mitoODEdata: Dynamical modelling of phenotypes in a genome-wide RNAi live-cell imaging assay

Gregoire Pau
pau.gregoire@gene.com

October 19, 2016

Contents

1   Introduction

The Mitocheck screen [2] is a time-lapse imaging assay that employed small-interfering RNAs (siRNAs) to test the implication of human genes in transient biological processes such as cell division or migration genome-wide. In this experiment, HeLa cells stably expressing core histone 2B tagged with green fluorescent protein (GFP) were seeded on siRNA-spotted slides, incubated for 18 h and imaged with automated fluorescence microscopy for 48 h. Video sequences of cell populations on each siRNA-spot were analysed by image segmentation, and at each frame, each individual cell was categorised into one of 16 morphological classes mostly related to cell division.

The `mitoODE` package implements a modelling by differential equations of cellular populations [1], to quantify the phenotypic effect induced by siRNA treatments in the Mitocheck screen. The package includes the code to fit any time course data to the model and the scripts used to generate the figures and results presented in the paper.

The `mitoODEdata` package, the experimental companion package of `mitoODE`, contains the screen data and methods to access the Mitocheck assay layout, siRNA annotation, time-lapse cell counts and the fitted phenotypes for each spot. Four cell types are reported: interphase (referred in the Mitocheck paper as: Interphase, Large, Elongated, Folded, Hole, SmallIrregular or UndefinedCondensed), mitotic (Metaphase, Anaphase, MetaphaseAlignment, Prometaphase or ADCCM), polynucleated (Shape1, Shape3, Grape) and apoptotic (Apoptosis).

2   Data access

Loading the package `mitoODEdata` loads the Mitocheck screen annotation variables `tab` and `anno` in the global environment. The object `tab` is a data frame containing spot metadata, including: `plate` (plate number), `replicate` (replicate number), `spot` (spot number within the plate), `qc` (original quality control from the paper), `type` (spot type) and `sirna` (spot siRNA ID). The object `anno` is a data frame containing the siRNA to gene mapping, including: `sirna` (siRNA ID), `ensembl` (target Ensembl gene ID), `hgnc` (target HGNC gene symbol), `entrez` (target Entrez gene ID), `genename` (target HGNC gene name).

```
> library("mitoODEdata")
> tab[1:5,]

  plate replicate spot id pr  ps   qc            type     sirna
1     1         2    2  1  1 102 1001 TRUE         marker MCO_0016401
2     1         2    2  2  2 102 1002 TRUE markerNeighbors MCO_0007359
3     1         2    2  3  3 102 1003 TRUE incenpNeighbors MCO_0007324
```

```
4     1         2   4  4 102 1004 TRUE         incenp MCO_0016401
5     1         2   5  5 102 1005 TRUE incenpNeighbors MCO_0006006

> anno[1:5,]

  ambion       sirna        ensembl hgnc1 entrez hgnc         genename
1  41652 MCO_0000001 ENSG00000004455   AK2    204  AK2 adenylate kinase 2
2  41815 MCO_0000002 ENSG00000004455   AK2    204  AK2 adenylate kinase 2
3 147152 MCO_0000003 ENSG00000004455   AK2    204  AK2 adenylate kinase 2
4 129427 MCO_0000004 ENSG00000140057   AK7 122481  AK7 adenylate kinase 7
5 129428 MCO_0000005 ENSG00000140057   AK7 122481  AK7 adenylate kinase 7
```

The functions getspot, getsirna and getanno allow simple conversions between HGNC gene symbols, Mitocheck siRNAs and spot IDs.

```
> getsirna(ann="CDH1")

[1] "MCO_0026105" "MCO_0026106" "MCO_0042725" "MCO_0046957" "MCO_0054655"
[6] "MCO_0054656"

> getspot(ann="FGFR2")

[1] 58916 59300 59684 64740 65124 65508 64551 64935 65319

> getanno(spot=1234, field=c("hgnc", "entrez", "genename"))

        hgnc entrez                                          genename
6991 SLC19A2  10560 solute carrier family 19 (thiamine transporter), member 2
```

The function readspot reads the time-course cell counts of a given spot ID. The output value is a matrix containing the number of cells of a given type (interphase i, mitotic m, polynucleated s and apoptotic a) per frame. The first image (e.g. row) was acquired 18 h after siRNA transfection and the following images were acquired every 30 minutes during 48 h. The function plotspot plots the cell count time series of a given spot ID.

```
> spotid <- getspot(ann="FGFR2")[1]
> y <- readspot(spotid)
> y[1:10,]

       i m s a
 [1,] 52 0 0 0
 [2,] 50 1 1 0
 [3,] 47 4 2 0
 [4,] 47 3 3 0
 [5,] 51 5 1 0
 [6,] 53 2 4 0
 [7,] 53 1 3 0
 [8,] 54 2 2 0
 [9,] 53 2 3 0
[10,] 50 3 6 0

> plotspot(spotid)
```
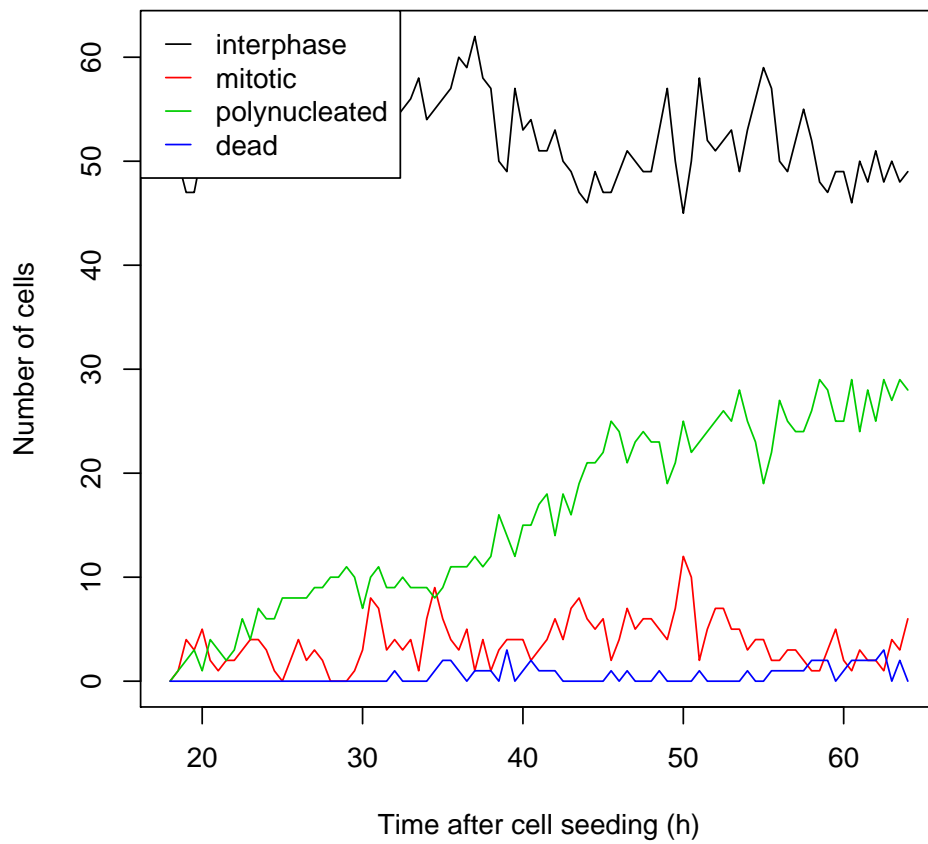
3 References

References

[1] Pau G, Walter T, Neumann B, Heriche JK, Ellenberg J, and Huber W (2013) Dynamical modelling of phenotypes in a genome-wide RNAi live-cell imaging assay. (submitted)

[2] Neumann B, Walter T, Heriche JK, Bulkescher J, Erfle H, et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464: 721–727.