

The bladderbatch data User's Guide

Jeffrey T. Leek

Modified: October 6, 2011 Compiled: November 15, 2011

Contents

1 Overview	1
------------	---

1 Overview

The `bladderbatch` package contains gene expression data on 57 samples from a bladder cancer study [1] which have been normalized with RMA and pre-processed according to a previously defined protocol [2]. The data are in an expression set object with `pData` including the variables “sample”, “outcome”, “batch”, and “cancer”. The first variable is the sample number, the second variable is the outcome as defined in the original study, the third variable is a batch variable defined based on the date the microarrays were processed and the cancer variable is a simplified outcome grouping all the cancers together. The data can be accessed as follows:

```
> library(bladderbatch)
> data(bladderdata)
> # Get the expression data
> edata = exprs(bladderEset)
> # Get the pheno data
> pdata = pData(bladderEset)
```

The data in this package are used as an example data set in the `sva` package.

References

- [1] Dyrskjot, L. and Kruhffer, M. and Thykjaer, T. and Marcussen, N. and Jensen, J. L. and Miller, K. and Orntoft, T. F., *Gene expression in the urinary bladder: a common carcinoma in situ gene expression signature exists disregarding histopathological classification*, Cancer Research 64:4040–4048.
- [2] Leek JT and Scharpf R and Corrada-Bravo H and Simcha D and Langmead B and Johnson WE and Geman D and Baggerly K and Irizarry IR. (2011) *Tackling the widespread and critical impact of batch effects in high-throughput data*, Nature Reviews Genetics 11:733–739.