Single Cell Sequening of the transcriptome

Joseph Powell

Computational and Single Cell Genomics, Institute for Molecular Bioscience

---

Summer Institute in Statistical Genetics

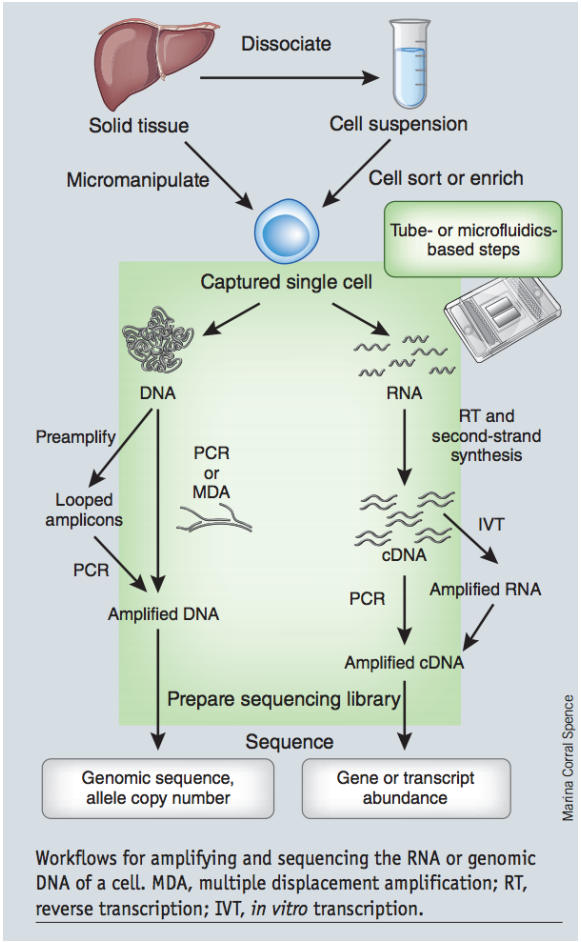
Brisbane, 14th Feb 2016

What is it?

scRNA-SEQ

Single cell sequencing examines the sequence information from individual cells with optimized NGS technologies, providing a higher resolution of cellular differences and a better understanding of the function of an individual cell in the context of its microenvironment

Overview



Workflows for amplifying and sequencing the RNA or genomic DNA of a cell. MDA, multiple displacement amplification; RT, reverse transcription; IVT, *in vitro* transcription.

Local context of the cell

Picking the right cell(s)

- Micromanipulation - precise but laborious way to target a single cell
- Microcapillaries - extract cell contents directly
- Tissue dissociated - to produce cell suspensions, which are easier to handle and allow cells expressing specific markers to be enriched with a cell sorter
- Instruments that trap very rare cells by their surface markers can also be used
- Flow sorting - based on markers or viability
- Direct from either suspended cell culture or primary cell-free tissue

PolyA vs non-PolyA

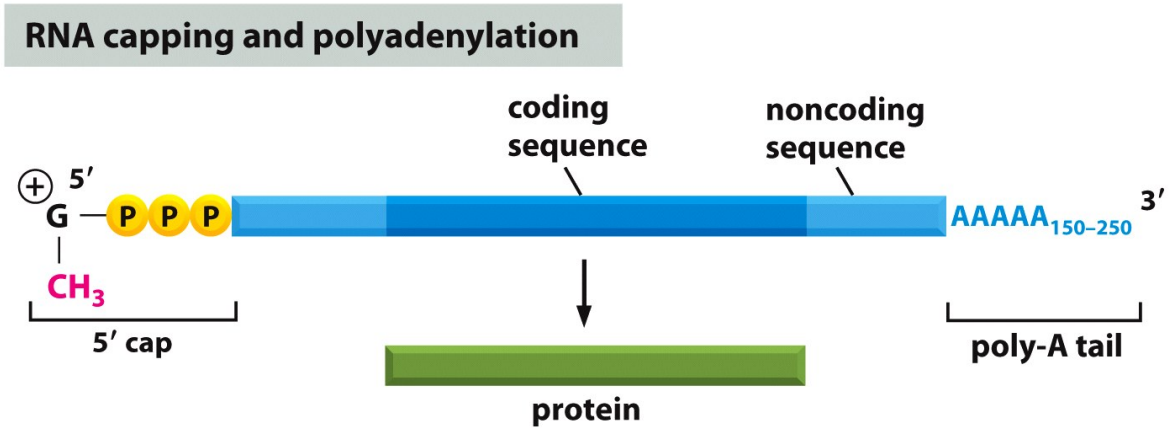


Figure 7-16a Essential Cell Biology 3/e (© Garland Science 2010)

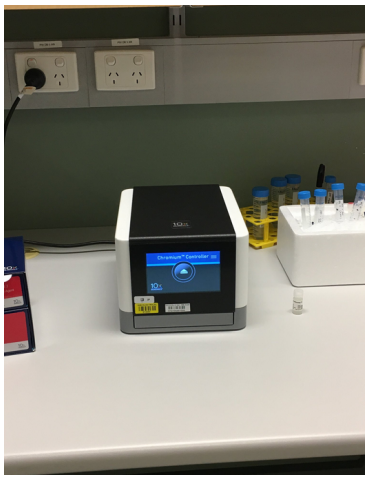
Poly(A)

Polyadenylation is the addition of a poly(A) tail to mRNA. The poly(A) tail consists of multiple adenosine monophosphates; in other words, it is a stretch of RNA that has only adenine bases. It is part of the process that produces mature messenger RNA (mRNA) for translation.

Focus on high-throughput methods

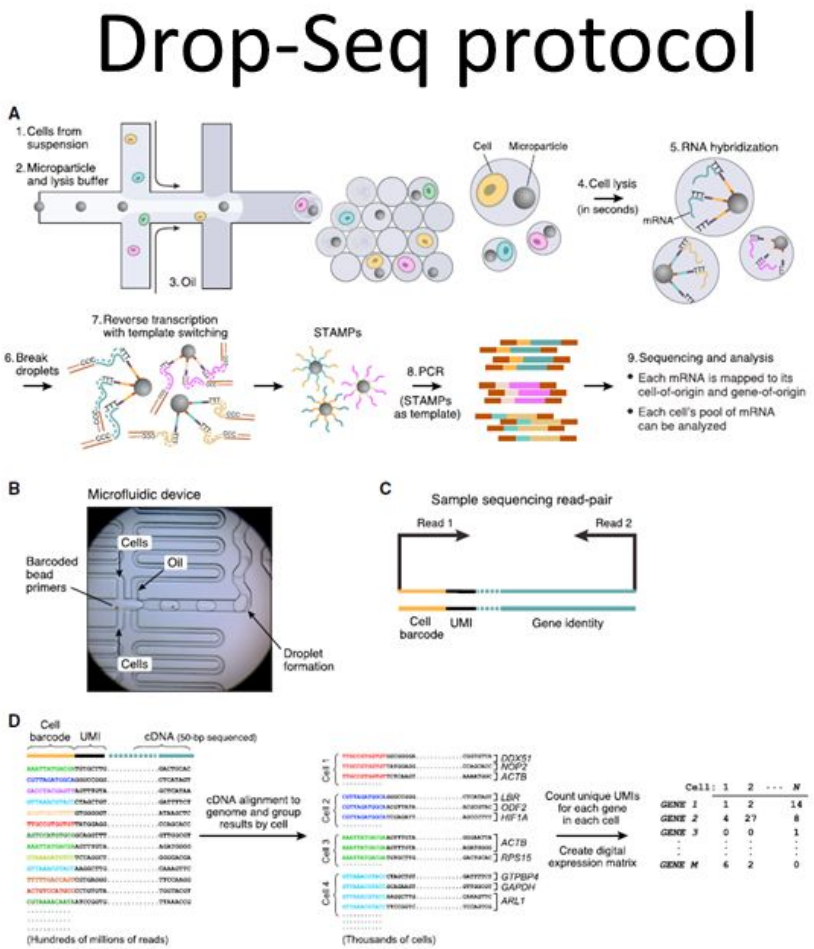
DropSeq

Chromium



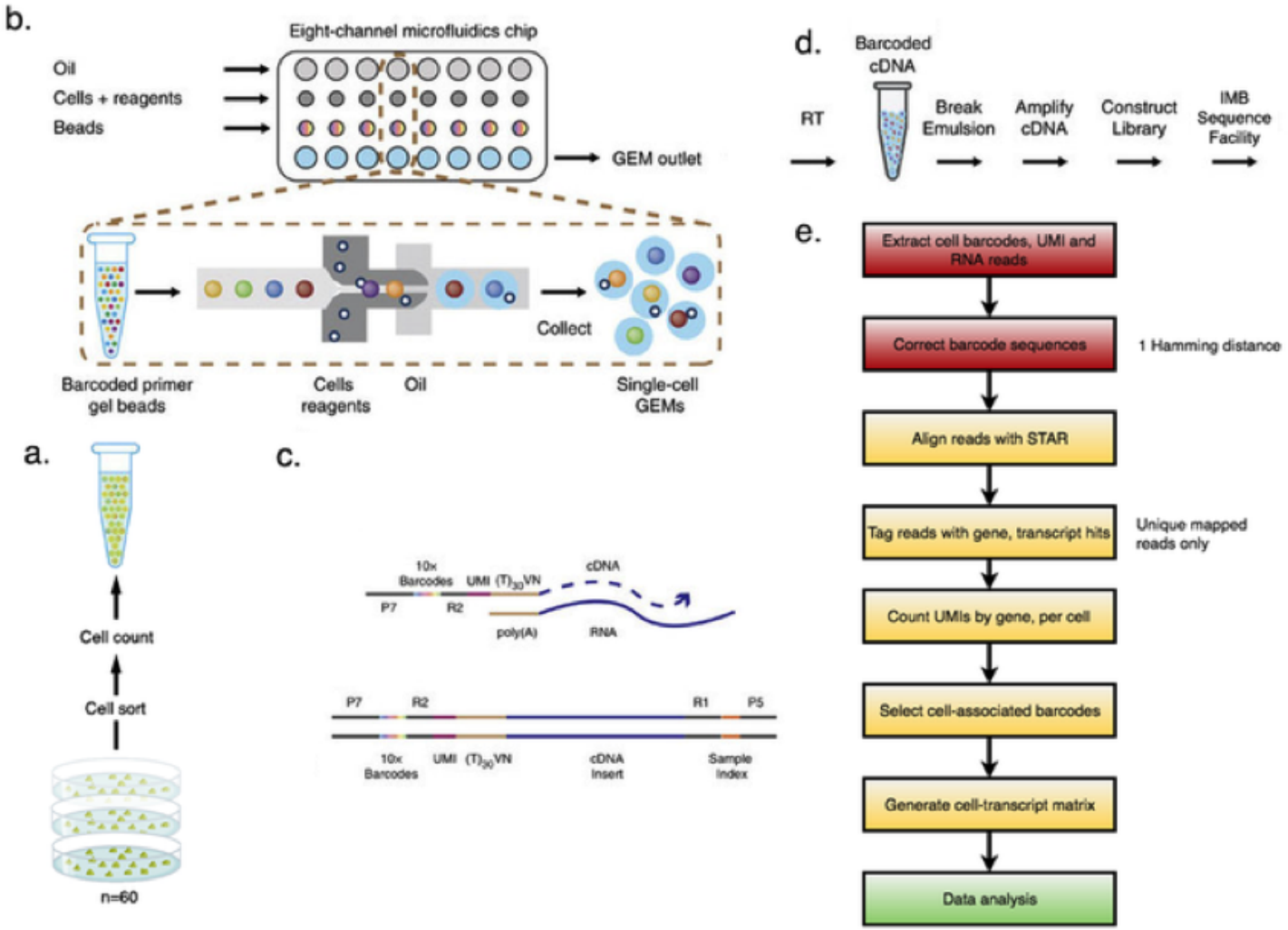
C1

DropSeq - Macosko et al. Cell, 2015



Macosko et al, 2015. http://dx.doi.org/10.1016/j.cell.2015.05.002

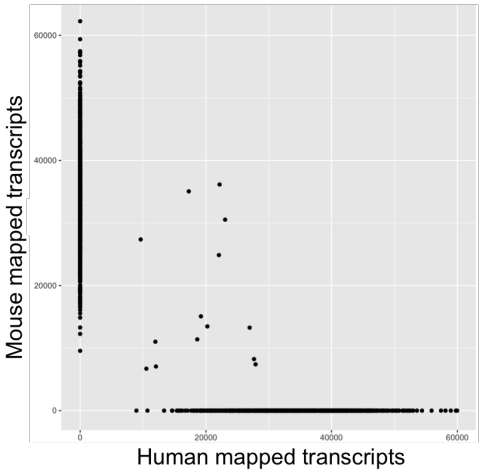
Chromium from 10X Genomics



Fluidigm C1


Enrich
Load & Capture
Wash & Stain
Isolate
Lyse, RT & Amplify
Prepare Library
Sequence
Analyze
C1 Single-Cell Auto Prep System
Any Illumina System

Doublet rate

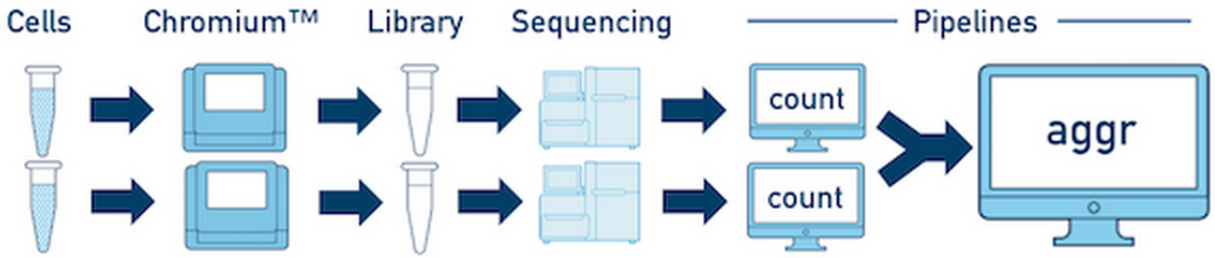


Multiple cells encapsulated in one droplet

Can estimated using a mix of house and human cells

Alignment, barcode assignment and UMI counting

- Demultiplexing - first base on the (8) bp sample index
- Demultiplexing - the on the 14bp GemCode barcode
- Extract the Unique Molecular Identifiers
- Remove and filter on quailty of barcode and UMI
- PCR duplicates marked if two sets of read pairs share a barcode, UMI tag, or gene ID



Bioinformatics

Normalisation

One of the first issues is the need for normalization due to amplification biases introduced by scarce amounts of starting RNA. Ideally, this should be addressed using a combination of experimental and computational methods. Often this is not practical.

Prominent features in single-cell RNA-seq data relative to bulk RNA-seq include an abundance of zeros, increased variability, and multi-modal expression distributions.

Pooling across cells to normalize single-cell RNA sequencing data with many zero counts

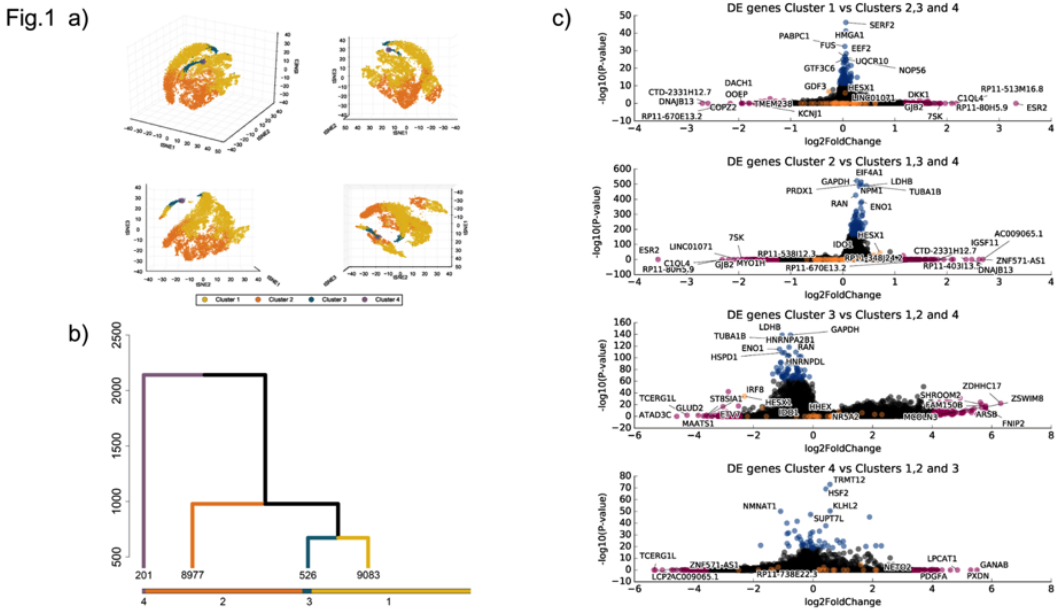
Aaron T. L. Lun ✉, Karsten Bach and John C. Marioni ✉

Genome Biology 2016 17:75 | DOI: 10.1186/s13059-016-0947-7 | © Lun et al. 2016
Received: 3 February 2016 | Accepted: 11 April 2016 | Published: 27 April 2016

Analysis

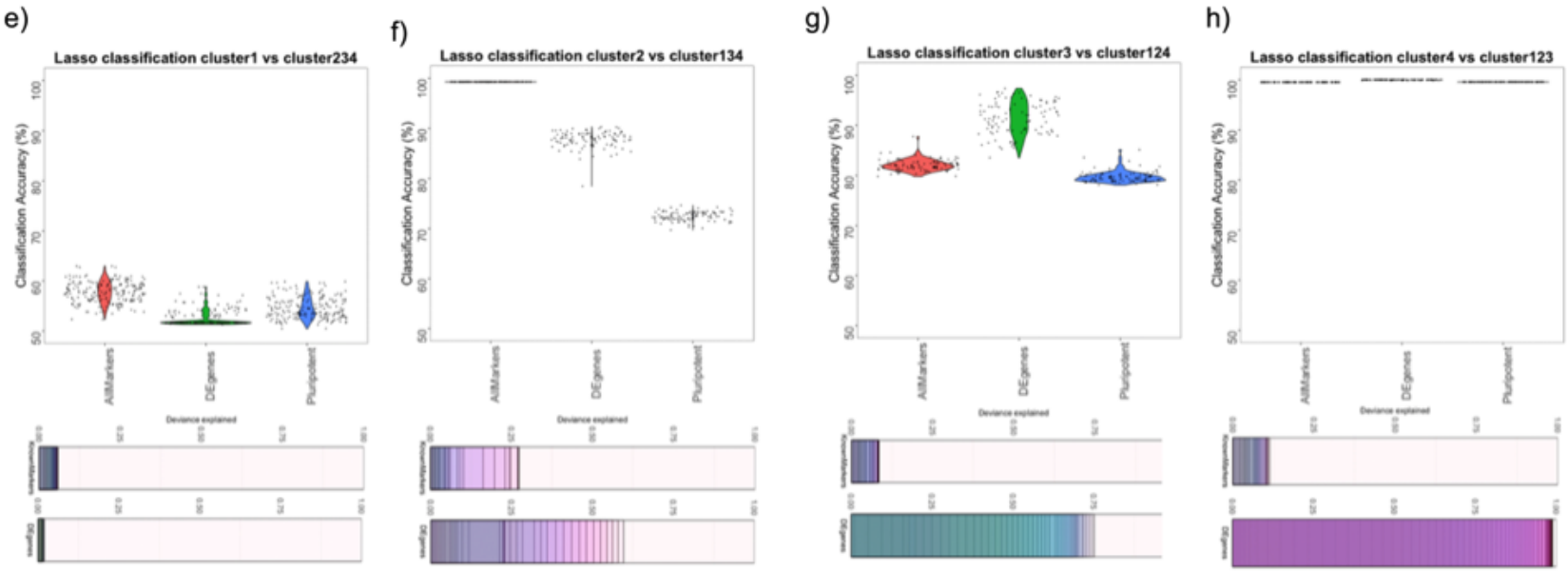
Deconvoluting heterogeneous cell populations

- PCA and t-distributed Stochastic Neighbor Embedding (tSNE) analysis
- Clustering
- Differential expression



Analysis

Building predictors of subpopulations of cells



Analysis

PseudoTime or lineage tracing - Trapnell et al. Nature Biotech 2014