A primer to Single Cell Transcriptomics

Bart Deplancke

credit: Sanger Institute

Dr. Petra Schwalie   Dr. Vincent Gardeux

One genome: diverse functional outputs
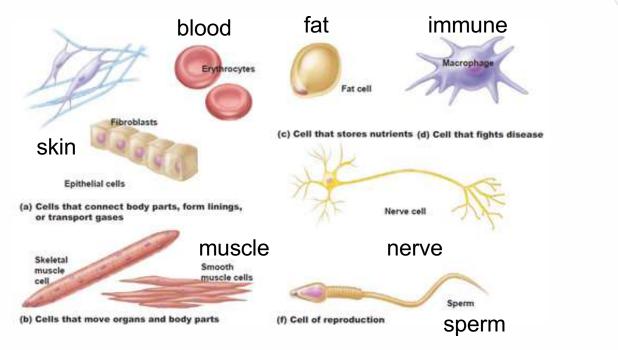
One genome gives rise to a multitude of different cell types with highly distinct morphologies & functions

~ 1 genome

Remarkable Cellular Diversity and Specialisation

Human body:
~ 100 trillion cells
~ 200 types of cells
complex
tissue & organ
functions

blood

Erythrocytes

fat

Fat cell

immune

Macrophage

skin

Fibroblasts

Epithelial cells

(a) Cells that connect body parts, form linings, or transport gases

(c) Cell that stores nutrients   (d) Cell that fights disease

Nerve cell

muscle

Skeletal muscle cell

Smooth muscle cells

nerve

(b) Cells that move organs and body parts

(f) Cell of reproduction

Sperm

sperm

anatomyandphysiology.com

One genome: diverse functional outputs

~ 1 genome

One genome gives rise to a multitude of different cell types with highly distinct morphologies & functions

Remarkable
Cellular Diversity
and Specialisation

Human body:
~ 100 trillion cells
~ 200 types of cells
complex
tissue & organ
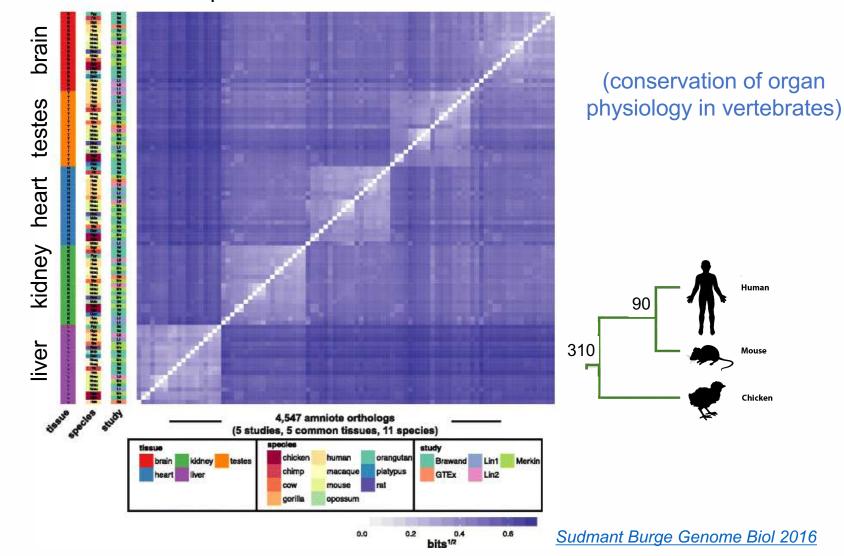functions

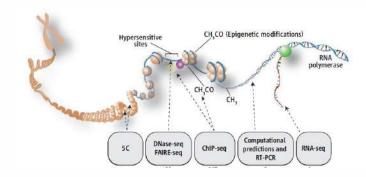Chan Zuckerberg Biohub ($600 M Initiative)

The Cell Atlas

One genome: diverse functional outputs

Tissue (cellular) identity is captured by gene expression, to the extent that
Gene expression patterns are more similar across homologous
tissues of different species than between diverse tissues of same

(conservation of organ physiology in vertebrates)

4,547 amniote orthologs
(5 studies, 5 common tissues, 11 species)

Sudmant Burge Genome Biol 2016

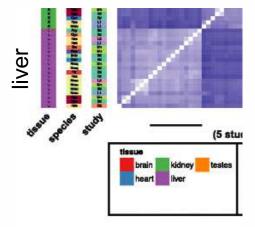# … Great advances through genomics, but …

Limitations: minimum starting material requirements
techniques applied on millions of cells

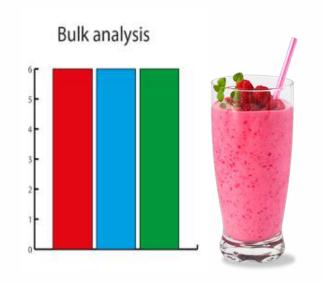*rare cell types & states* (e.g. transitions) cannot be analysed
confounding (measurements = averages => false correlations, masked differences)

! each sample is an AVERAGE!
no idea of the underlying values in single cells,
of the heterogeneity of the tissue

Single-cell genomics

Bulk analysis

Single cell transcriptome analysis

Macaulay IC, Voet T (2014) PLoS Genet
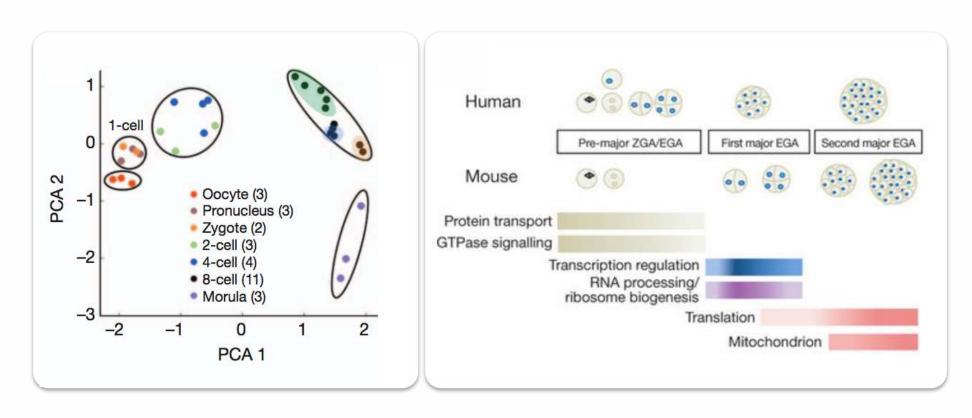
RARE CELL TYPES (e.g. early development, stem cells, circulating tumor cells)
HETEROGENEITY (e.g. tissue composition, cancer, temporal processes)
GENE REGULATORY NETWORKS (non-confounded correlations)
SINGLE-CELL PHENOMENA* (gene expression stochasticity, mono-allelic expression)

* *see also review CoulonLarsonNatRevGenet2013*

Rare cell types

Genetic programs in human and mouse early embryos

Transcriptome groups cells according to their developmental stage
Each developmental stage is characterised by a small nr. of functional modules of co-expressed genes
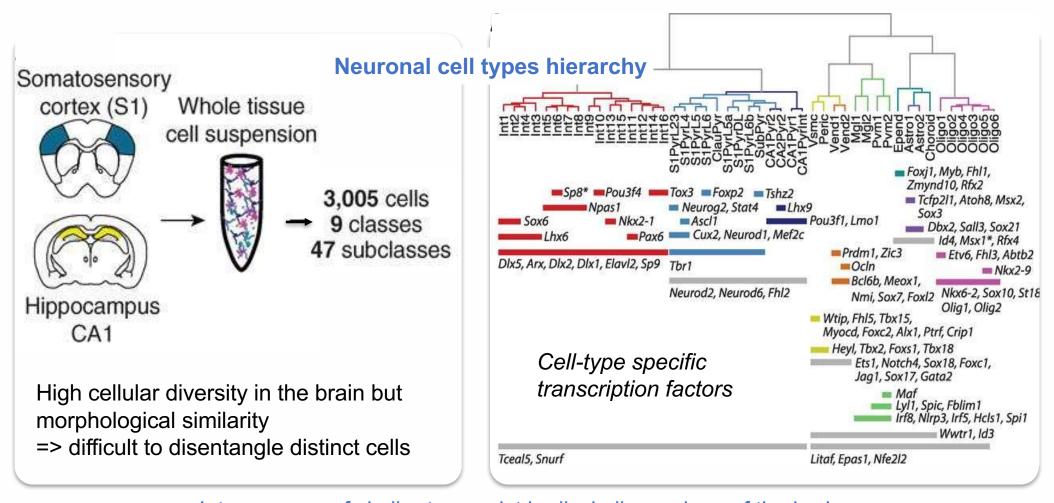
Identification of conserved key members of human & mouse networks
The temporal developmental pattern is different mouse-human

Xue et al. (2013) Nature
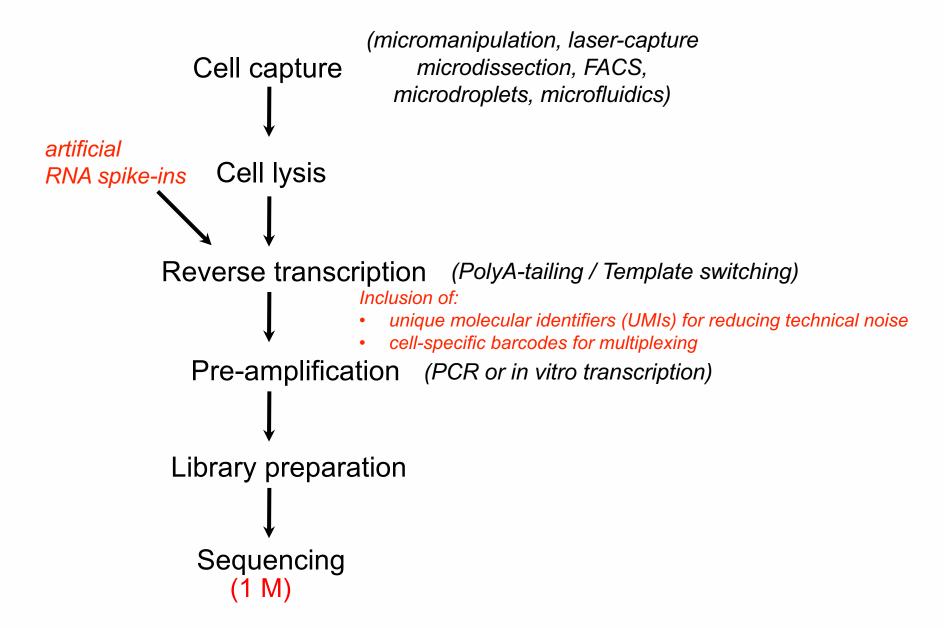
Tissue composition
Mapping out cell types in the mouse cortex & hippocampus

Neuronal cell types hierarchy

Cell-type specific transcription factors

High cellular diversity in the brain but morphological similarity
=> difficult to disentangle distinct cells

Interneurons of similar type exist in dissimilar regions of the brain
Identification of oligodendrocytes subtypes
Microglia associated with blood vessels distinguished from perivascular macrophages

Zeisel et al. (2015) Science
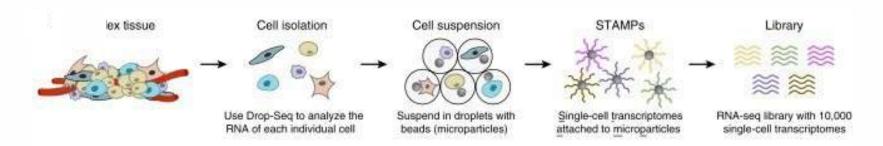
scRNA-seq experimental workflow

Cell capture   *(micromanipulation, laser-capture microdissection, FACS, microdroplets, microfluidics)*

*artificial
RNA spike-ins*   Cell lysis

Reverse transcription   *(PolyA-tailing / Template switching)*

Inclusion of:
*   *unique molecular identifiers (UMIs) for reducing technical noise*
*   *cell-specific barcodes for multiplexing*

Pre-amplification   *(PCR or in vitro transcription)*

Library preparation

Sequencing
(1 M)

DROP-seq for high throughput scRNA-seq
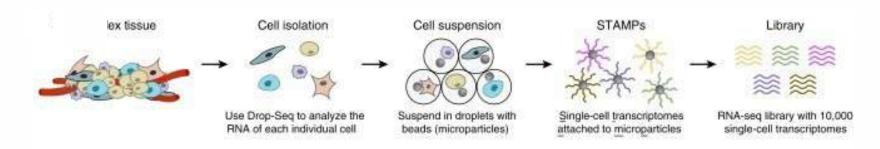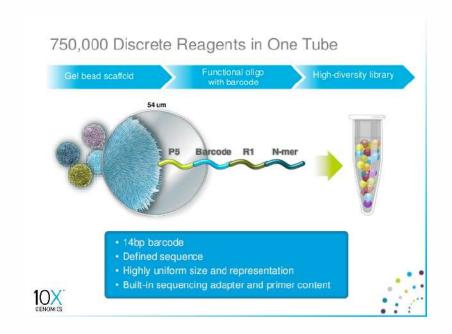
MacoskoMcCarrollCell2015

DROP-seq for high throughput scRNA-seq

MacoskoMcCarrollCell2015

10X Genomics (Commercial Solution)

- Hydrogel bead dissolves in droplet → uniform distribution of oligo's
- RT in drop
- Overall, higher data quality compared to DROP-seq

Reverse transcription

mRNA

AA…AAA

UMI-based 3' tag workflow (DROP-seq)

Reverse transcription

mRNA

AA...AAA

TT..TTT | UMI | Cell_ID

cDNA

unique per mRNA   *unique per cell*

UMI-based 3' tag workflow (DROP-seq)

Reverse transcription

mRNA

AA…AAA

TT..TTT | UMI | Cell_ID

unique per mRNA　　*unique per cell*

cDNA

AA…AAA

CCC

cDNA

Incorporation of CCC by RT

UMI-based 3' tag workflow (DROP-seq)

Reverse transcription

mRNA

AA…AAA

TT..TTT | UMI | Cell_ID

cDNA

unique per mRNA   *unique per cell*

AA…AAA

CCC

cDNA

Incorporation of CCC by RT

+ GGG

GGG
CCC

cDNA

Template switching

UMI-based 3' tag workflow (DROP-seq)

Reverse transcription

mRNA

AA…AAA

| TT..TTT | UMI | Cell_ID | |

cDNA

unique per mRNA   *unique per cell*

AA…AAA

CCC

cDNA

Incorporation of CCC by RT

+   GGG

GGG

CCC

cDNA

Template switching

Preamplification (PCR)

UMI-based 3' tag workflow (DROP-seq)

Reverse transcription

mRNA

AA...AAA

cDNA

| TT..TTT | UMI | Cell_ID | |

unique per mRNA   *unique per cell*

AA...AAA

CCC

cDNA

Incorporation of CCC by RT

+ GGG

GGG

CCC

cDNA

Template switching

Preamplification (PCR)

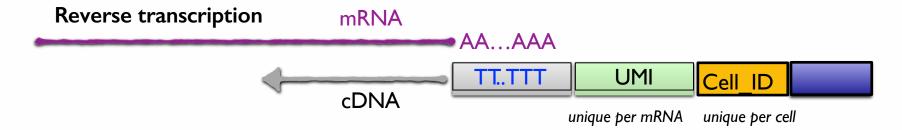**Tagmentation-based Library Preparation**

UMI-based 3' tag workflow (DROP-seq)

Reverse transcription

mRNA
AA…AAA

TT..TTT | UMI | Cell_ID

cDNA

unique per mRNA   *unique per cell*

AA…AAA
CCC
cDNA

Incorporation of CCC by RT

+ GGG

GGG
CCC
cDNA

Template switching

Preamplification (PCR)

reads/tags: 3'end of initial mRNA + barcodes

Tagmentation-based Library Preparation

Read2 (61 bp)
RNA sequence

Read1 (20 bp)
barcodes

Sequencing

UMI-based 3' tag workflow (DROP-seq)

MacoskoMcCarrollCell2015

"then a miracle occurs"

Cell Class	Percentage of Retina (Jeon et al., 1998) (%)	Percentage of Cell Population in Drop-Seq (%)
Rod photoreceptors	79.9	65.6
Cone photoreceptors	2.1	4.2
Muller glia	2.8	3.6
Retinal ganglion cells	0.5	1.0
Horizontal cells	0.5	0.6
Amacrine cells	7.0	9.9
Bipolar cells	7.3	14.0
Microglia	—	0.2
Retinal endothelial cells	—	0.6
Astrocytes		0.1

Today: The path from short sequenced "reads" to cell types?

Deplancke lab

Daniel Alpern
Roel Bevers
Marjan Biocanin
Virginie Braman
Johannes Bues
Wanze Chen
Riccardo Dainese
Teresa Didonna
Michael Frochaux

Vincent Gardeux
Maria Litovchenko
Gerard Llimos
Magali Masson (Admin)
Antonio Mereiles Filho
Rachana Pradhan
Julie Russeil
Petra Schwalie
Magda Zachara

Wolfrum lab

Christian Wolfrum
Nassila Akchiche
Hua Dong

ETHzürich

Lausanne GTF:
Keith Harshman
Gene Expression Core Facility (GECF):
Bastien Mangeat
FACS Facility:
Miguel Garcia, Loic Tauzin
Center for Phenogenomics:
Xavier Warot,
Raphael Doenlen

SIB
Vital-IT
High Performance
Computing Center

AgingX
Systems Genetics
Approach to
the Biology of Aging

EMBO   HFSP   FEBS   EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Smart-seq2
Picelli et al., Nature Prot., 2014

Cell lysis (Steps 1–8)

Poly(A)$^+$ RNA

AAAAAAAAAA

Oligo(dT) primer

Reverse transcription
and terminal transferase (Steps 9–11)

LNA-containing TSO

rGrG+G

AAAAAAAAAA

C C C

Template switching
by reverse transcriptase (Steps 9–11)

ISPCR primers

CCC

ISPCR primers

PCR preamplification of cDNA (Steps 12–14)
PCR cleanup (Steps 15–26)

GGG

CCC

Tagmentation (Tn5) (Steps 28–31)

GGG

CCC

Gap repair, enrichment PCR
and PCR purification (Steps 32–36)

P5 primer   i5 index

i7 index   P7 primer

Sequencing (Steps 37–41)

i5 index seq   Read 1 seq   Read 2 seq
                Sequencing-ready fragment → i7 index seq