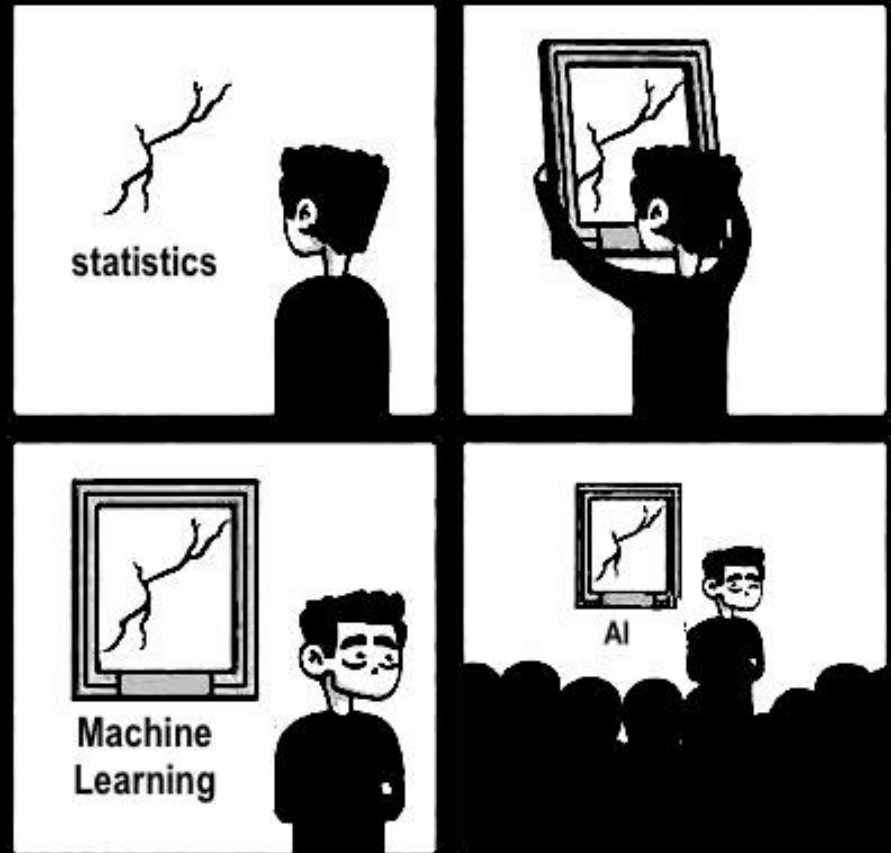




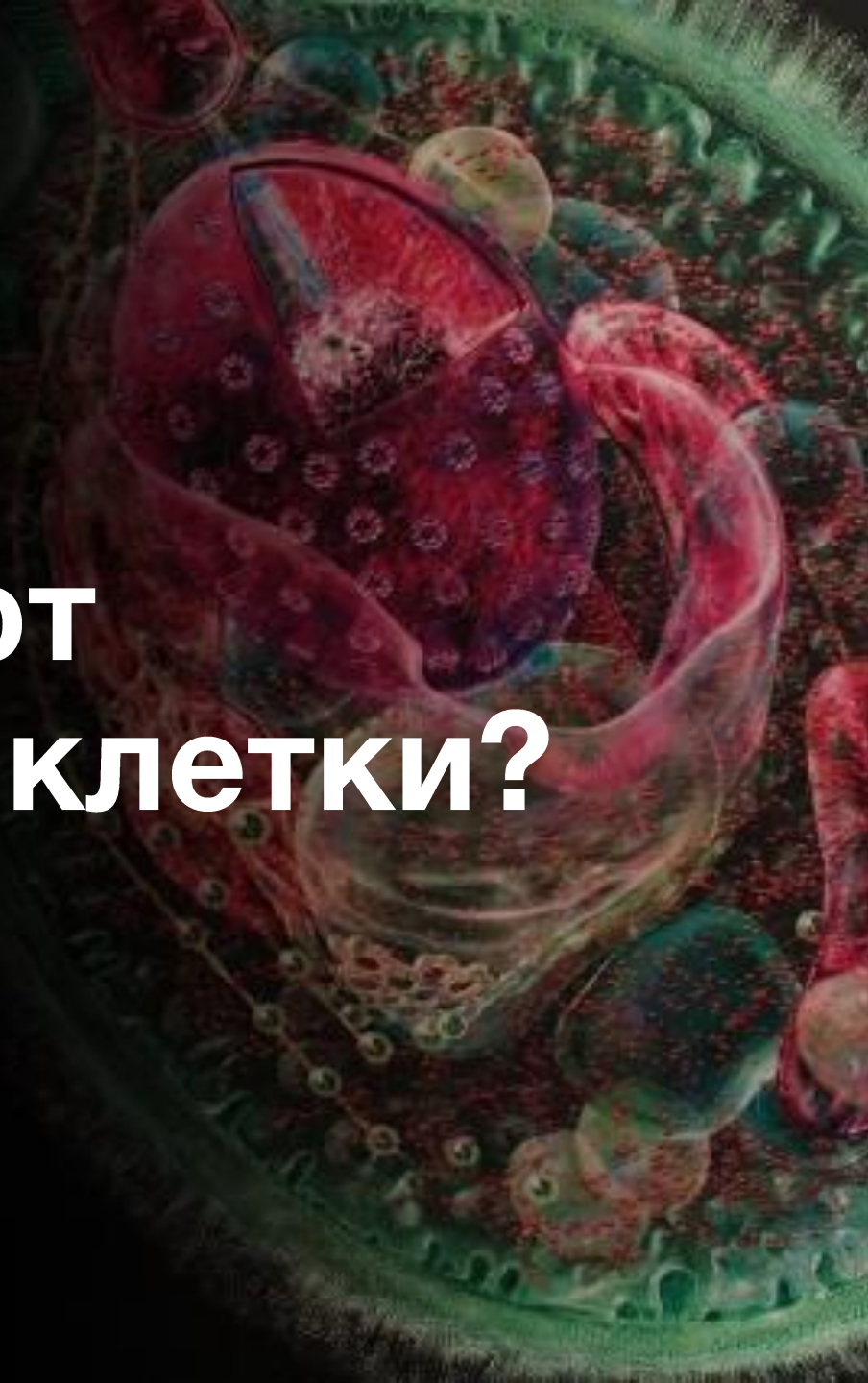
Чем занимаются биоинформатики?

Качественный анализ данных предполагает уверенное понимание того, откуда данные взялись и как они устроены.



Биоинформатики занимаются анализом биологических и медицинских данных — с пониманием их специфики.

**Как изучают
отдельные клетки?**



Живой организм
СОСТОИТ ИЗ КЛЕТОК.



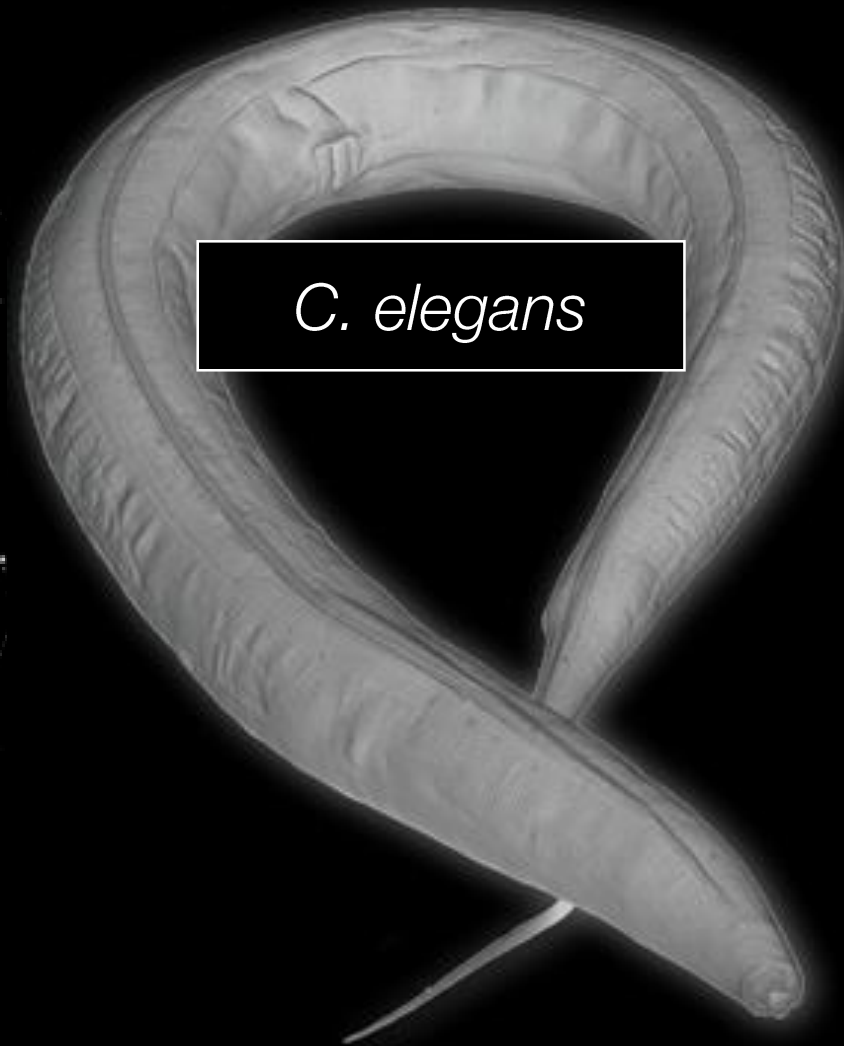
Рисунок *Dicyeta* sp. — сильно упрощённого животного. На рисунке отчётливо прорисованы клетки.

Одна клетка



Эвглена

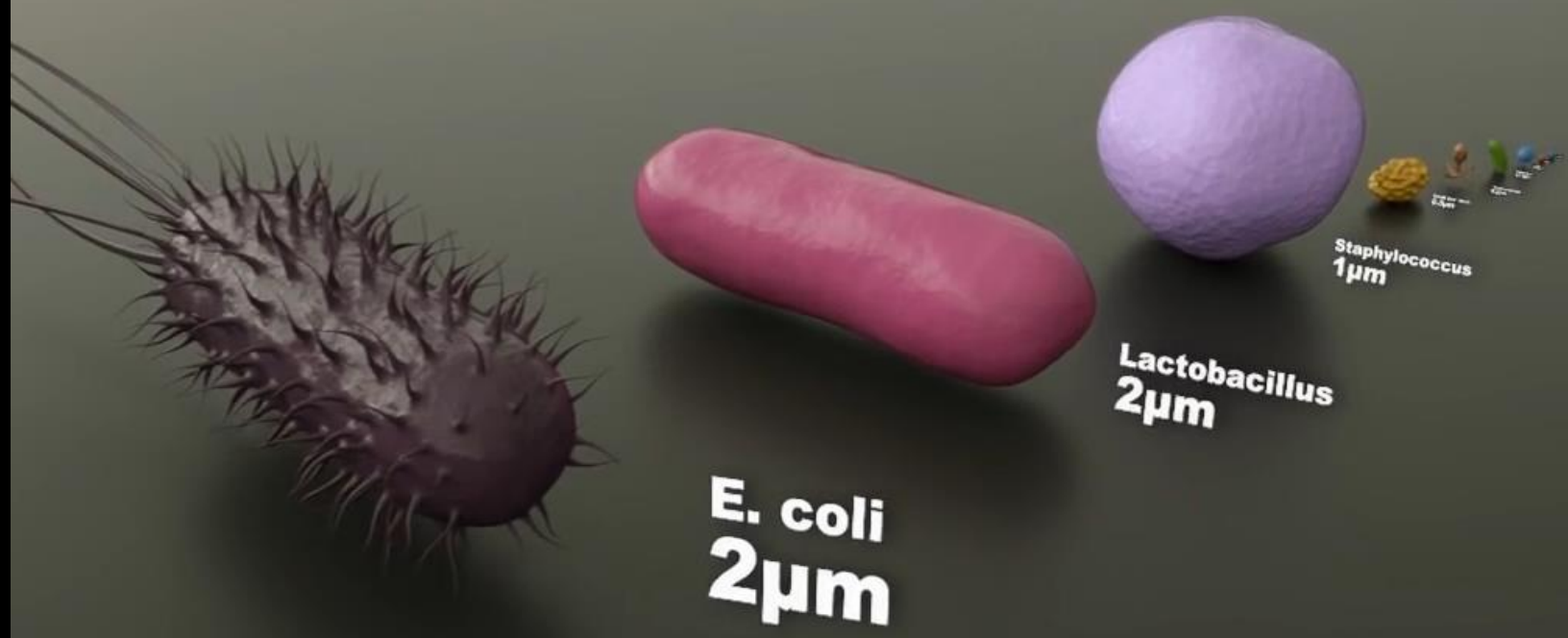
Много клеток



C. elegans

Клетки многоклеточного организма различны по размеру, форме и функциям.

Álvaro Gracia Montoya
MBS
November 2017

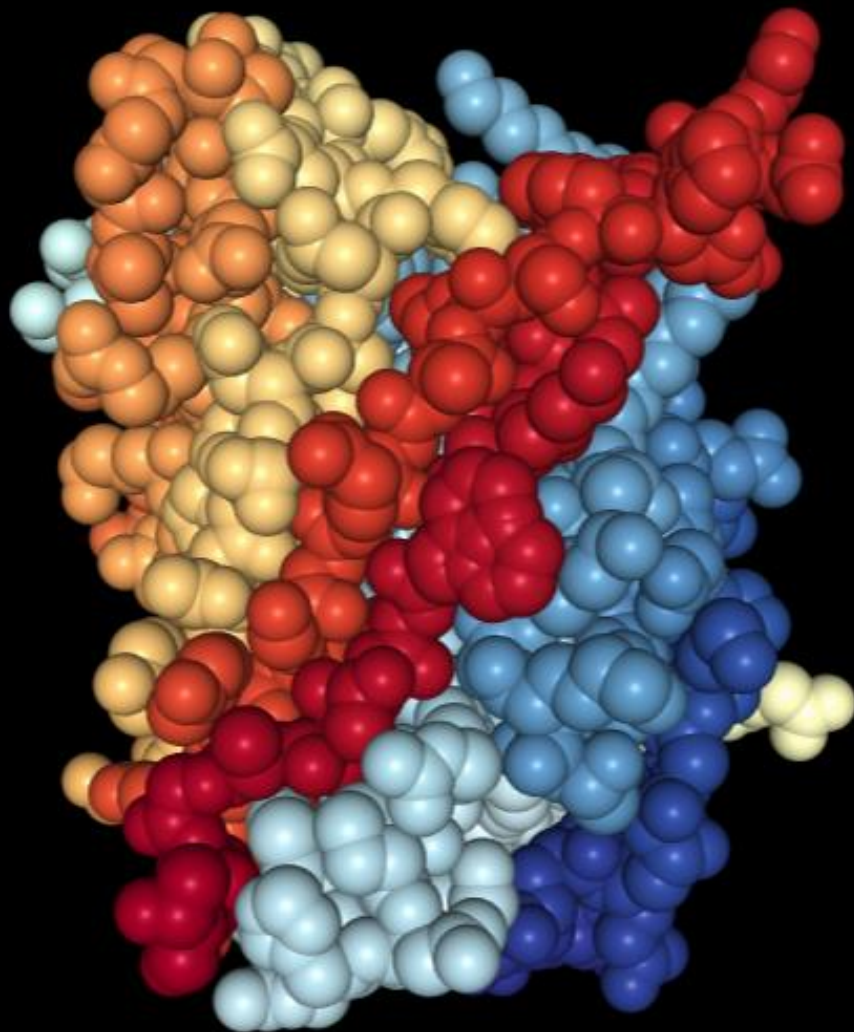


Клетка определяется её химическим составом. Основные химические компоненты клетки — это

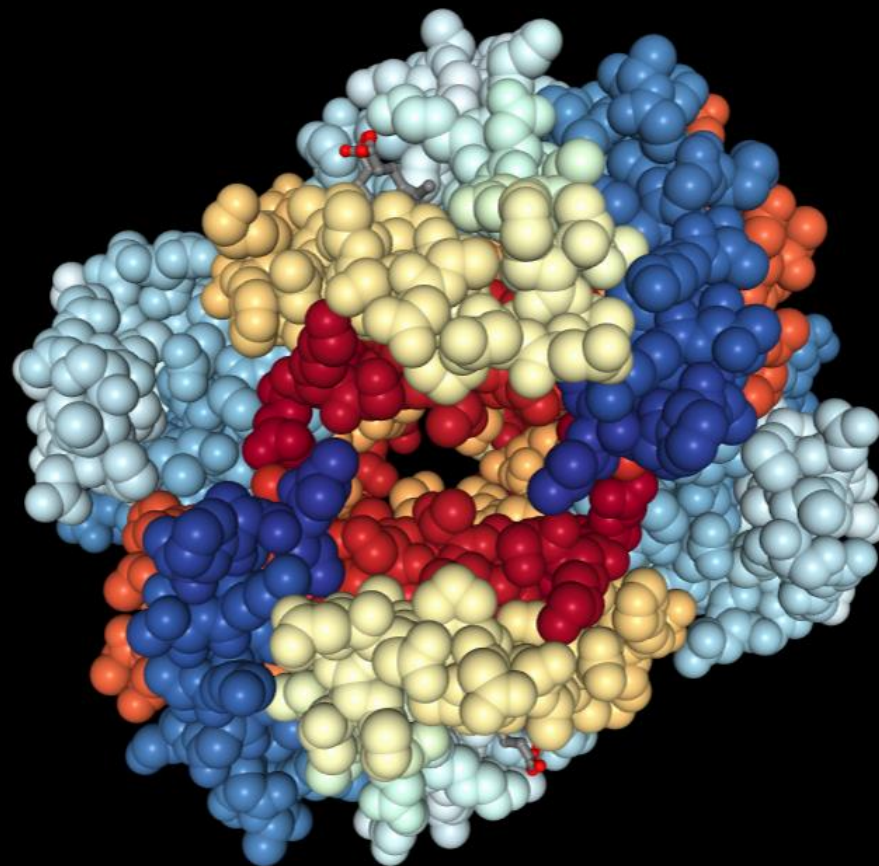
- **белки,**
- жиры,
- углеводы и
- нуклеиновые кислоты.

В первую очередь клетка определяется тем, какие **белки** в ней содержатся.

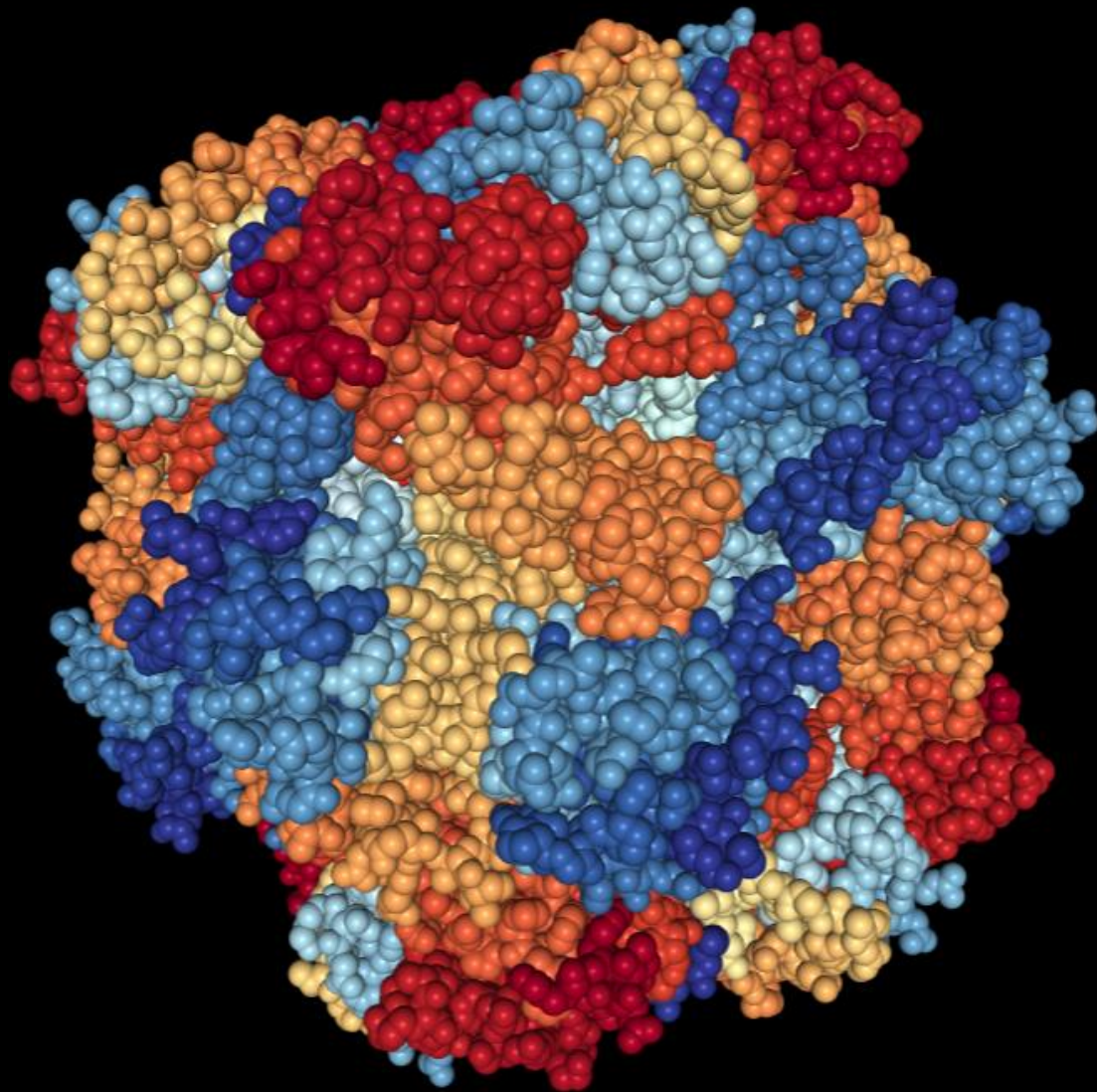
GFP —



Гемоглобин —



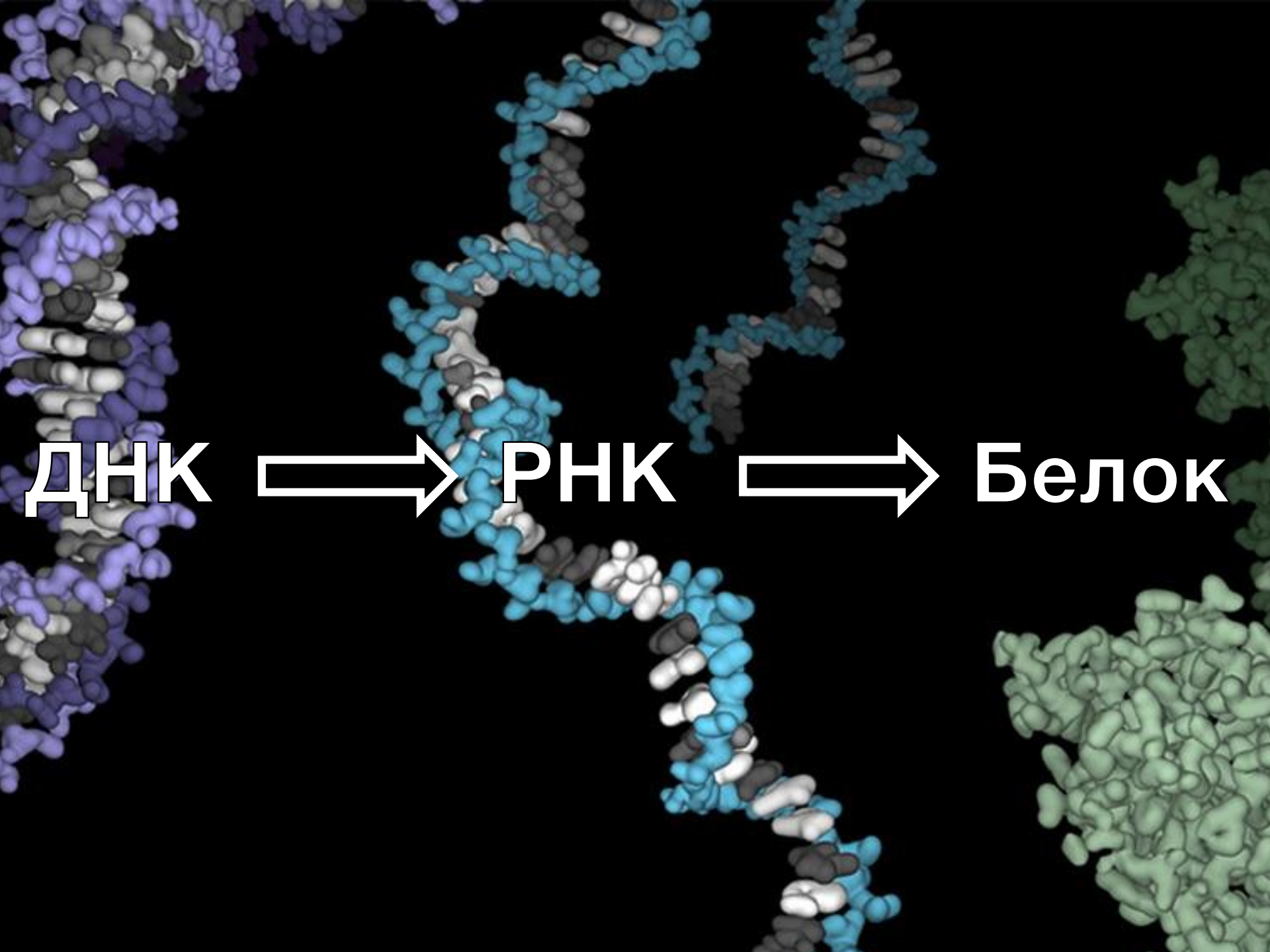
Рубиско —



В разных клетках содержатся разные белки. Белки — это паспорт клетки.

Мы не умеем определять, сколько и каких белков есть в каждой клетке.

Зато мы умеем определять, сколько
и каких РНК есть в клетке.



ДНК



РНК



Белок

Весь этот процесс называется
экспрессией гена.

Таким образом, клетка характеризуется набором чисел — это значения экспрессий каждого из генов.

Клетка	Ген 1	Ген 2	Ген 3	...
---------------	--------------	--------------	--------------	------------

№1	20	56	0	...
-----------	----	----	---	-----

№2	0	17	120	...
-----------	---	----	-----	-----

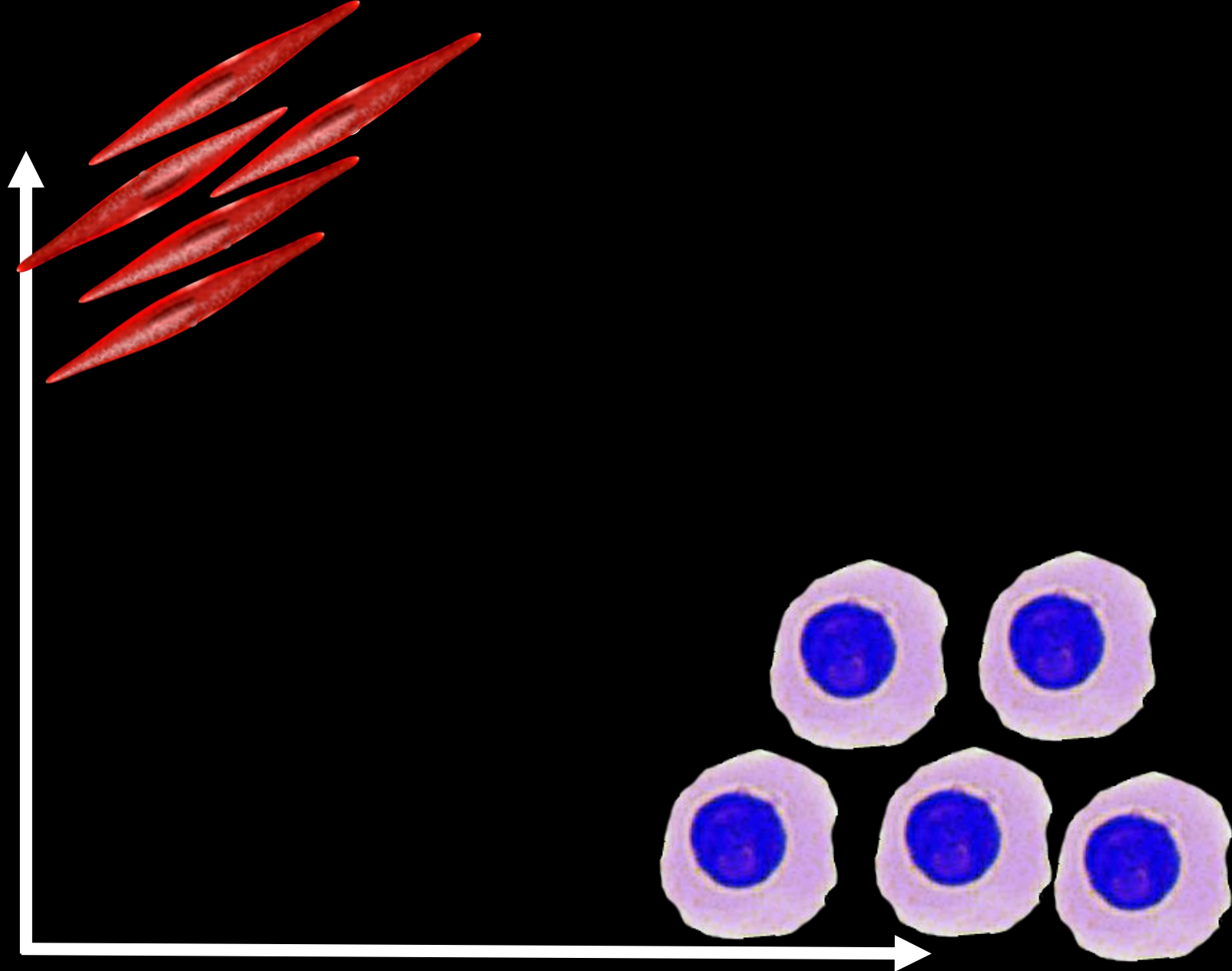
№3	2	16	25	...
-----------	---	----	----	-----

...
-----	-----	-----	-----	-----

Сколько разных типов клеток мы наблюдаем?

Попробуем посмотреть на ситуацию попроще — у нас есть два гена (гемоглобин и миозин) и десять клеток.

МИОЗИН



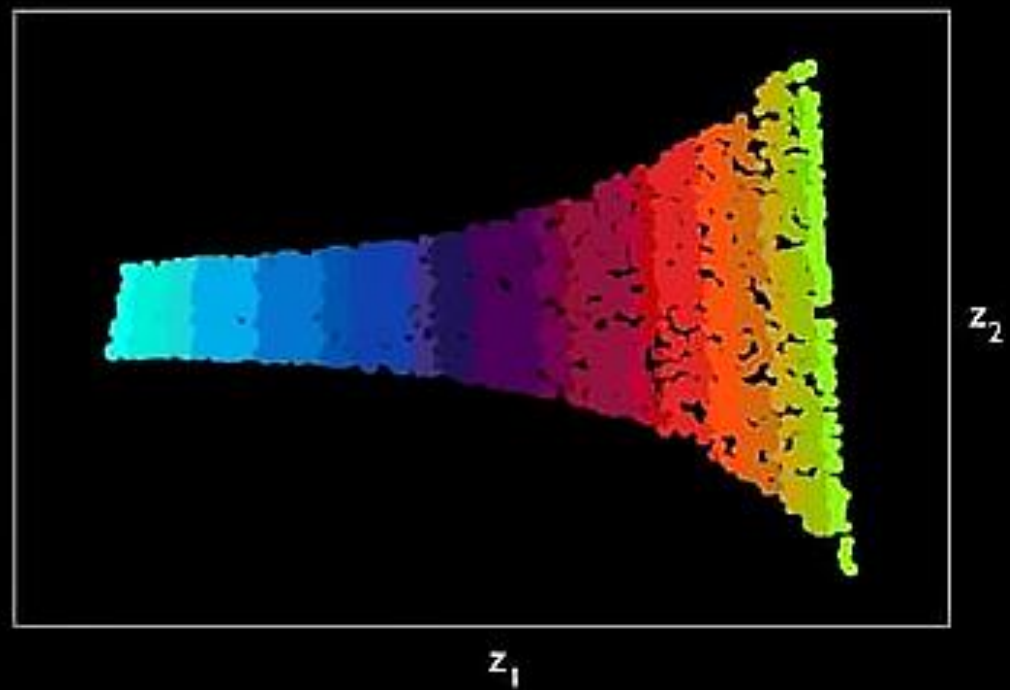
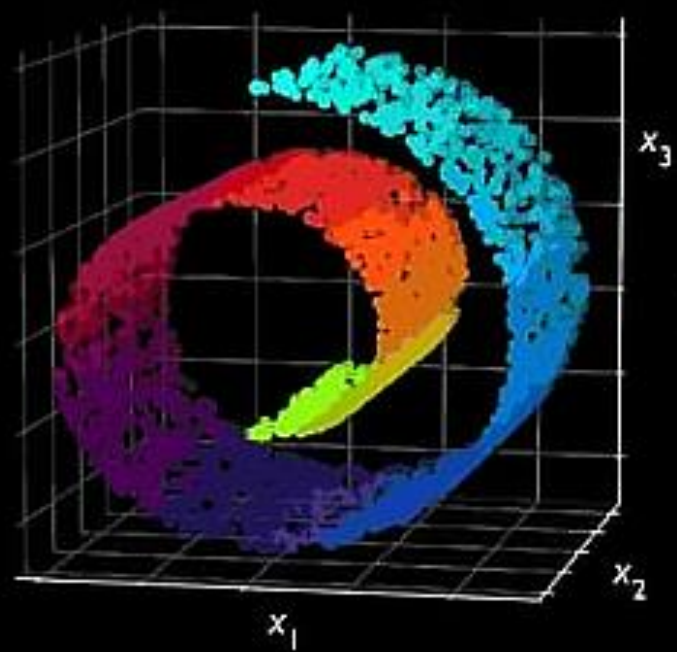
ГЕМОГЛОБИН

Ситуация с тремя генами ясна. А если их четыре? Или четыре тысячи?

Формально мы изучаем набор из **M** точек (M — число клеток) в **N**-мерном пространстве (N — число генов).

Как их сгруппировать? Подход простой — находим «расстояния» между точками по теореме Пифагора, после чего объединяем самые близкие друг с другом.

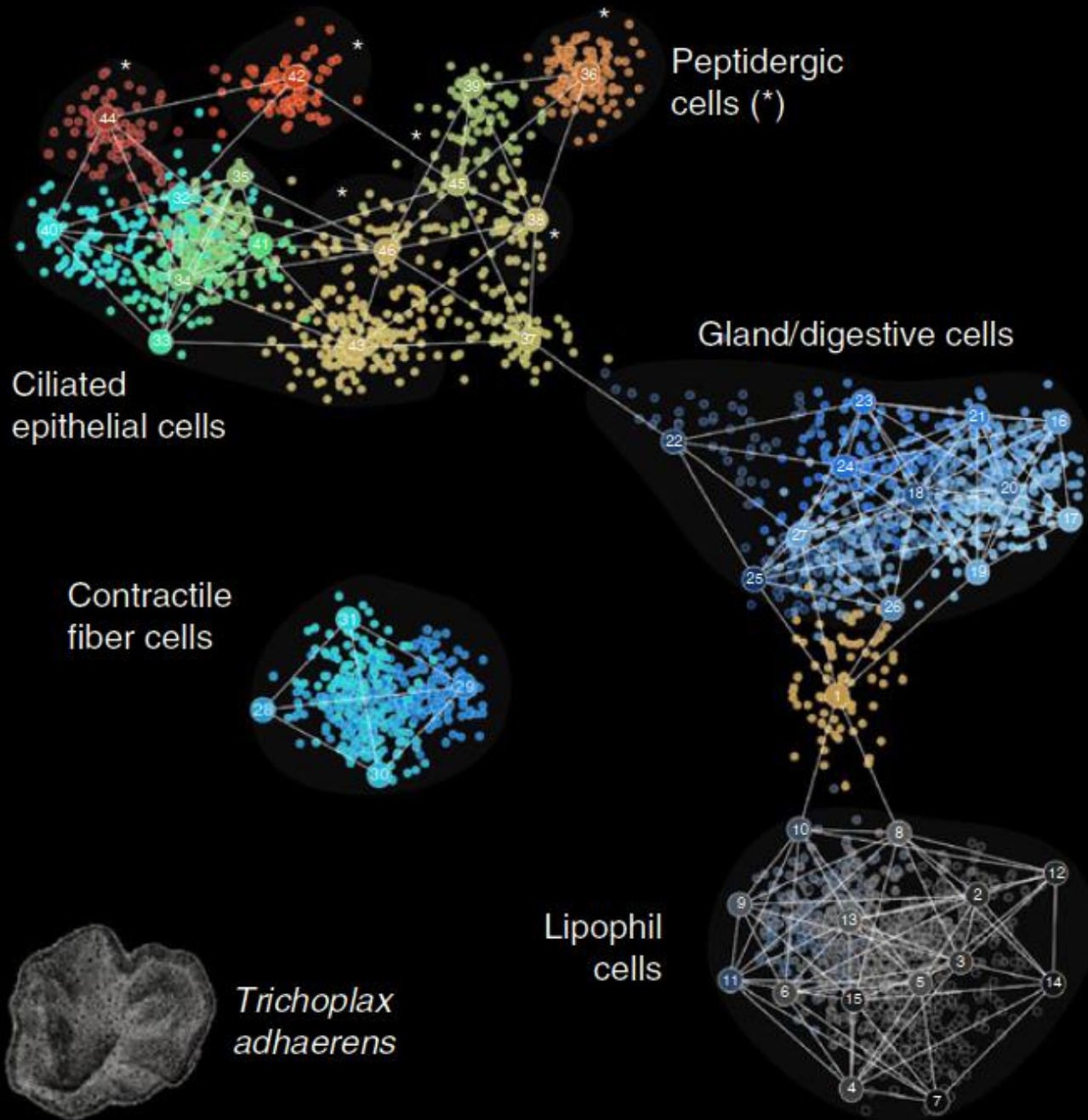
Порой для изучения данных таких высоких размерностей сначала производят процедуру понижения размерности.



Сайт с примерами

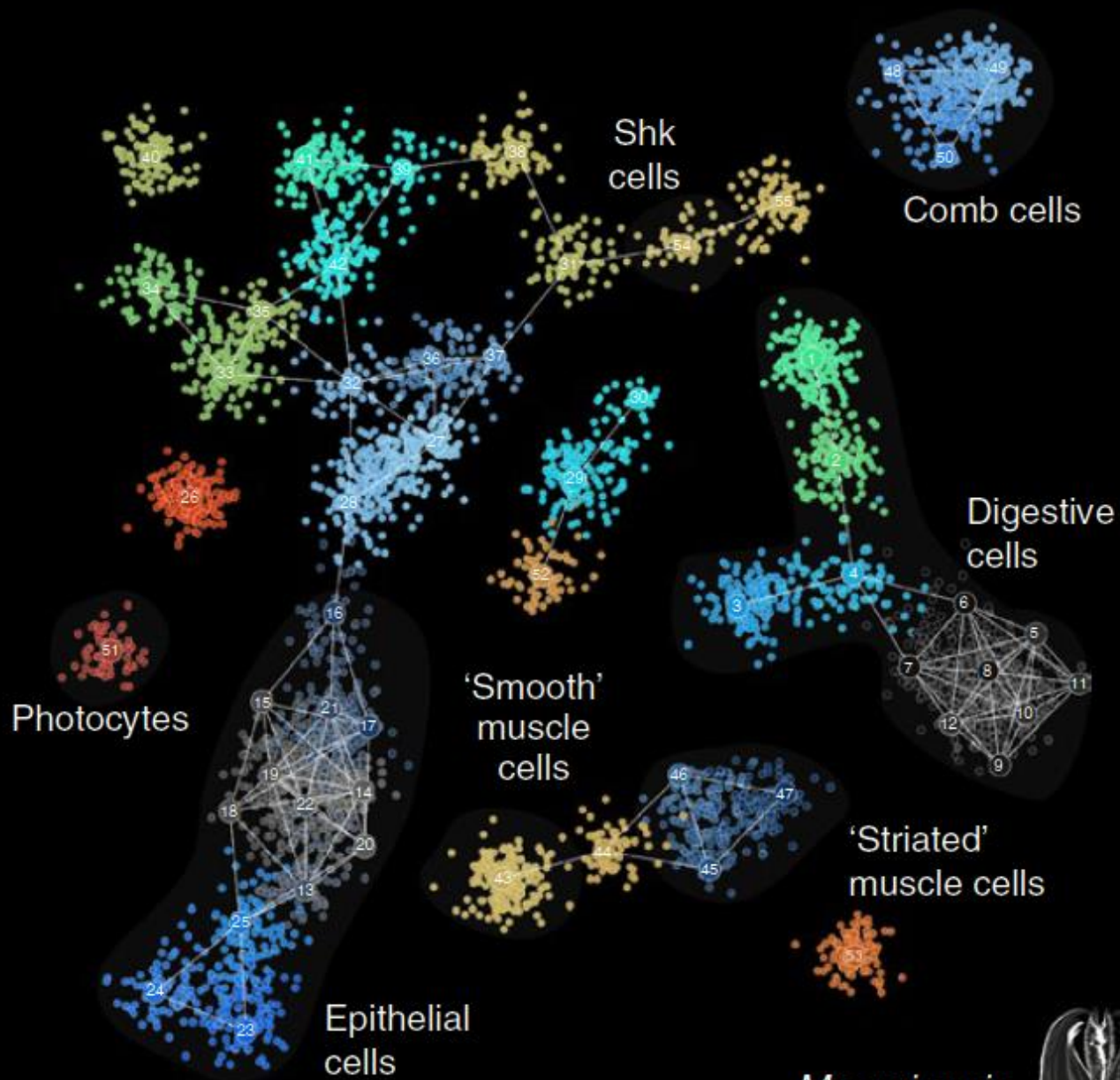
Знакомьтесь, перед вами
трихоплакс.







А это гребневик.



Mnemiopsis leidyi



Что ещё можно изучить при помощи этой методики?