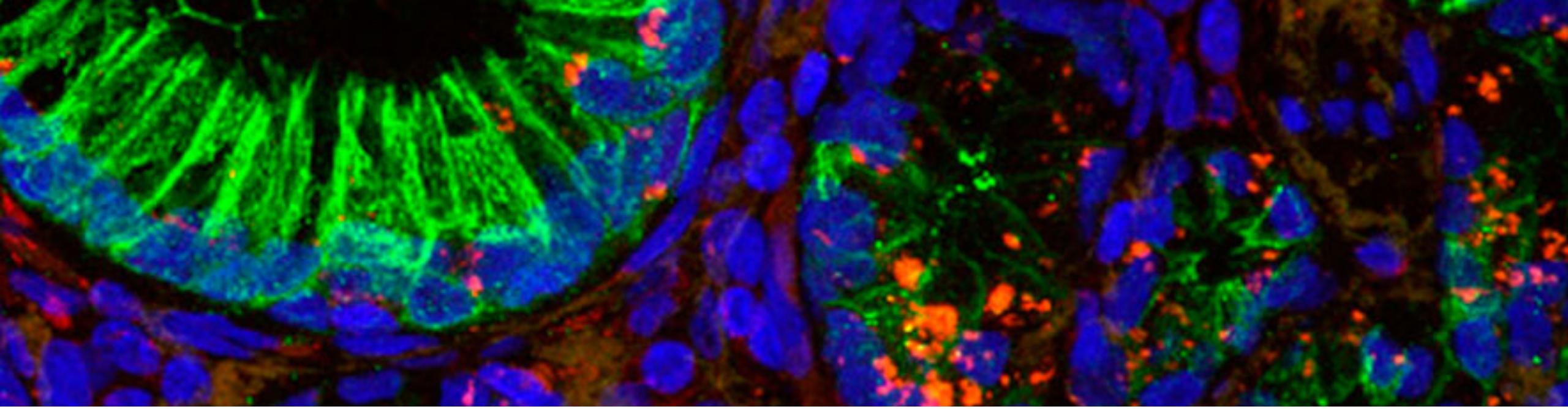


HOW MACHINE LEARNING HELPS CANCER RESEARCH

Evelina Gabasova
University of Cambridge

DNA

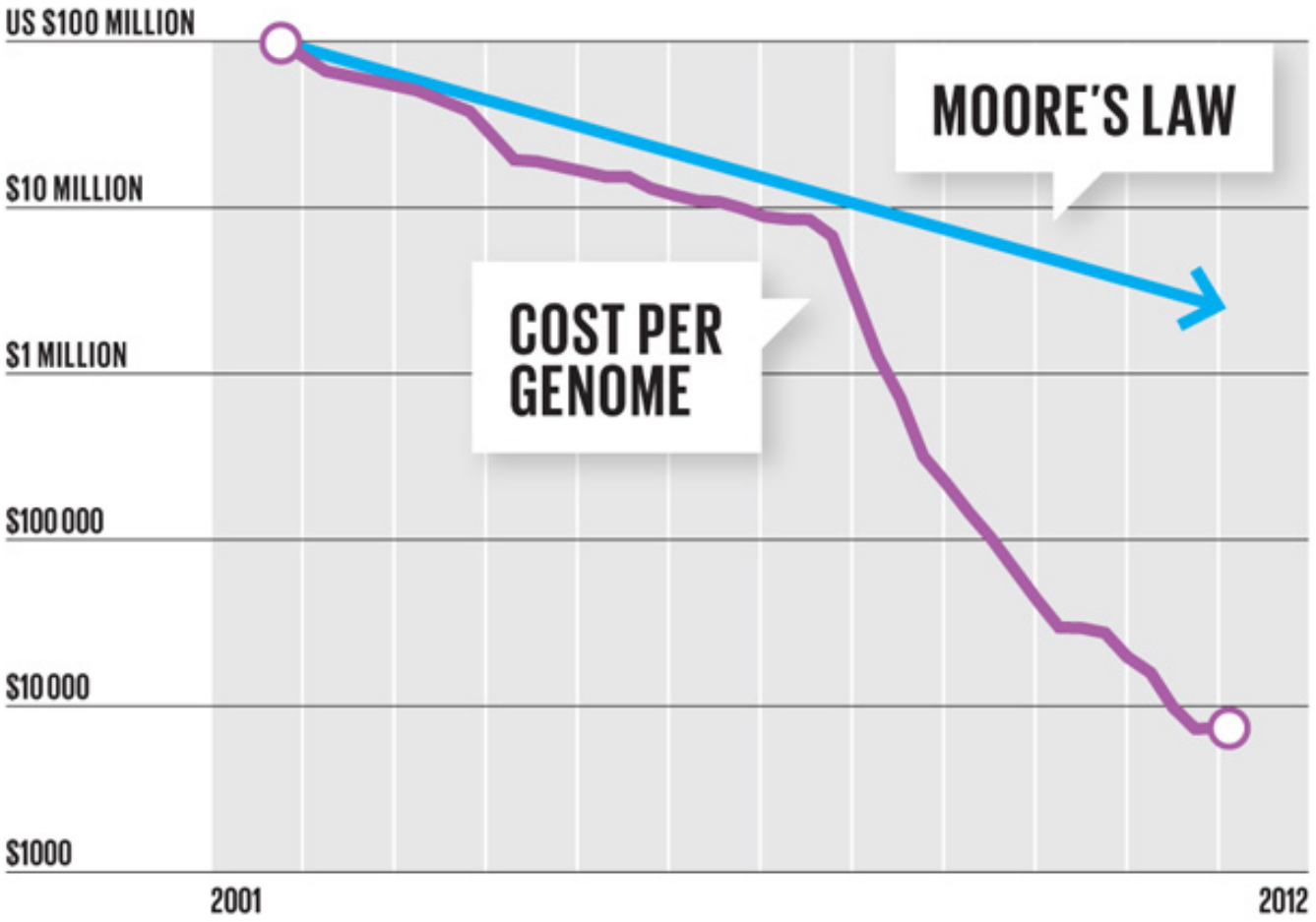
A

T

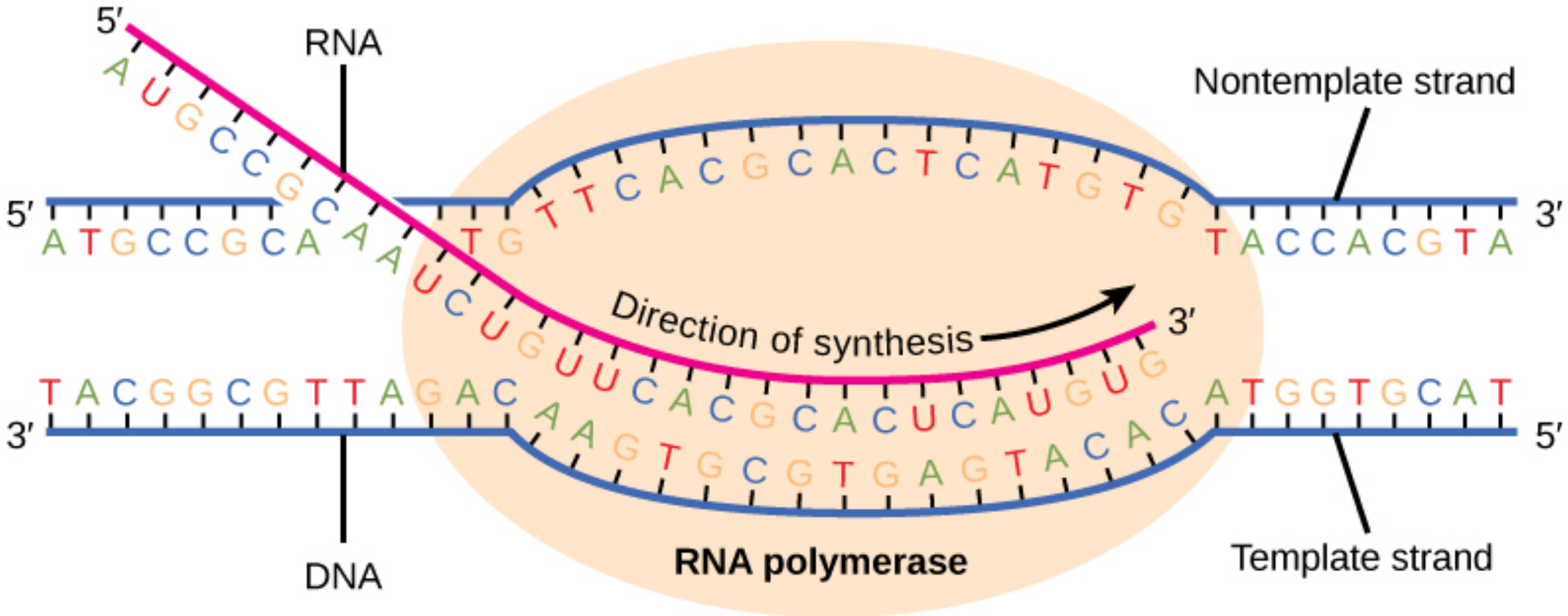
C

G

DNA sequencing



DNA and genes

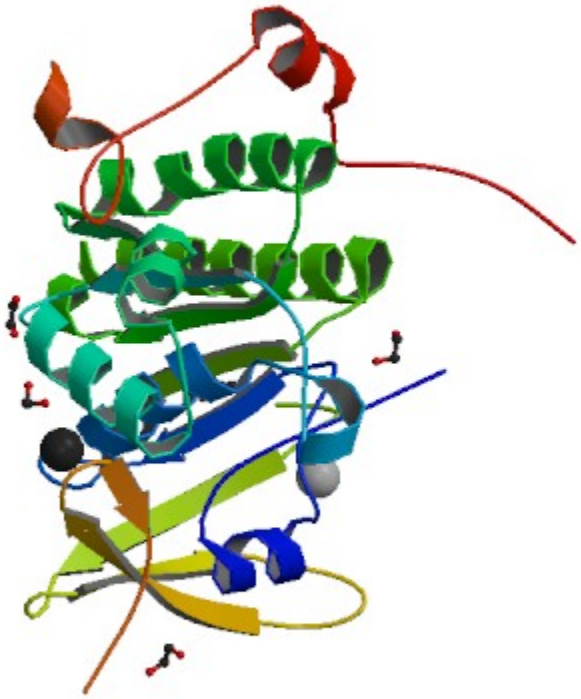


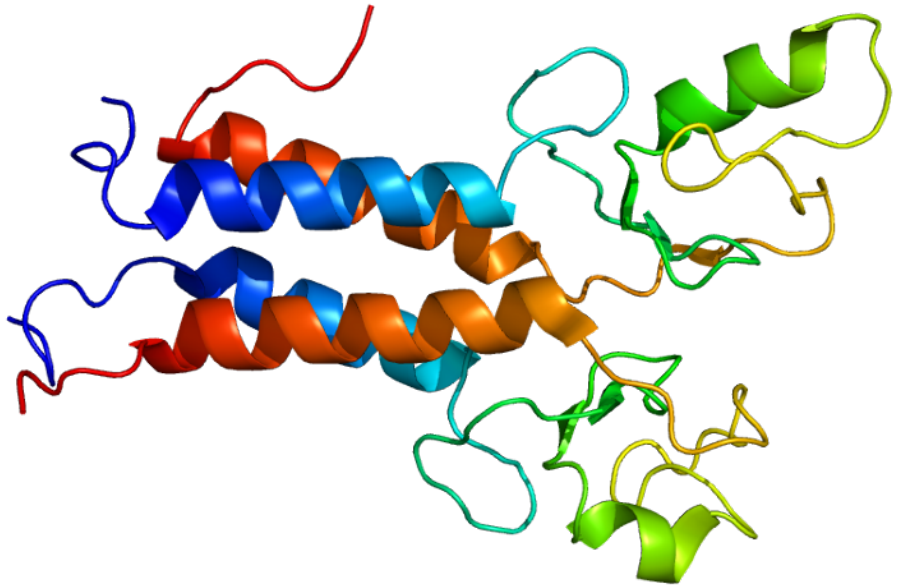
Cancer

- Genetic mutations
- Oncogenes and tumour suppressors

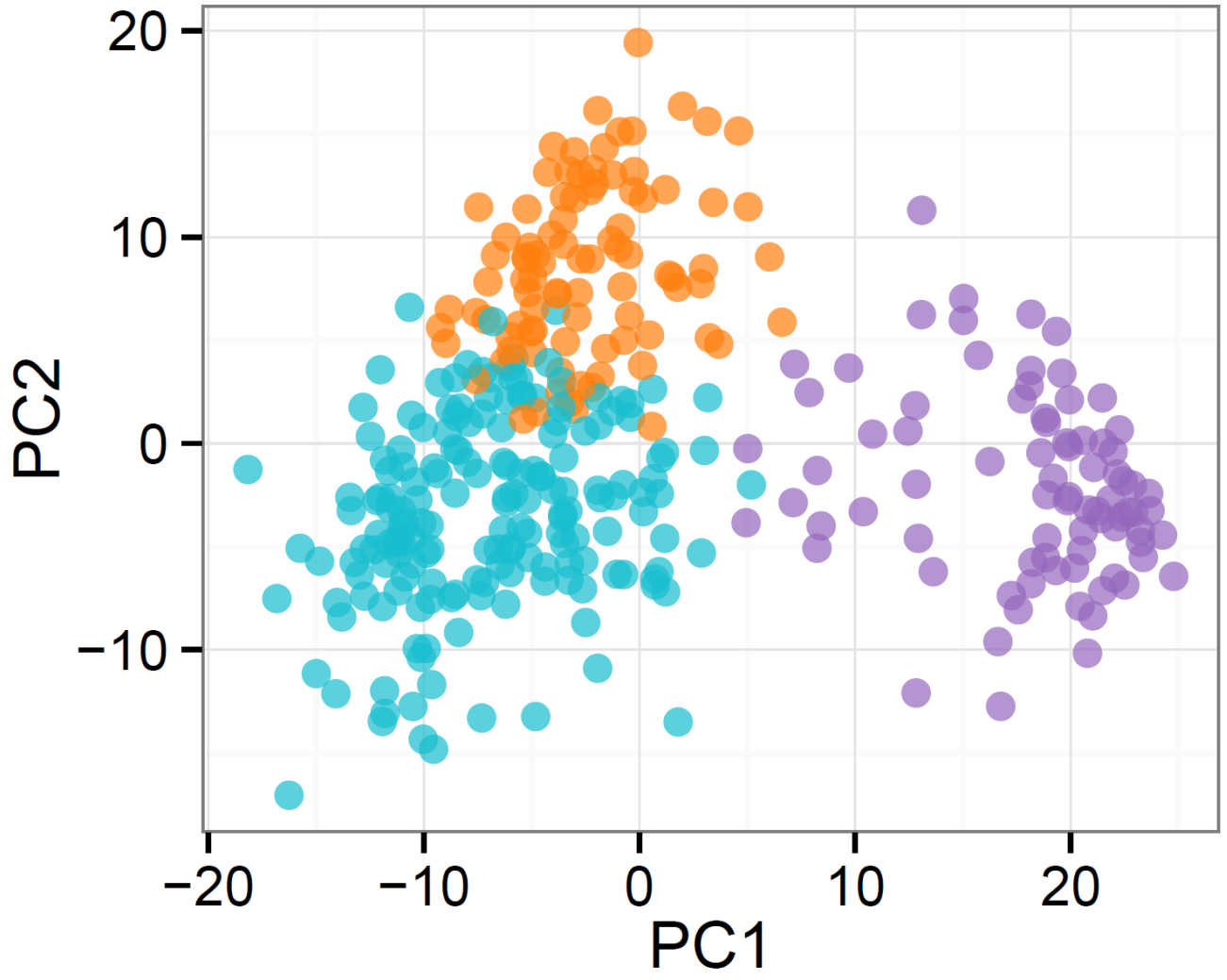
BRCA1 and BRCA2 are chromosome guardians

Cancer is not a single disease





CLUSTERING

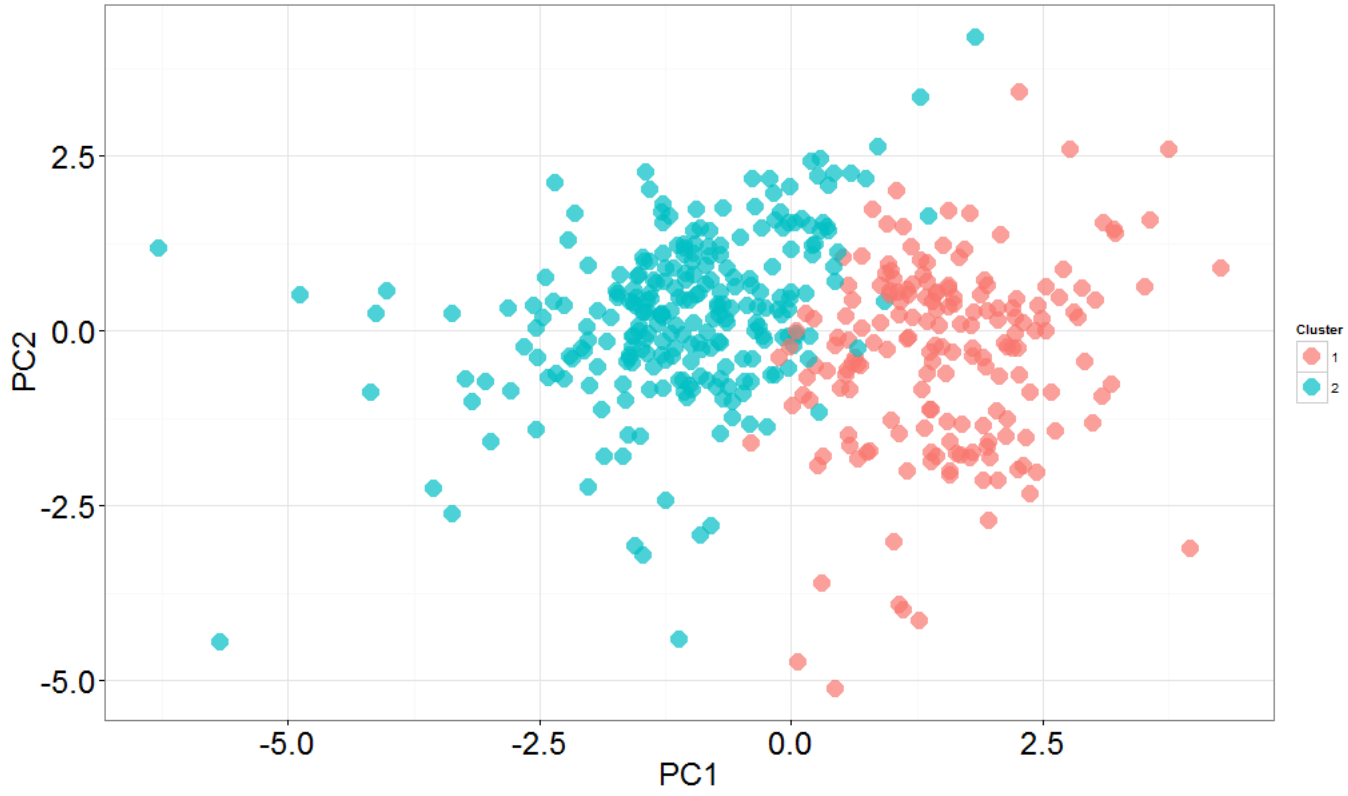


Example: clustering customers

440 wholesale customers

Annual spending on

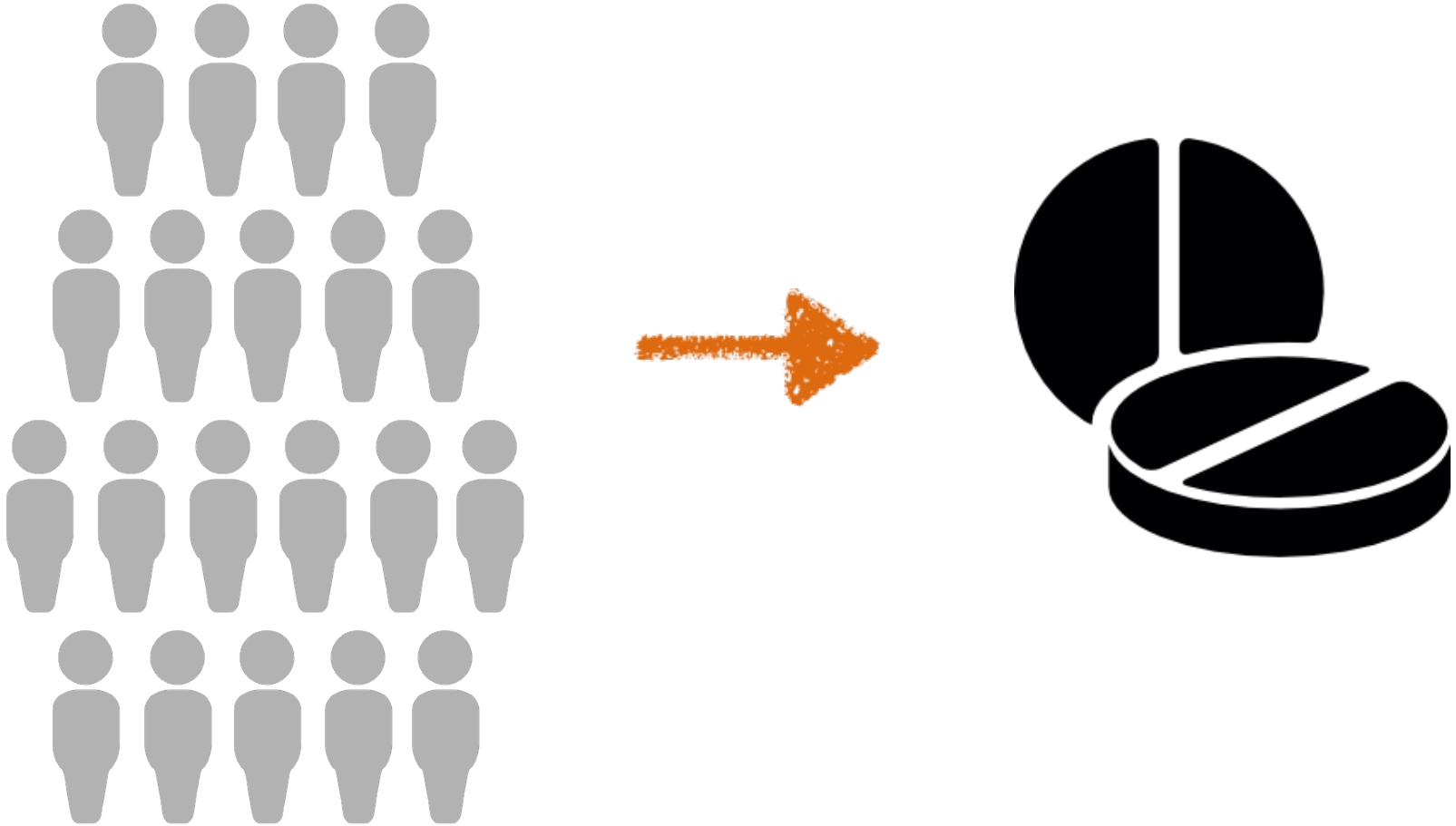
- Fresh produce
- Milk products
- Grocery products
- Frozen products
- Detergents and paper
- Delicatessen



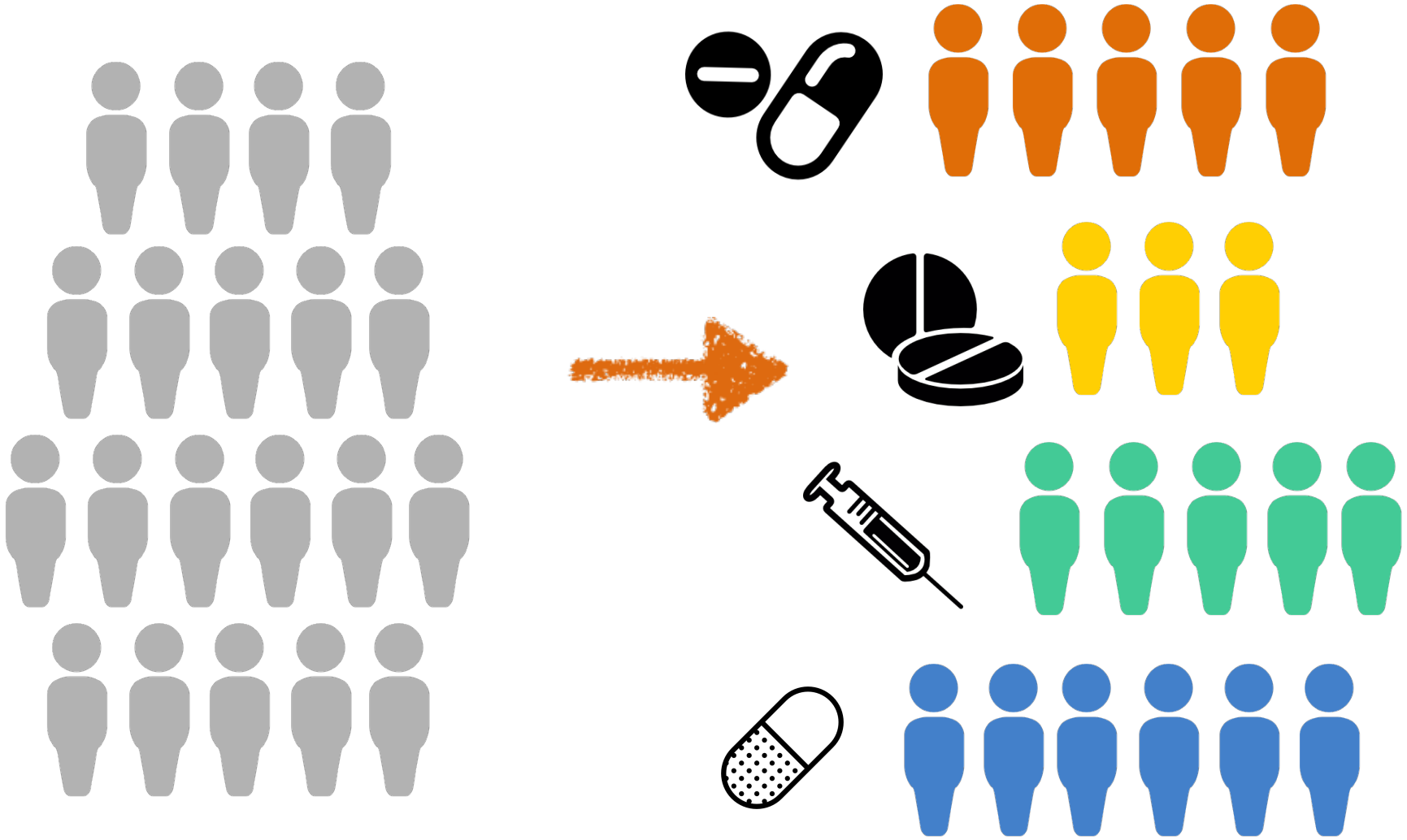
Customers  Genes

Products  Expression

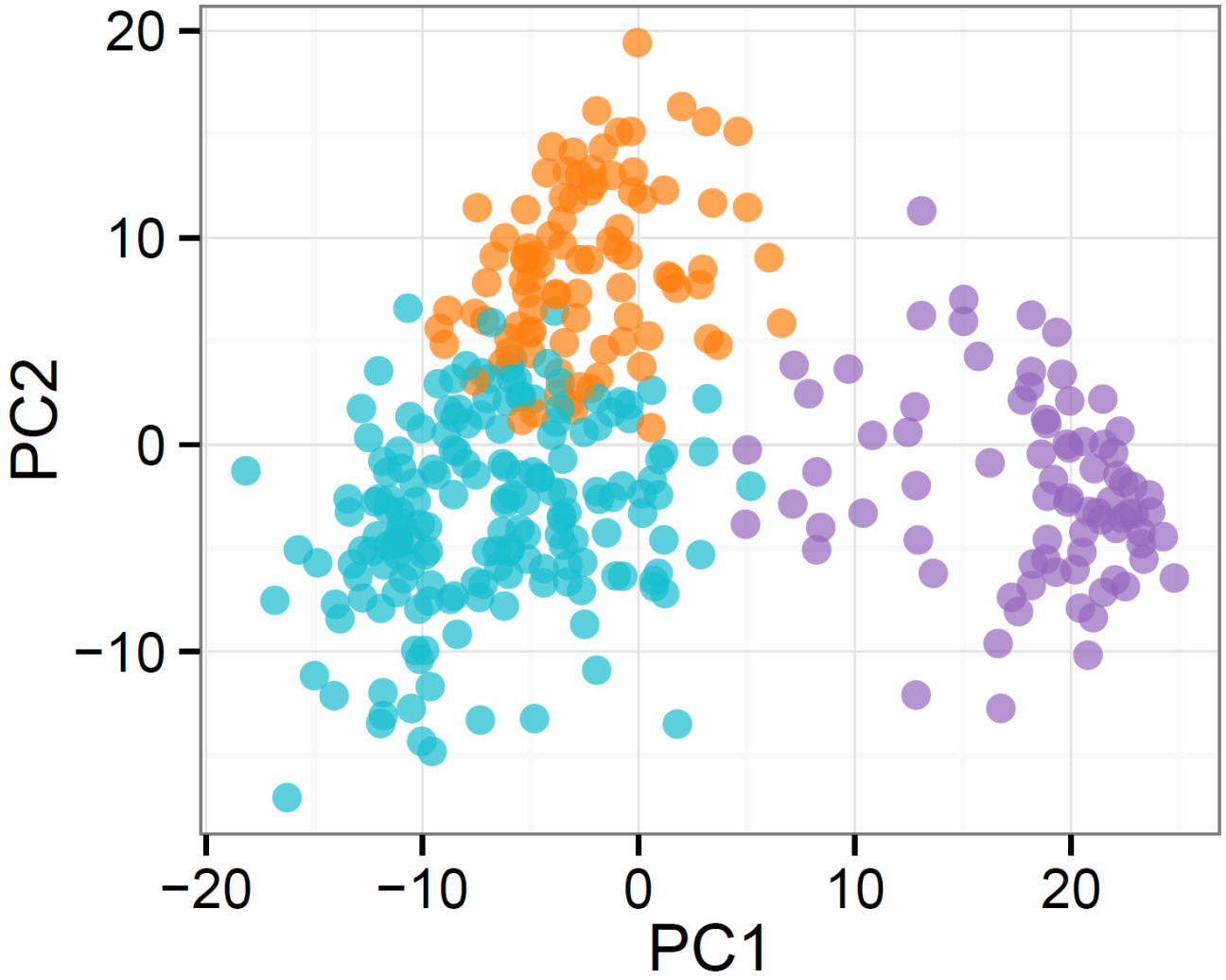
Conventional medicine



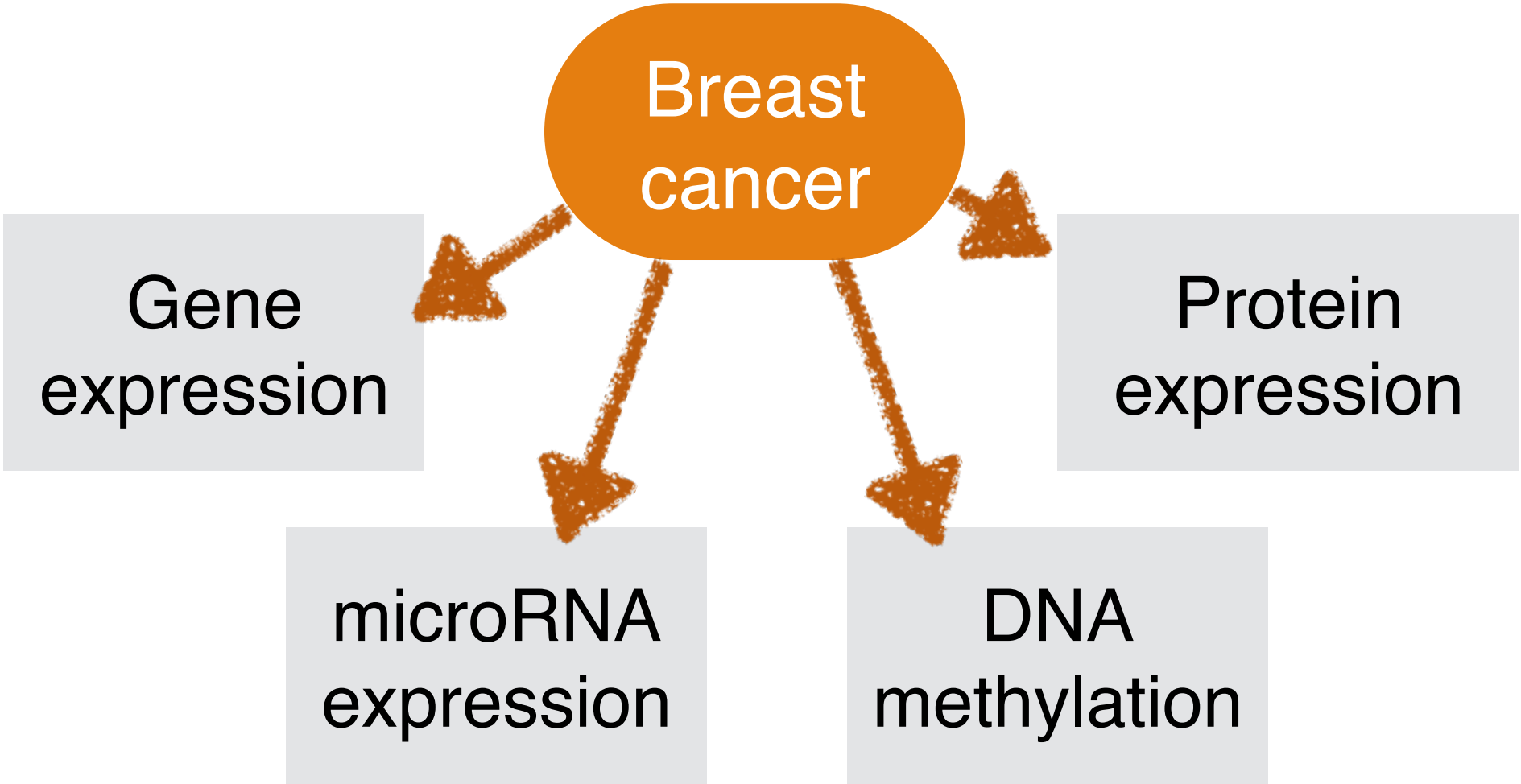
Precision medicine

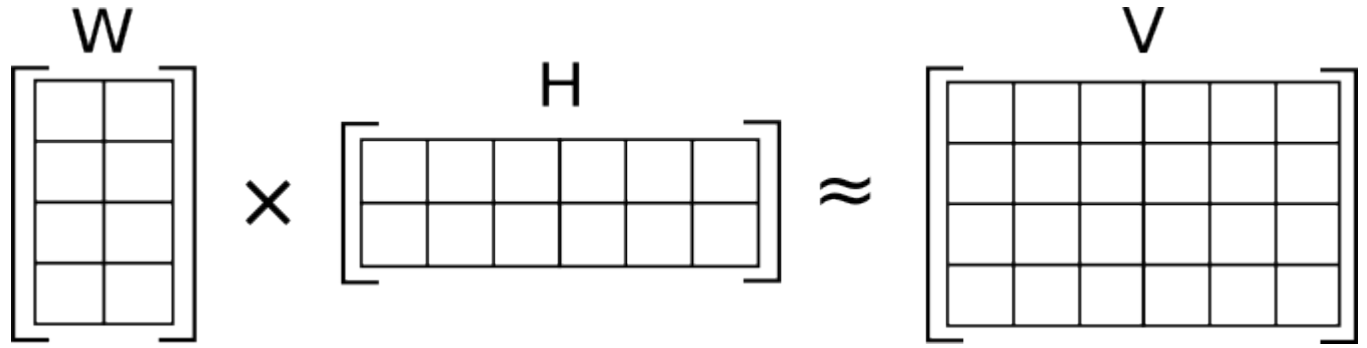


Clustering in cancer research



Integrative clustering



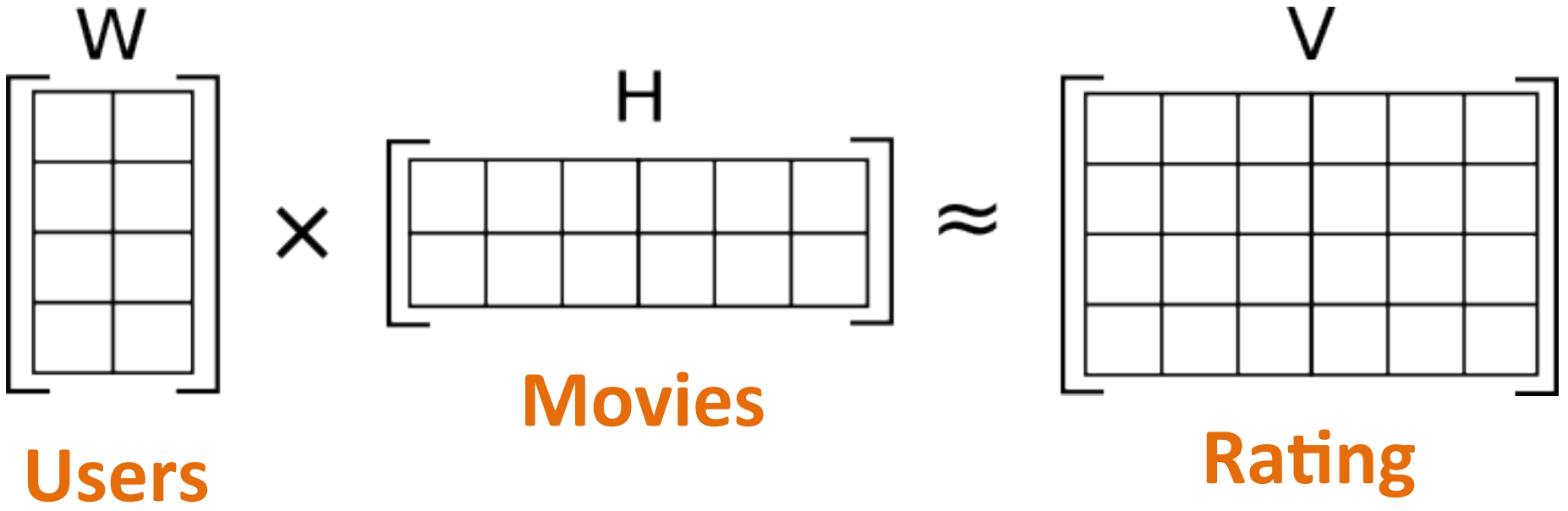


Collaborative filtering

Example: the Netflix prize

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie 6	Movie 7
Alice	★★★★★		★★★★	★★★★	★★★		
Bob		★★★		★★★★	★★		
Carol	★★					★★★★★	★★★★
Dave			★★★★★		?		

Matrix factorization

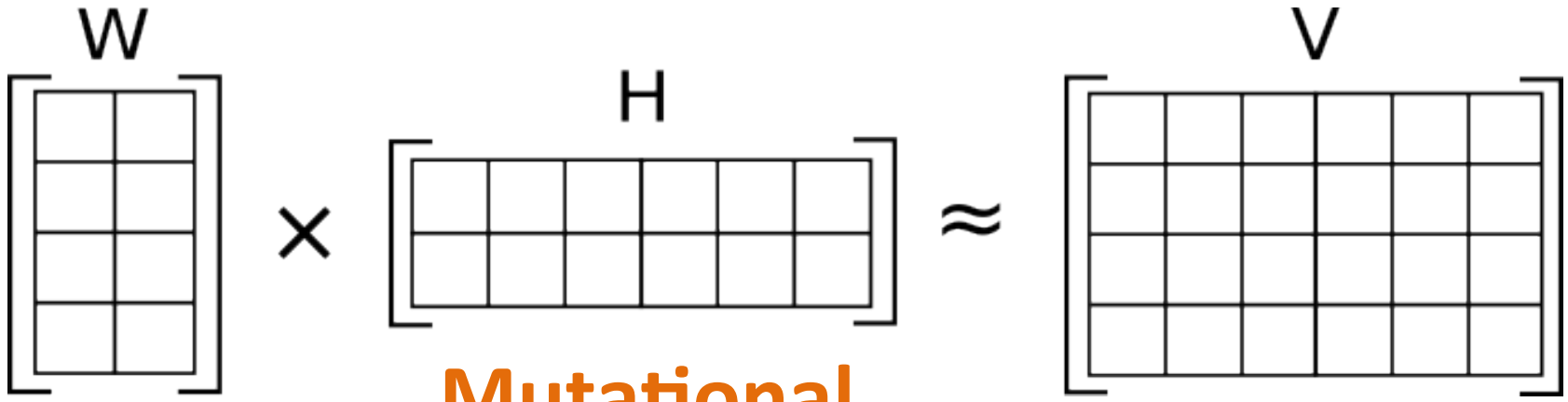


Users  Patients

Ratings  Mutations

Matrix factorization in cancer research

	C → A	C → G	C → T	T → A	T → C	T → G
Alice	5	0	0	1	2	0
Bob	1	2	0	0	0	1
Carol	3	1	0	2	0	3
Dave	0	0	2	3	1	1

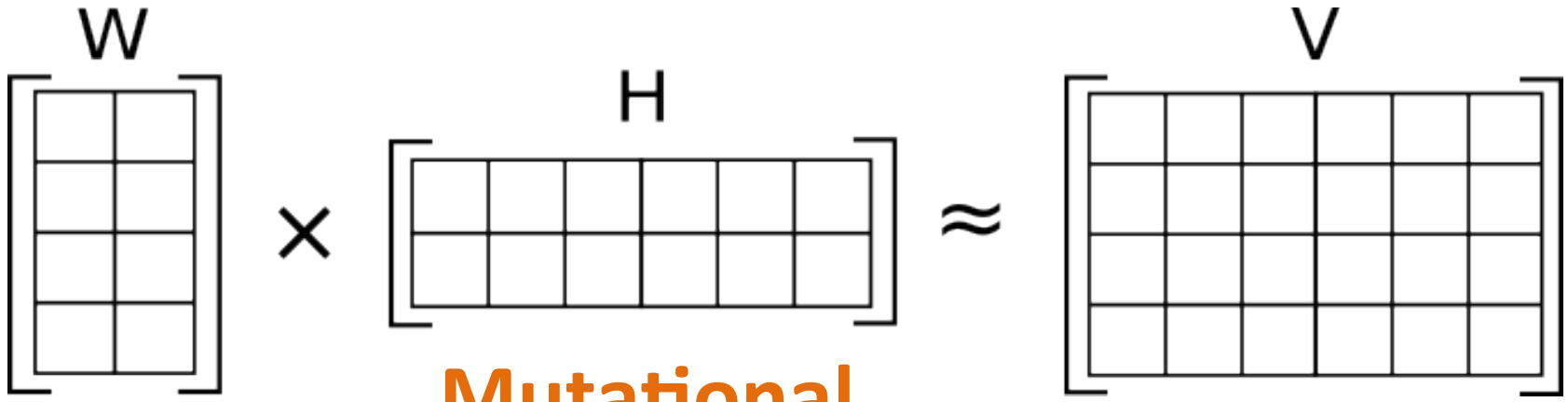


Patients

Mutational signatures

Matrix factorization in cancer research

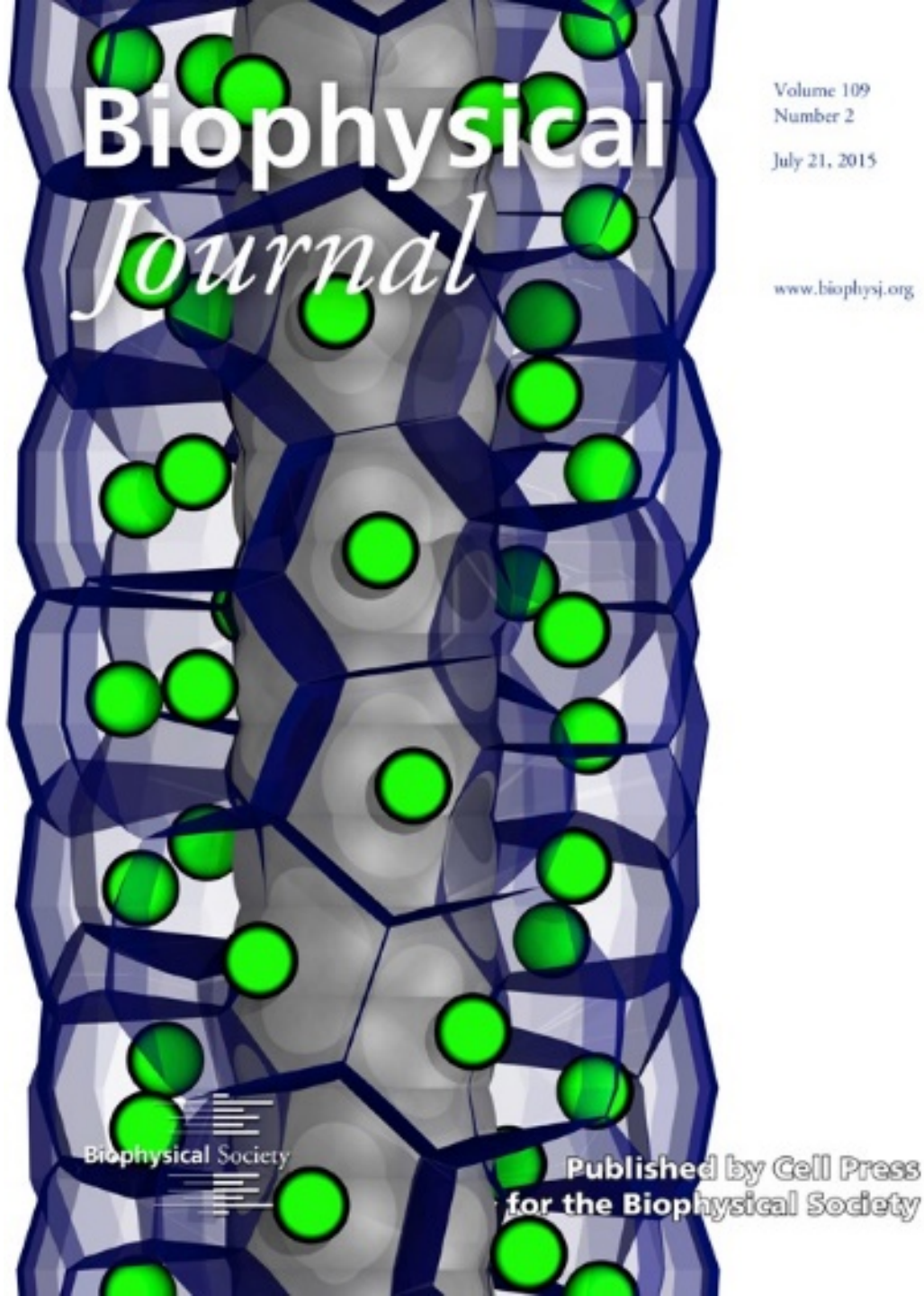
	C → A	C → G	C → T	T → A	T → C	T → G
Alice	5	0	0	1	2	0
Bob	1	2	0	0	0	1
Carol	3	1	0	2	0	3
Dave	0	0	2	3	1	1



Patients

Mutational signatures

Proving system stability



Theorem proving : SAT

$$(A \vee \neg B) \wedge (\neg A \vee B)$$

Theorem proving : SAT

$$(A \vee \neg B) \wedge (\neg A \vee B)$$

$$A = \text{true}$$

$$B = \text{true}$$

Theorem proving : SMT

$$(A \vee \neg B) \wedge (\neg A \vee B)$$

$((a > 3) \vee (b < 1))$ $((a < 5) \vee (b = 0))$

Theorem proving : SMT

Satisfiability modulo theories

$$(A \vee \neg B) \wedge (\neg A \vee B)$$

$((a > 3) \vee (b < 1))$ $((a < 5) \vee (b = 0))$

$a = 4$

$b = 0$

Software verification

Preconditions

Postconditions

Loop conditions

Assertions

…

SMT formulas

Z3 theorem prover

Software verification

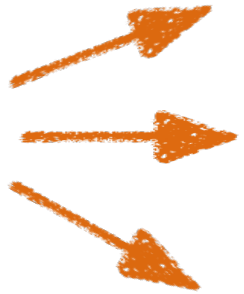
```
static void Swap(int[]! a, int i, int j)
requires 0 <= i && i < a.Length;
requires 0 <= j && j < a.Length;
{
    int temp;
    temp = a[i];
    a[i] = a[j];
    a[j] = temp;
}
```

Spec#

Proving stability of biological systems

Proteins
Genes
Receptors
…

Variables



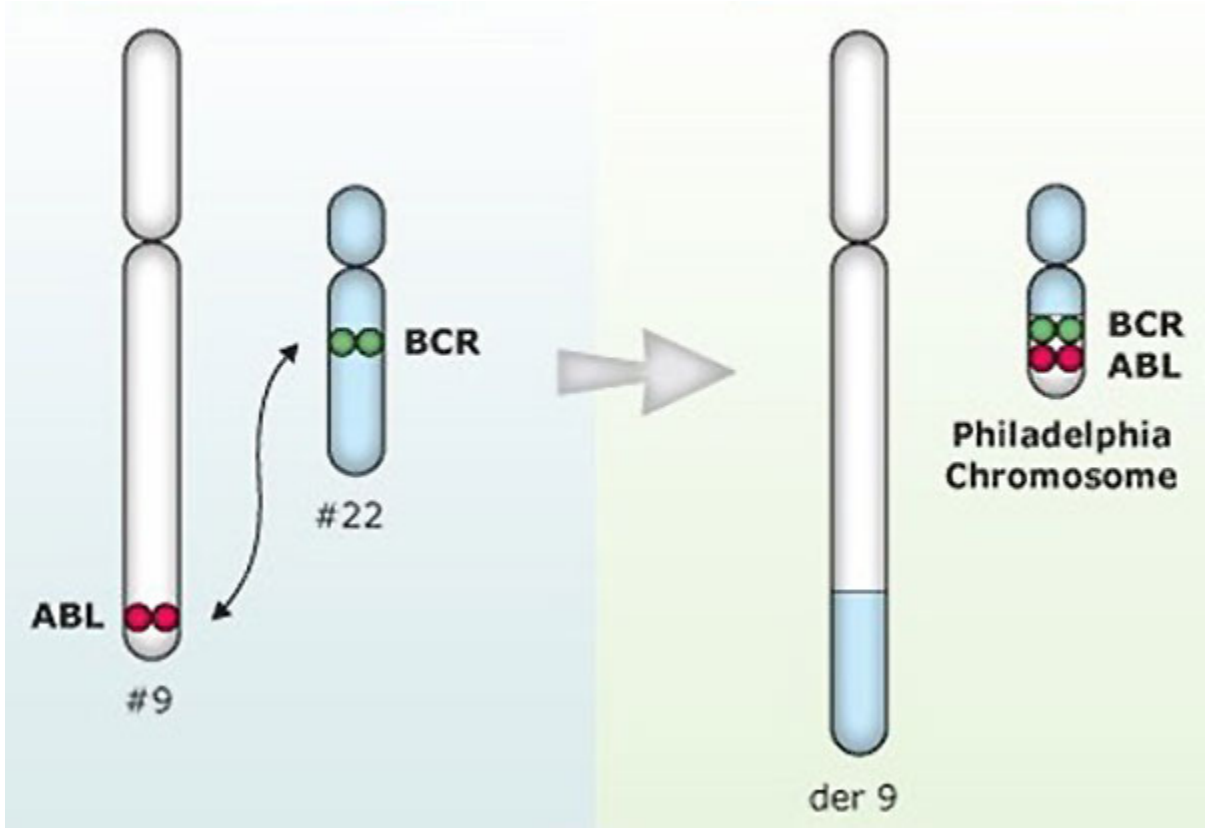
v

$v + 1$ if $v < T(v)$

v if $v = T(v)$

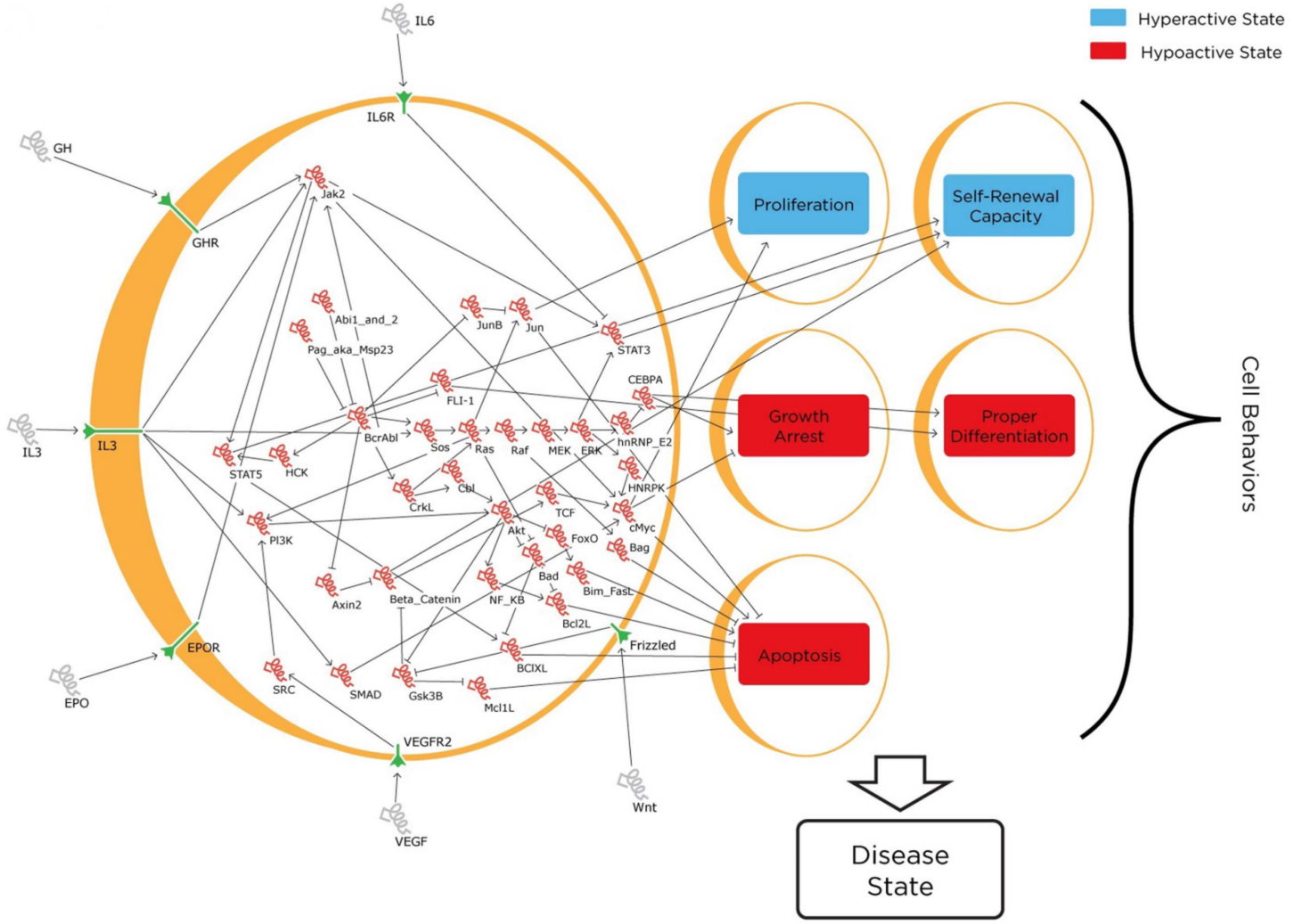
$v - 1$ if $v > T(v)$

Chronic myeloid leukemia



Proving stability of biological systems

Chronic myeloid leukemia



Machine learning is not just
for targeted advertising
or
algorithmic trading

@evelgab

evelinag

evelina@evelinag.com

evelinag.com

Links

- MRC Cancer Unit, University of Cambridge http://www.mrc-cu.cam.ac.uk/
- Z3 code samples https://github.com/evelinag/cancer-research-and-ML
- Big Data: Astronomical or Genomical? http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002195
- Signatures of mutational processes in cancer http://www.nature.com/nature/journal/v500/n7463/full/nature12477.html
- Proving stabilization of biological systems http://research.microsoft.com/en-us/um/cambridge/projects/terminator/biocheck11-final.pdf
- Bio Model Analyser http://biomodelanalyzer.research.microsoft.com/
- Drug target optimization in Chronic Myeloid Leukemia http://research.microsoft.com/apps/pubs/default.aspx?id=226356
- Z3 theorem prover https://github.com/Z3Prover