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

Gregoire Pau
pau.gregoire@gene.com

May 3, 2016

Contents

1   Introduction

The Mitocheck assay is a genome-wide time-lapse imaging screen that employed small-interfering RNAs (siRNAs) to test the implication of human genes in transient biological processes such as cell division or migration [2]. siRNAs are double-stranded RNA molecules, implicated in the RNA interference pathway, that are used to disrupt the expression of specific genes. The screen consisted of 206592 time-lapse experiments, using 51766 different siRNA constructs and targeting 17293 human genes. Experiments were organised in slides, each containing 384 siRNA spots. Each slide contained 7 to 8 negative and 8 to 11 positive control spots. The control siRNAs were: siScrambled, a non-targeting negative control; siKIF11, targeting the gene KIF11, which encodes a kinesin needed for centrosome segregation; siCOPB1, targeting an essential protein binding to the Golgi vesicle and siINCENP, targeting a centromere-associated protein coding gene required for proper chromosome segregation and cytokinesis. To allow the chromosomes to be imaged by fluorescence microscopy, a cervix carcinoma HeLa cell line stably expressing GFP-tagged H2B histone was used. Cells were filmed starting 18 h after seeding for a duration of 48 h, with an imaging rate of one per 30 min. After acquisition, cell nuclei were segmented, quantified and classified into one of 16 morphological classes by a fully automated algorithm.

The `mitoODE` package implements a population-level dynamic model to represent the temporal evolution of dividing cells [1]. By fitting cell counts in four transient cellular states, our model yielded parameters that quantify the dynamic effects of siRNA treatments on cell population levels. Model parameters allows reliable estimation of the penetrance and time of four disruption events of the cell cycle: quiescence, mitosis arrest, polynucleation and cell death. 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 considered: 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 Methods

2.1 Ordinary differential equation model

We consider four cellular states: interphase (I), mitotic (M), polynucleated (P) and dead (D). We modelled the number of cells $n_p(t)$ of state p, at time t with the system $\mathcal{S}$ of differential equations, depicted in Fig. 1:

$$
\begin{aligned}
\dot{n}_I(t) &= -\big(k_{IM}(t) + k_D(t)\big)n_I(t) + 2k_{MI}(t)n_M(t) \\
\dot{n}_M(t) &= k_{IM}(t)n_I(t) - \big(k_{MI}(t) + k_D(t) + k_{MP}(t)\big)n_M(t) \\
\dot{n}_P(t) &= k_{MP}(t)n_M(t) - k_D(t)n_P(t) \\
\dot{n}_D(t) &= k_D(t)n_I(t) + k_D(t)n_M(t) + k_D(t)n_P(t),
\end{aligned}
$$

with:

$$
\begin{aligned}
k_{IM}(t) &= \alpha_{IM}^0 - \alpha_{IM}/\big(1 + \exp(\tau_{IM} - t)\big) \\
k_{MI}(t) &= \alpha_{MI}^0 - \alpha_{MI}/\big(1 + \exp(\tau_{MI} - t)\big) \\
k_{MP}(t) &= \alpha_{MP}/\big(1 + \exp(\tau_{MP} - t)\big) \\
k_D(t) &= \alpha_D/\big(1 + \exp(\tau_D - t)\big)
\end{aligned}
$$

where $\dot{n}_p(t)$ is the time derivative of $n_p(t)$, with the initial conditions $n_I(0) = (1 - \omega_0)n_0$, $n_M(0) = \omega_0 n_0$, $n_P(0) = n_D(0) = 0$, and n_0 is the number of cells at seeding time $t = 0$. In agreement with observations of untreated cells, the mitotic index at seeding time is set to $\omega_0 = 0.05$, the basal interphase-to-mitosis transition rate to $\alpha_{IM}^0 = 0.025$ h^{-1} and the basal mitosis-to-interphase transition rate $\alpha_{MI}^0 = 0.57$ h^{-1}. To account for contamination of spots by normal cells, due to untransfected cells moving into the spot region, we assumed that cell counts are a mixture of two independent, growing cell populations: a treated population, modelled by $\mathcal{S}$ with parameters $\{(1-\mu)n_0,\ \alpha_{IM},\ \alpha_{MI},\ \alpha_{MP},\ \alpha_D,\ \tau_{IM},\ \tau_{MI},\ \tau_{MP},\ \tau_D\}$ and an untreated population, modelled by $\mathcal{S}$ with parameters $\{\mu n_0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0\}$. In total, 10 parameters are required to model a cell count time course: the initial cell number at seeding time n_0, the normal contamination fraction parameter μ and 8 transition parameters $\{\alpha_{IM},\ \alpha_{MI},\ \alpha_{MP},\ \alpha_D,\ \tau_{IM},\ \tau_{MI},\ \tau_{MP},\ \tau_D\}$.

2.2 Estimation of model parameters

Model parameters $\theta = \{n_0,\ \mu,\ \alpha_{IM},\ \alpha_{MI},\ \alpha_{MP},\ \alpha_D,\ \tau_{IM},\ \tau_{MI},\ \tau_{MP},\ \tau_D\}$ are estimated by penalised least squares regression of time-course cell count data, minimising the cost function:

$$
J(\theta) = \frac{1}{\#\mathcal{T}} \sum_{t \in \mathcal{T}} \sum_{p \in \mathcal{P}} \big(y_p(t) - n_p(t, \theta)\big)^2 + \lambda \sum_{k \in \mathcal{K}} \alpha_k^2, \tag{1}
$$

where $\mathcal{T}$ is the set of observed time points, $y_p(t)$ the observed number of cells of state p at time t, λ a constant to weigh the penalty term and the set of penalised parameters $\mathcal{K} = \{\mu,\ \alpha_{IM},\ \alpha_{MI},\ \alpha_{MP},\ \alpha_D\}$. Given the parameters, the ODE system is integrated using the Runge-Kutta fourth-order method. Minimisation of the penalised criterion J was achieved with the Levenberg-Marquardt algorithm, with the `nls.lm` pacakge, applying a positivity constraint to the components of θ.

3 Modelling time course cell count data

The `mitoODE` package depends on the `mitoODEdata` experiment data package, containing 206592 time-course cell counts from the Mitocheck time-lapse genome wide siRNA screen. The following example plots the time-course cell counts of a given spot and loads them in the matrix `y`. The matrix contains the number of cells of a given type (interphase `i`, mitotic `m`, polynucleated `s` and apoptotic `a`) per image. The first image was acquired 18 h after siRNA transfection and the following images were acquired every 30 minutes during 48 h.

```
> library("mitoODE")
> spotid <- 1000
> plotspot(spotid)
> y <- readspot(spotid)
> y[1:5,]
```
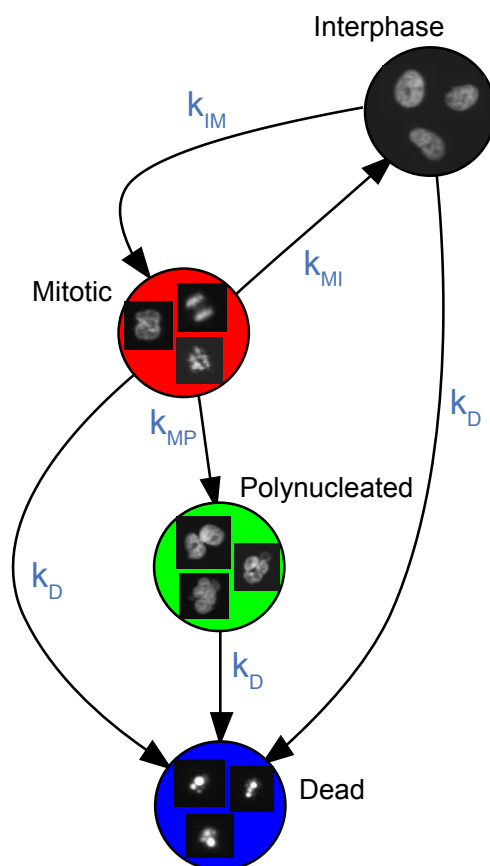
Figure 1: A differential equation model to quantify temporal phenotypes. Cell populations could enter and leave states with four different transition rates: k_{IM}, k_{MI}, k_{MP} and k_D.

```
      i m s a
[1,] 65 3 1 0
[2,] 63 4 2 1
[3,] 65 5 1 1
[4,] 70 3 0 0
[5,] 71 1 0 2
```

To fit the time-course cell counts, we first need to set the constant parameters of the model. This is a named vector with the elements: `g.kim` (in our model, $\alpha_{IM}^0 = 0.025$ h^{-1}), `g.kmi` ($\alpha_{MI}^0 = 0.57$ h^{-1}), `g.mit0` (the mitotic index at seeding time, $\omega_0 = 0.05$), and `p.lambda` (the regularization parameter, $\lambda = 4$).

```
> pconst <- c(g.kim=0.025, g.kmi=0.57, g.mit0=0.05, p.lambda=4)
```

The function `getp0` generates an initial condition vector of 10 parameters: him (in our model, α_{IM}), hmi (α_{MI}), hmp (α_{MP}), ha (α_D), tim (τ_{IM}), tmi (τ_{MI}), tmp (τ_{MP}), ta (τ_D), mu (μ) and i0 (n_0).

```
> p0 <- getp0()
> p0

him hmi hmp  ha tim tmi tmp  ta  mu  i0
  0   0   0   0  42  42  42  42   0  40
```

The function `fitmodel` fits the time-course cell counts to the model, using the initial condition parameters p0 and the constant parameters pconst. Results are the fitted parameters and some fitting statistics, including: `score` (the cost function J), `rss` (the residual sum of squares) and `pen` (the penalty term).

3

```
> pp1 <- fitmodel(y, p0, pconst)
> round(pp1, 2)

  him   hmi   hmp    ha   tim   tmi   tmp    ta    mu    i0 score   rss   pen
-0.03  0.00  0.02  0.01 33.34 61.29 18.44 54.74  0.00 42.96 16.48 16.27  0.22
 imax  mmax  smax  amax
94.42  3.28  1.23  5.43
```

The function `plotfit` displays the fitted data.
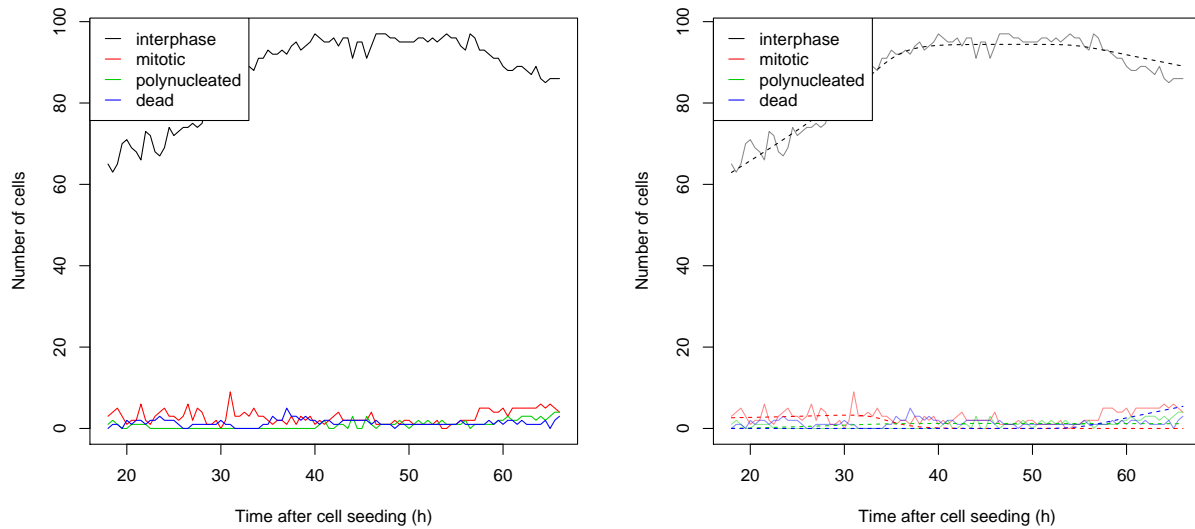
```
> plotfit(spotid, pp1)
```


Figure 2: Time course cell counts (left) and fitted data (dotted lines, right) of the Mitocheck spot ID 1000.

The model is sensitive to initial conditions. To decrease the risk of finding local minima, the data be fitted several times using the argument `nfits`, adding to the initial conditions some Gaussian noise of standard devision `sd`. The following example uses 4 fits to find a better fit (i.e. with a lower `score`) than `pp1`.

```
> set.seed(1)
> pp2 <- fitmodel(y, p0, pconst, nfits=4, sd=0.1)
> round(pp2, 2)

  him   hmi   hmp    ha   tim   tmi   tmp    ta    mu    i0 score   rss   pen
-0.03 -0.57  0.03  0.00 35.27 37.48 33.09 52.68  0.00 42.36 12.96 10.45  2.52
 imax  mmax  smax  amax
96.16  3.92  2.58  3.31
```

4   Reproducing figures

The following functions reproduce the plots shown in the paper [1]. See the paper for details.

```
> loadFittedData()
> figure1()
> figure2()
> figure3a()
> figure3b()
> figure4()
```
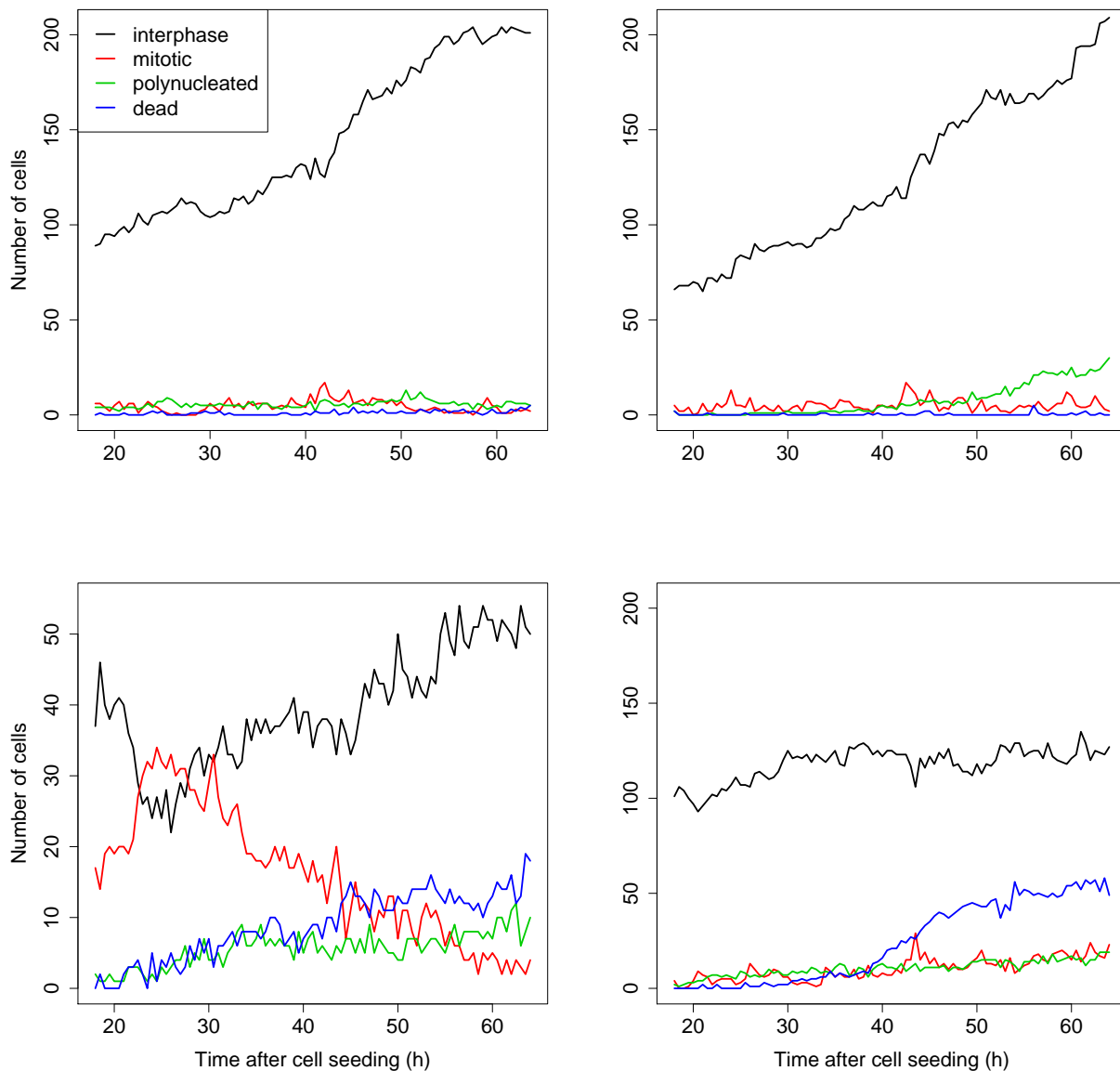
Figure 3: Time course of cell counts in four different siRNA spots. The negative control siRNA siScrambled led to normal cell growth in two replicated spots **(top)**. siKIF11 led to an early accumulation of mitotic cells, while at later times many of the arrested cells went into cell death **(bottom-left)**. siCOPB1, which targets an essential Golgi-binding protein, caused cell death, but no mitotic phenotypes **(bottom-right)**. Y-axis scales were changed to accommodate the different dynamics of the phenotypes.

5   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.
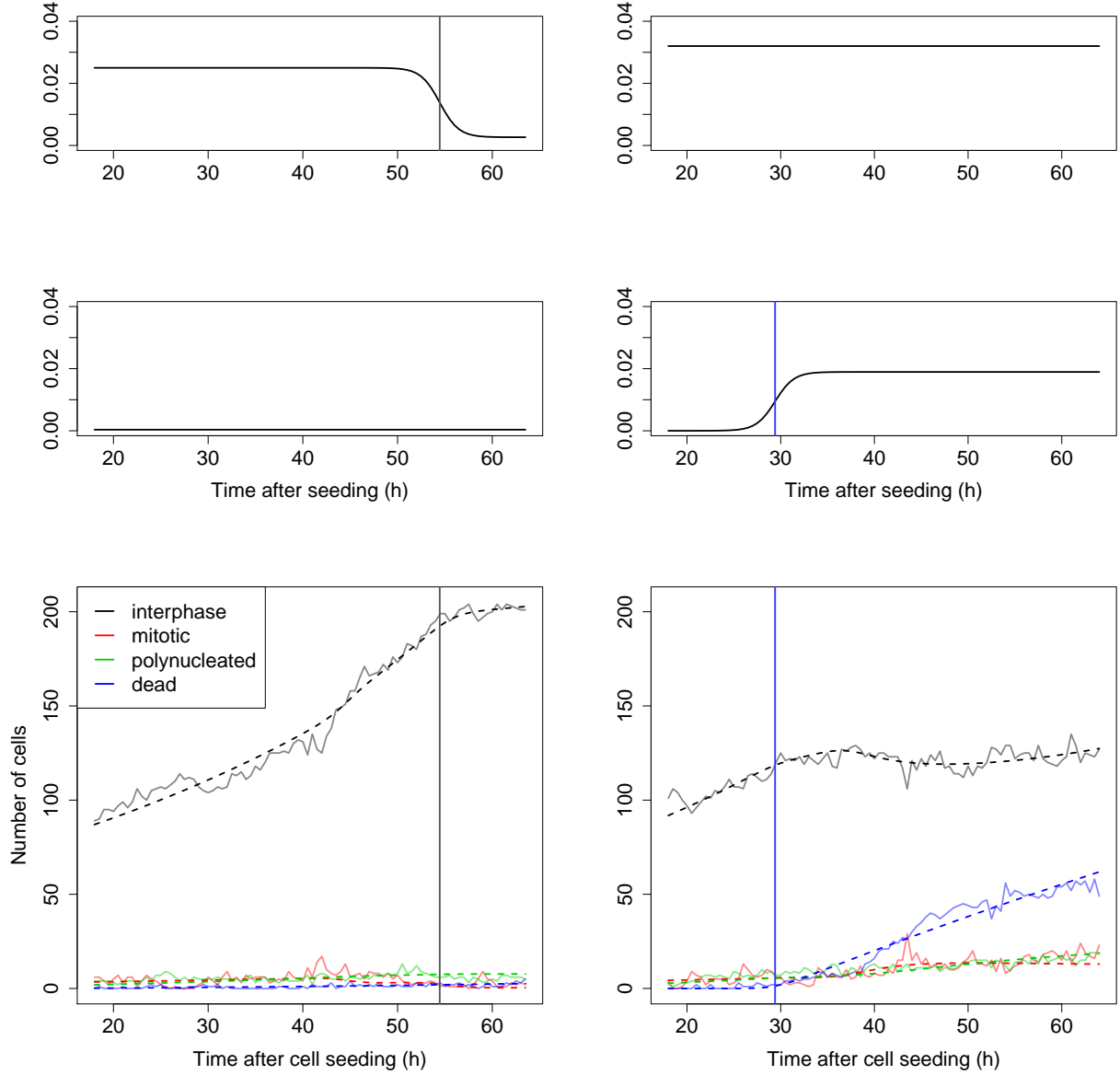
Figure 4: Transition rates vary over time and are modelled by parametric sigmoid functions. In the negative control spot **(left)**, the interphase-to-mitosis transition rate k_{IM} inflects at $\tau_{\mathrm{IM}} = 54.5$ h, as the capacity of the spot to support a growing number of cells becomes limiting. The death transition rate k_{D} remains constant and null. In the siCOPB1 spot **(right)**, the death transition rate k_{D} inflects at $\tau_{\mathrm{D}} = 29.4$ h, indicating the time of cell death.
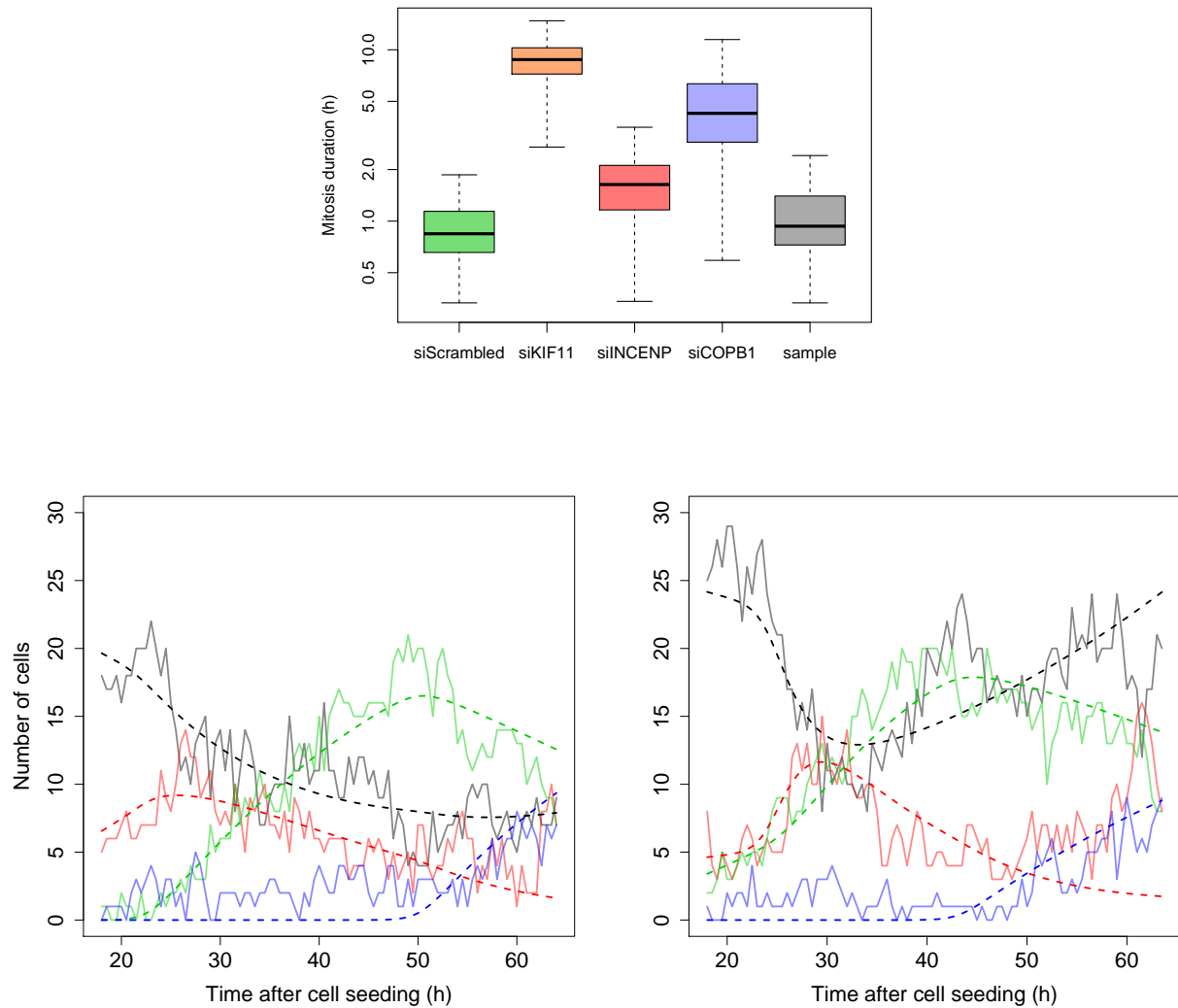
Figure 5: Box plots of cell death penetrance α_D estimated in control and sample wells **(top)**. Lethal phenotypes induced by siKIF11 and siCOPB1 had significantly higher cell death penetrance than the negative siScrambled control spots. Cell count time courses of two spots containing an siRNA (MCO_0026491) targeting the uncharacterised gene C3orf26 **(bottom)**. Induction of cell death started at $\tau_D = 30.8$ h and 33.8 h.
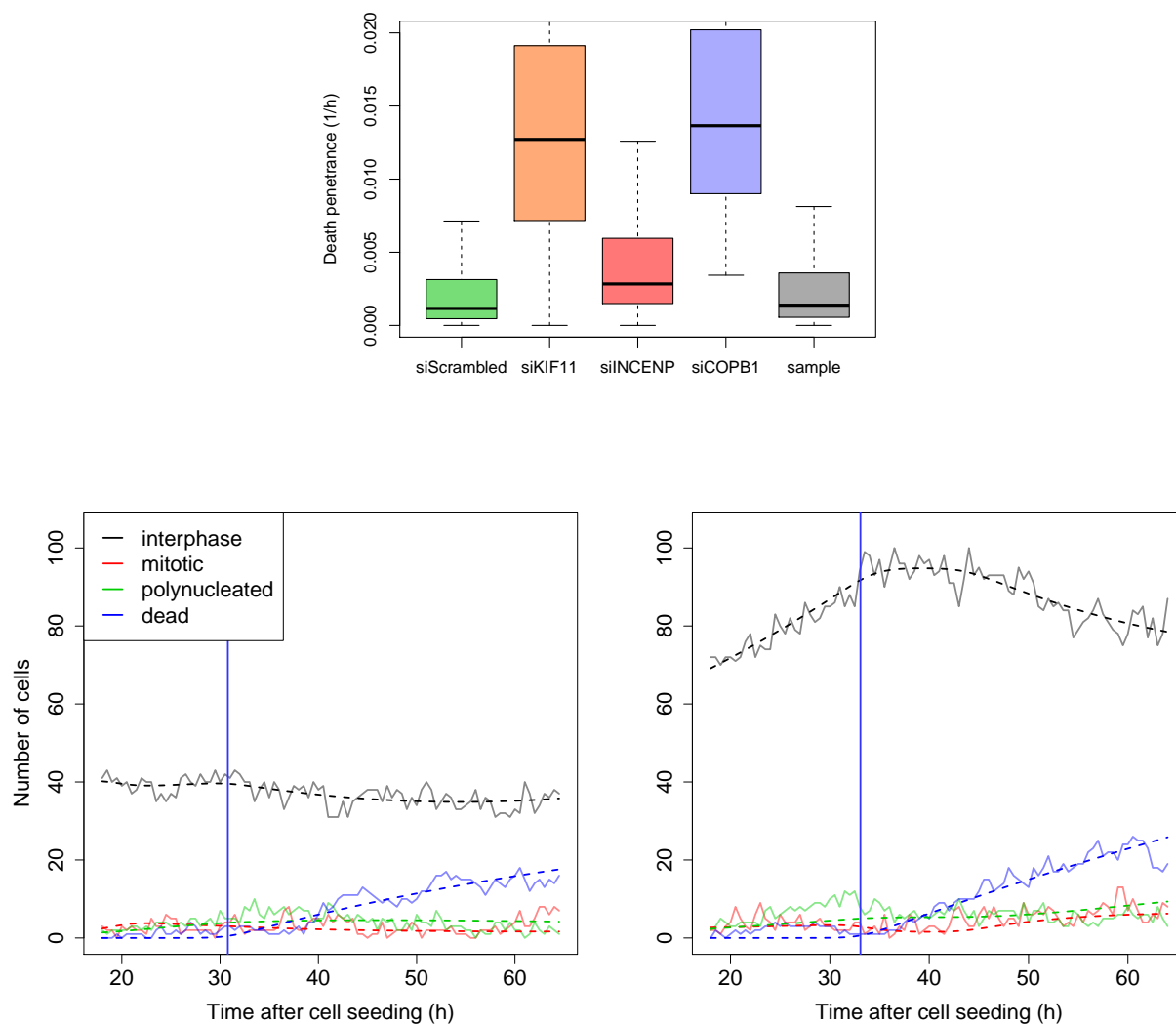
Figure 6: Box plot of mitosis duration estimated in control and sample wells **(top)**. Prometaphase arrest caused by siKIF11 led to a significantly higher mitosis duration than cells in negative control spots. Cell count time courses of two spots containing an siRNA (MCO_0020444) targeting the poorly characterised gene CCDC9 **(bottom)**. The overall high proportion of mitotic cells, compared to the negative controls illustrated in Fig. 1b, is indicative of a longer mitosis duration, estimated at 5.7 h.
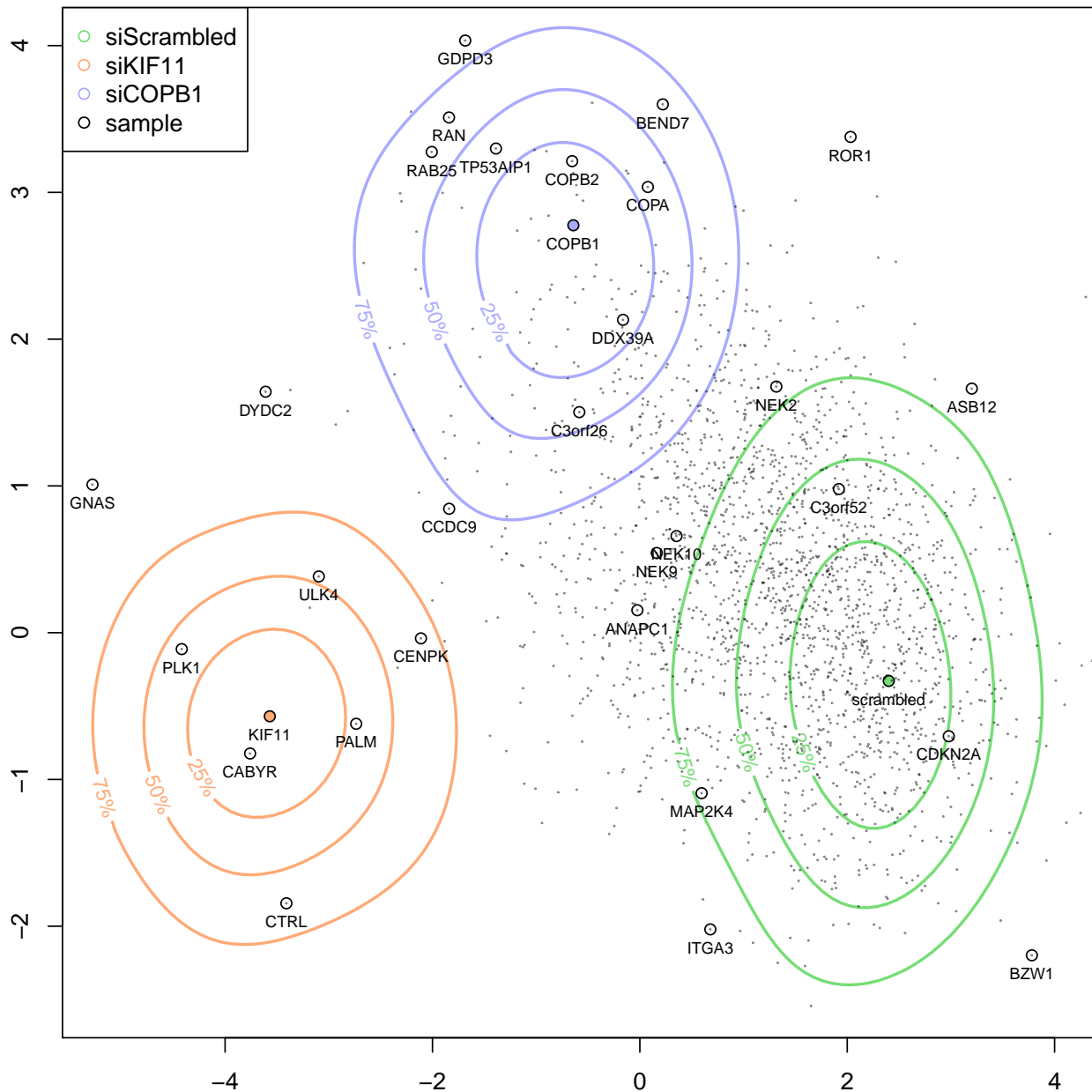
Figure 7: Phenotypic map of the Mitocheck screen. Phenotypic profiles of all siRNA treatments were derived from fitted parameters and projected on two dimensions using linear discriminant analysis between the siScrambled, siCOPB1 and siKIF11 control spots. Color contour lines denote the quantiles of control experiments that fall into the corresponding phenotypic region. The green phenotypic region, centered on the non-targeting negative control siScrambled, is representative of a normal growth phenotype. The orange phenotypic region, centered on siKIF11, is representative of prometaphase arrest followed by apoptosis, while the blue region, centered on siCOPB1, is representative of cell death without mitotic arrest. Shown are the 2190 siRNAs that induced a disruption of the cell cycle, or increased the duration of the interphase or mitosis phases. A selection of siRNA gene targets are highlighted on the map.