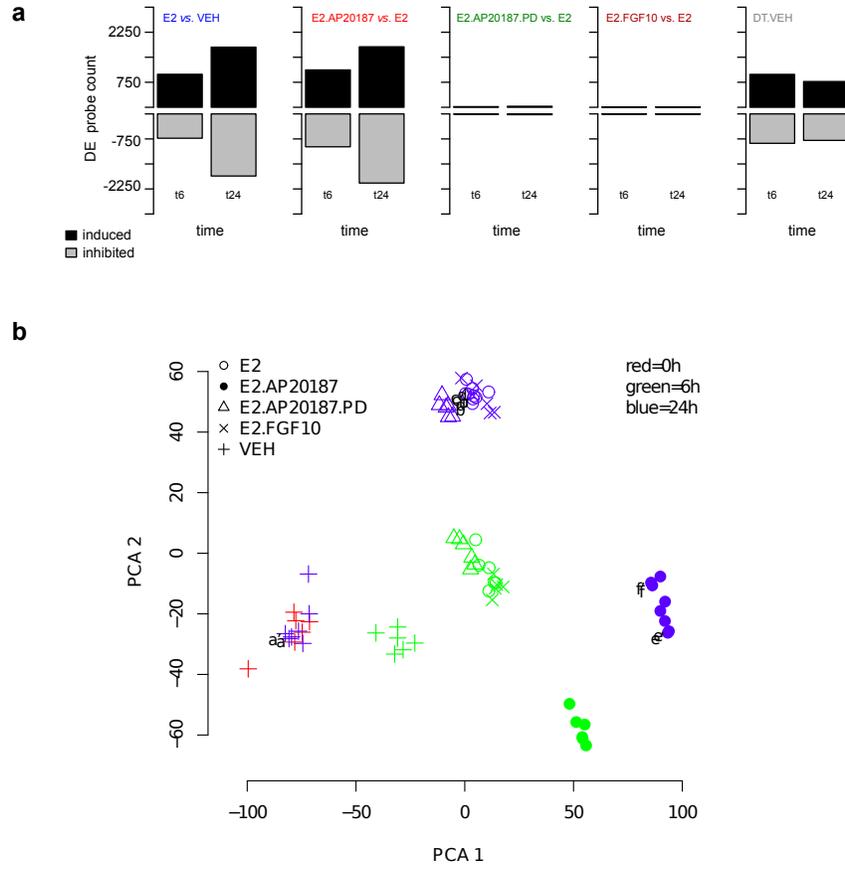
Figure 2: **iF2 construct perturbation experiments.** Summary of the differential expression analysis showing significant gene counts for P<0.01. (a) Bar charts depicting the number of probes significantly deregulated at each time point after stimulation. (b) PCA analysis of variation observed for the DE genes (n=7647) in the microarray analysis. The letters a-f and a'-f' indicate technical repeats included in the microarray experiments.
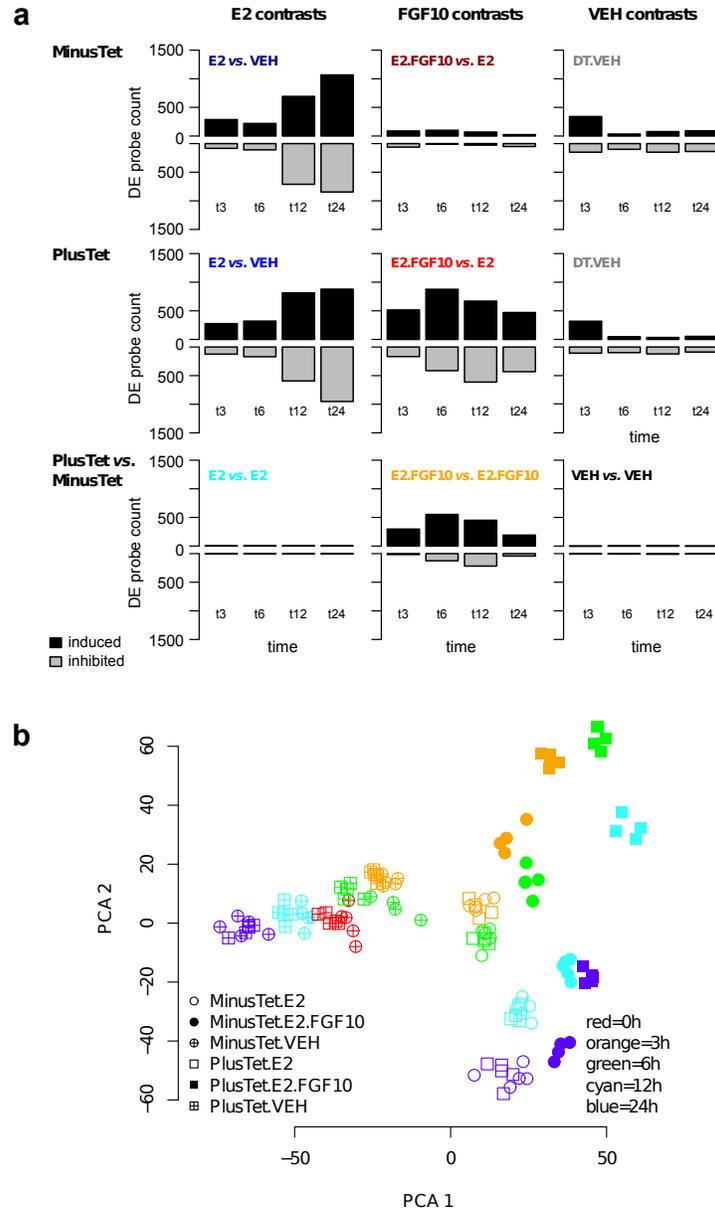
Figure 3: **FGFR2b perturbation experiments.** Summary of the differential expression analysis showing significant gene counts for P<0.01. (a) Bar charts depicting the number of probes significantly deregulated at each time point after stimulation. (b) PCA analysis of variation observed for the DE genes (n=2519) in the microarray analysis.
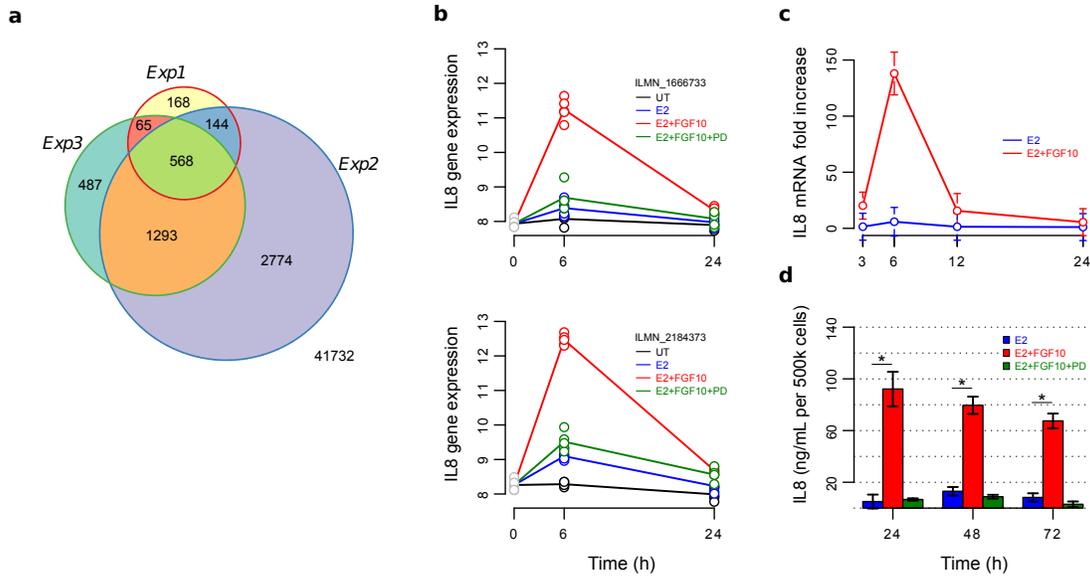
Figure 4: **Follow-up on differentially expressed genes.** (a) Venn diagram depicting the overlap between the genes deregulated after FGFR2 signalling in the experimental systems *Exp1-3*. Each list of FGFR2 regulated genes was derived as a contrast between the FGFR2 stimulus with estradiol versus estradiol only treatment to obtain the FGFR2 specific response. (b-d) Confirmation of gene expression microarray response by RT-PCR and protein expression (see Fletcher et al. [1] for additional details).

6 Session information

```
R version 4.3.0 RC (2023-04-18 r84287)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 22.04.2 LTS

Matrix products: default
BLAS:   /home/biocbuild/bbs-3.18-bioc/R/lib/libRblas.so
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods
[7] base

other attached packages:
[1] Fletcher2013a_1.37.0 limma_3.57.0

loaded via a namespace (and not attached):
 [1] compiler_4.3.0       gplots_3.1.3         formatR_1.14
 [4] futile.logger_1.4.3  tools_4.3.0          lambda.r_1.2.4
 [7] Biobase_2.61.0       futile.options_1.0.1 KernSmooth_2.23-20
[10] grid_4.3.0           BiocGenerics_0.47.0  caTools_1.18.2
[13] VennDiagram_1.7.3    bitops_1.0-7         gtools_3.9.4
```

References

[1] Michael NC Fletcher, Mauro AA Castro, Suet-Feung Chin, Oscar Rueda, Xin Wang, Carlos Caldas, Bruce AJ Ponder, Florian Markowetz, and Kerstin B Meyer. Master regulators of FGFR2 signalling and breast cancer risk. *Nature Communications*, 4:2464, 2013.

[2] Gordon K. Smyth. Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420. Springer, New York, 2005.

[3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.