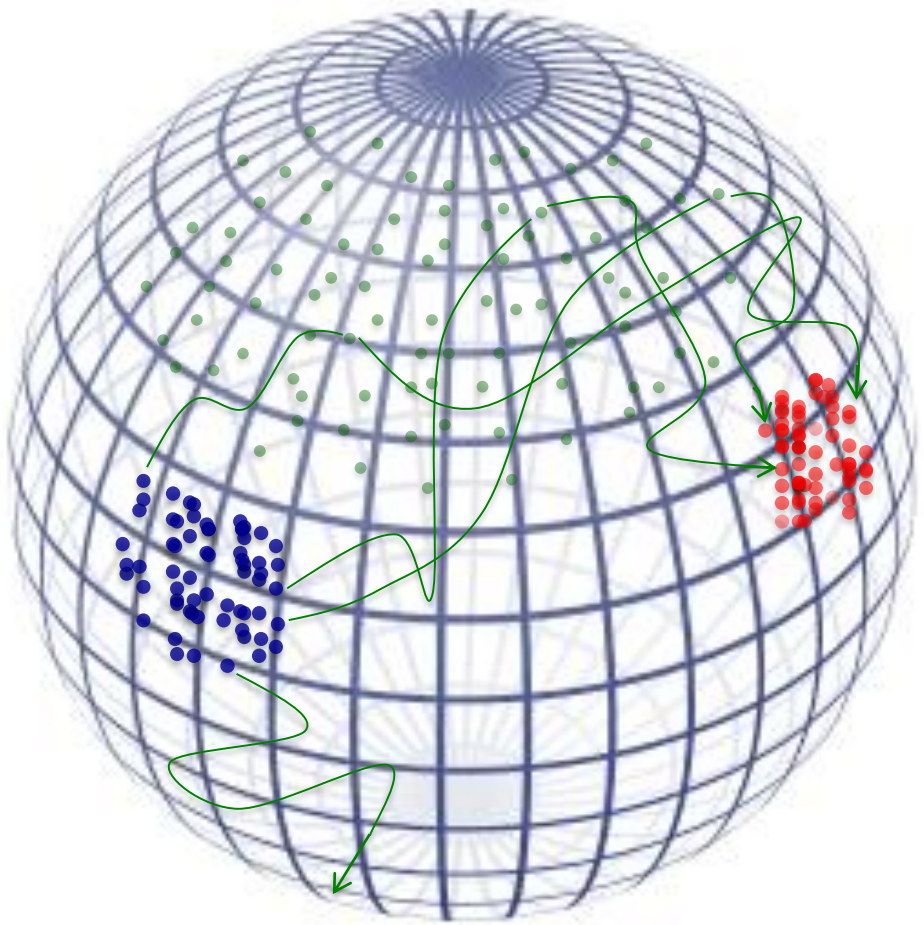
Dimensionality reduction; why?

Olivier Gandrillon

June, 2022

Our view: cells as dots dynamically moving in a multidimensional space



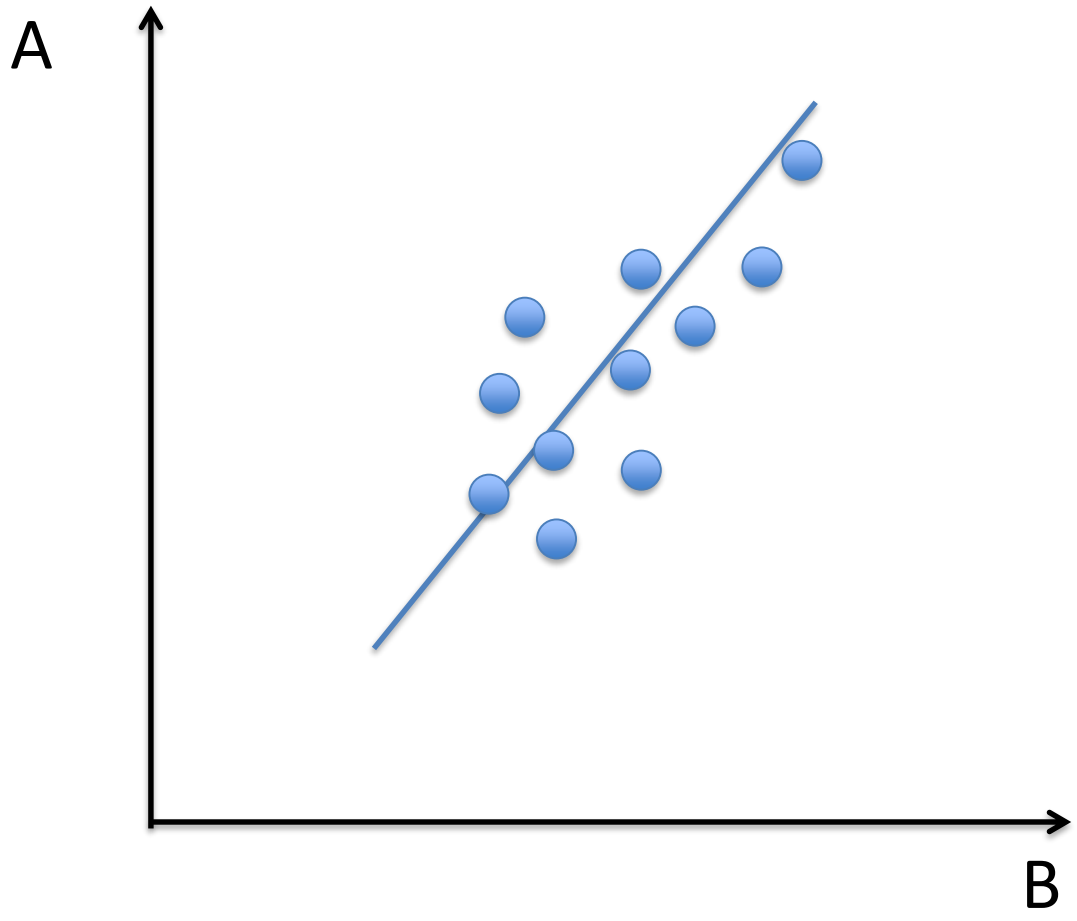
Richard *et al.* (2016). *PLoS biology* Huang *et al.* (2010). *PLoS biology*

What is the size of the relevant « gene expression space »

In typical RNAseq experiment, initial dimension is in the tenth of thousand (the number of genes one can detect)

But it could be much larger (number of post-translationally modified proteins)

The typical case: linear correlation

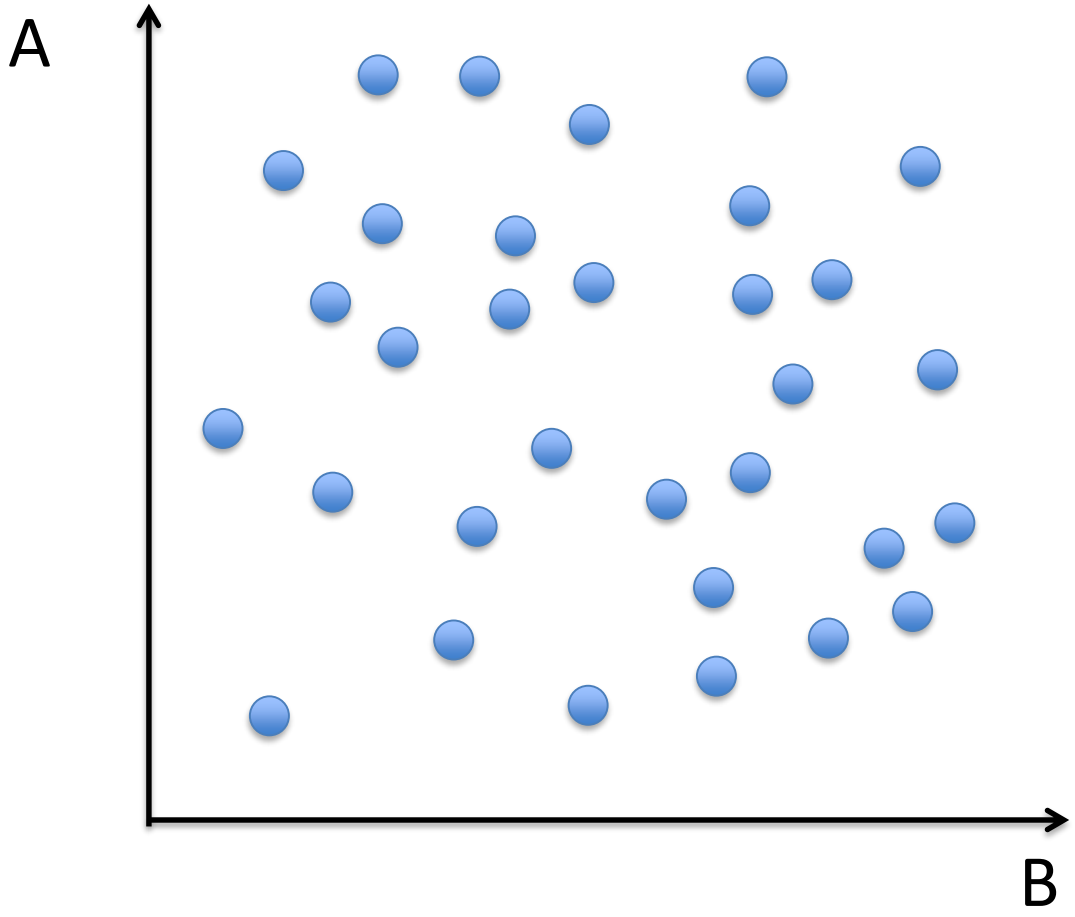


In such a case, the “relevant” manifold has a dimension of 1.

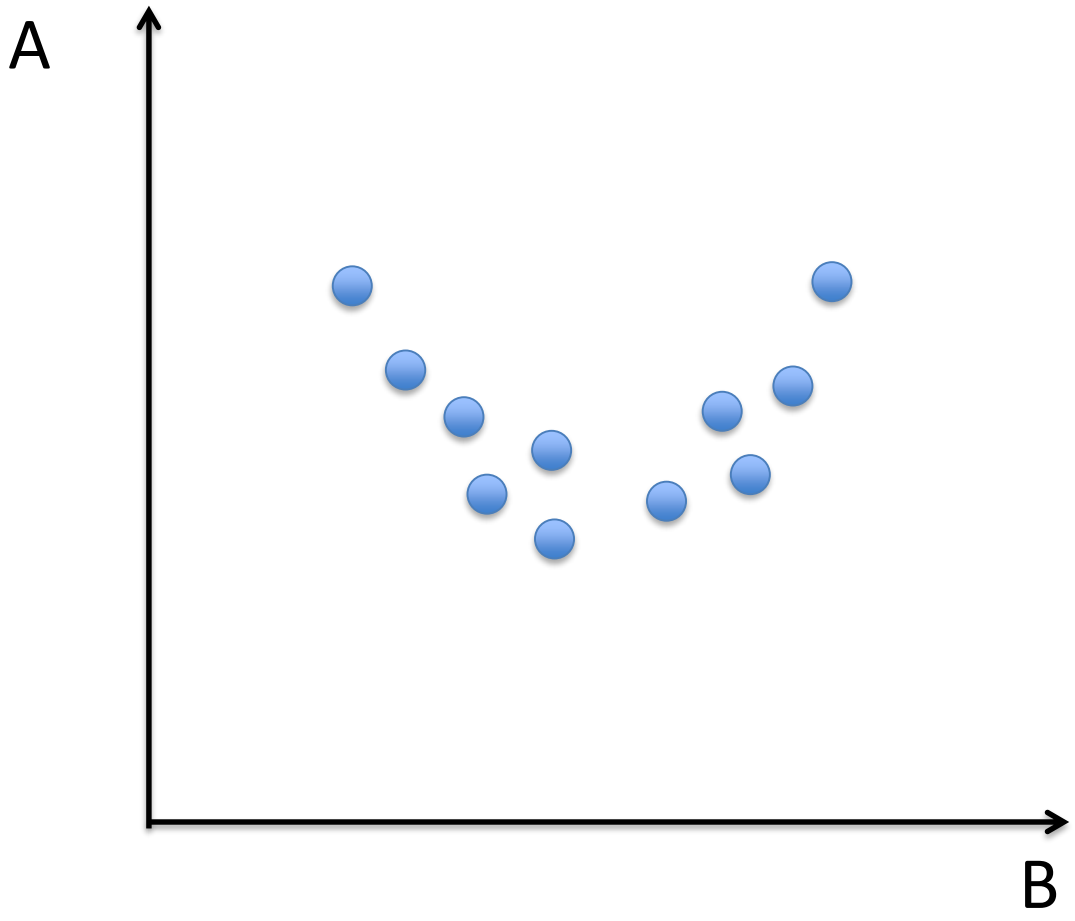
(In French: manifold: variété)

A topological space that locally resembles Euclidean space near each point.

The never happennig case :



The untypical case: non-linear correlation



From:

PROBLEMS AND PARADIGMS

Prospects & Overviews

BioEssays

www.bioessays-journal.com

Gravity Constraints Drive Biological Systems Toward Specific Organization Patterns

Commitment of cell specification is constrained by physical cues

Mariano Bizzarri, Maria Grazia Masiello, Alessandro Giuliani, and Alessandra Cucina*

« we can argue that the “phenotypic landscape” the system can freely explore does *not include infinite possibilities* (configurations). As anticipated by Waddington, only a discrete number of stable attractors can be reached. »

IOP Publishing
Reports on Progress in Physics
Rep. Prog. Phys. 78 (2015) 036602 (51pp)
doi:10.1088/0034-4885/78/3/036602

Review Article

The unforeseen challenge: from genotype-to-phenotype in cell populations

Erez Braun

Department of Physics and Network Biology Research Laboratories, Technion, Haifa 32000, Israel

E-mail: Erez@physics.technion.ac.il

Received 27 June 2014, revised 4 December 2014
Accepted for publication 18 December 2014
Published 26 February 2015

"From the physics perspective, the enormous reduction from the huge combinatorial space of microscopic degrees of freedom of intracellular molecular interactions to the minute number of stable cell states, is one of the most remarkable and fascinating properties of the living cell. This phenomenon should inspire a theoretical framework describing this class of dynamical systems."

How comes there is an underlying manifold?

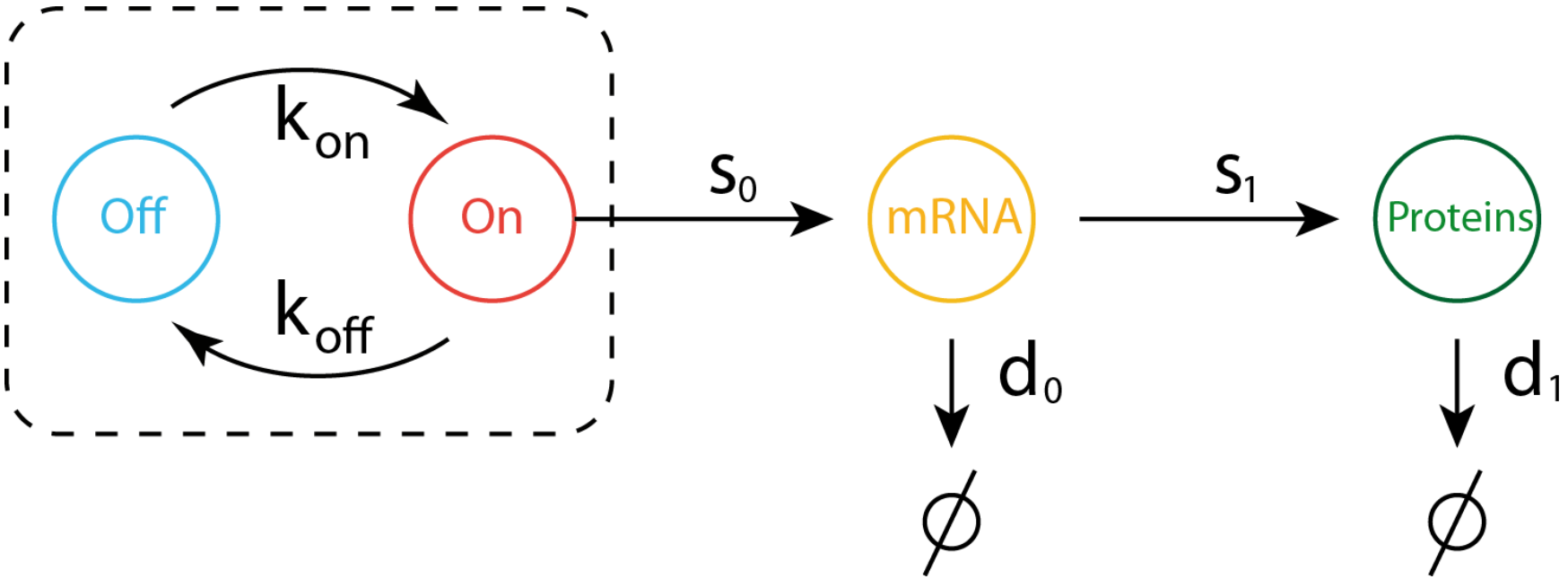
Evolution-selected constraints
(in all flavours and shapes):

1. Internal constraints (structure of the network, metabolism, epigenetics, …)

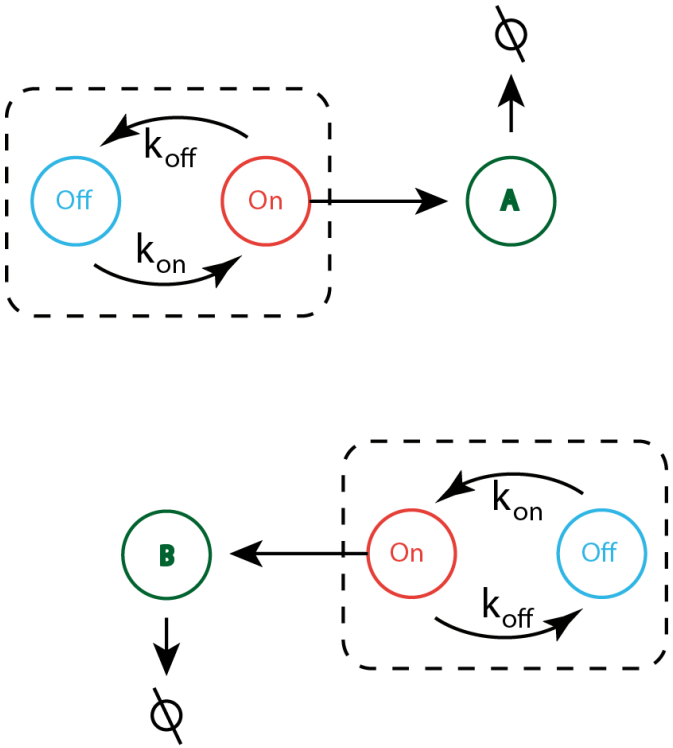
2. External constraints (physical constraints like gravity or mechanical constraints,.)

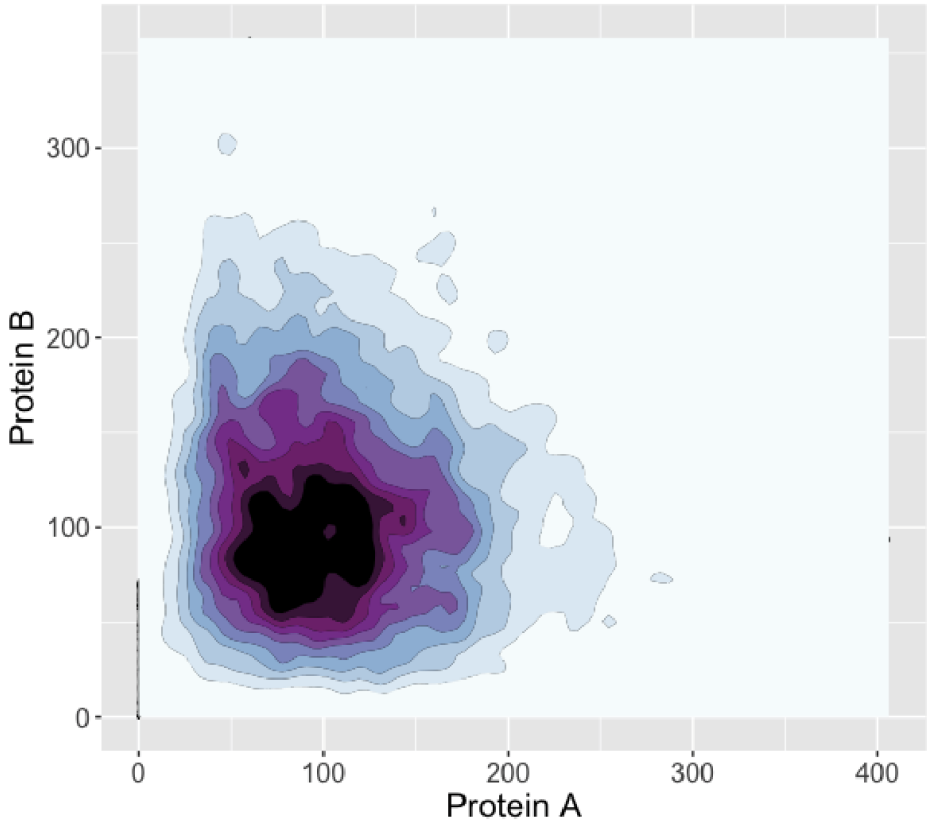
Structure of the network

Let’s assume a 2 state model for gene expression at the single cell level

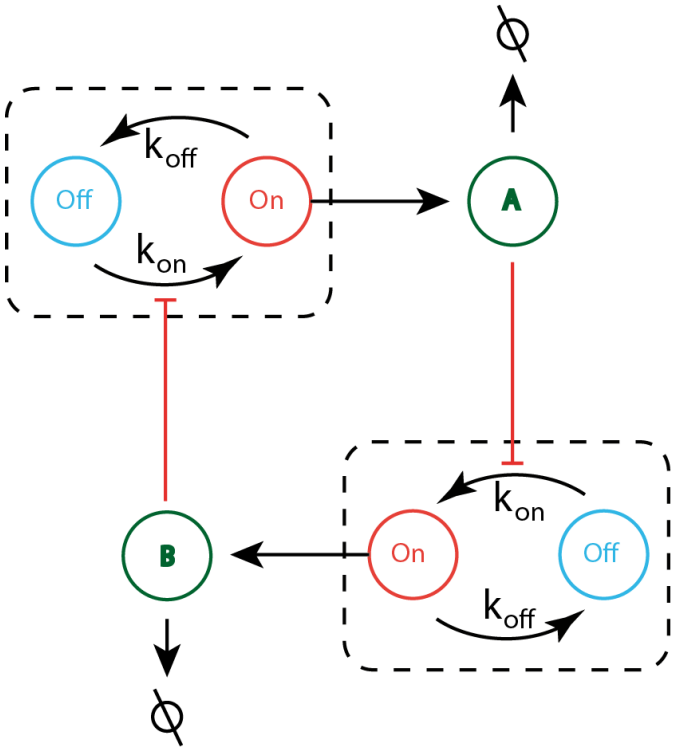


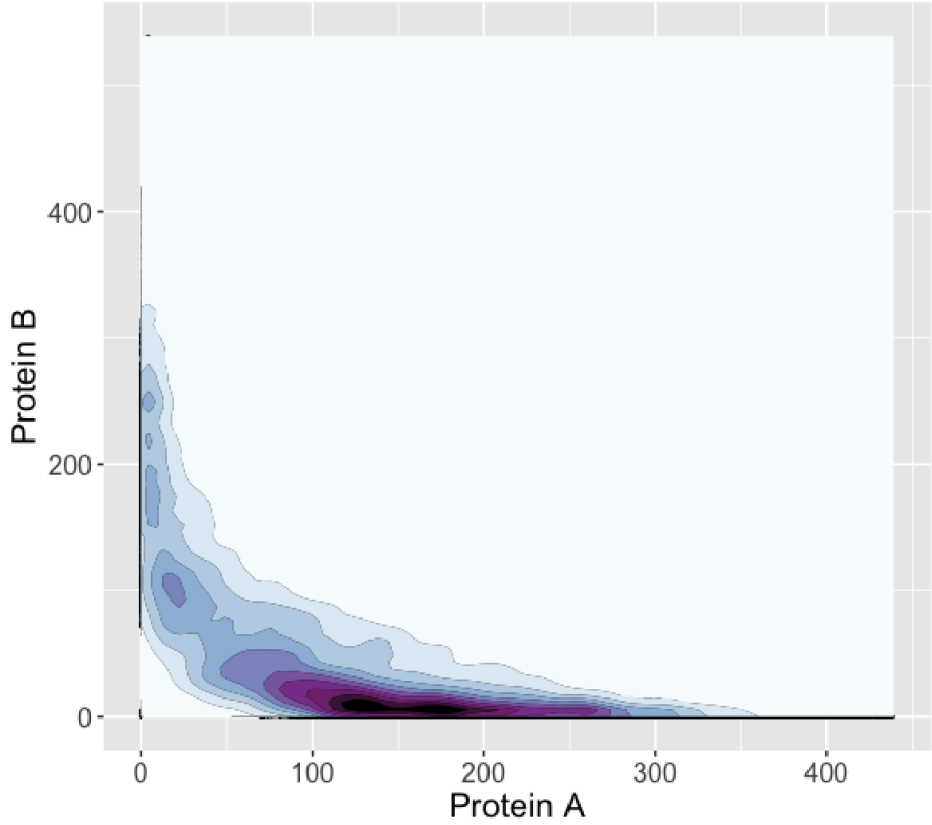
The case where two gene live independently



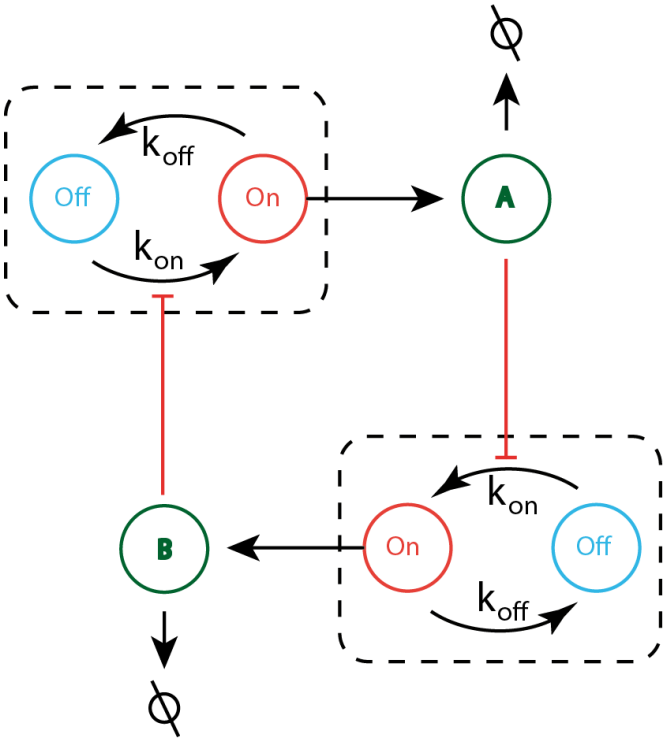


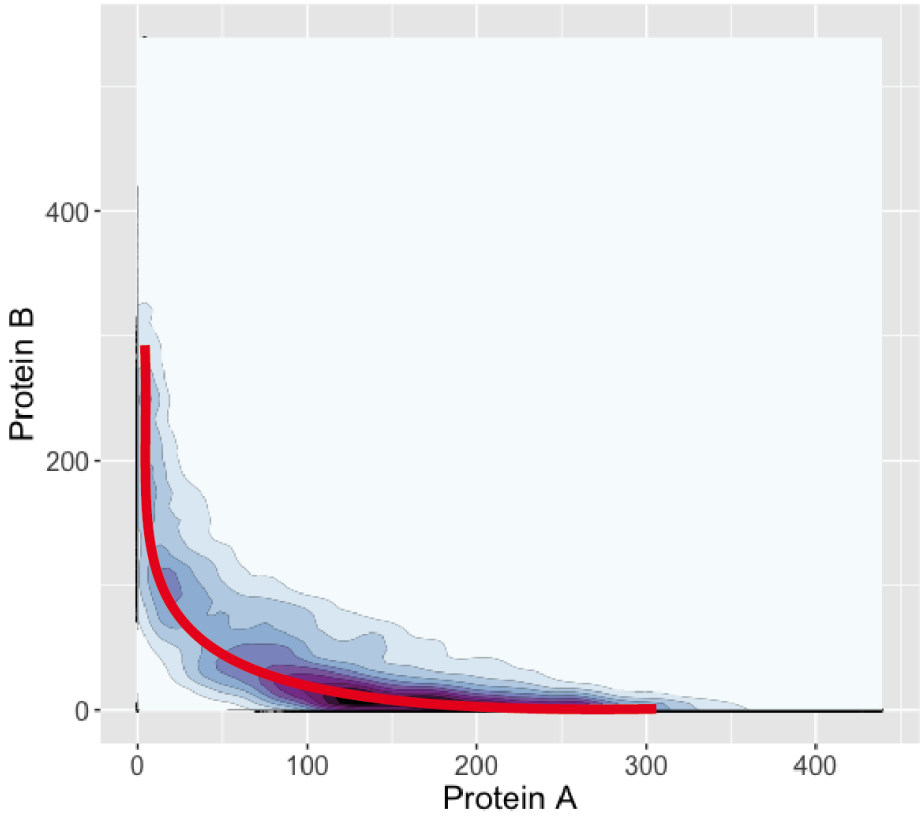
The case where two gene *do not* live independently (toggle switch)





One would like to catch the relevant information using a 1D curve.





The first thing everyone in scRNAseq scene does is reducing the dataset dimensionality by

1. Applying a filter that selects “the most variant genes”

and

2. Applying a “dimensionality reduction technique”

“In the single-cell sequencing field, low-dimensional projections are **more** than a visualization method: they are ubiquitous tools for **inference and discovery**. Transcriptomics workflows convert large data arrays to human-parsable visuals; these visuals are then used to explore gene expression and validate cell type relationships, under the assumption that they **represent** the underlying data well enough to draw conclusions. However, the embedding procedures involve several **distortive** steps, which should be recognized and questioned.”

Gorin et al. bioRxiv