

Single-cell epigenomics

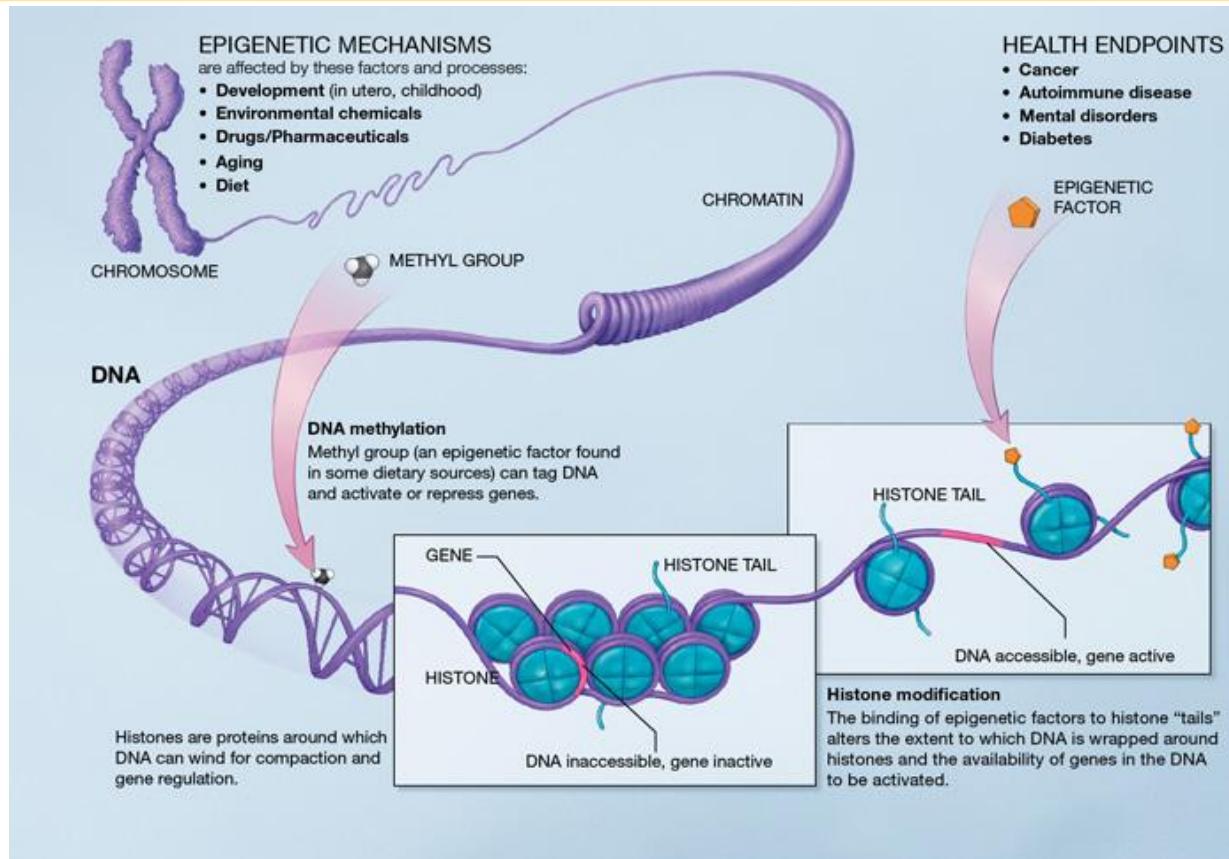
Ecole thématique 2019 | sincellTE

Morgane Thomas-Chollier,
Nicolas Servant, Nathalie Lehmann

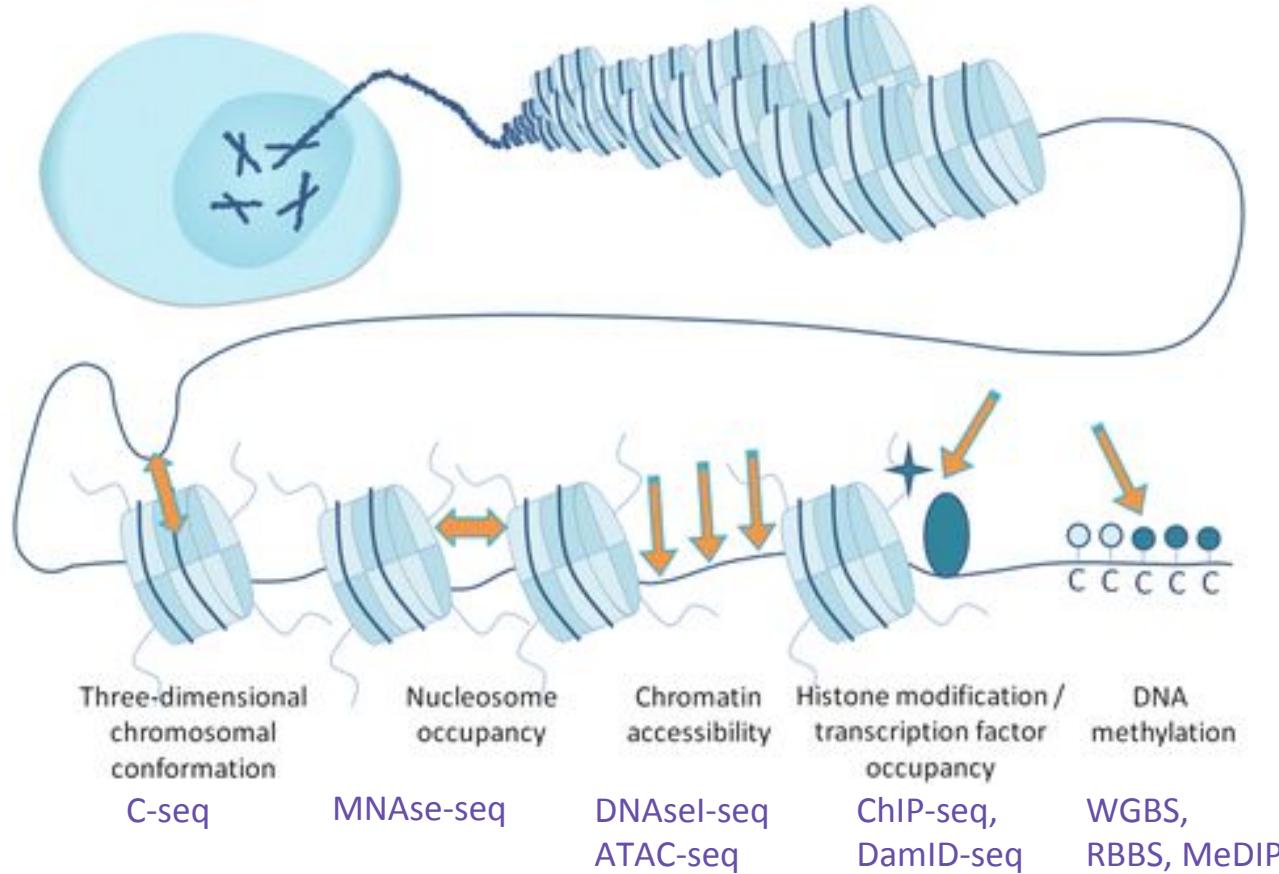
- Profiling the epigenome
- Overview of single-cell epigenomics
- Single-cell technologies : state of the art
- Combining multiple assays (Multi omics)

What are the techniques to study the epigenome (bulk) ?

What is the epigenome ?

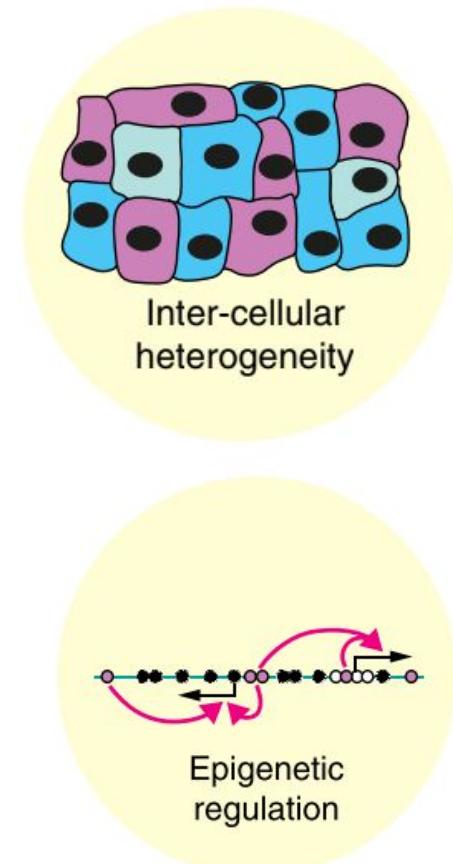


Common assays to study epigenome (bulk)



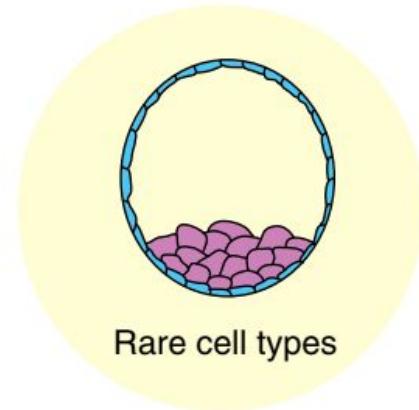
Why single-cell ?

- Study cell-to-cell variation chromatin states
- How transcriptional and epigenetic heterogeneity are related ?
- Is it possible to predict changes in transcription levels with epigenetic marks (and vice versa) ?
- Is it possible to define a rare cell population by its epigenetic signature ?
- Follow cells differentiation at different time points



Single-cell : advantages and challenges

- Investigate the epigenome in rare (precious) samples
- Analyse beyond the average of pooled cells : contribution of each cell to the average signal



Single-cell : advantages and challenges

- scRNA-seq is well-established but not single-cell epigenomic analyses
- Major problem : usually based on whole genome amplification (WGA), which introduces biases
- Low read coverage across the genome (i.e high sparsity)

Single-cell epigenomics in real life

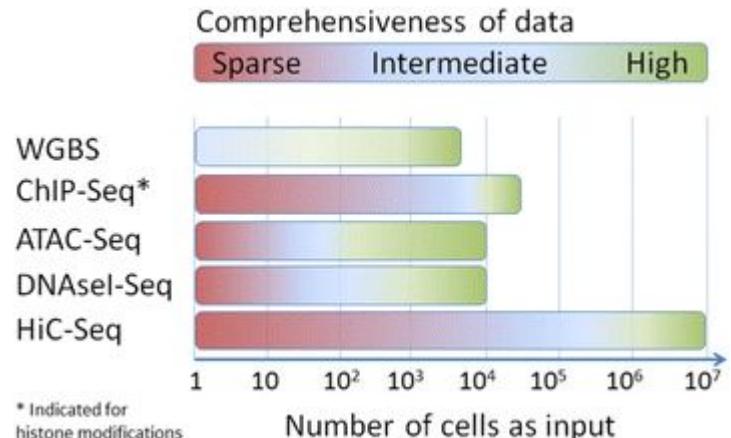
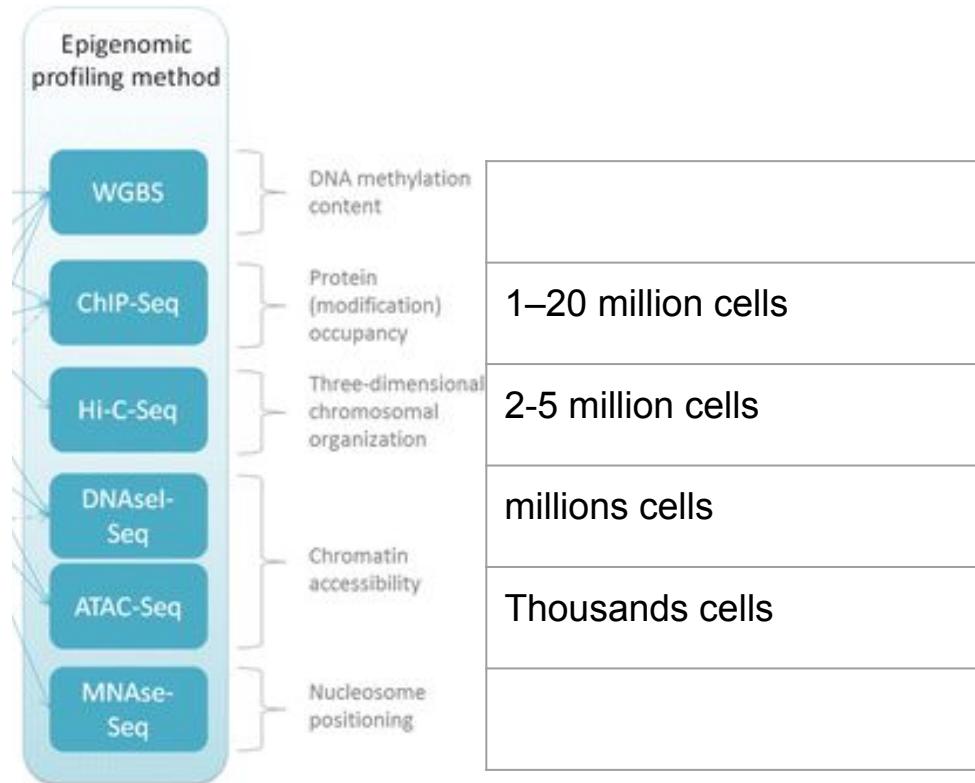
Not yet available for all labs

No standardized protocols

No standardized way to analyse data



Starting material in bulk : not all methods are equal !

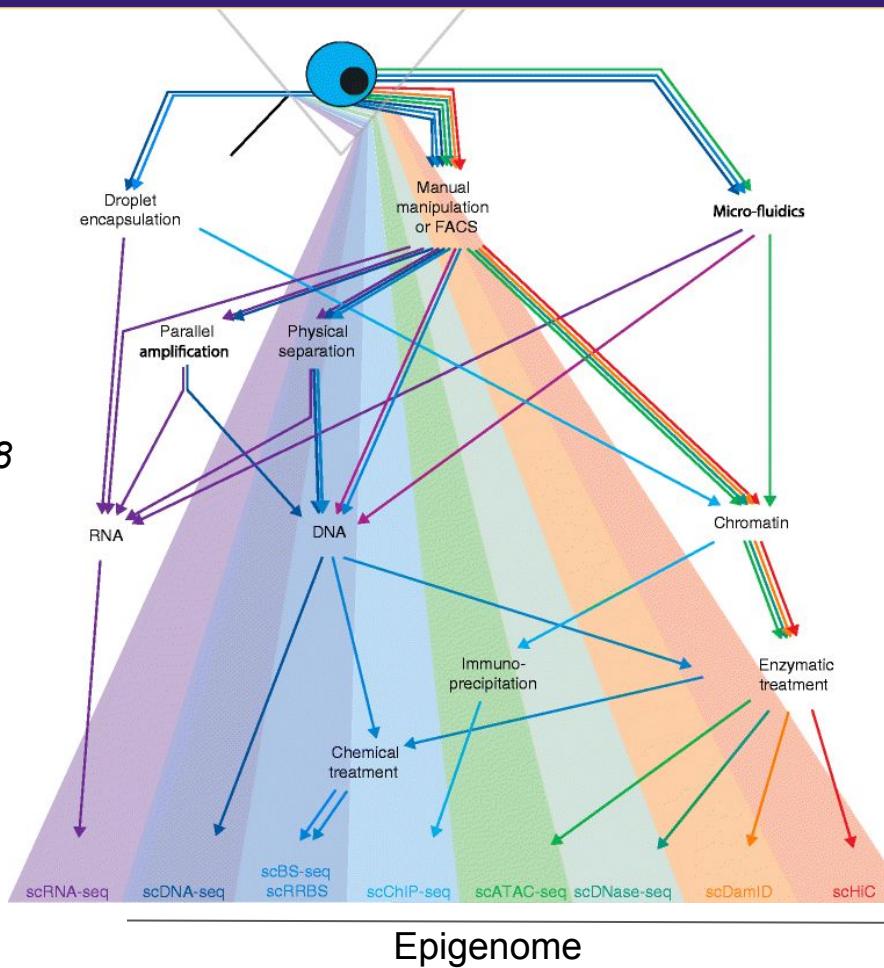


Dirks, Clin Epigenet, 2016

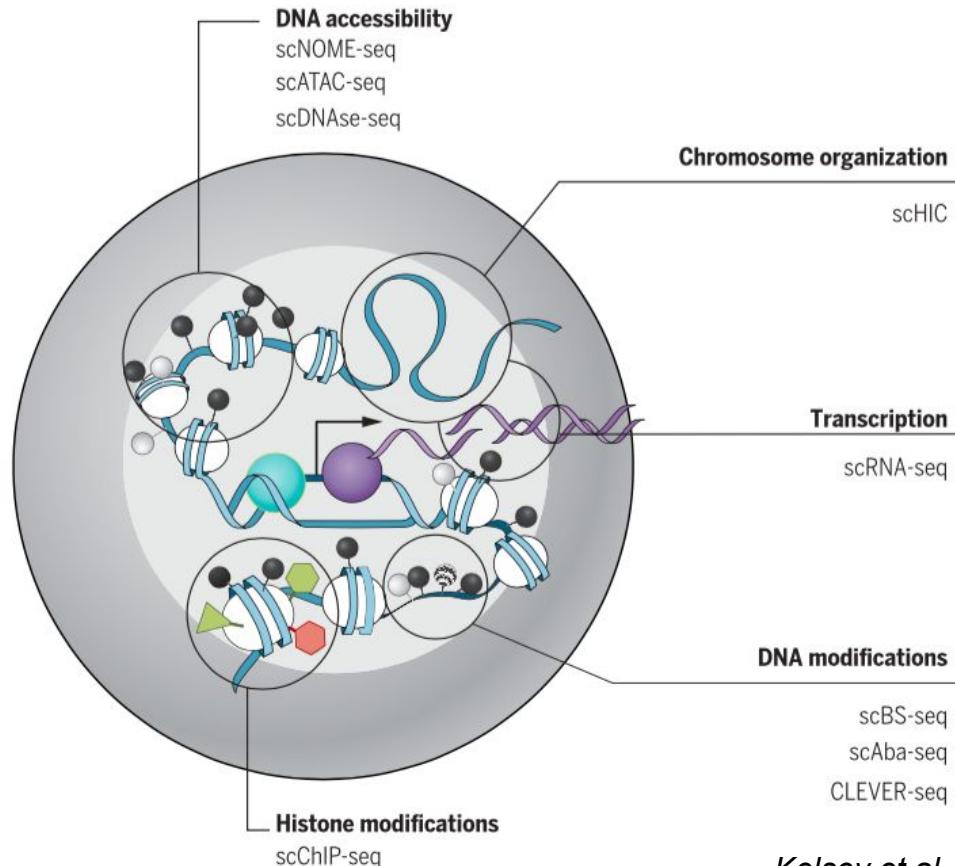
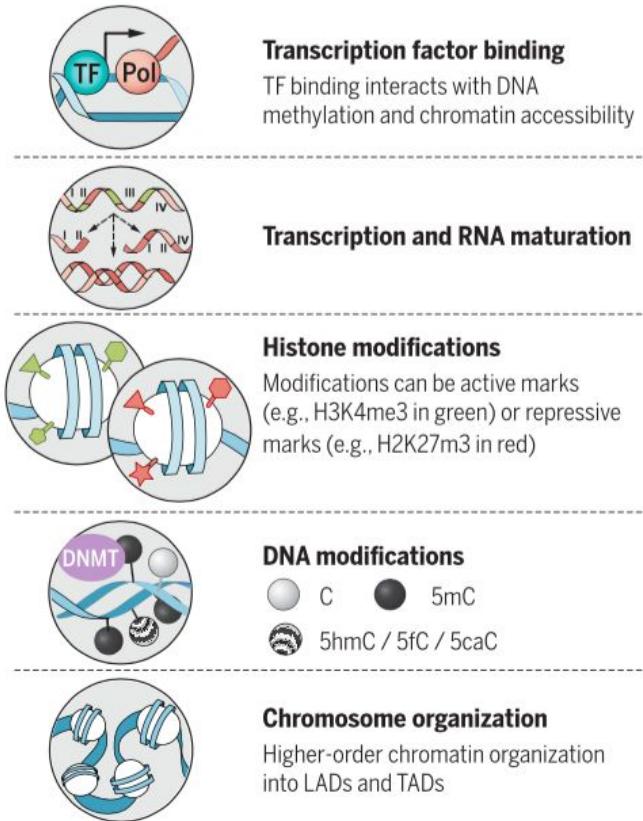
- Profiling the epigenome
- Overview of single-cell epigenomics
- Single-cell technologies : state of the art
- Combining multiple assays (Multi omics)

What are the techniques that have been adapted to single-cell ?

Overview of single-cell techniques



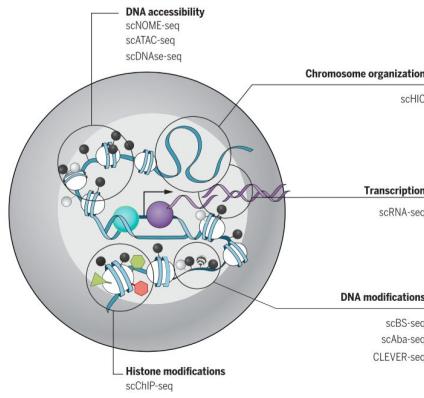
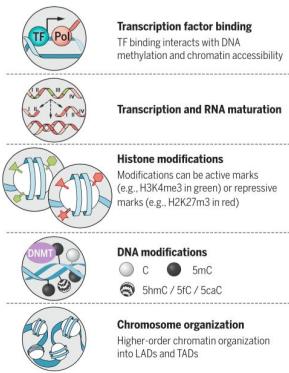
Overview of sc epigenomics techniques



Summary of sc epigenomics technologies

Assays	Epigenetic markers	Strategies	Coverage per cell	Throughput	References
scRRBS	5mC	RRBS	1.5 million CpG sites	1–100	(Guo et al., 2013; Guo et al., 2015a)
scBS-seq	5mC	PBAT (five rounds of preamplification)	3.7 million CpG sites	1–100	(Smallwood et al., 2014; Clark et al., 2017)
scWGBS	5mC	PBAT (one round of preamplification) with 3' tagging	1.0 million CpG sites	1–100	(Farlik et al., 2015)
scWGBS (Gravina)	5mC	PBAT (two rounds of preamplification)	2.2 million CpG sites	1–100	(Gravina et al., 2016)
snmC-seq	5mC	PBAT (one round of preamplification) with 3' tagging	5% of the genome	100–1000	(Luo et al., 2017)
sci-MET	5mC	in nuclei Tn5 transposition for combinatorial cellular indexing	0.8% of the genome	>1000	(Mulqueen et al., 2017)
scAba-seq	5hmC	Aba-seq	44,000 5hmC sites	100–1000	(Mooijman et al., 2016)
CLEVER-seq	5fC	Selectively labeling of 5fC that induces C-to-T conversion in PCR amplification, combined with MALBAC	60–80% of the genome	1–100	(Zhu et al., 2017)
scMAB-seq	5fC/5caC	Combining MAB-seq with scRRBS or scBS-seq	Depending on scRRBS or scBS-seq used	1–100	(Wu et al., 2017)
sciATAC-seq	Chromatin accessibility	Combining ATAC-seq with combinatorial cellular indexing	450 DHSs	>1000	(Cusanovich et al., 2015)
scATAC-seq	Chromatin accessibility	Combining ATAC-seq with microfluidics	5,000 DHSs	100–1000	(Buenrostro et al., 2015)
scDNase-seq	Chromatin accessibility	DNase-seq	38,000 DHSs	1–100	(Jin et al., 2015)
Drop-ChIP	Histone modification	Combining drop-based microfluidics and DNA barcoding with ChIP-seq	1,000 H3K4me2 peaks	>1000	(Rotem et al., 2015)
scDamID	Nuclear-lamina interactions	DamID	450,000 Dam/DpnI sites	1–100	(Kind et al., 2015)
scHi-C	3D chromatin architecture	Combining in nuclei Hi-C with manipulation of individual nuclei	1,000 contacts	1–100	(Nagano et al., 2013)
snHi-C	3D chromatin architecture	Phi29 amplification after proximal ligation	400,000 contacts	1–100	(Flyamer et al., 2017)
sciHi-C	3D chromatin architecture	Combining in nuclei Hi-C with combinatorial cellular indexing	9,000 contacts	>1000	(Ramani et al., 2017)

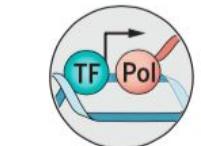
- Profiling the epigenome
- Overview of single-cell epigenomics
- Single-cell technologies : state of the art
- Combining multiple assays (Multi omics)



Let's look in more details at some assays

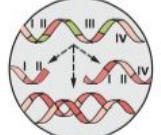
1. Bulk approach
2. Adaptation to single-cell
3. Bioinformatics analyses

Overview of sc epigenomics techniques

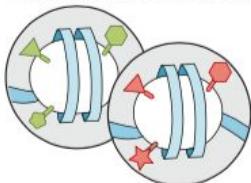


Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility



Transcription and RNA maturation



Histone modifications

Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27m3 in red)



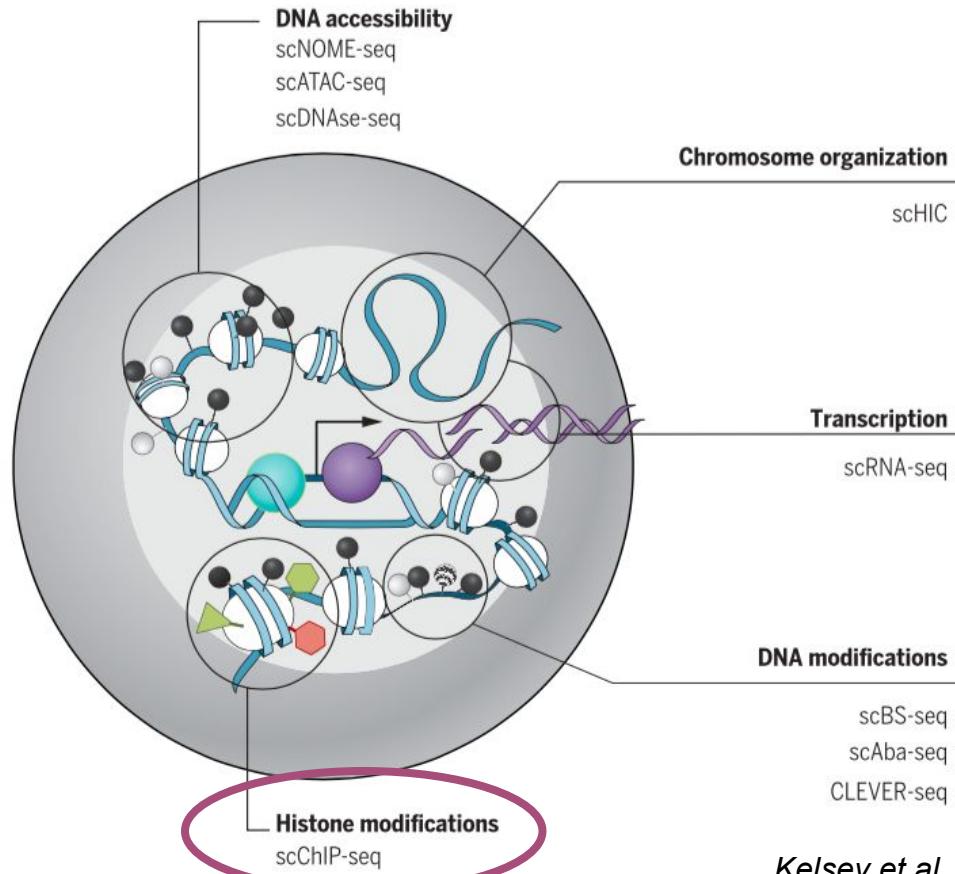
DNA modifications

- C
- 5mC
- 5hmC / 5fC / 5caC

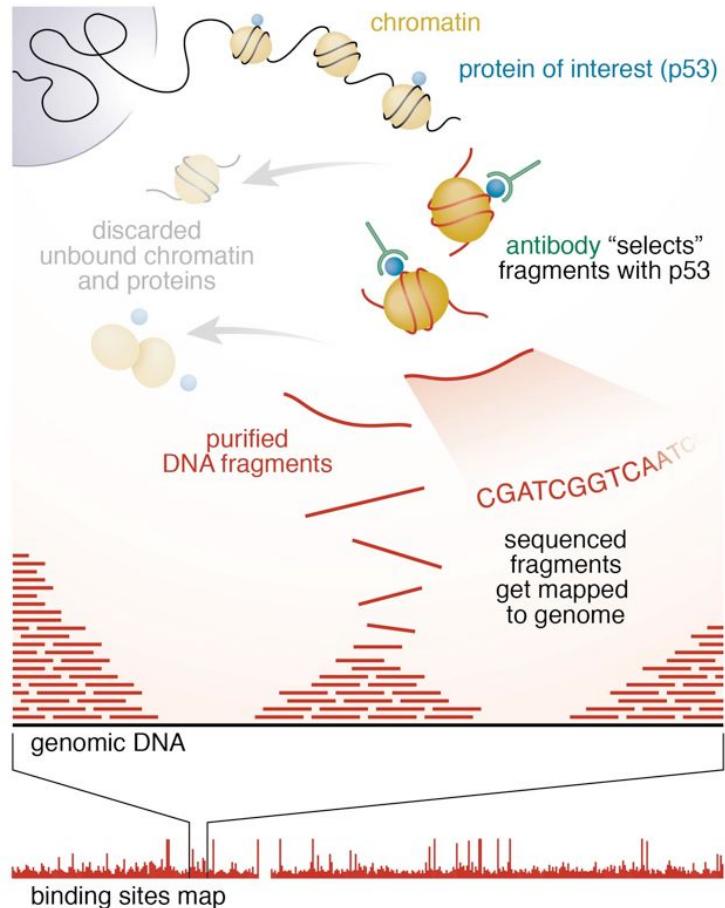


Chromosome organization

Higher-order chromatin organization into LADs and TADs



ChIP-seq profiles transcription factor binding and histone modifications



ChIP-seq uses immunoprecipitation to profile genome-wide DNA regions on which particular proteins are bound

ChIP-seq : low input rather than single-cell



Briefings in Functional Genomics, 2017, 1–7

doi: 10.1093/bfgp/elx037
Review paper

How low can you go? Pushing the limits of low-input ChIP-seq

John Arne Dahl and Gregor D. Gilfillan

Corresponding author. Gregor D. Gilfillan, Department of Medical Genetics, Oslo University Hospital and University of Oslo, 0450 Oslo, Norway. E-mail: gregor.gilfillan@medisin.uio.no

Optimized ChIP-seq protocols => 100-500 cells

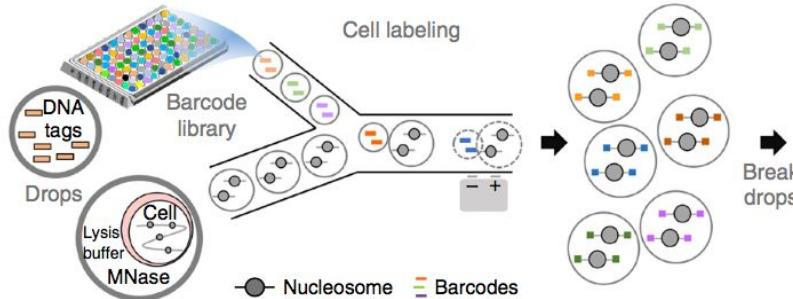
scChIP-seq/Drop-ChIP: proof-of-concept

Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state

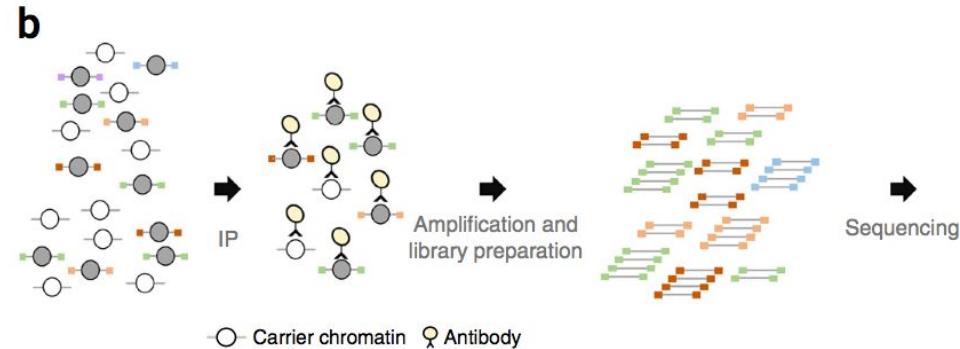
Nature Biotechnology 2015

Assaf Rotem^{1,2,7}, Oren Ram^{2-4,7}, Noam Shores^{2,7}, Ralph A Sperling^{1,6}, Alon Goren⁵, David A Weitz¹ & Bradley E Bernstein²⁻⁴

a



b

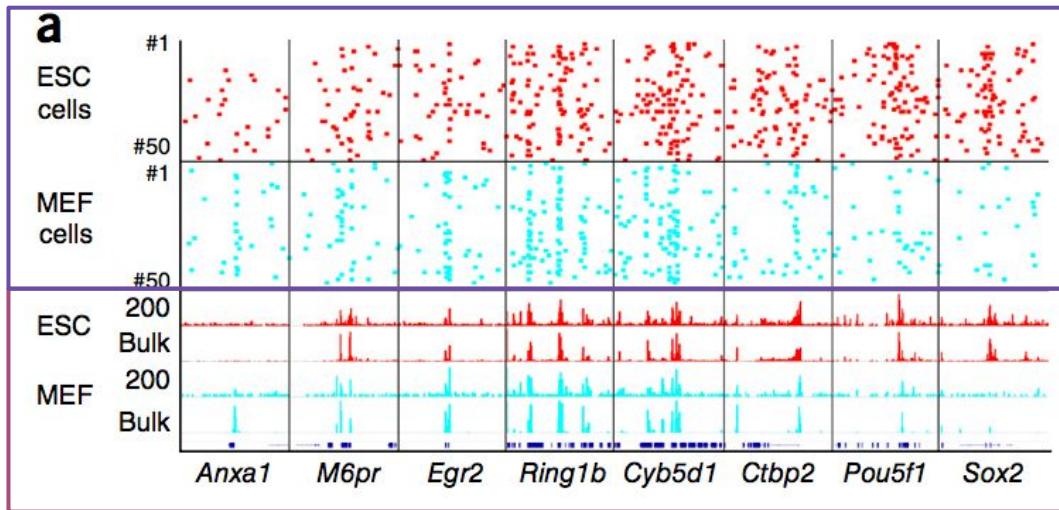


Histone modifications (H3K4me3/H3K4me2)

scChIP-seq/Drop-ChIP: proof-of-concept

Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state

Assaf Rotem^{1,2,7}, Oren Ram^{2-4,7}, Noam Shoresh^{2,7}, Ralph A Sperling^{1,6}, Alon Goren⁵, David A Weitz¹ & Bradley E Bernstein²⁻⁴



1 row = 1 cell => *reads*

Bulk or pooled cells => *peaks*

scChIP-seq : 4 years later



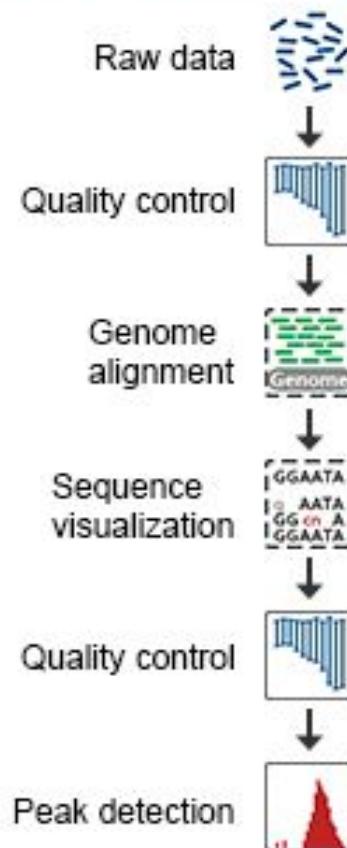
Team Céline Vallot

A poster for the Single-cell symposium held at Institut Curie, Paris, France, on December 12-13, 2018. The poster features a blue and purple background with a stylized Eiffel Tower and various scientific illustrations. It lists the organizers, keynote speakers, and speakers, along with logos for PSL, IONpath, AKOYA Biosciences, 10X Genomics, MACS, ttplabtech, and Fluidigm.

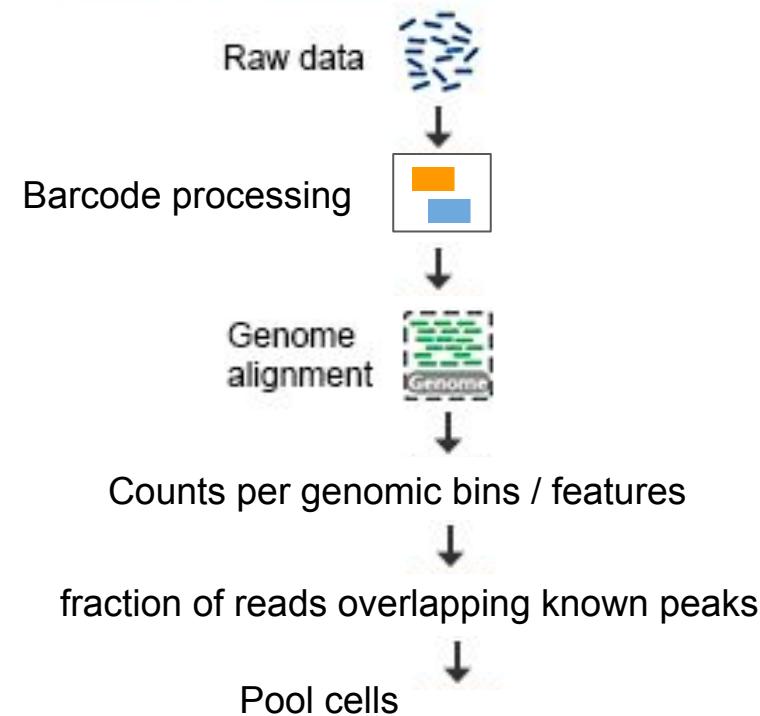
Histone modifications (H3K27me3 H3K4me3)

Bioinformatics analyses

Bulk ChIP-seq

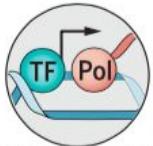


Single-cell



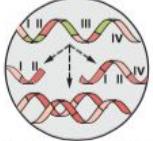
- Profiling the epigenome
 - Overview of single-cell epigenomics
 - Single-cell technologies : state of the art
 - scChIP-seq/Drop-ChIP
-
- Only one proof-of-concept (2015) and an announcement (2019)
 - Low coverage (1000 uniquely mapped reads/cell), but pooling cells allows to recover peaks.
 - Not applied to transcription factors.
 - Custom microfluidic device to manufacture.

Overview of sc epigenomics techniques

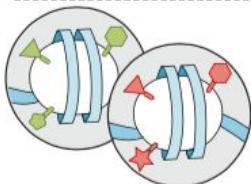


Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility



Transcription and RNA maturation



Histone modifications

Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27m3 in red)



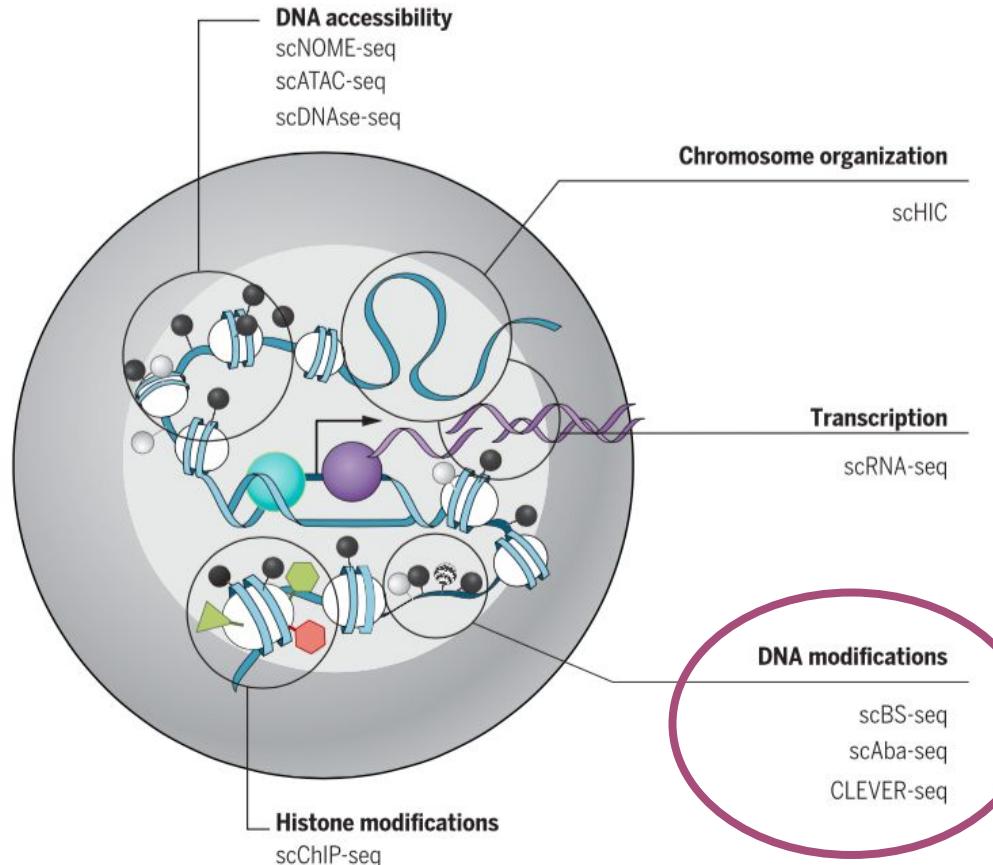
DNA modifications

- C
- 5mC
- 5hmC / 5fC / 5caC



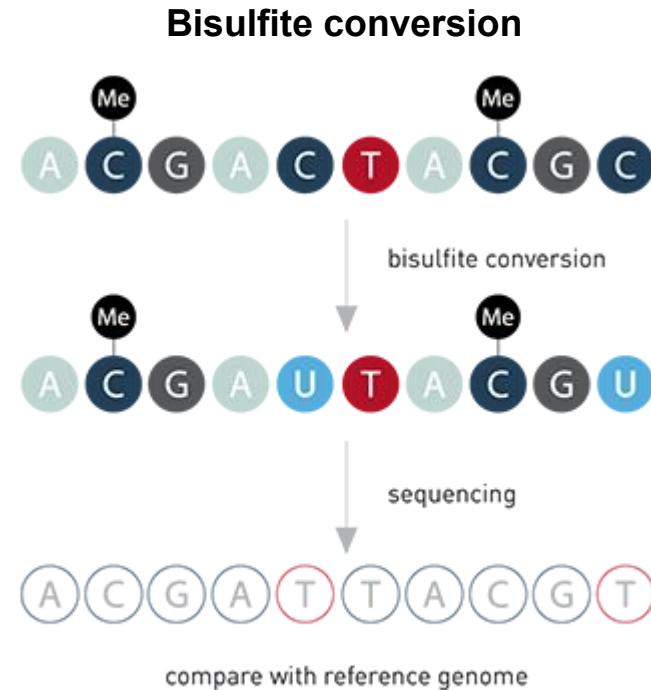
Chromosome organization

Higher-order chromatin organization into LADs and TADs



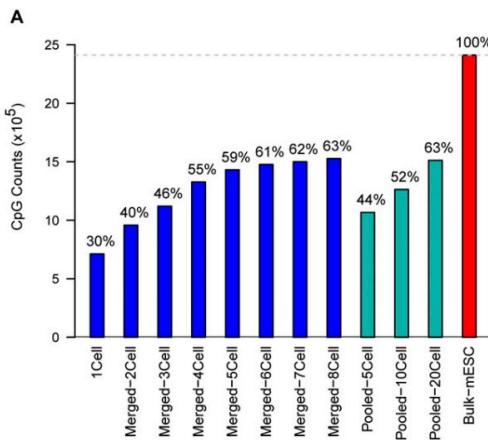
Methylation (bisulfite conversion)

DNA methylation is the process in which a methyl group is added to the 5' position of cytosines in the DNA, which mainly occurs within the context of CpGs



Single cell methylation

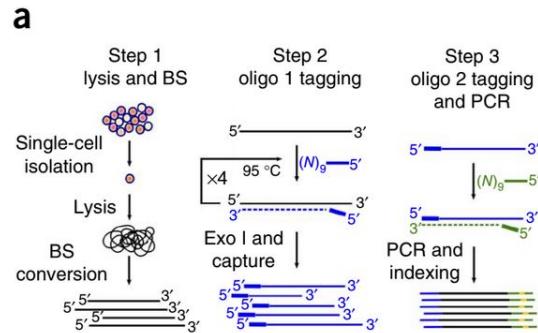
scRRBS
(<10 cells, 1–2 million CpGs (10% of total CpGs))



Guo, *Genome Research*, 2013

in a single tube, no purification prior to bisulfite conversion

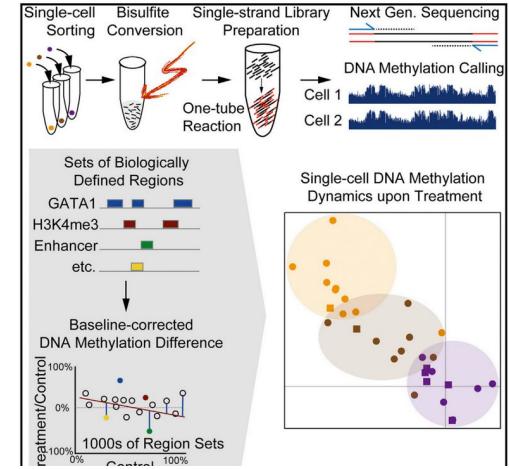
scBS-seq
(<40 cells, 3.7 million CpGs: 17.7%)



Smallwood, Nature, 2014

optimized protocol : bisulfite treatment causes extensive DNA damage, they use post-bisulfite adaptor ligation (to avoid DNA damage between the two ligated adaptors that can interfere with PCR amplification).

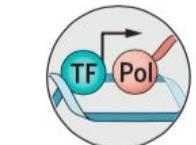
scWGBS
(80 cells, 1.4 million CpG)



Farlik, Cell, 2015

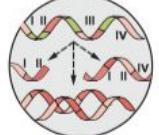
- Profiling the epigenome
 - Overview of single-cell epigenomics
 - Single-cell technologies : state of the art
 - scChIP-seq/Drop-ChIP
 - sc methylation
- low coverage (4M reads/cell)
- <20% all CpG recovered per cell, but pooling dozen of cells allows to recover most CpGs.
- Recent improvements to increase the number of cells + read coverage to obtain more CpGs (Mulqueen, Nature Biotech, 2018)

Overview of sc epigenomics techniques

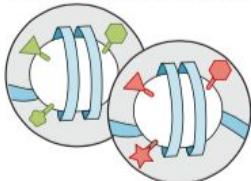


Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility



Transcription and RNA maturation



Histone modifications

Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27m3 in red)



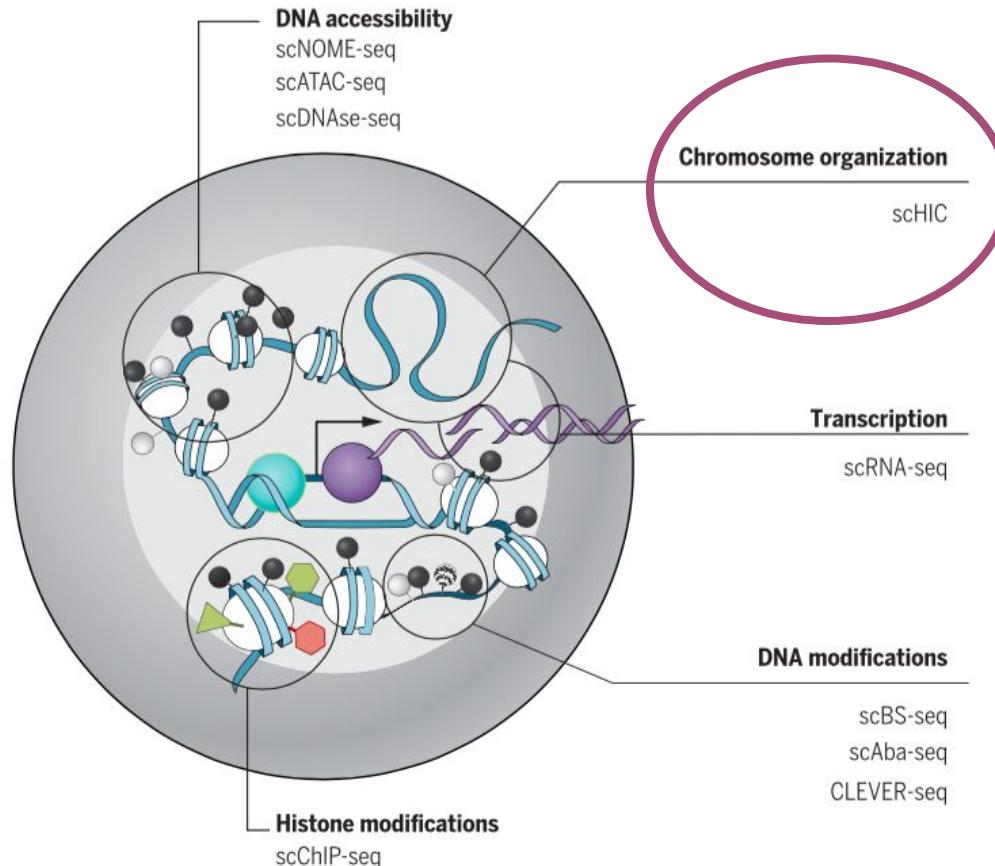
DNA modifications

- C
- 5mC
- 5hmC / 5fC / 5caC

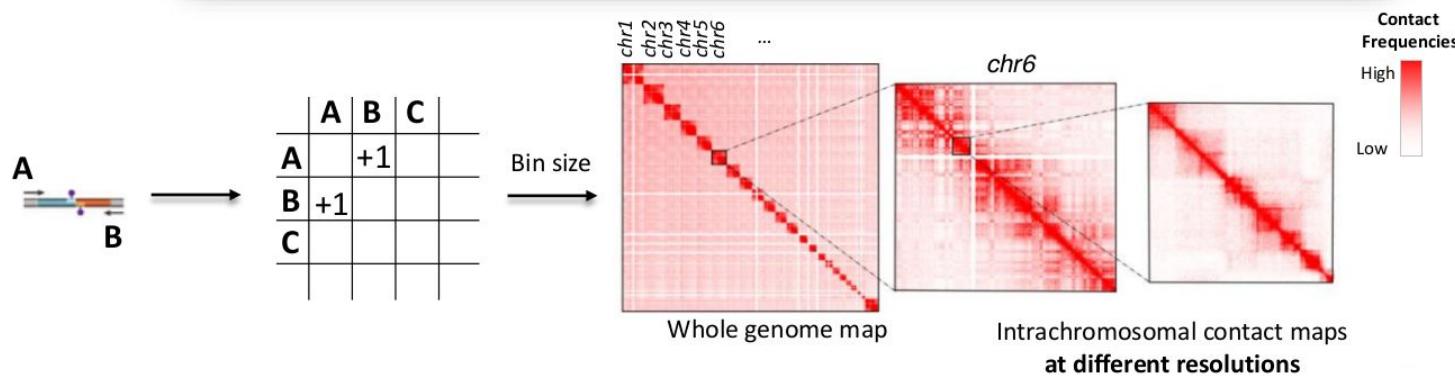
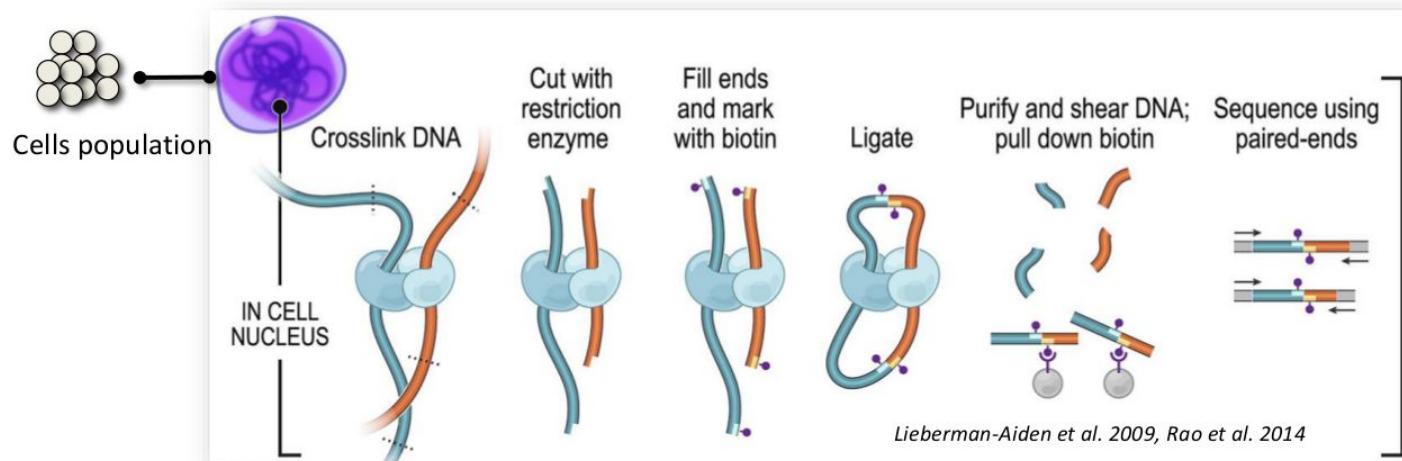


Chromosome organization

Higher-order chromatin organization into LADs and TADs



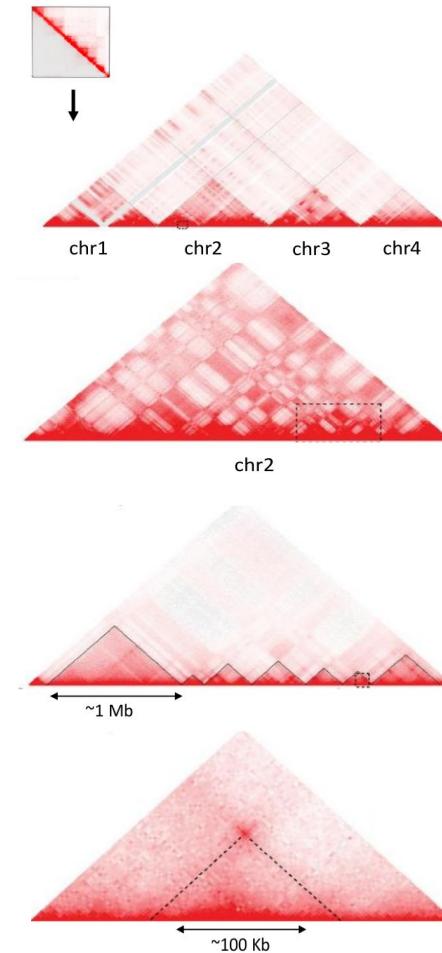
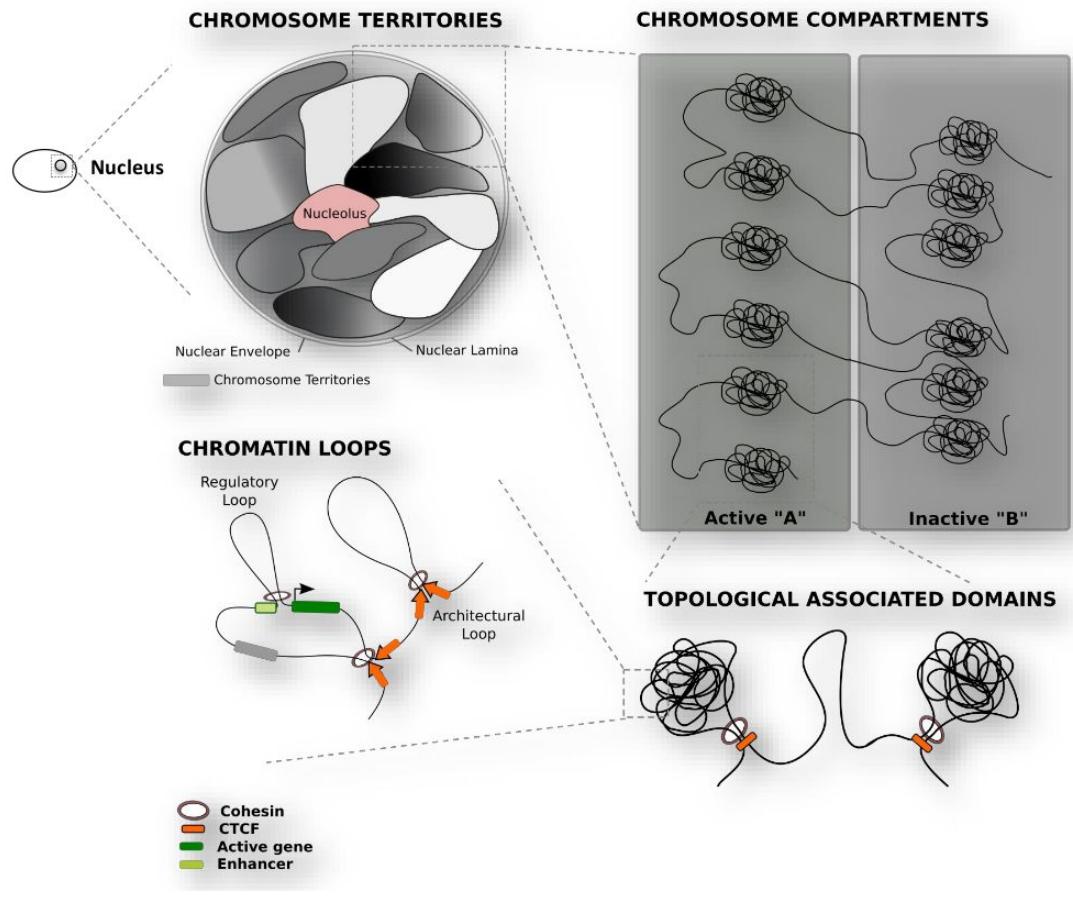
Hi-C captures the 3D organization of chromosomes



Hi-C based protocols

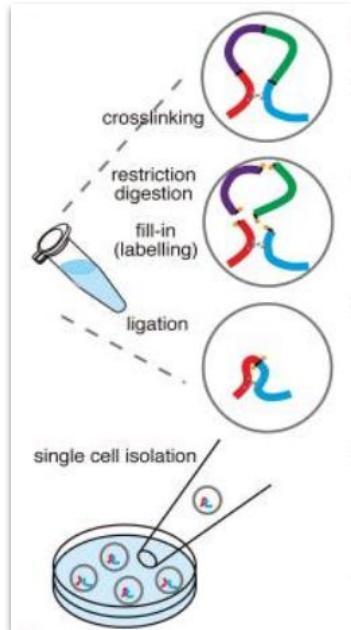
Method	Main features	References
Hi-C	For mapping whole-genome chromatin interaction in a cell population; proximity ligation is carried out in a large volume	Lieberman-Aiden <i>et al.</i> (2009)
TCC	Similar to Hi-C, except that proximity ligation is carried out on a solid phase-immobilized proteins	Kalhor <i>et al.</i> (2011)
Single-cell Hi-C	For mapping chromatin interactions at the single-cell level	Nagano <i>et al.</i> (2013)
In situ Hi-C	Proximity ligation is carried out in the intact nucleus	Rao <i>et al.</i> (2014)
Capture-C	Combines 3C with a DNA capture technology ; equivalent to high-throughput 4C	Hughes <i>et al.</i> (2014)
Dnase Hi-C	Chromatin is fragmented with Dnase I; proximity ligation is carried out on a solid gel	Ma <i>et al.</i> (2015)
Targeted Dnase Hi-C	Combine Dnase or in situ Dnase Hi-C with a capture technology	Ma <i>et al.</i> (2015)
Micro-C	Chromatin is fragmented with micrococcal nuclease	Hsiech <i>et al.</i> (2015)
In situ DNase Hi-C	Chromatin is fragmented with Dnase I; proximity ligation is carried out in the intact nucleus	Deng <i>et al.</i> (2015)
Capture-Hi-C	Combines 3C with a DNA capture technology ; equivalent to high-throughput 5C	Mifsud <i>et al.</i> (2015)
HiChIP	Detecting genome-wide chromatin interaction mediated by a particular protein ; equivalent to ChAI-PET	Mumbach <i>et al.</i> (2016)

What can we do with bulk Hi-C ?



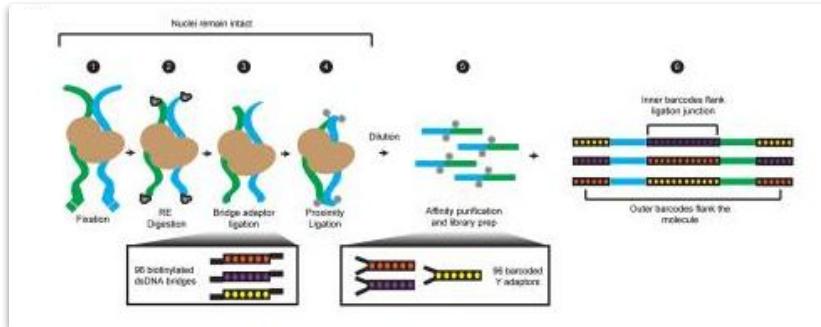
Single-cell Hi-C

Dilution scHi-C
(<100 cells, 10-20K contacts per cell)

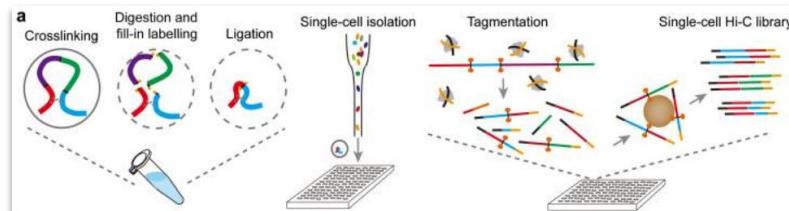


Nagano et al. 2013

In situ scHi-C
Massively Parallel
(~1K-10K cells, ~400K contacts per cell)

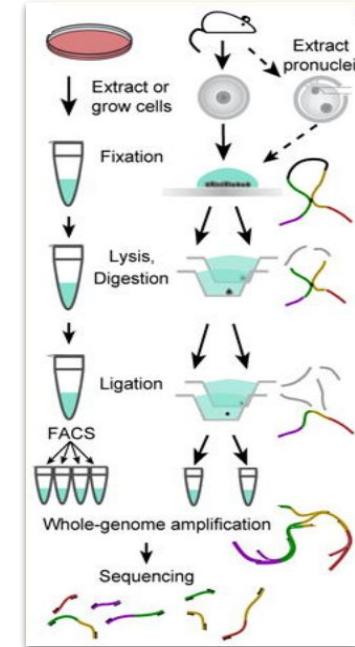


Ramani et al. 2017



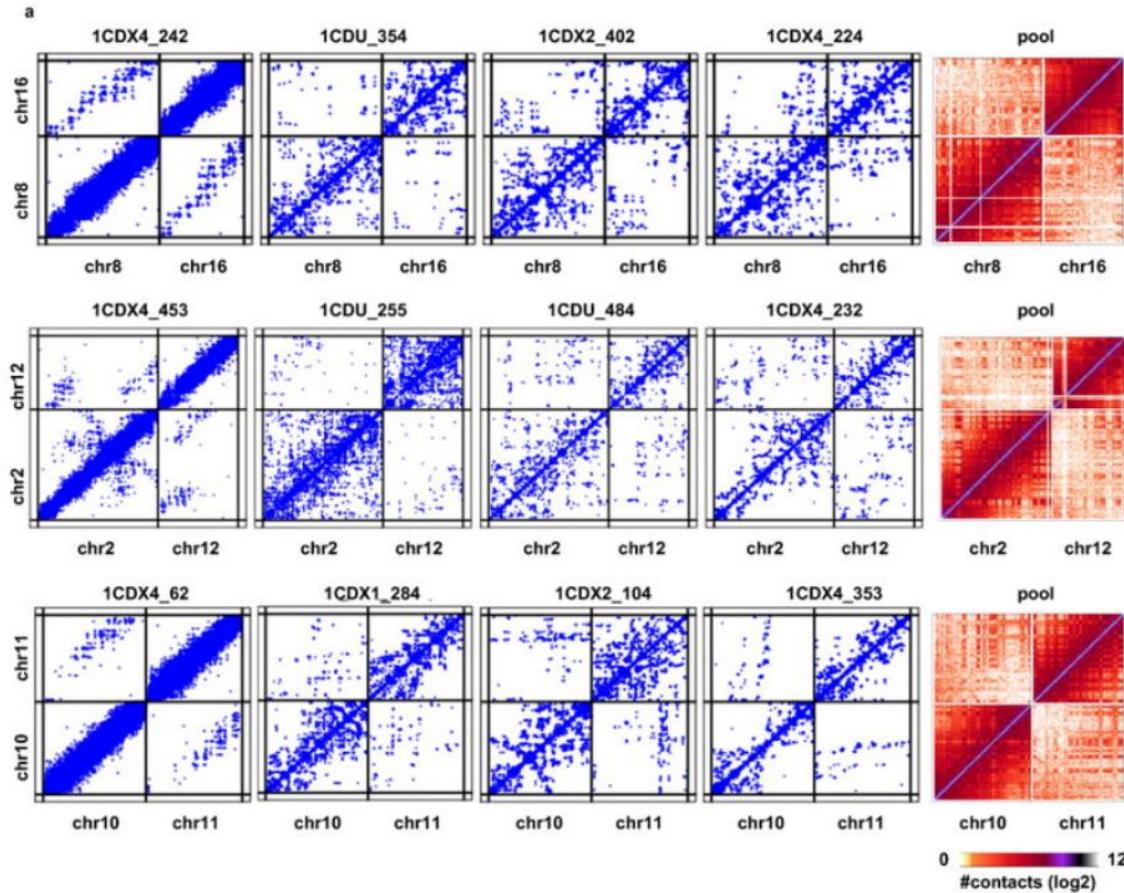
Nagano et al. 2017

WGA instead of PCR
No biotin pull-down !!
(~250 cells, 400K - 1M contacts per cell)

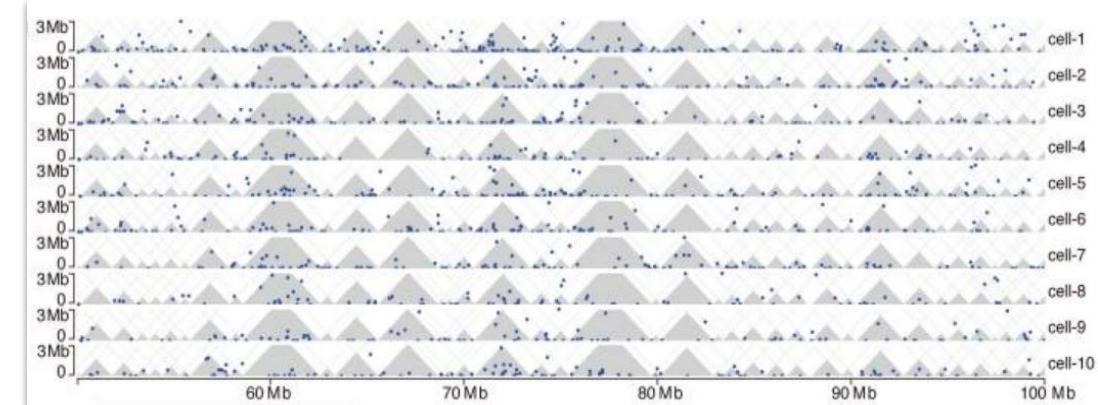


Flyamer et al. 2017

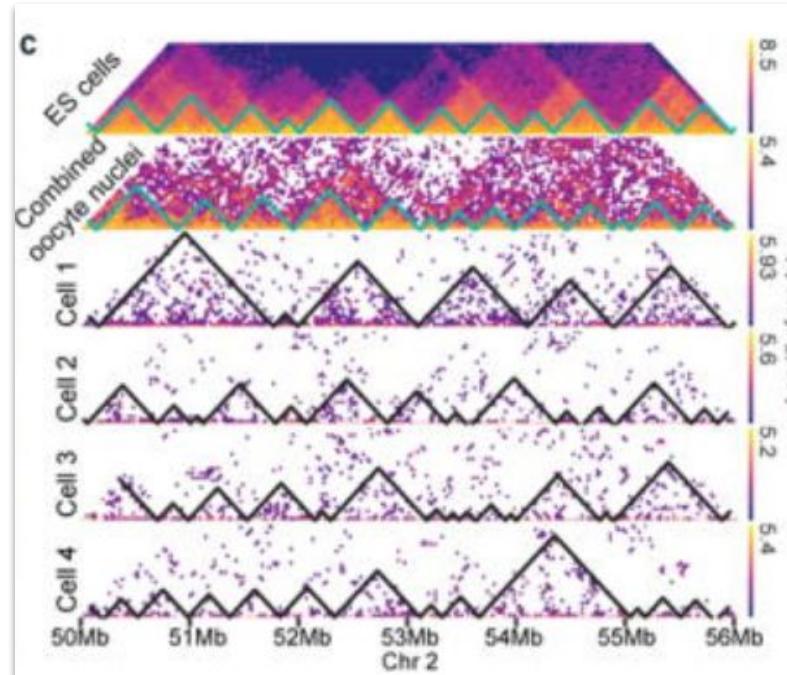
What does scHi-C look like ?



What does scHi-C look like ?



Nagano et al. 2013



Flyamer et al. 2017

Unsupervised analysis of scHi-C data

Unsupervised embedding of single-cell Hi-C data

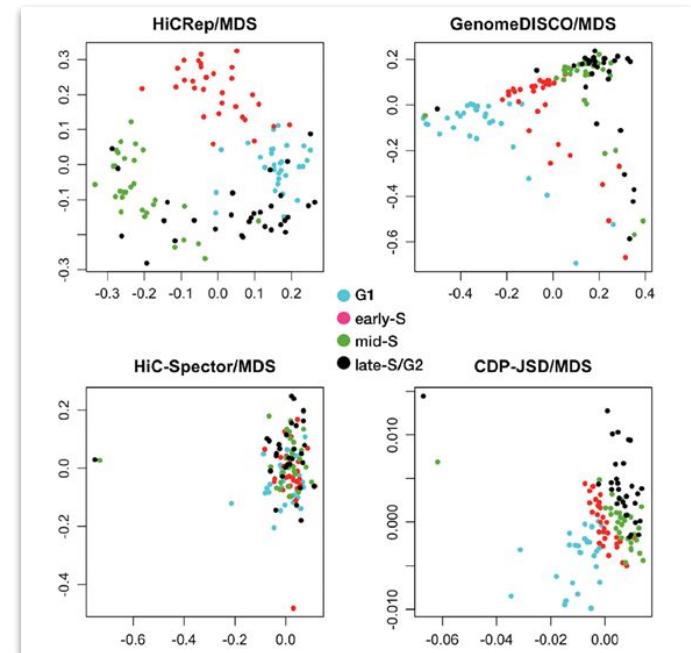
Jie Liu ✉, Dejun Lin, Galip Gürkan Yardımcı, William Stafford Noble ✉

Bioinformatics, Volume 34, Issue 13, 1 July 2018, Pages i96–i104,

<https://doi.org/10.1093/bioinformatics/bty285>

Published: 27 June 2018

- Unsupervised methods developed for scRNA-seq cannot be applied to scHiC
- Comparison of 4 distance-based metrics
- HiCRep + MDS seems to perform well



Bioinformatics analysis of scHi-C data

Overall **the same bioinformatics processing can be applied to both bulk Hi-C and scHi-C** (ie. from raw sequencing reads to raw contact matrices), with a few exception depending on the protocol ;

- Demultiplexing / adapter trimming
- In the context of WGA, careful duplicates removal must be applied
- *Flyamer et al.* uses a double digestion protocol ...

Standard normalization method (such as iterative correction) cannot be applied due to the extreme level of data sparsity

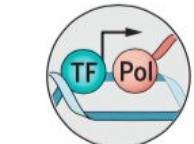
→ **No normalization method for scHi-C**

TADs or A/B compartments calling are still usually do on pooled scHi-C maps, although some methods seem to be adapted (with a few changes) as the number of contacts increases ...

- Profiling the epigenome
- Overview of single-cell epigenomics
- Single-cell technologies : state of the art
 - scChIP-seq/Drop-ChIP
 - sc methylation
 - scHi-C

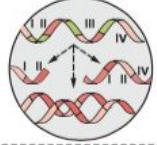
- The scHi-C protocol is not yet stable
- Although the data are much more sparse, the first processing level looks the same
- Need for new methodological developments in the field (normalization, domains calling, etc.)

Overview of sc epigenomics techniques

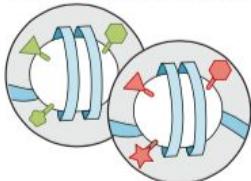


Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility



Transcription and RNA maturation



Histone modifications

Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27m3 in red)



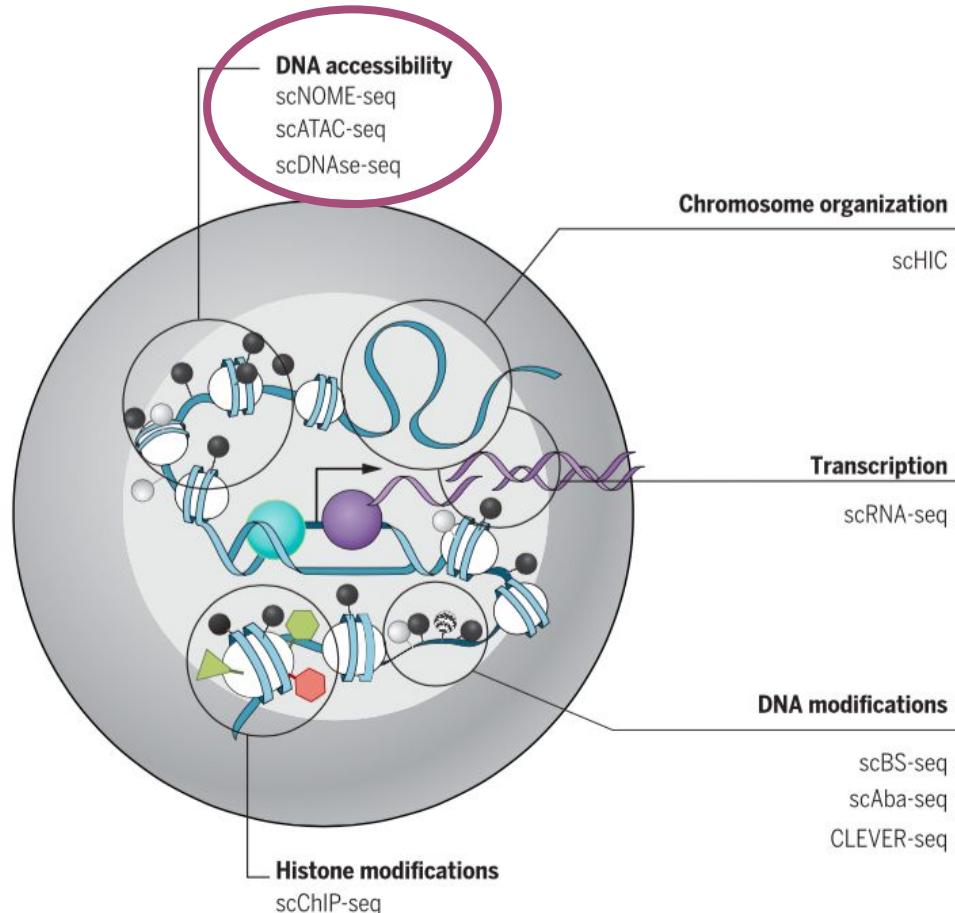
DNA modifications

- C
- 5mC
- 5hmC / 5fC / 5caC



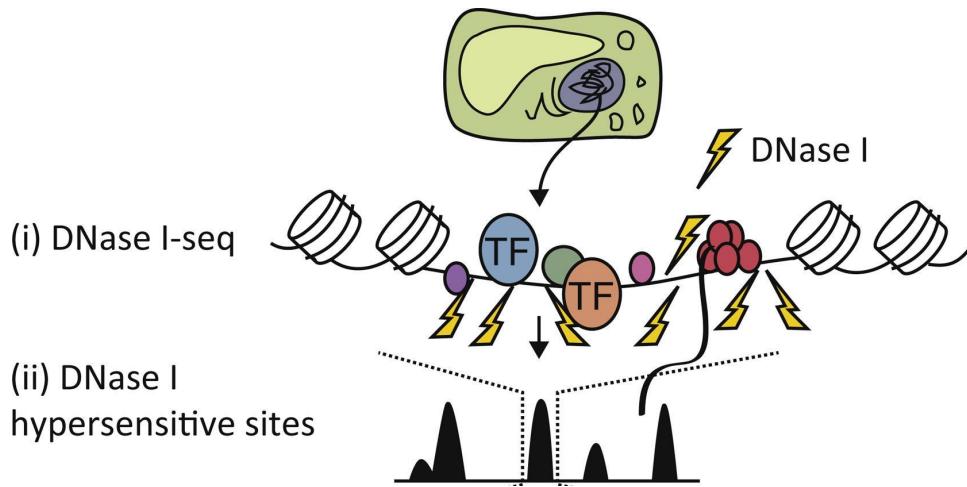
Chromosome organization

Higher-order chromatin organization into LADs and TADs

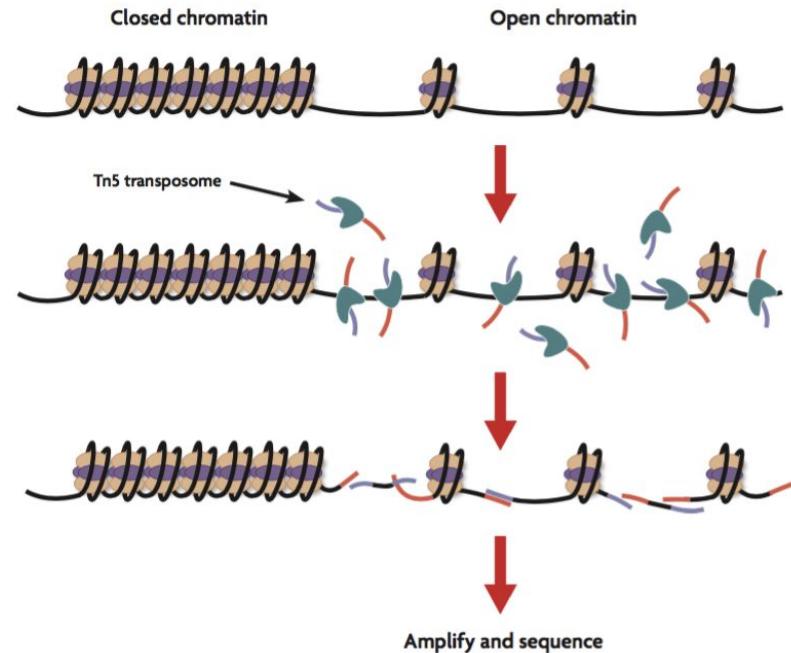


Assays to detect open chromatin

DNase-seq

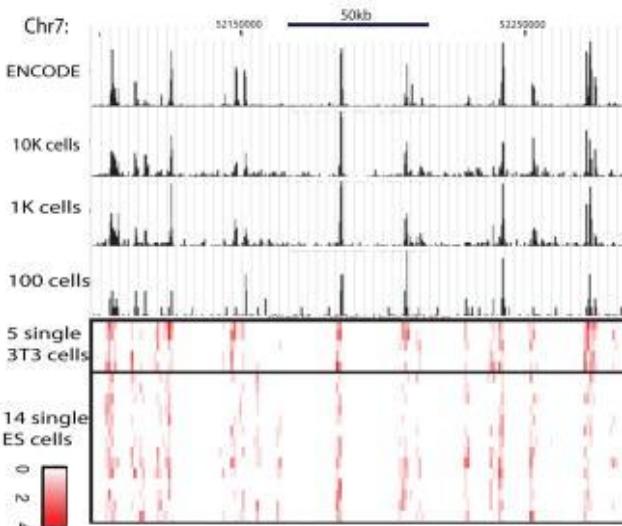


ATAC-seq



Single-cell open chromatin

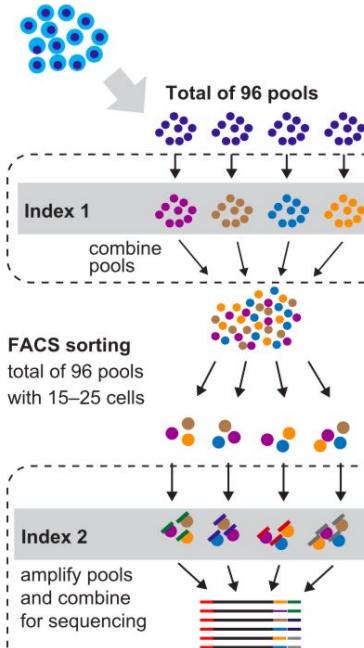
scDNase-seq / pico-seq
(<20 cells, 38.000 DHS per cell)



Jin, *Nature*, 2015

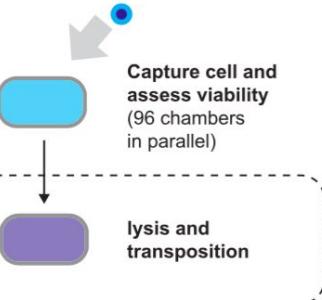
sciATAC-seq
(> 1000 cells, 450 DHS per cell)

scATAC-seq
based on cellular indexing
(Cusanovich et al, *Science*, 2015)



scATAC-seq
(100-1000 cells, 5,000 DHS per cell)

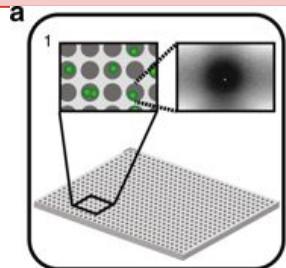
scATAC-seq
based on microfluidics
(Buenrostro et al, *Nature*, 2015)



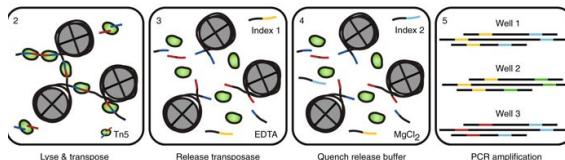
Pott et al, *Genom Biol*, 2015

Single-cell open chromatin

nano-well μATAC-seq
(2000 cells, ? DHS per cell)



5184 wells
Dispense & image
Select wells with 1 single living cell (microscopy)

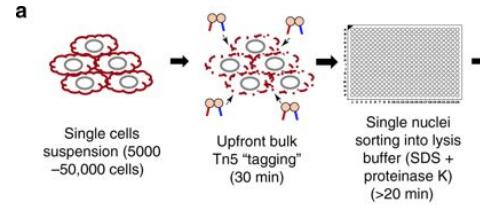


Lysis, transposition, PCR amplification in wells

Mezger et al, Nat Comm, 2018

Chen et al, Nat Comm, 2018

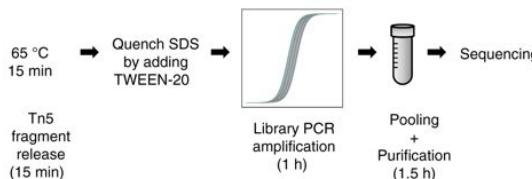
Plate-based scATAC-seq
(3648 cells, ? DHS per cell)



Standard 96/384-well plates
Tn5 tagging in bulk BEFORE sorting

droplet scATAC-seq
10x Genomics

Available end 2018



10X Genomics - scATAC-seq

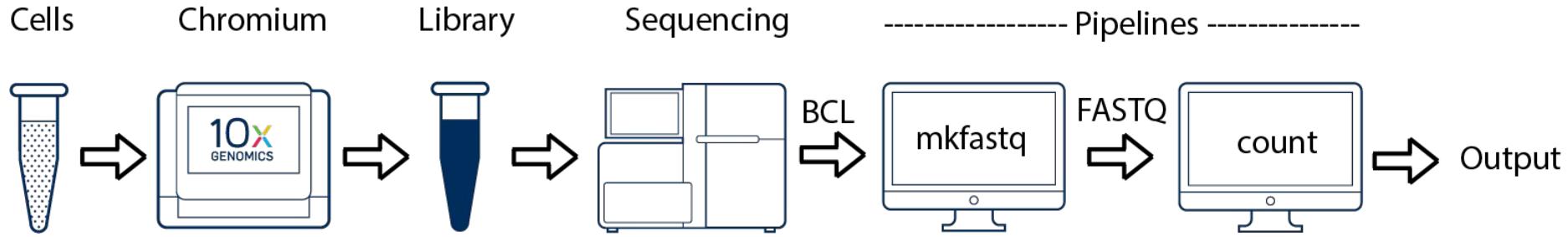
Our Solution Features

- Detect open chromatin regions in single cells with enriched signals in Transcription Start Sites (TSS) and regulatory regions
- Profile 500 to 10,000 nuclei per channel with our scalable, high-throughput solution
- Recover up to 65% of nuclei loaded on chips
- Investigate cell lines, primary cells, fresh, and cryopreserved samples with confidence
- Analyze your single cell epigenetic data with our turnkey analysis software and interactive visualization tools



[View the Workflow.](#)

Cell Ranger ATAC



Include a peak calling algorithm and generate a peak-barcode matrix
Dimension reduction using LSA (Latent Semantic Analysis)

LSA is a kind of PCA applied on a document-text matrix (which describes the frequency of terms that occur in a collection of documents) + tSNE

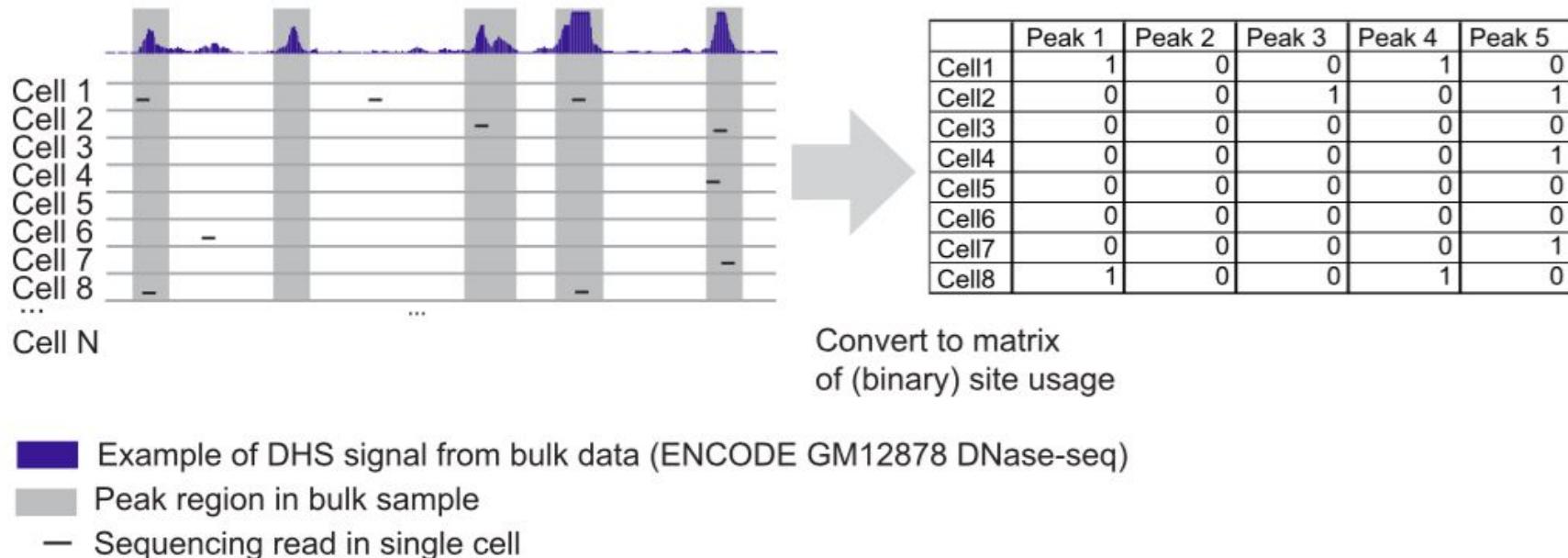
	I	like	hate	databases
D1	1	1	0	1
D2	1	0	1	1



	Barcode 1	Barcode 2	Barcode 3
Peak A	1	0	3
Peak B	0	1	0

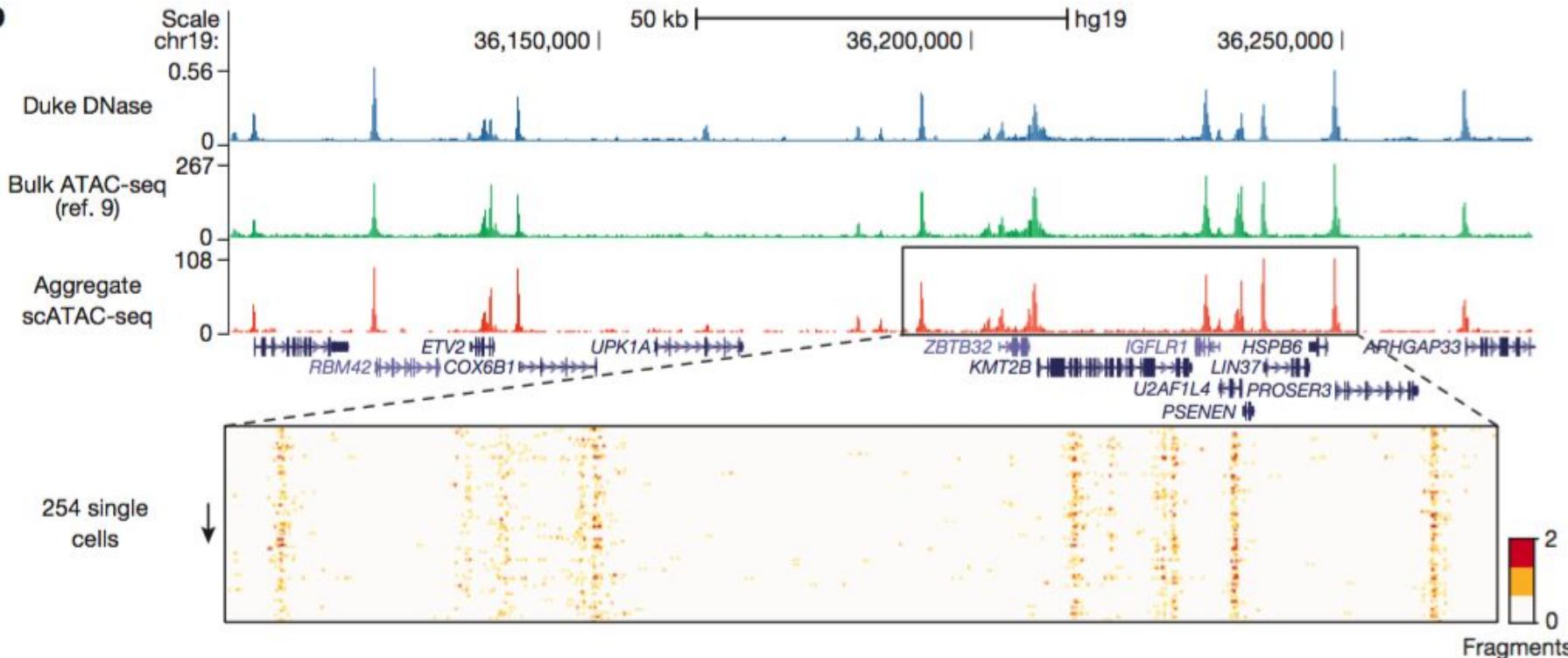
System requirements
16 cores
128 Go RAM
1 To free disk

Data processing in scATAC



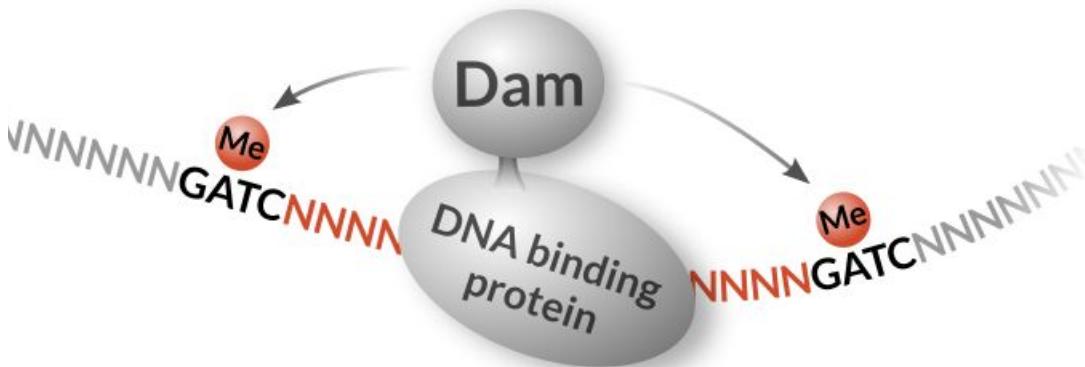
Aggregation of scATAC gives similar results to bulk data

b



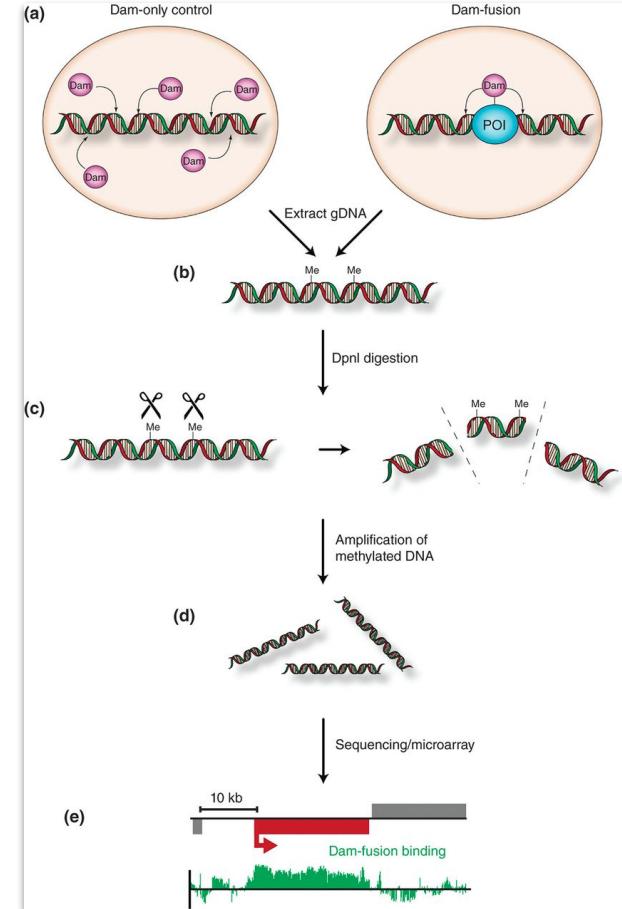
- Profiling the epigenome
 - Overview of single-cell epigenomics
 - Single-cell technologies : state of the art
 - scChIP-seq/Drop-ChIP
 - sc methylation
 - scHi-C
 - sc open chromatin
- Most popular techniques in single-cell epigenomics
- Low coverage, cell aggregation usually done for analysis (“bulk-like”)

One word about DamID-seq



DamID allows the identification of protein-binding sites in living cells without the need for crosslinking or immunoprecipitation.

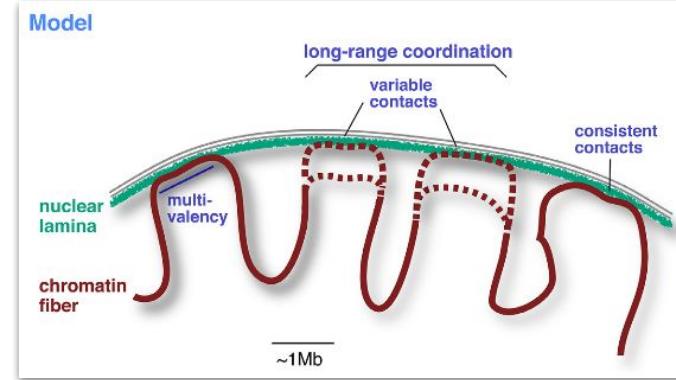
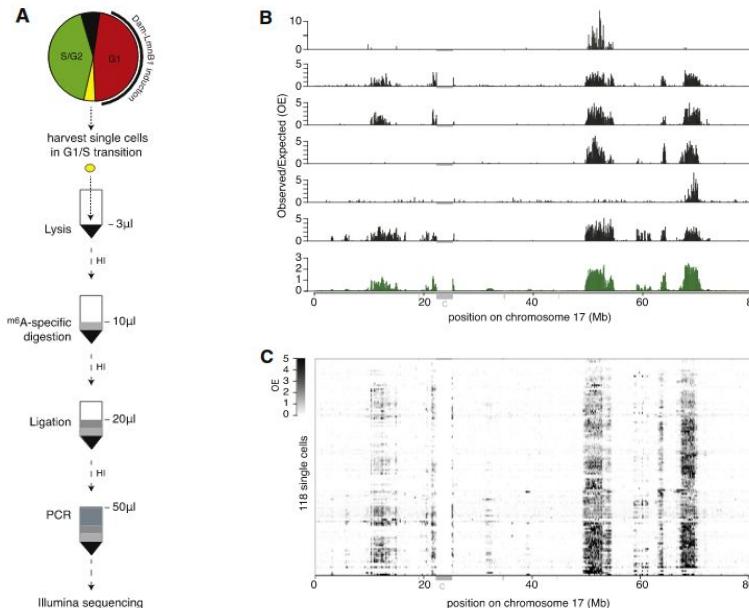
More information @ http://research.nki.nl/vansteensellab/DamID_FAQ.htm



Single-cell Dam-ID-seq

Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells

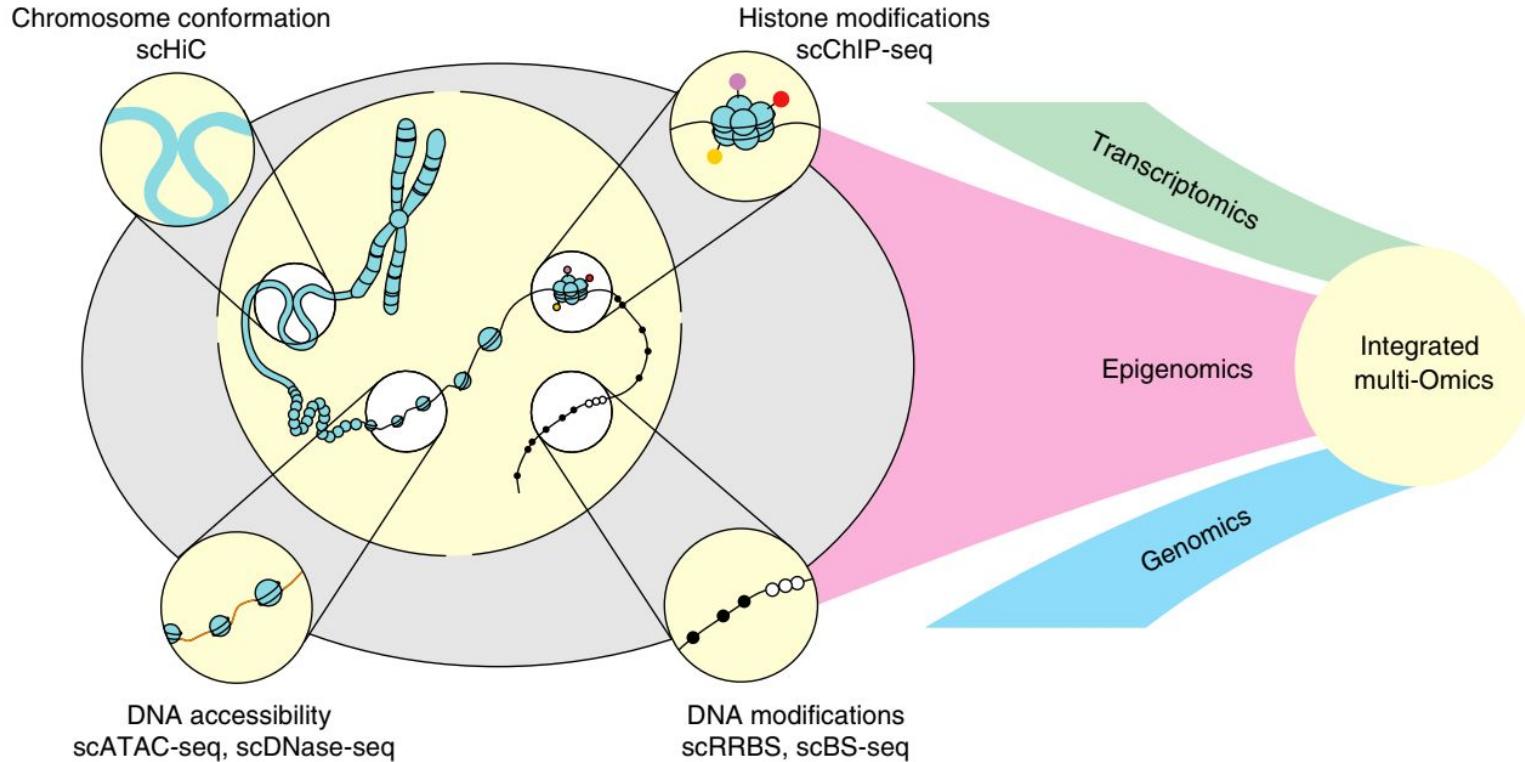
Jop Kind,^{1,2,*} Ludo Pagie,¹ Sandra S. de Vries,¹ Leila Nahidiazar,³ Siddharth S. Dey,² Magda Bienko,⁷ Ye Zhan,⁴ Bryan Lajoie,⁴ Carolyn A. de Graaf,^{1,10} Mario Amendola,¹ Geoffrey Fudenberg,⁵ Maxim Imakaev,⁶ Leonid A. Mirny,^{6,8} Kees Jalink,³ Job Dekker,⁹ Alexander van Oudenaarden,² and Bas van Steensel^{1,*}



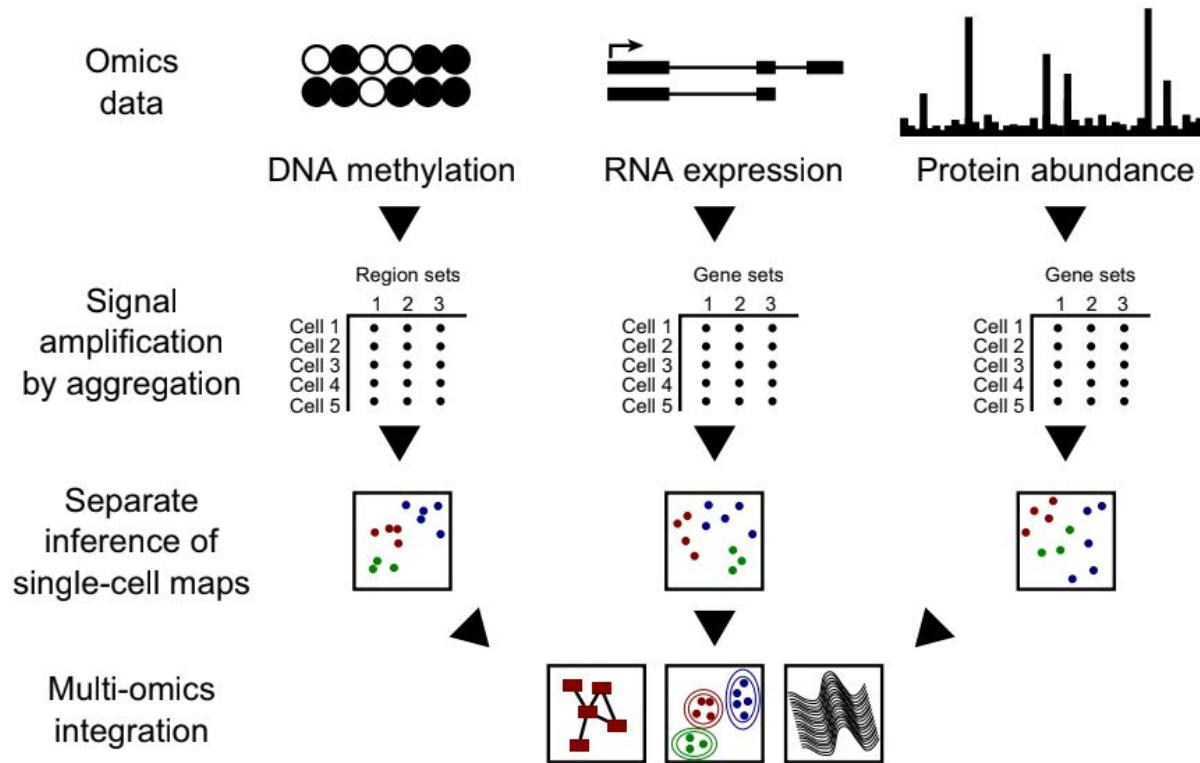
The DamID protocol seems to be applicable at the single cell level with minor modifications, although a matched control is not possible.

- Profiling the epigenome
- Overview of single-cell epigenomics
- Single-cell technologies : state of the art
 - scChIP-seq/Drop-ChIP
 - sc methylation
 - scHi-C
 - sc open chromatin
- Combining multiple assays (Multi omics)

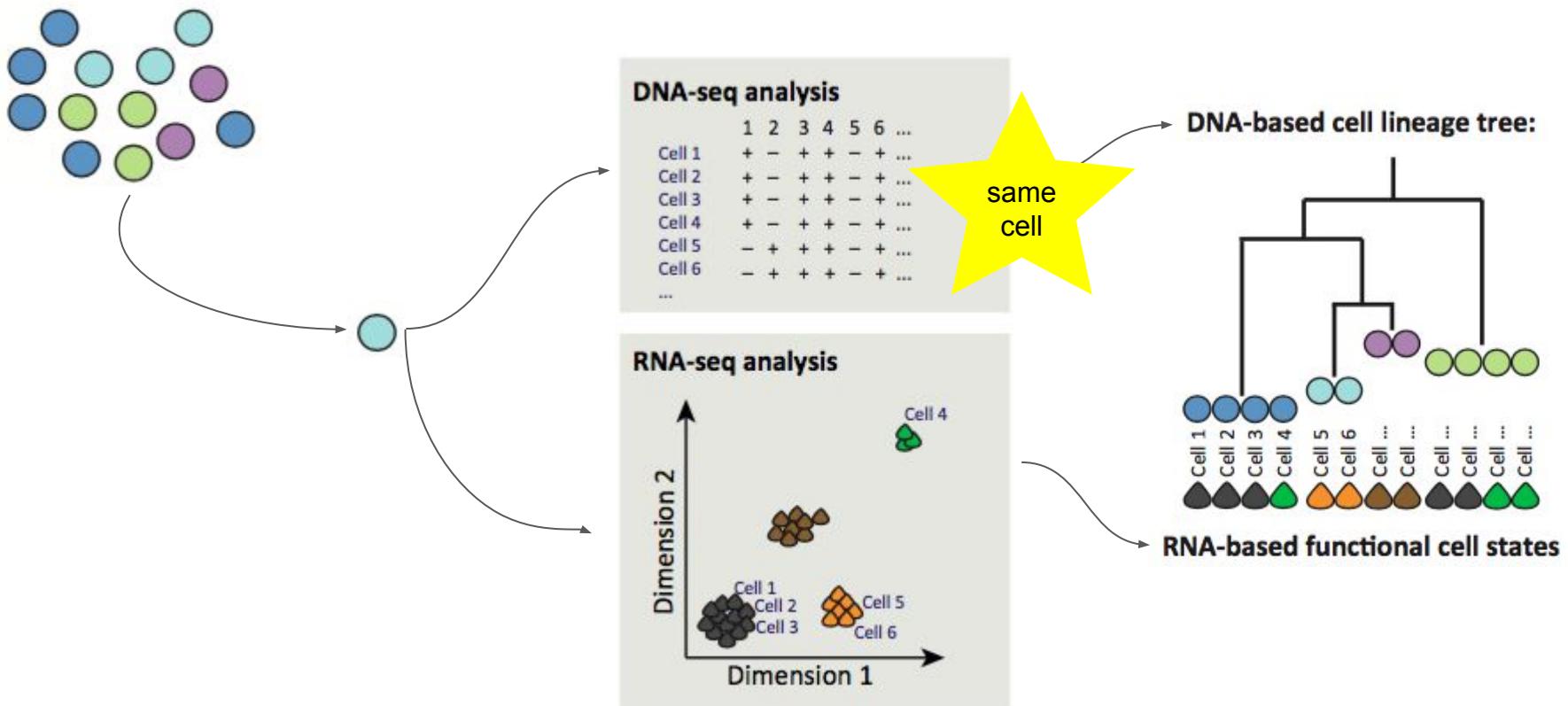
And what about sc multi-omics ?



And what about sc multi-omics ?



And what about sc multi-omics ?



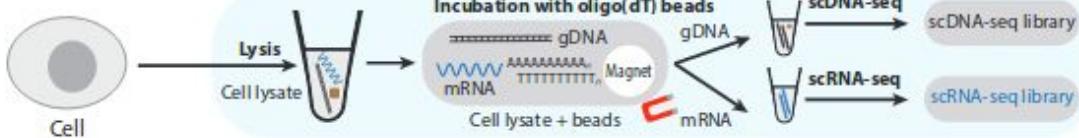
Inspired from Macaulay, Trends in Genetics, 2017

And what about sc multi-omics ?

RNA	Chromosome accessibility	scCAT-seq (<i>Liu et al, Nat Com, 2019</i>) sci-CAR (<i>Cao et al, Science, 2018</i>)	
RNA	Methylome	scMT-seq (<i>Youjin Hu et al, Genome Biol, 2016</i>) scM&T-seq (<i>Angermueller et al, Nat Methods, 2016</i>)	
RNA	Methylome	DNA	scTrio-seq (<i>Yu Hou et al, Cell Research, 2016</i>) scGEM (<i>Cheow et al, Nat Methods, 2016</i>)
RNA	Methylome	Chromosome accessibility	scNMT-seq (<i>Clark et al, Nat Com, 2018</i>) https://www.biorxiv.org/content/10.1101/519207v1
RNA	DNA	G&T-seq (<i>Macaulay et al, Nat Methods, 2015</i>) DR-seq (<i>Dey et al, Nat Biotech, 2015</i>)	
DNA	Methylome	Chromosome accessibility	scCOOL-seq (<i>Guo et al, Cell Research, 2017</i>) scNOME-seq (<i>Pott, eLife, 2017</i>)
RNA	TF binding	scCC-seq (<i>Moudgil et al, bioRxiv, 2019</i>)	
RNA	Protein	CITE-seq (<i>Stoeckius et al, Nat Methods, 2017</i>) REAP-seq (<i>Peterson et al, Nat Biotech, 2017</i>)	
Protein	Chromosome accessibility	Pi-ATAC (<i>Chen et al, Nat Com, 2018</i>)	

And what about sc multi-omics ?

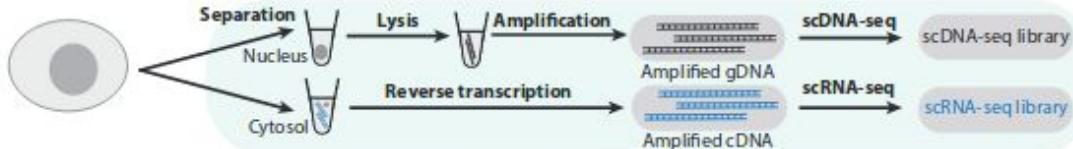
Separation of DNA and RNA (e.g., G&T-seq and scM&T-seq)



Advantages and limitations

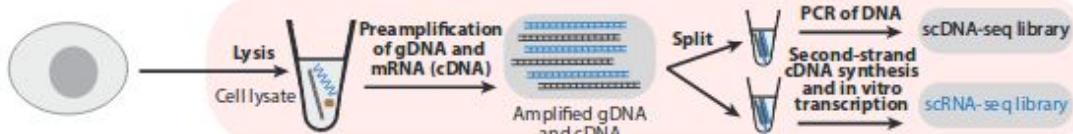
- scDNA-seq and scRNA-seq protocols are easily modified
- Amenable to bisulfite sequencing
- Potential loss of molecules

Separation of nucleus and cytoplasm (e.g., scTrio-seq and scMT-seq)



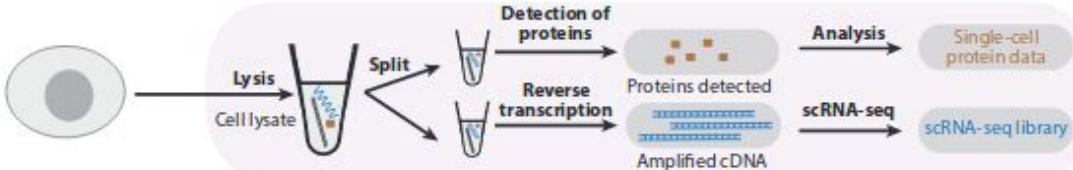
- scDNA-seq and scRNA-seq protocols are easily modified
- Amenable to bisulfite sequencing
- Loss of cytoplasmic and all nuclear mRNA molecules

Process without separation (e.g., DR-seq)



- Minimal risk of loss of molecules (as preamplified in the first vessel)
- cDNA contamination in gDNA library
- Not amenable to bisulfite sequencing

Splitting of cell lysate, such as protein and RNA (e.g., PEA)



- Amenable to RNA and protein measurements in parallel
- Significant loss of molecules

Reviews multi-omics

Review

Single-Cell Multi-omics: An Engine for New Quantitative Models of Gene Regulation

Jonathan Packer¹ and Cole Trapnell^{1,*}

2018

REVIEWS

2019



Integrative single-cell analysis

Tim Stuart¹ and Rahul Satija^{1,2*}

SPECIAL SECTION

SINGLE-CELL GENOMICS

2018

REVIEW

Single-cell epigenomics: Recording the past and predicting the future

Gavin Kelsey,^{1,2,*†} Oliver Stegle,^{3,4,*†} Wolf Reik^{1,2,5†}

Annual Review of Genomics and Human Genetics
Single-Cell (Multi)omics
Technologies

2018

Lia Chappell,^{1,*} Andrew J.C. Russell,^{1,*}
and Thierry Voet^{1,2}

Take-home messages

- When

2013 - 2015: first assays for each technology

2016 - 2018: increased number of cells, multi-omics

- Feasibility

Technology not widely accessible

10x Genomics scATAC-seq kit or common plate-based protocol likely to provide wider access



- Bioinformatics analysis

Many studies still rely on **pooled cells** analysis. Due to low coverage, matrices are often **binarized**. Used in combination with low-input/bulk complementary data

Will likely change in the upcoming years !

Further readings

- Reviews
 - Fiers et al, *Briefings in Functional Genomics*, 2018
 - Pang-Kuo Lo and Qun Zhou, *J Clin Genom*, 2018
 - Kelsey et al, *Science*, 2017
 - Wen et al, *Molecular Aspects of Medicine*, 2017
 - Ortega et al, *Clinical and Translational Medicine*, 2017
 - Clark et al, *Genome Biology*, 2016
 - Pott et al, *Genome Biology*, 2015
 - Hyun et al, *Methods*, 2015
- Selected research papers
 - Cusanovich et al, *Nature*, 2018
 - Luo et al, *Science*, 2017
 - Lake et al, *Nature*, 2017
 - Corces et al, *Nature*, 2016