Case study, Fev 10

# Sameer’s data of Western blot.

The measurements are replicated at different dates and with different transfer time. Each experiment is then scanned at different intensities to improve dynamical range. None of these dimensions are consistent across measured kinases.

The raw data are then average across the intensities (with proper scaling). The replicates are treated the same even if the transfer time or the date differs. The intermediate datacube is therefore 3D with the last dimension comprising the replicates (irrelevant of the date and transfer time).

The final datacube is a 2D matrix of cell lines and kinases.

In this case, I included only two kinases (Akt pS473 and beta actin); there is no control.

# Jérémie’s data: single cell time course.

The measurements are time course of single cells after stimulation with TRAIL and an additional perturbation (usually the doses of TRAIL are the same between ‘TRAIL only’ and ‘TRAIL+ perturbation’). Cells are treated at different doses (with 0 being the control) and a fluorescent reporter is measured. In addition, another reporter (constant over time) is measured: it reflects the level of expression of a protein. The cells are tracked (with a tracking number) and then stored in a matrix for each condition (combination of TRIAL/perturbation). The number of tracked cell lines differs from one condition to another, but the time interval is generally the same (or we can enforce it by adding ‘NaN’ to complete the matrix).

The cell lines in the control case are then proceeded to obtain a median background (a unique trajectory) that is then subtracted on all trajectories across the different conditions. The usual analyses on these trajectories are to extract a few properties (like slope, time of death) or to integrate the trajectories. Therefore, the final datacube are 3D matrices with the last dimension having different number of elements (TRAIL doses; perturbation or no; single cells – variable)

# Benes’ data: GI50 curves.

Cell lines are seeded and then treated with 7 different concentrations of drug (4 orders of magnitude). The number of cell lines is measured at seeding and after either 3, 5 or 7 days of treatment. The actual measured signal need to be converted to a cell count using a standard curve (specific for each cell line). Multiple replicates are made for the different doses/drugs/treatment duration.

The datacube is complete and regular with 5 dimensions: cell line, drug, treatment duration, doses, and replicate. The measures are used to plot the survival curves and a few metrics are derived (GI50, Amax …). We finally obtain a few datacube with 3 dimensions (cell line, drug, treatment duration) for each metric.

I still have to sort out a few details to convert the data from signal to cell lines and properly match the different replicates and treatment duration. For now, I send you the .xls spreadsheet, but I will generate a .txt file as discussed as soon as I sort out these issues.