



«Анализ транскриптомных данных»

Лекция #10. Коррекция батч-эффекта

Серёжа Исаев

аспирант **ФБМФ МФТИ**
аспирант **MedUni Vienna**

Содержание курса

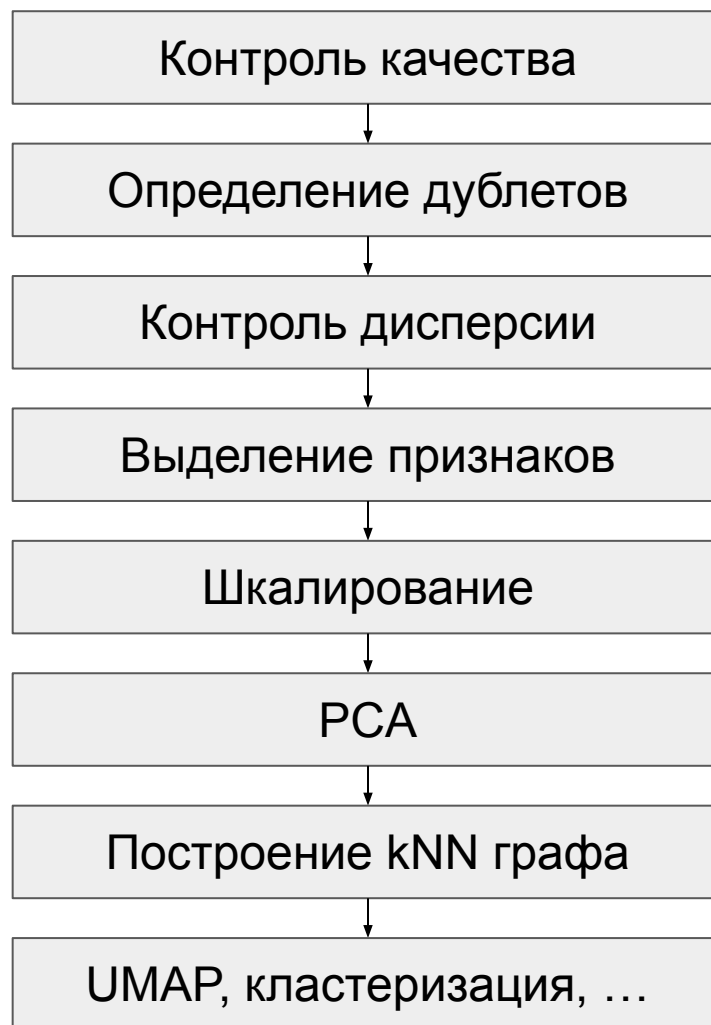
1. Bulk RNA-Seq:

- a. экспериментальные подходы,
- b. выравнивания и псевдовыравнивания,
- c. анализ дифференциальной экспрессии,
- d. функциональный анализ;

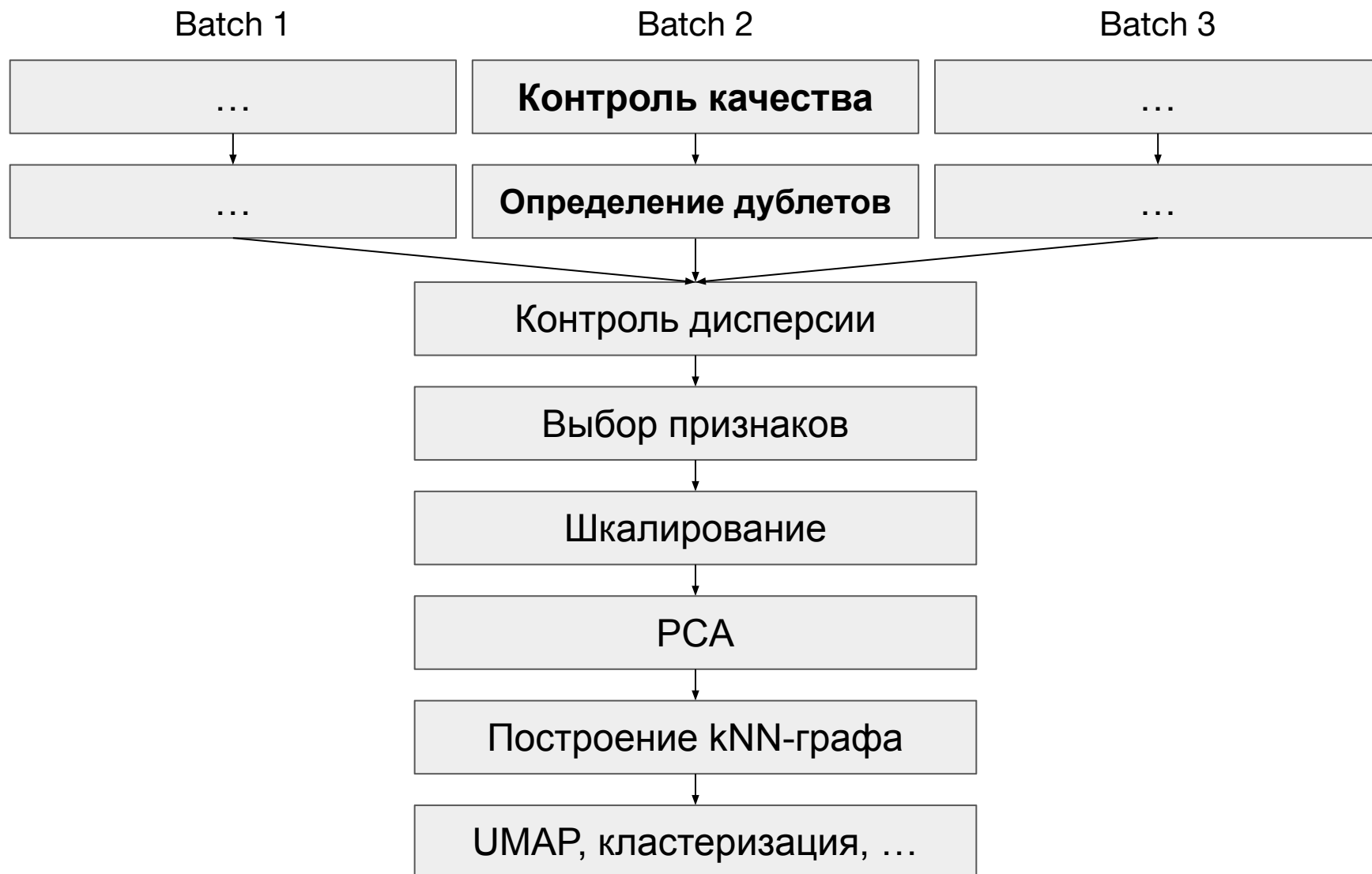
2. Single-cell RNA-Seq:

- a. экспериментальные подходы,
- b. отличия от процессинга bulk RNA-Seq,
- c. методы снижения размерности,
- d. кластера и траектории,**
- e. мультимодальные омики одиночных клеток.

Стандартный процессинг датасета



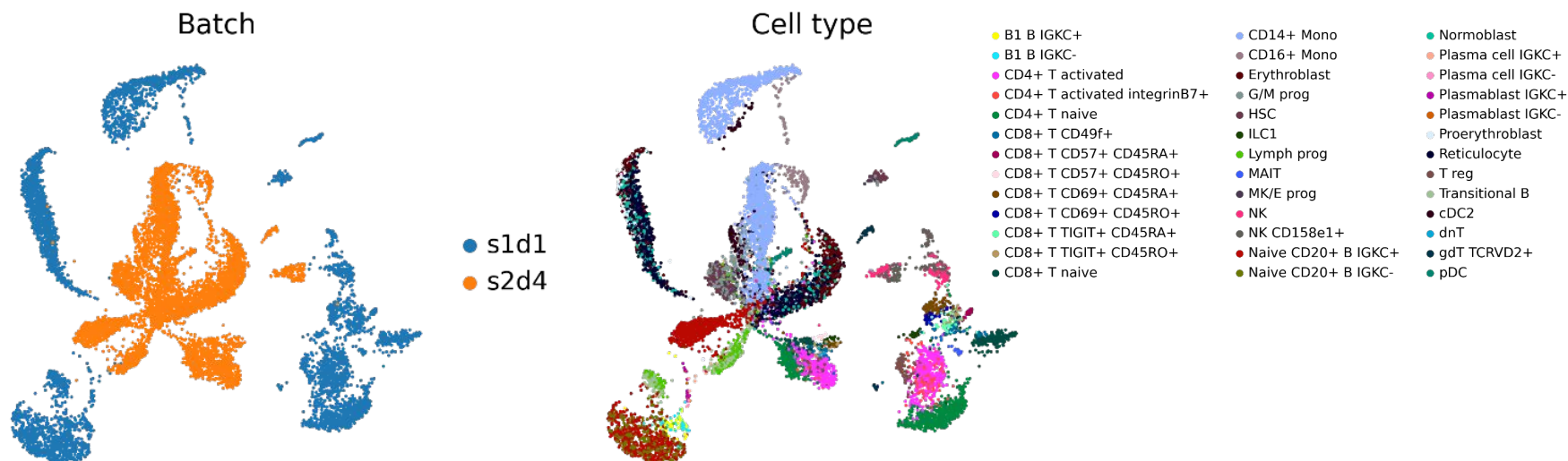
Процессинг нескольких датасетов



Батч-эффект: как он виден?

Батч-эффект проявляется в том, что клетки кластеризуются в первую очередь по принципу того, из какого образца они пришли, и только потом — по типам

Батч-эффект затрудняет дальнейшие кластеризацию, определение типов клеток, реконструкцию траектории и прочее



Причины батч-эффекта

Биологические:

- внутривидовая вариабельность, а именно
 - пол (некоторые ткани особенно чувствительны к разнице между полами — например, надпочечники),
 - болезнь / норма (в болезни могут появляться новые, нетипичные для этой ткани типы клеток и состояния клеток),
 - генетическая изменчивость,
- межвидовая вариабельность;

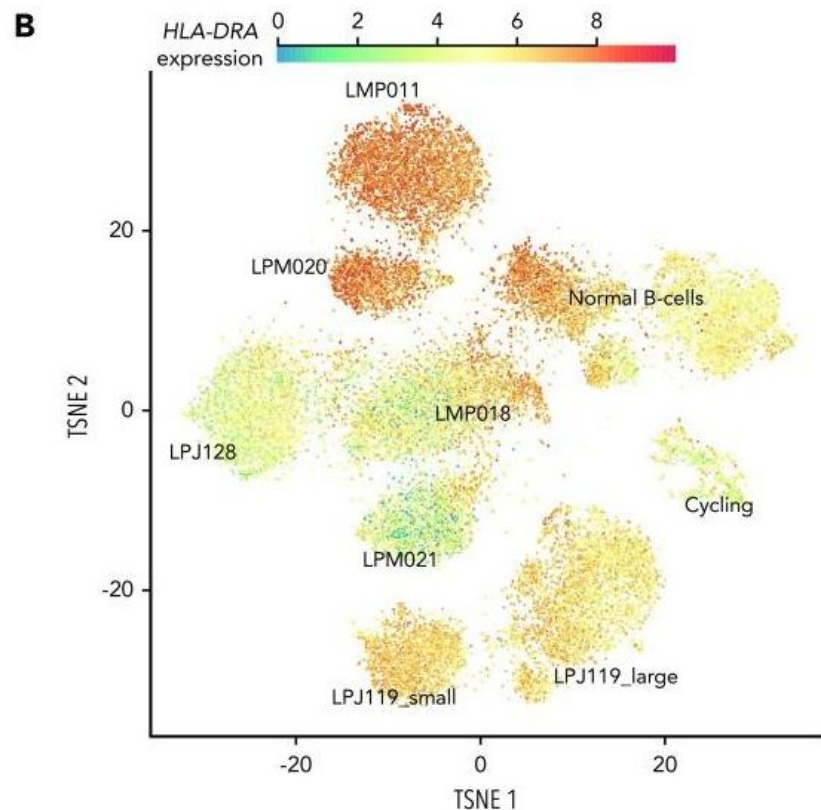
Технические:

- качество подготовки образцов,
- способ подготовки образцов (single cell / single nucleus),
- способ подготовки библиотек,
- ...

Не любая разница — это батч

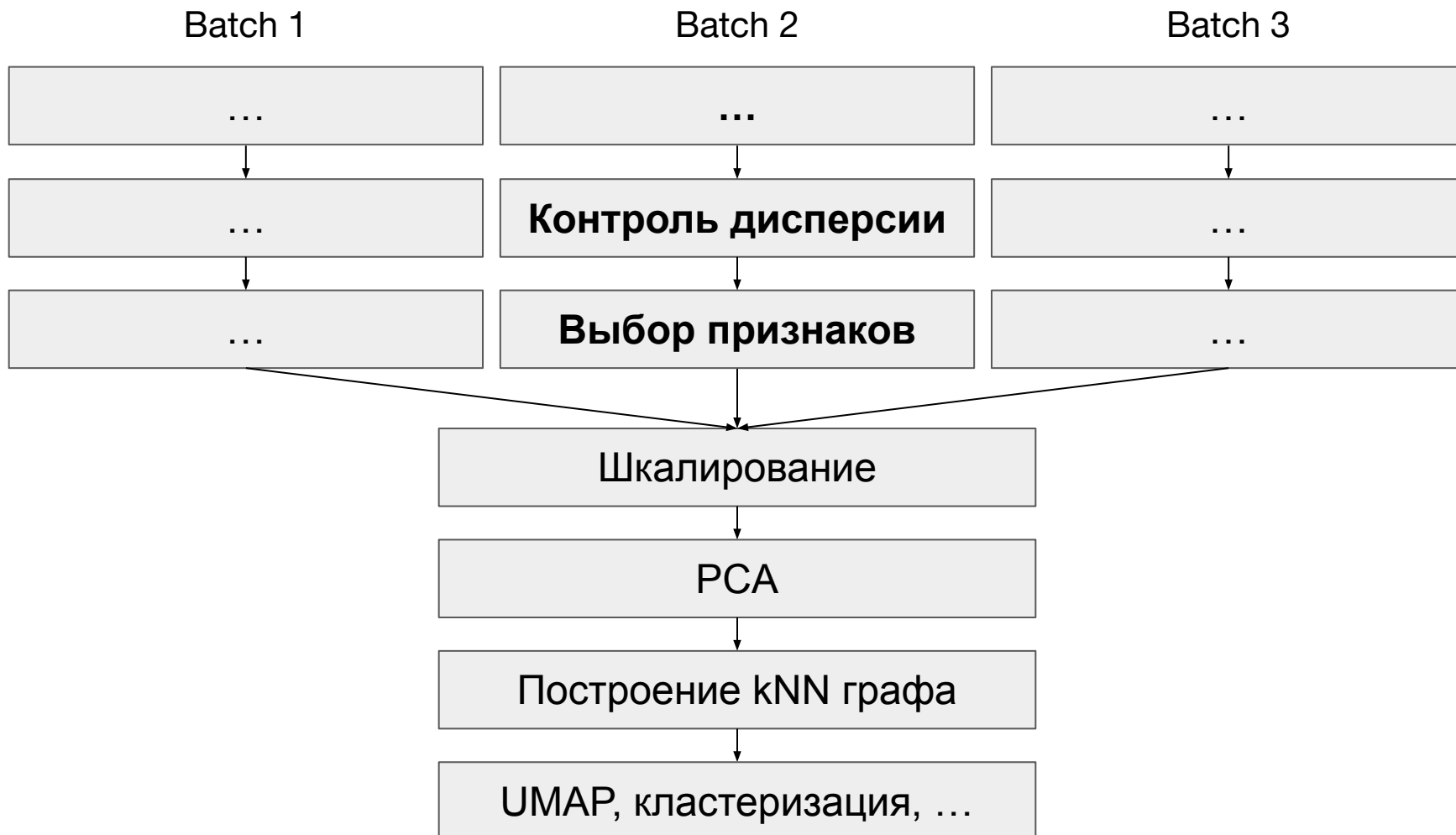
Некоторые типы клеток в целом и не должны быть похожи между различными пациентами — например, в случае раковых клеток

Всегда необходимо чётко осознавать, где вы ожидаете увидеть разницу, а где — нет

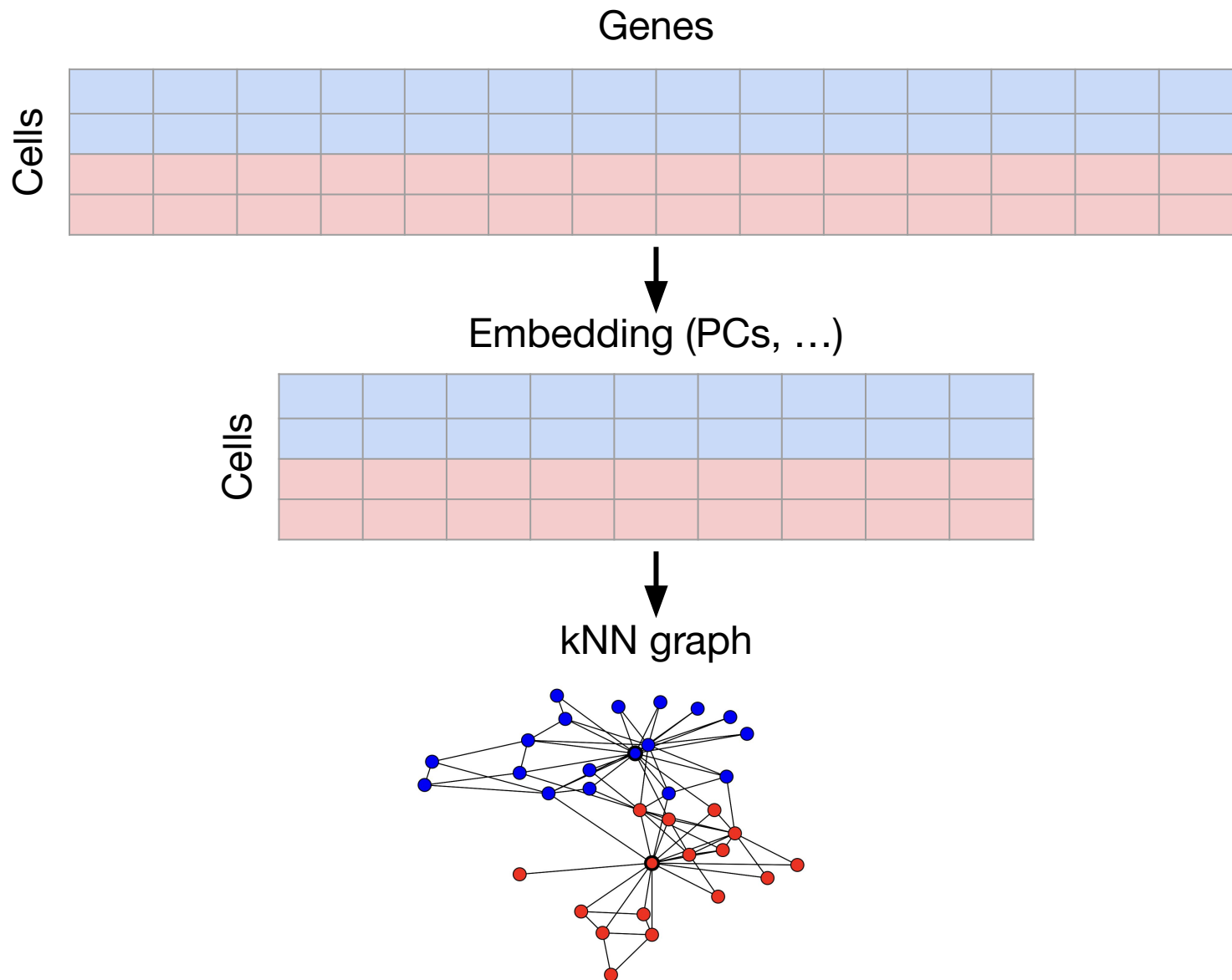


Andor et al., **Blood**, 2019

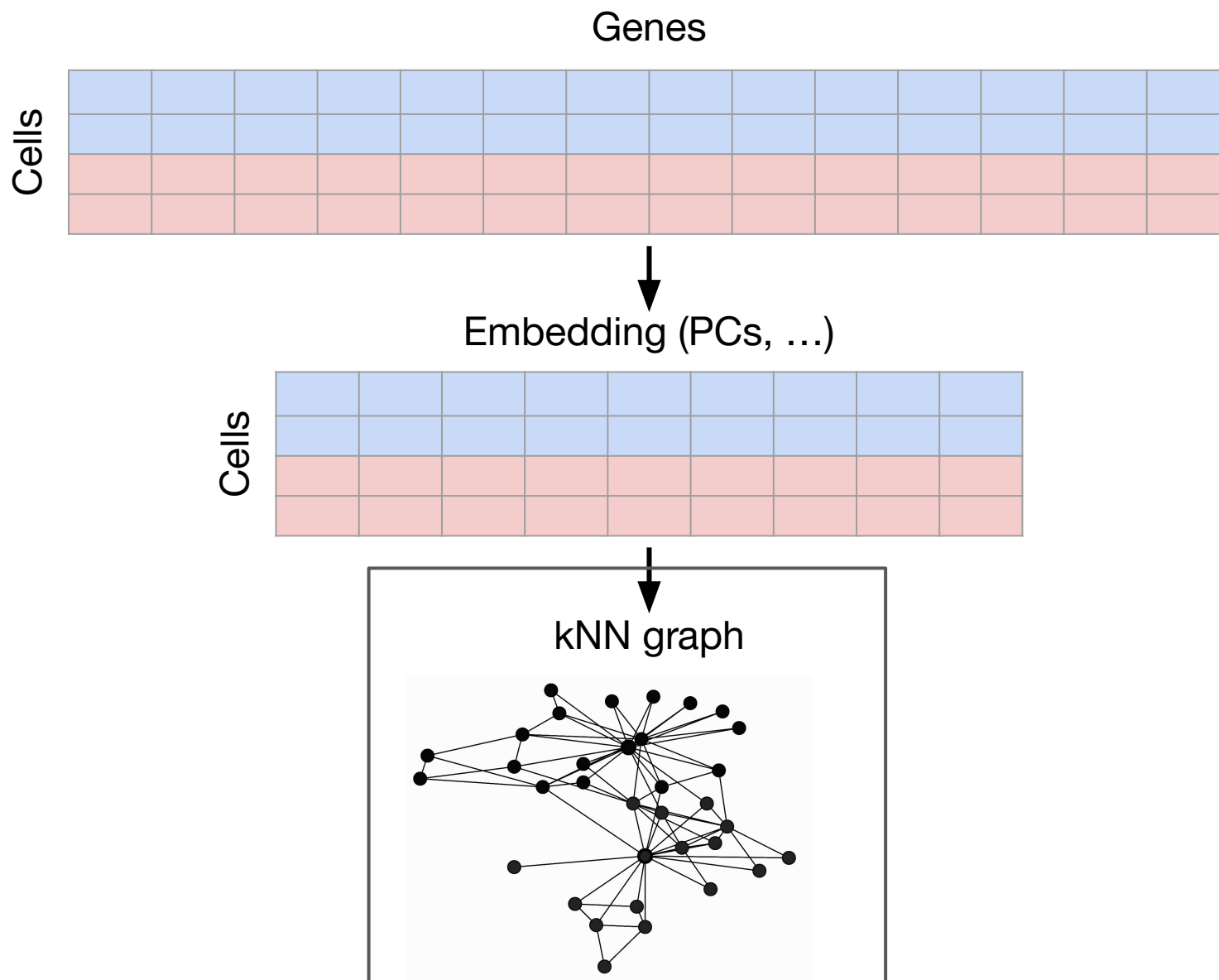
Выделение HVG в каждом из батчей



Принципиальные шаги преобразования данных

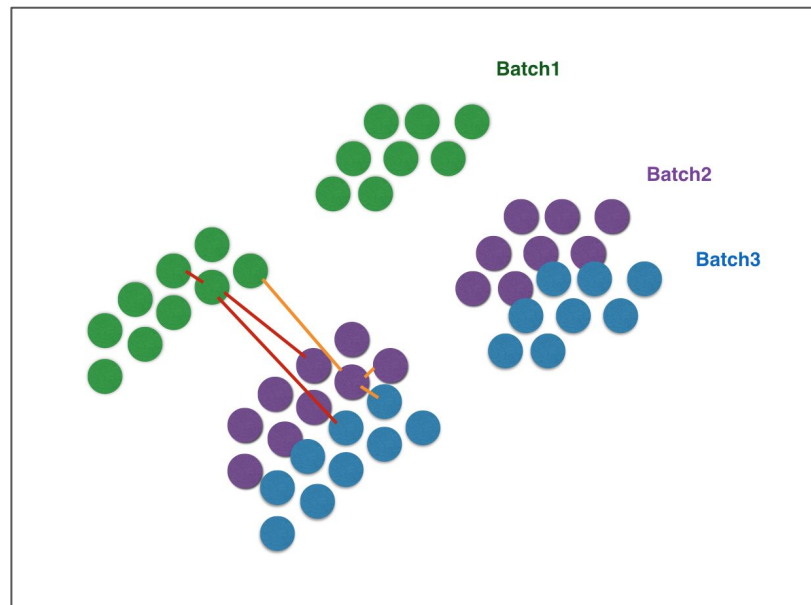
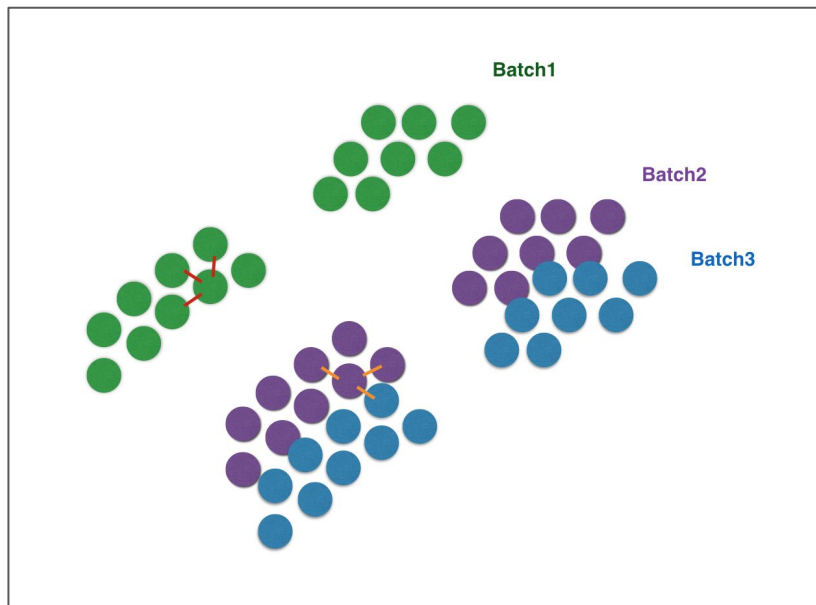


Совместное построение графа (BBKNN, conos)

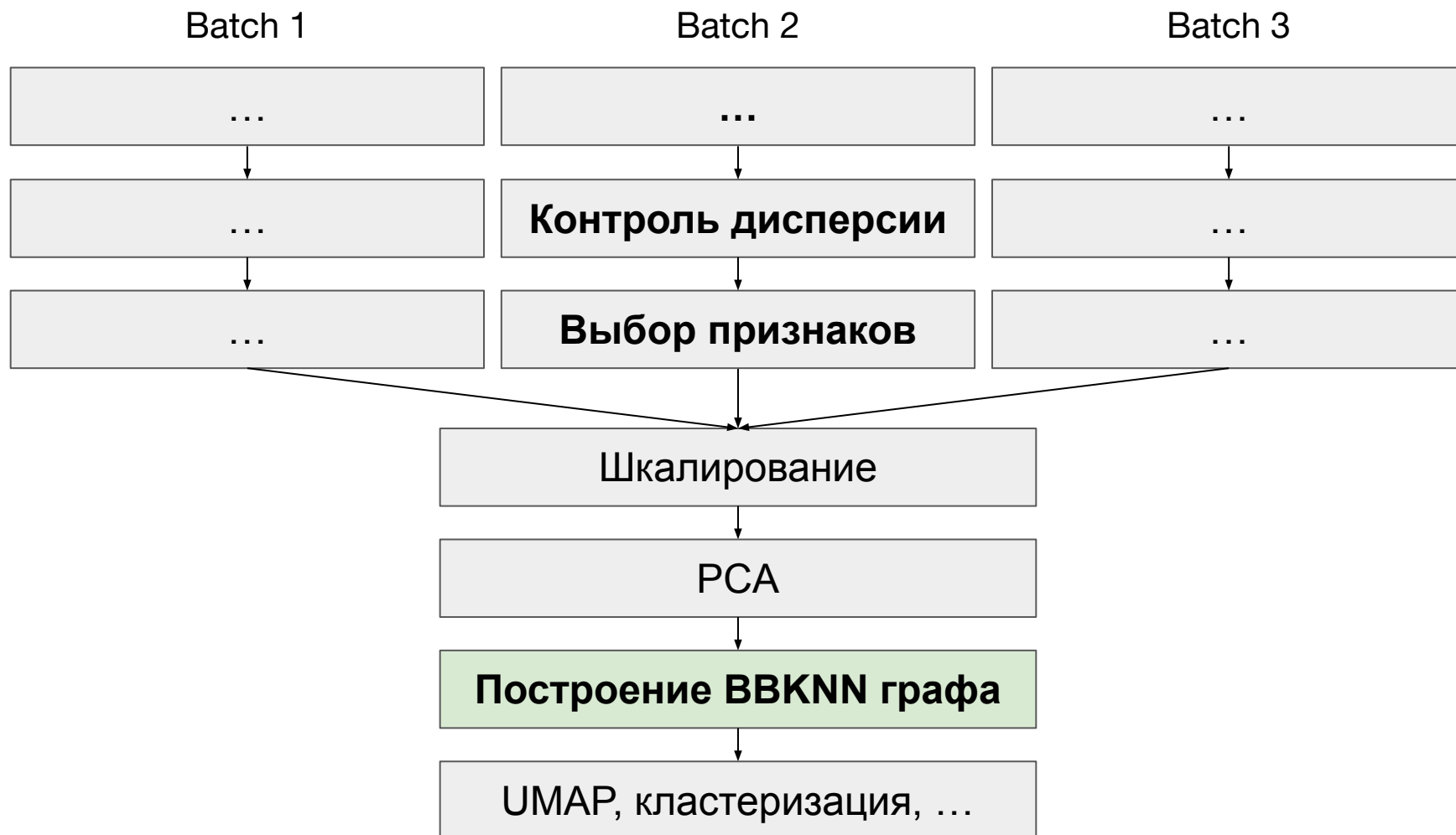


Batch balanced kNN граф

Вершины bbkNN соединяются рёбрами так, чтобы у каждой клетки оказались соседи из всех батчей, после этого граф очищается от False Positive связей при помощи определения связности клеток, соединённых ребром



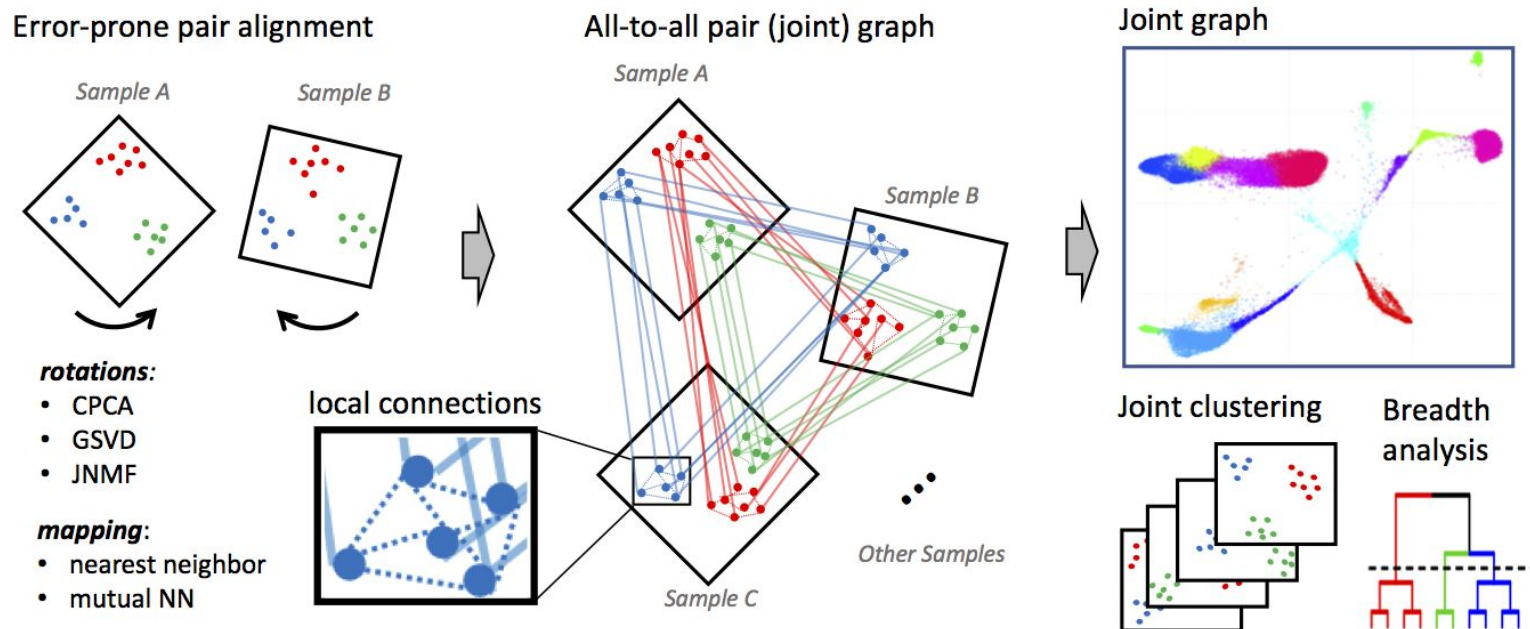
Batch balanced kNN граф



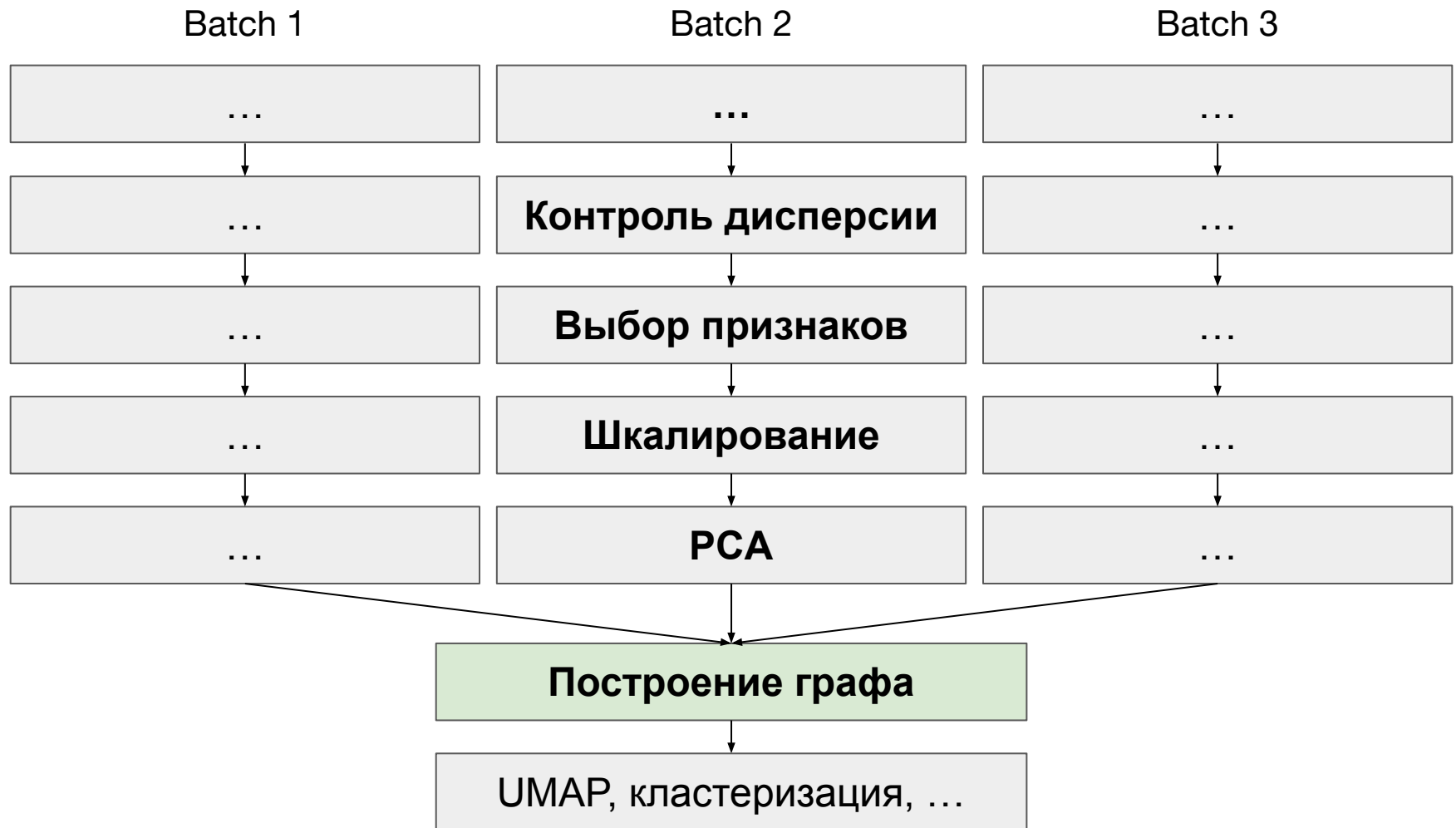
conos

conos попарно сопоставляет клетки из каждой пары батчей, после чего объединяет полученные рёбра и получает итоговый граф

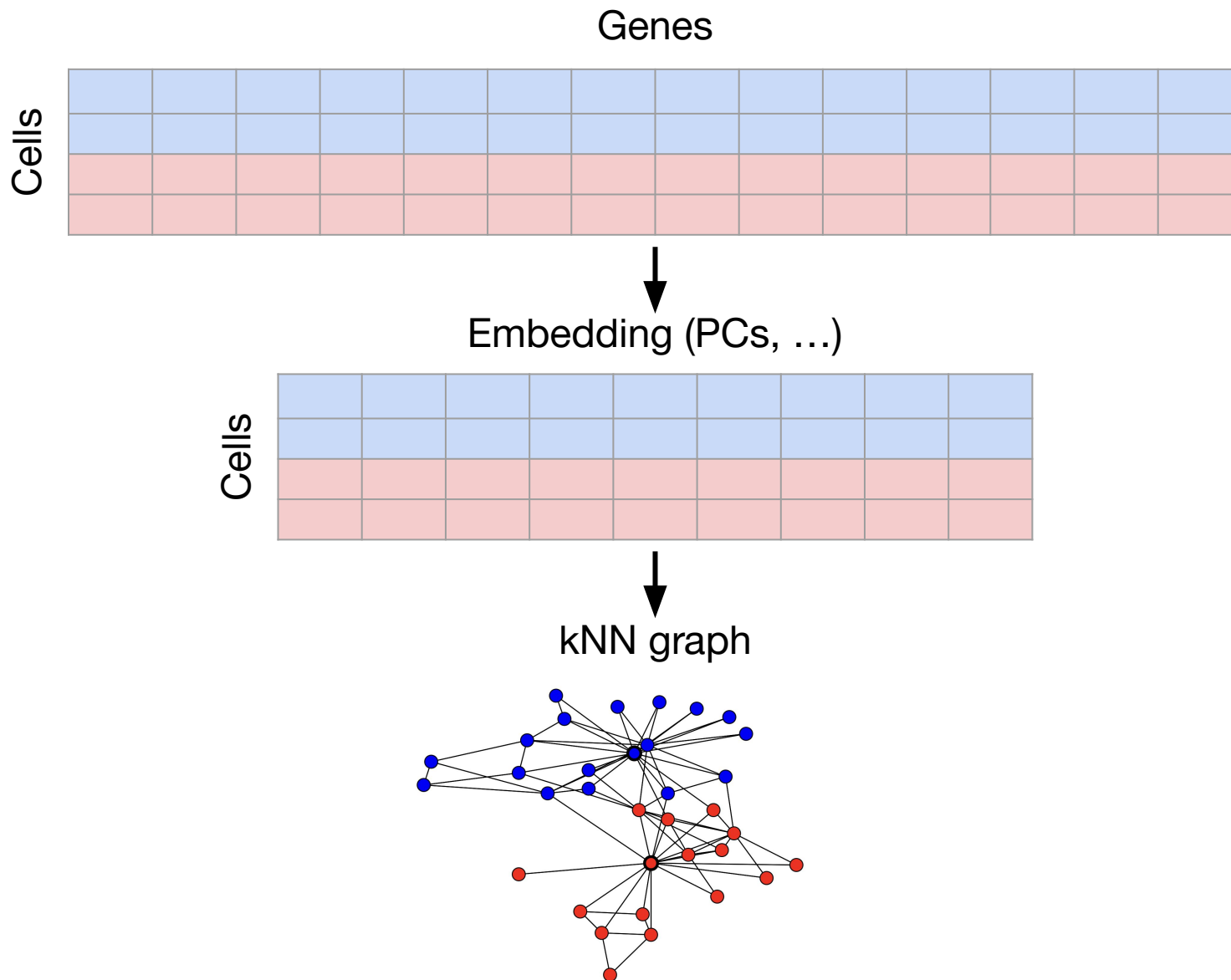
Важно понимать, что граф, который образуется в результате, — это не kNN!



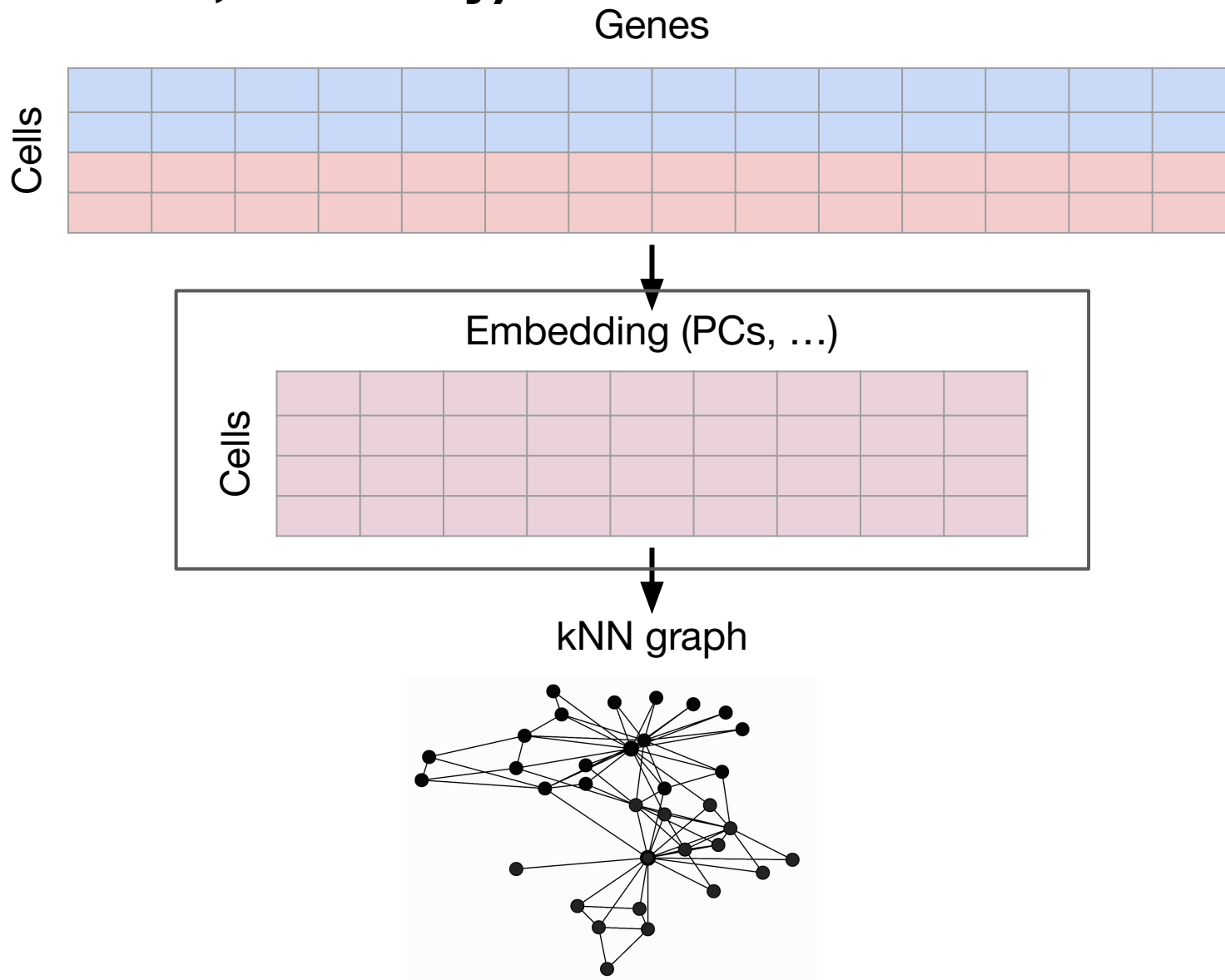
conos



Принципиальные шаги преобразования данных



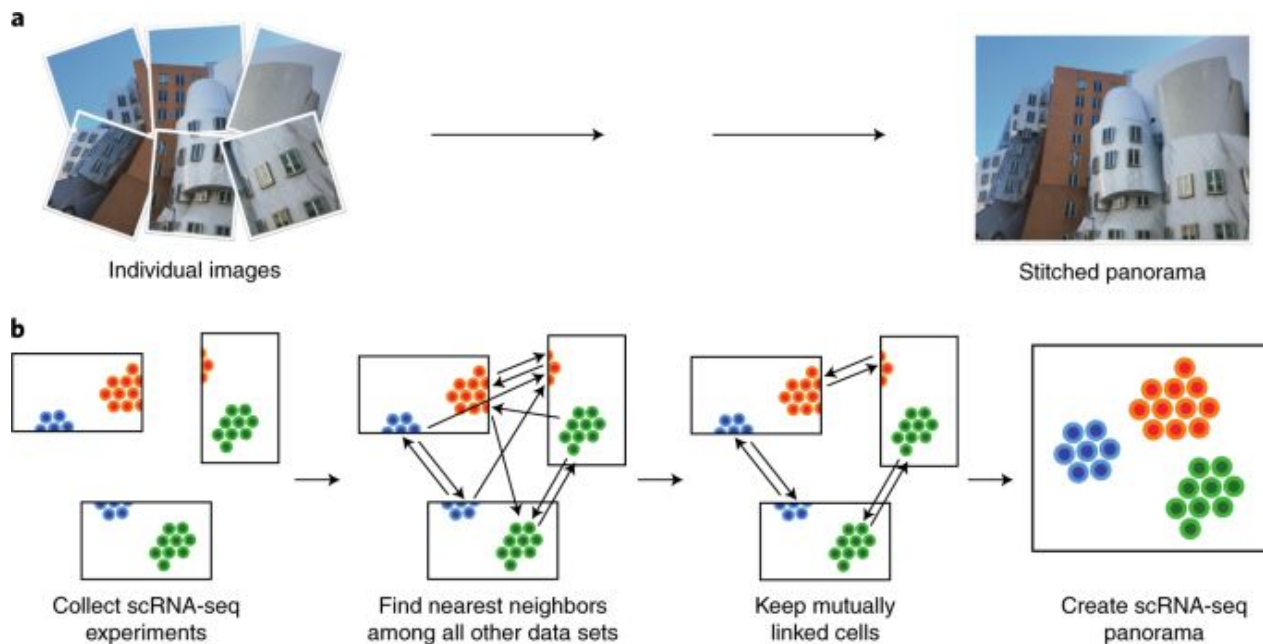
Батч-скорректированное представление (Scanorama, Harmony)



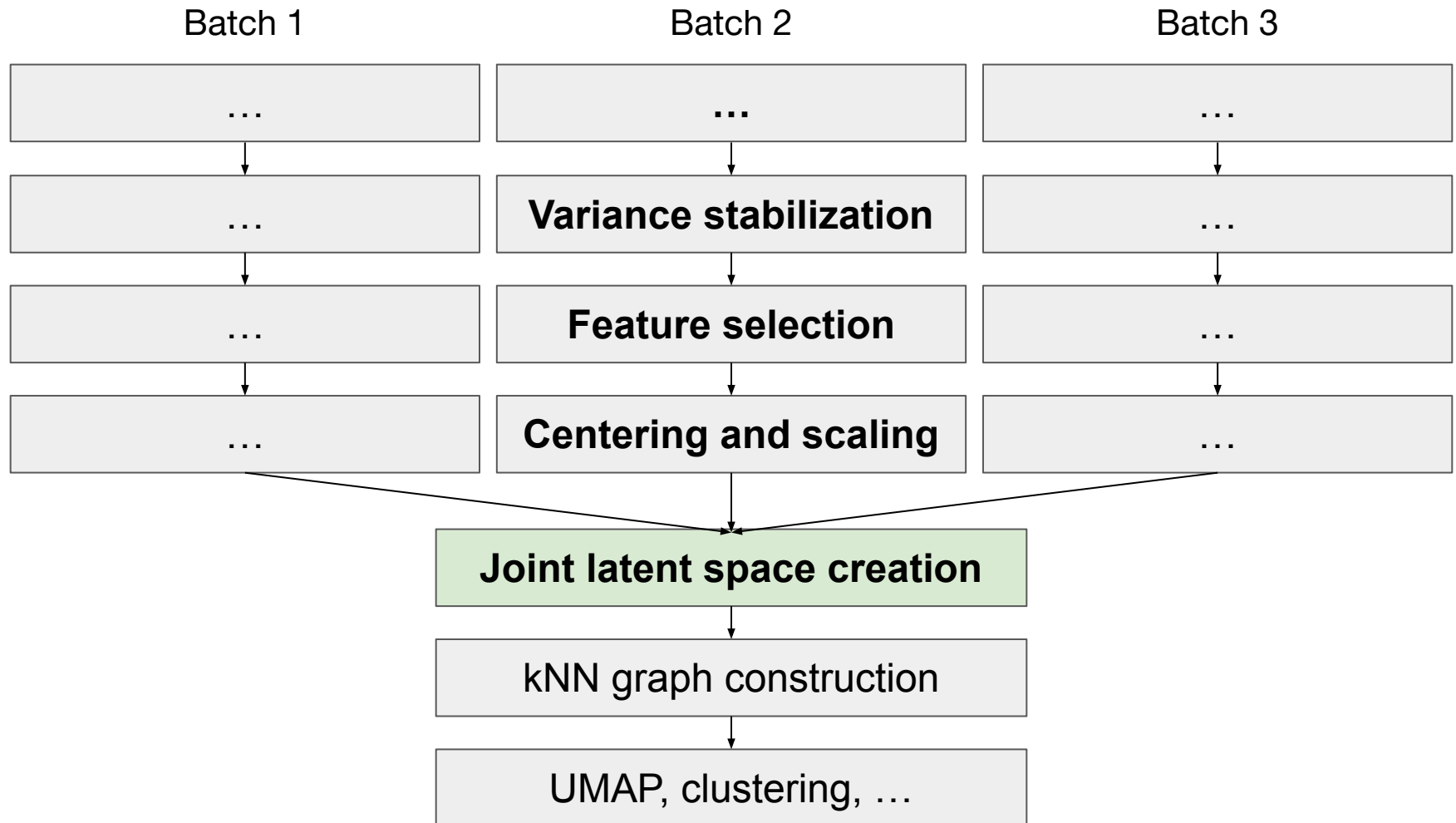
Scanorama

Умеет устранять батч как на уровне эмбединга, так и на уровне экспрессий

Основан на алгоритмах соединения различных частей изображений в единую целую

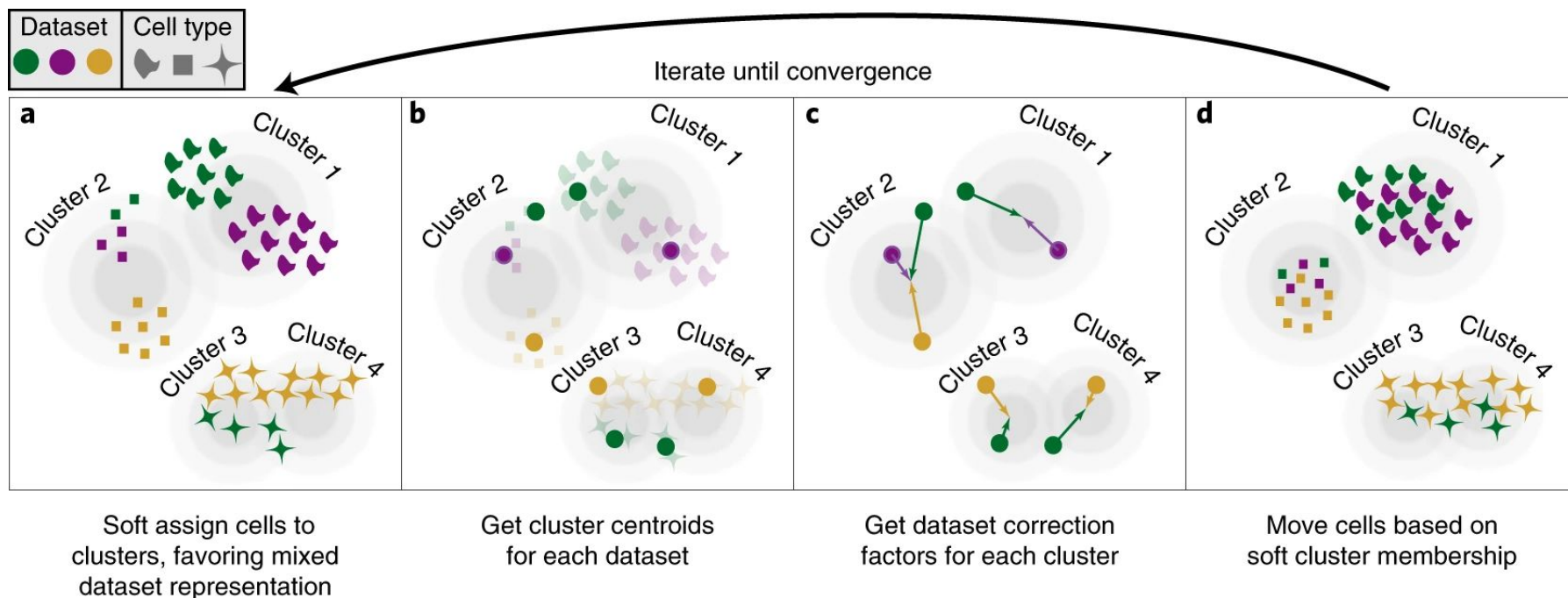


Scanorama

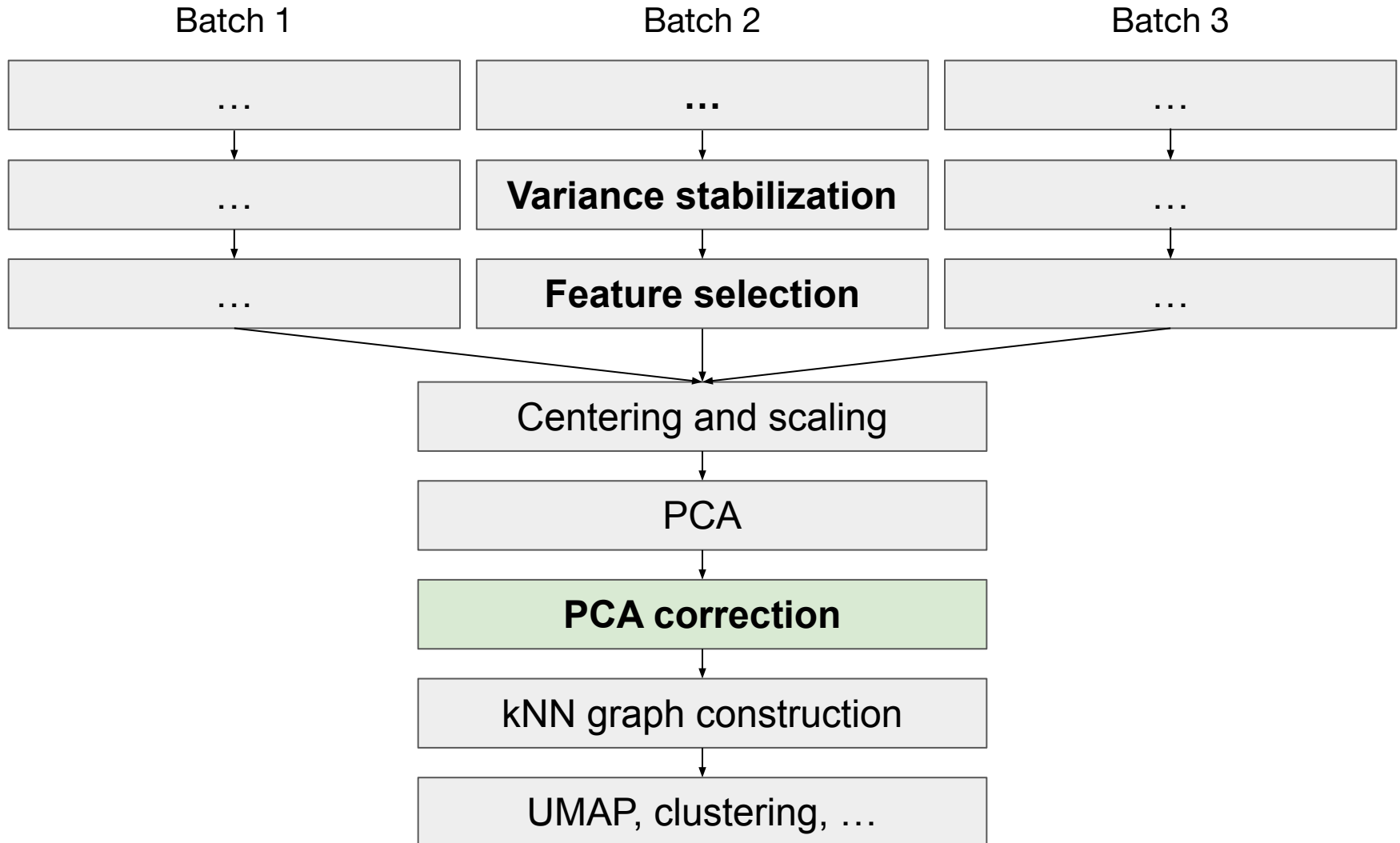


Harmony

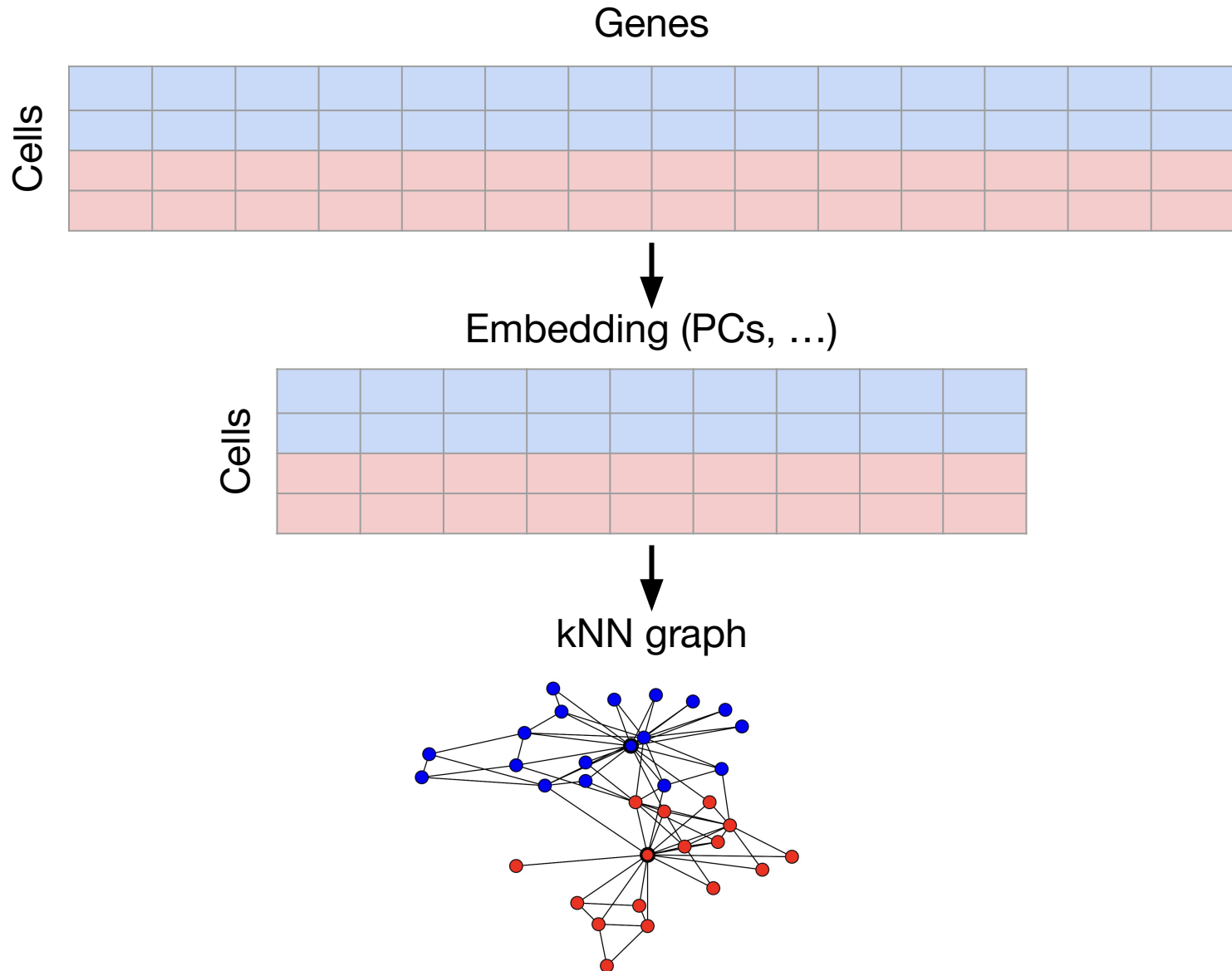
Работает быстро и эффективно; имплементирован как в Python, так и в R



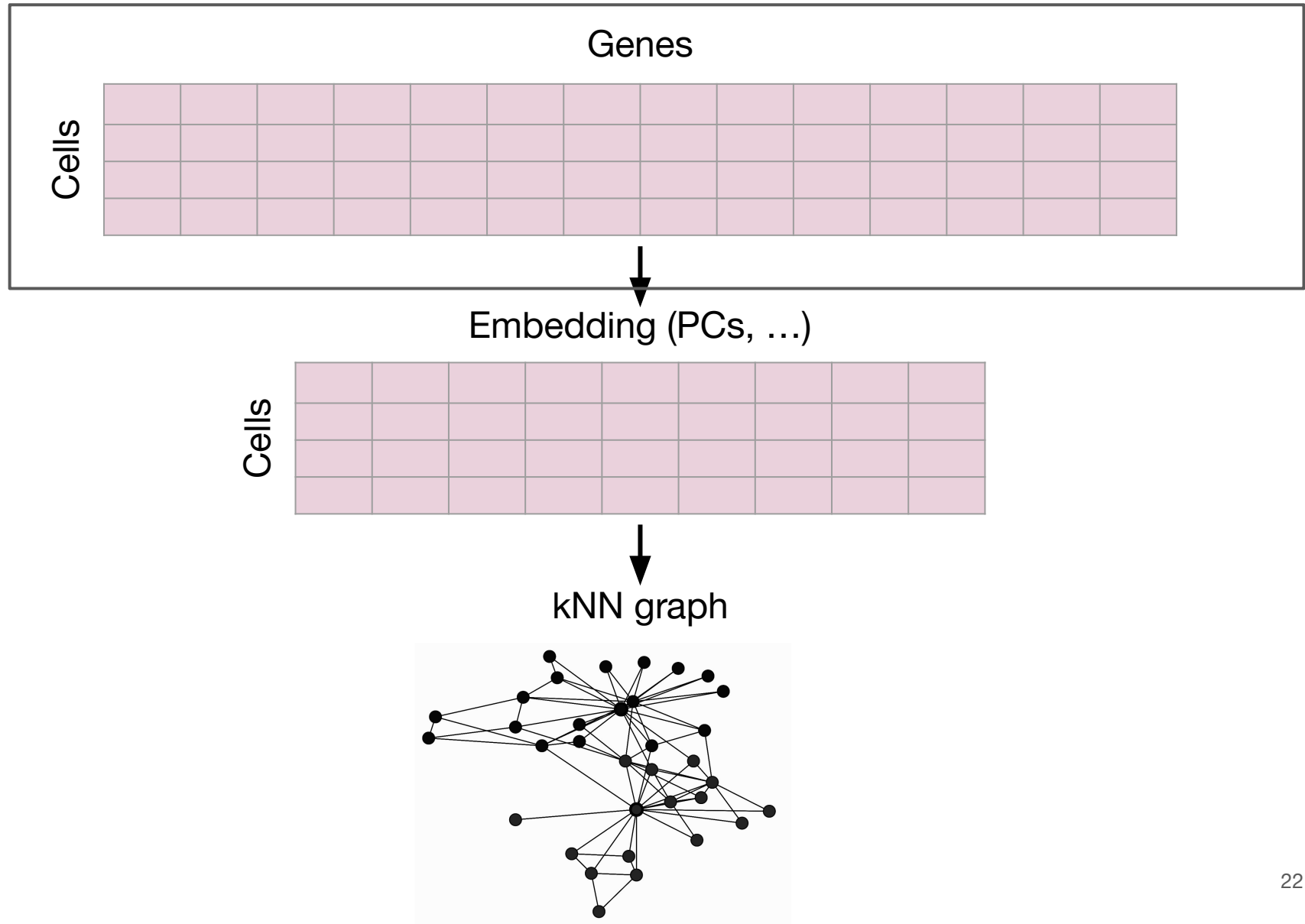
Harmony



Principal steps of the data transformation

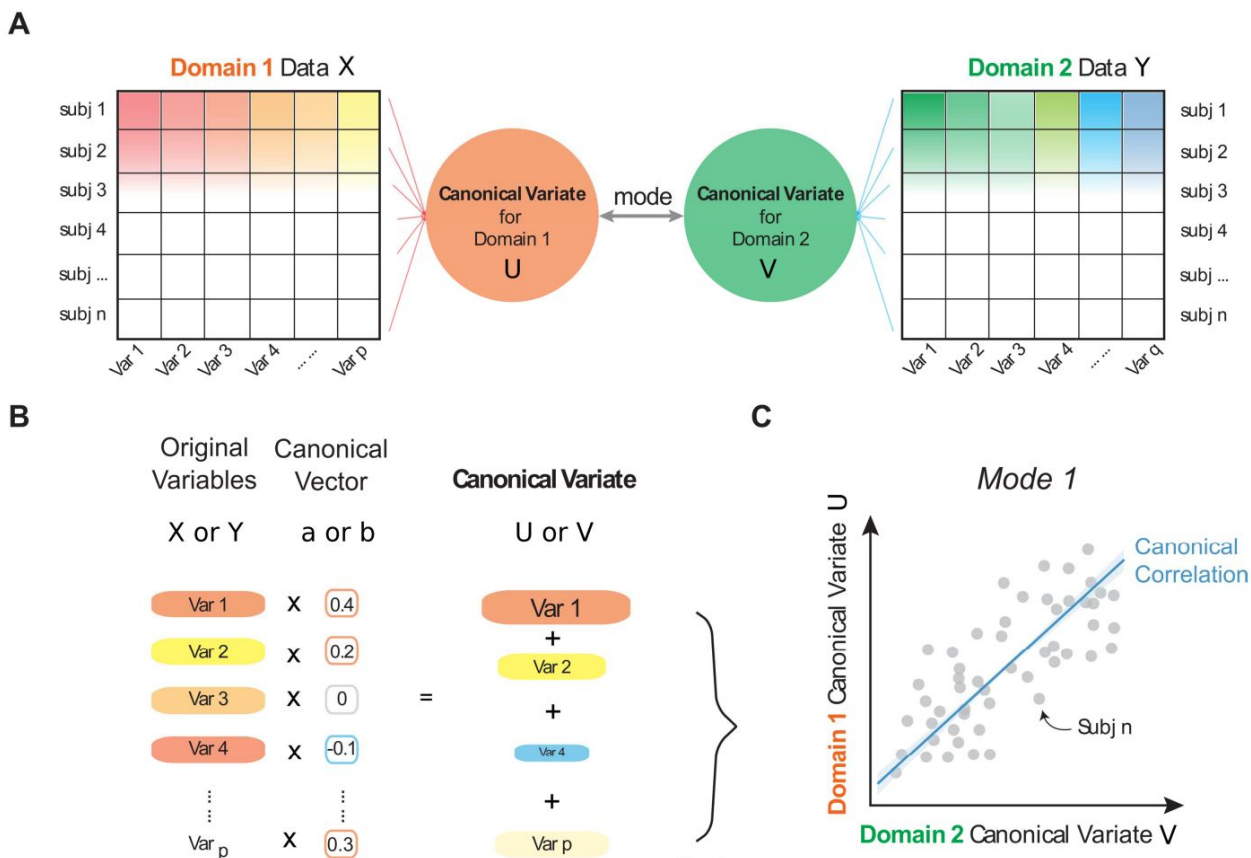


Expression correction (Seurat)

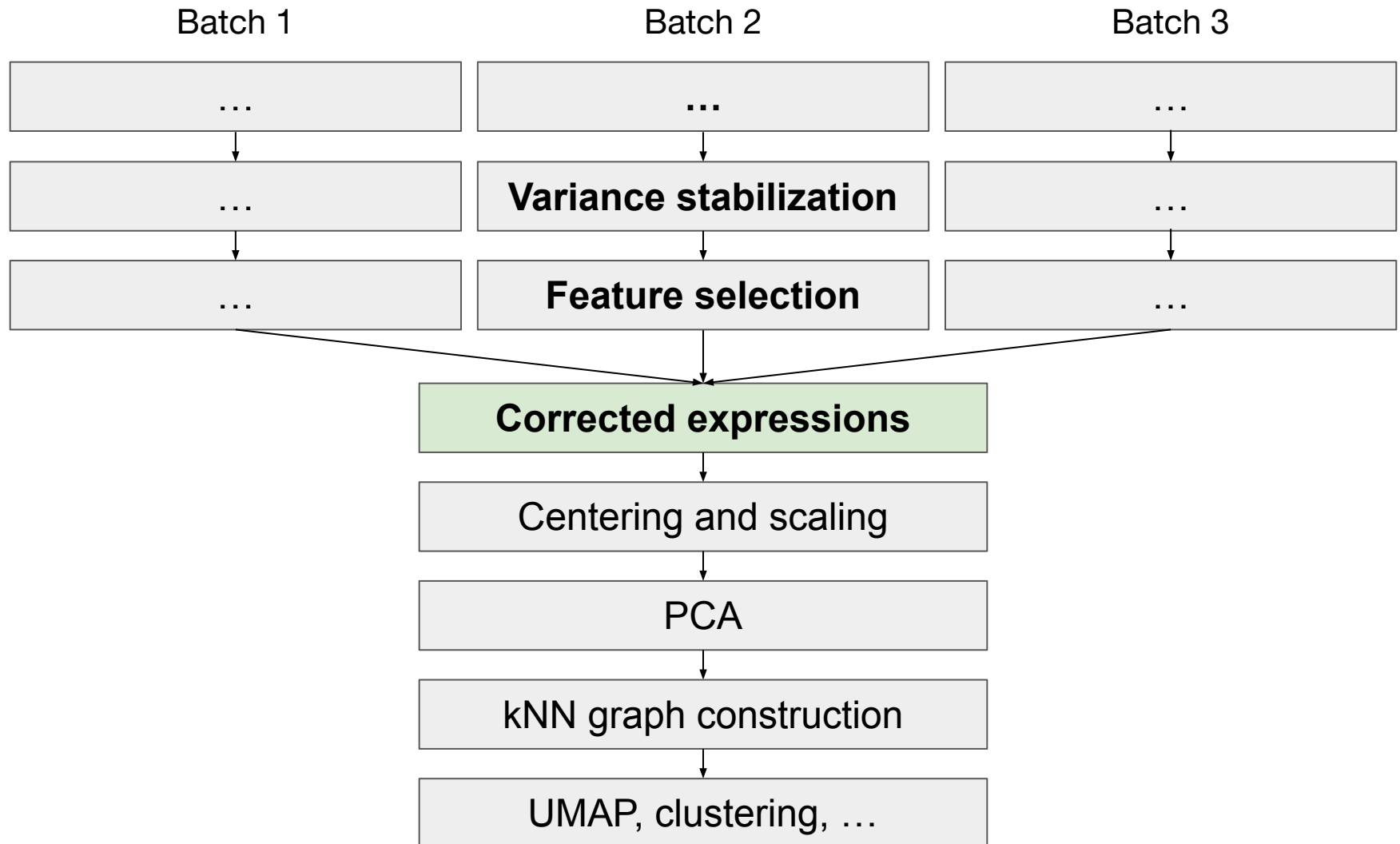


Seurat CCA

Устраняет батч-эффект достаточно эффективно, но требует очень много оперативной памяти и времени



Seurat



Сравнение существующих методов



OPEN

Benchmarking atlas-level data integration in single-cell genomics

Malte D. Luecken¹, M. Büttner¹, K. Chaichoompu¹, A. Danese¹, M. Interlandi², M. F. Mueller¹, D. C. Strobl¹, L. Zappia^{1,3}, M. Dugas⁴, M. Colomé-Tatché^{1,5,6} and Fabian J. Theis^{1,3,5}

Не так давно вышла статья, в которой провели сравнительный анализ нескольких десятков методов батч-коррекции, и результаты там показали, что лучше всего с задачей справляется метод **scanorama**

Но(!) у этого бенчмаркинга есть ряд проблем

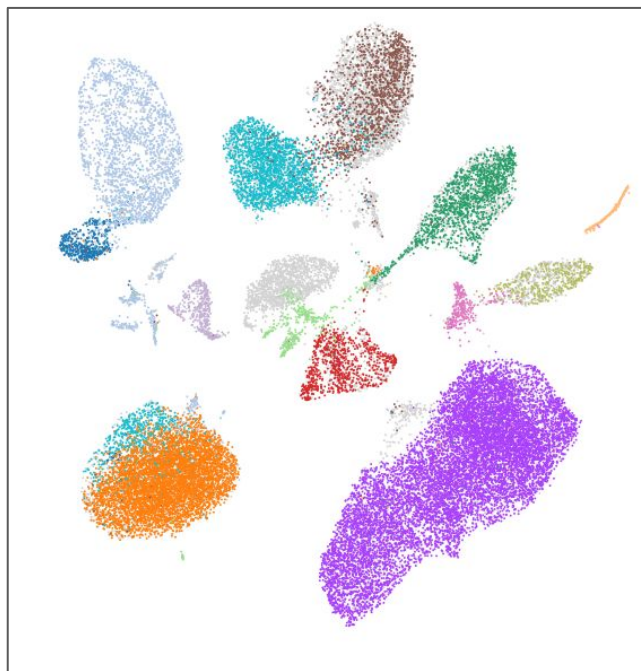
Как выбрать метод батч-коррекции?

1. Ориентируйтесь на задачу исследования. Чем вы готовы пожертвовать, а что бы вам лучше сохранить?
2. Используйте тот метод, с которым вы хорошо знакомы и которым вы хорошо умеете пользоваться
3. Если ваш метод батч-коррекции выдаёт что-то очень неправдоподобное, то используйте другой метод

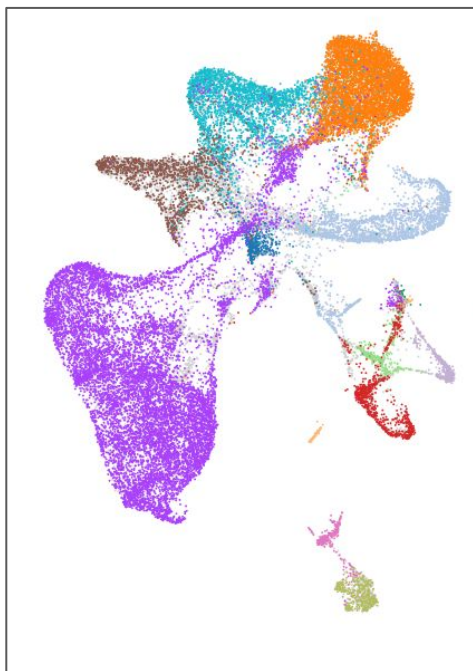
На практике обычно в самом начале используют несколько методов, после чего сравнивают результаты, которые ими получаются

Пример из жизни: мезенхима тимуса

Harmony



Seurat CCA



Conos

