

# FY2023 Current Trends in Bioinformatics

## Lecture 2

*Long non-coding RNAs: Genomic Junk or Regulatory Treasure?*

**Jordan RAMILOWSKI, PhD**

Bioinformatics Associate Professor



23-10-2023

# Jordan RAMIŁOWSKI

## Yokohama City University 2020→

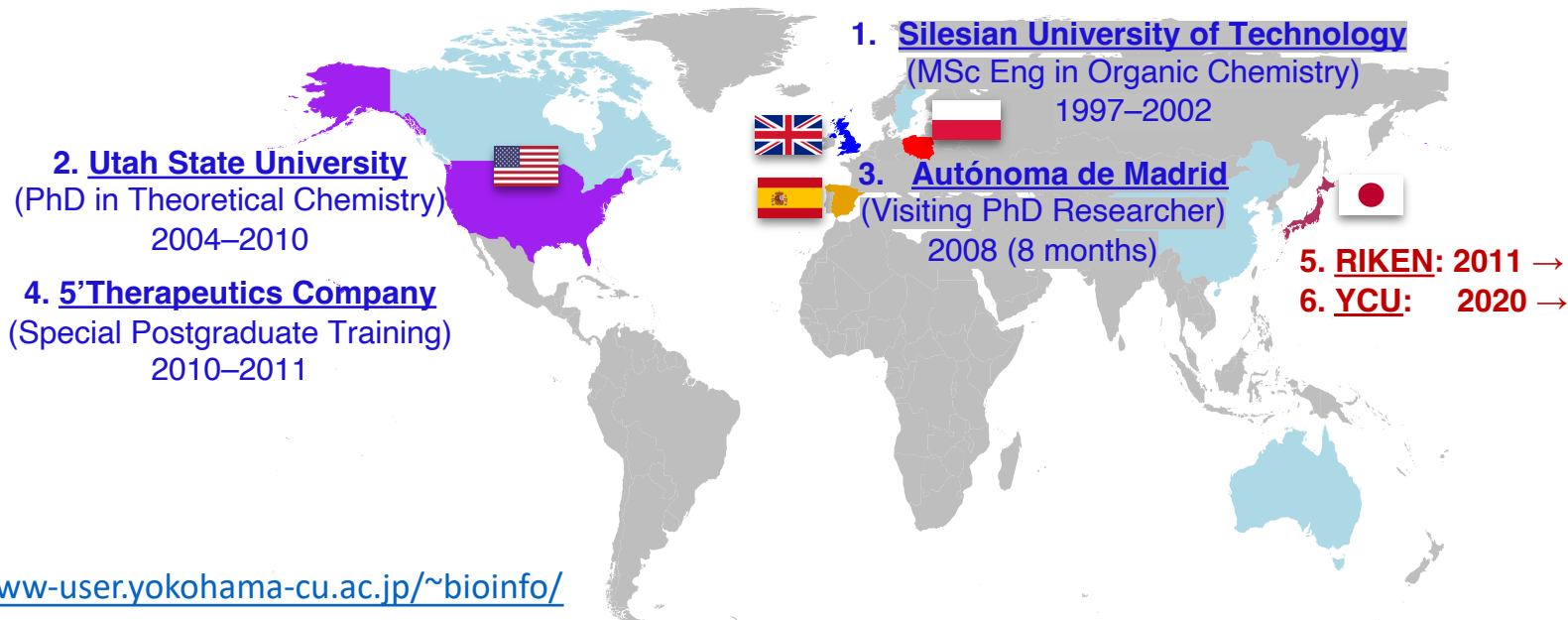
Bioinformatics Associate Professor (バイオインフォマティクス 准教授)

[long non-coding RNAs, cell differentiation, cell-cell communication, enhancers, and more]

## RIKEN Yokohama IMS 2011→

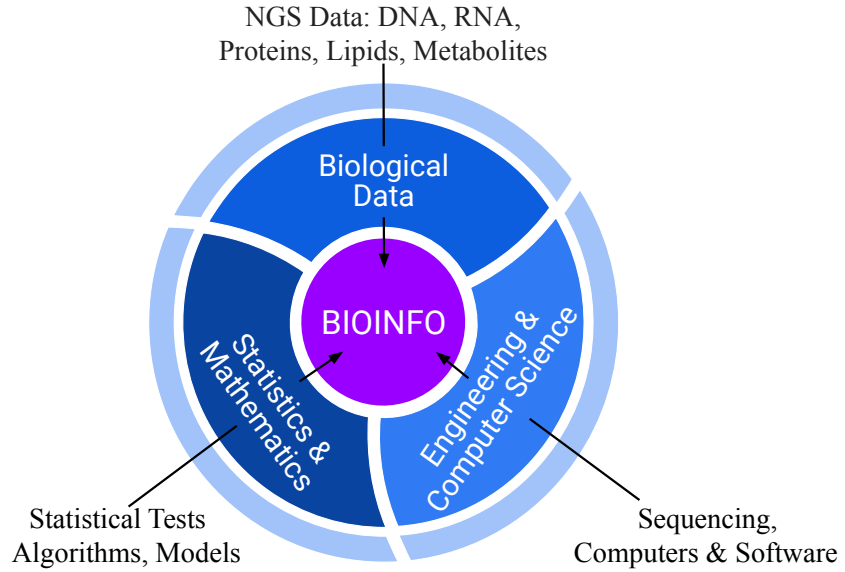
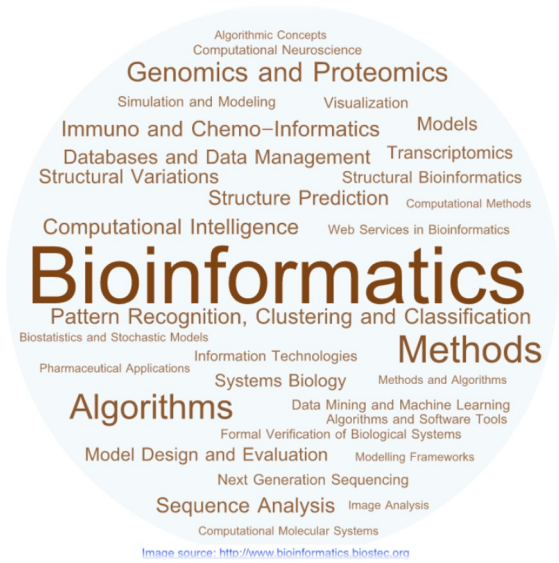
Postdoctoral Researcher/Research Scientist 2011—2020

Visiting Scientist 2020→



# Bioinformatics: Definition & Utility

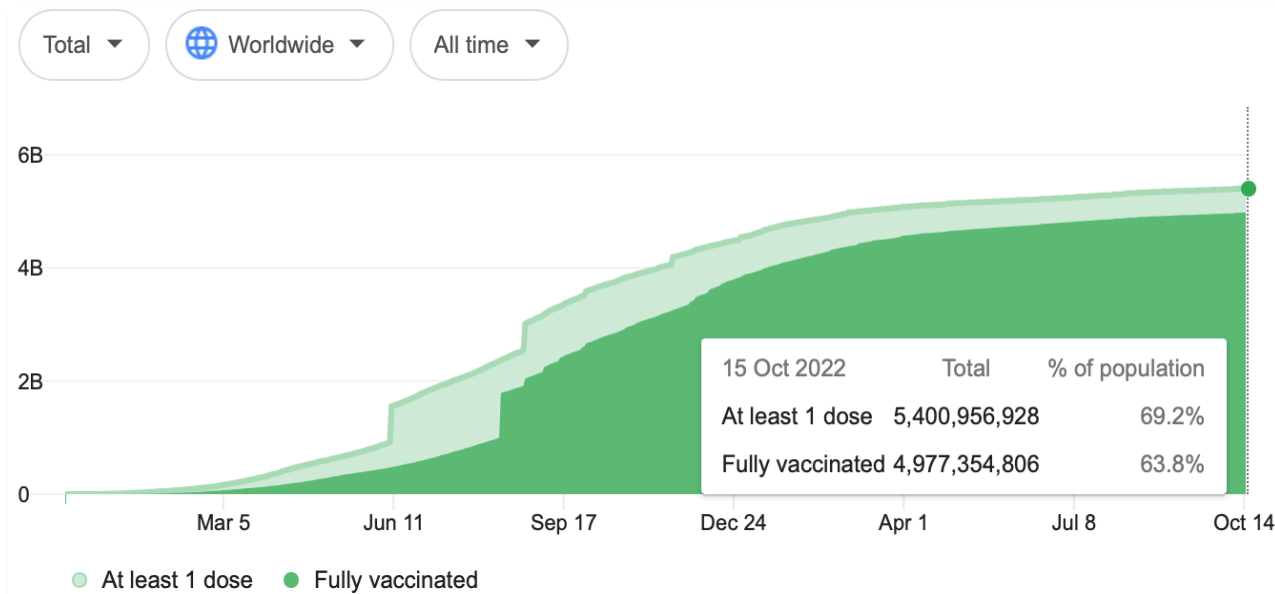
**Bioinformatics:** broadly defined study of living systems (**bio**) using computational (**informatics**) tools



**Bioinformatics (BI): applied science** that interprets biological data using development and advancements in other sciences to **solve many important medical and biological problems.**

# COVID-19 Pandemics

COVID-19 (2019-2022) ~ 6.5M (0.1%) global population died  
*Spanish flu* (1918-1920) ~50.0M (2.5%) global population died



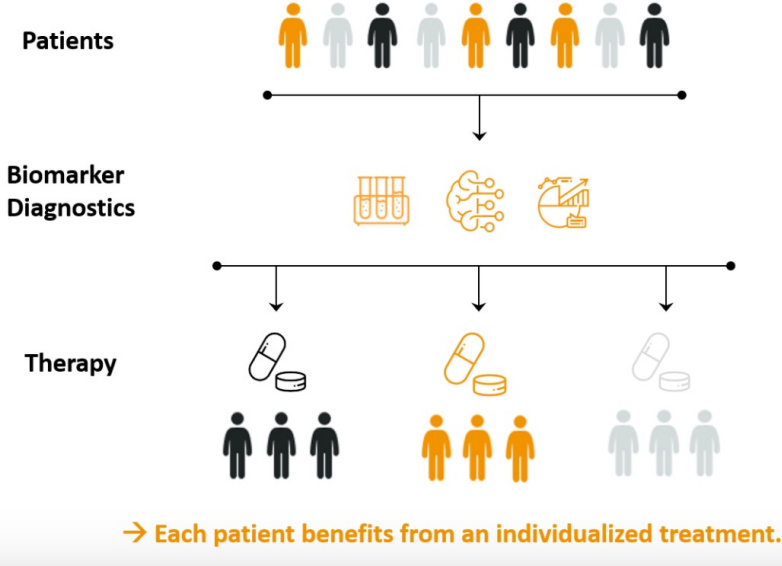
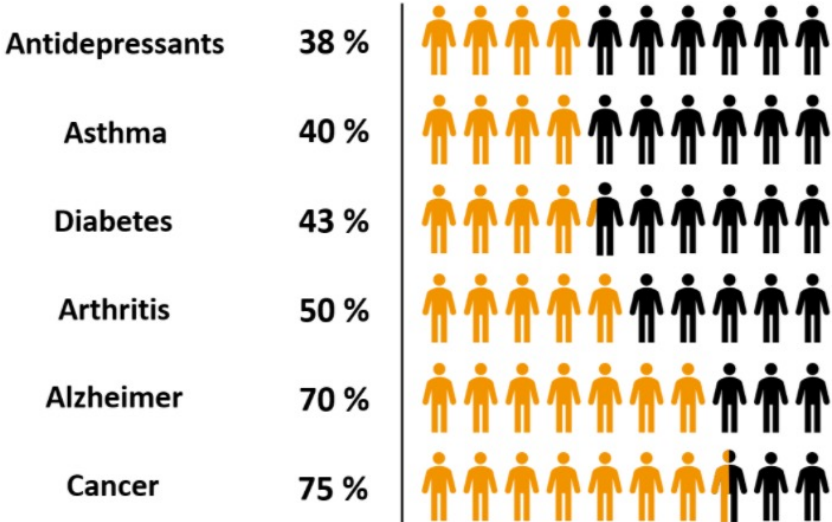
**Bioinformatics was crucial in vaccine and anti-COVID 19 drugs development!**

# Development of individual & effective treatments

**Current:** drugs designed using data from selected having many side effects and are often ineffective.

**Goal:** diagnosis of individual patients & treatment with most effective drugs and without side effects.

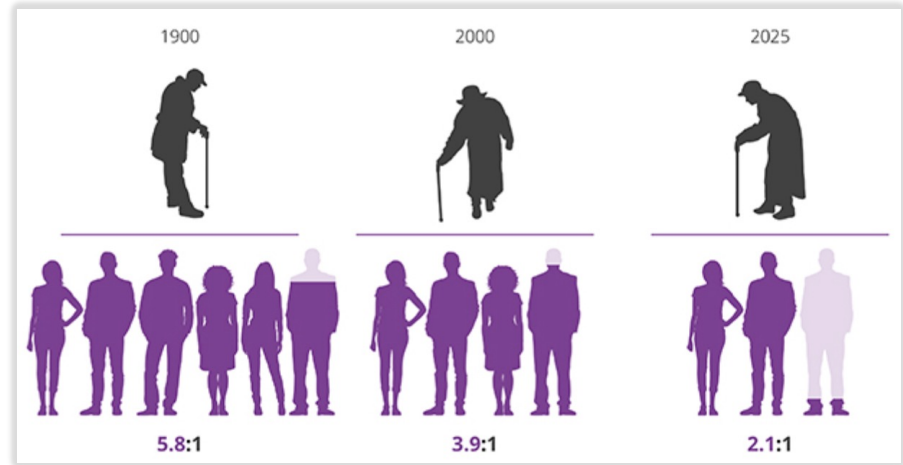
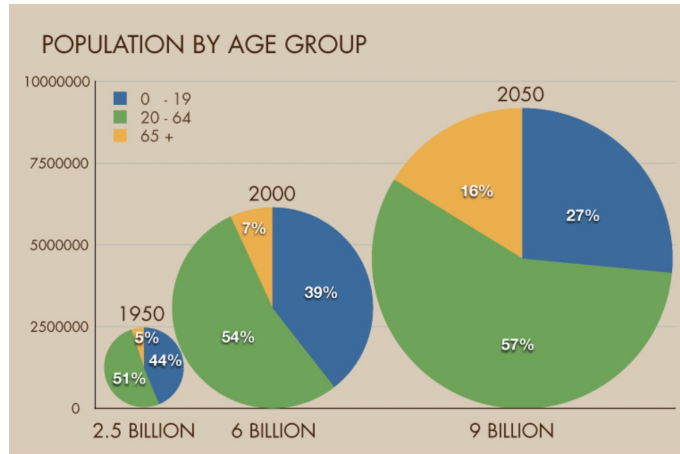
## Ineffective Drugs:



**Bioinformatics is playing a major role in sequencing and analyzing individual genomes and in research developing personalized drugs and therapies!**

# Helping with increasing and ageing population

Total world population is constantly growing and many developed societies are aging, negatively affecting economic growth . This forces governments to increase the retirement age, while keeping 'older' workers healthy and productive.



**Bioinformatics is playing a major in helping to develop strategies to help handle increasing world population and is important in ageing research!**

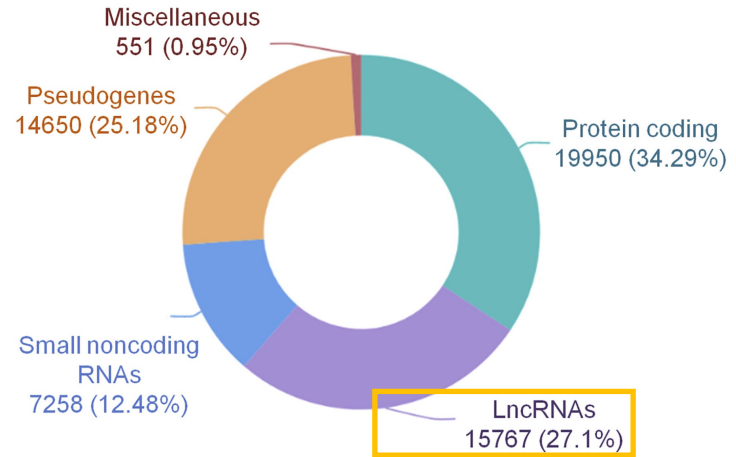
# Today we will talk about long non-coding RNAs...

**lncRNAs : transcripts longer than 200 & without protein coding potential.**

## Features of long non-coding RNAs

- transcribed mostly by RNA polymerase II
- many are capped, spliced and polyadenylated
- **very abundant in mammalian genomes**
- **poor sequence conservation across species**
- **nearby by the same mRNA across species**
- **lowly expressed & highly cell type specific**

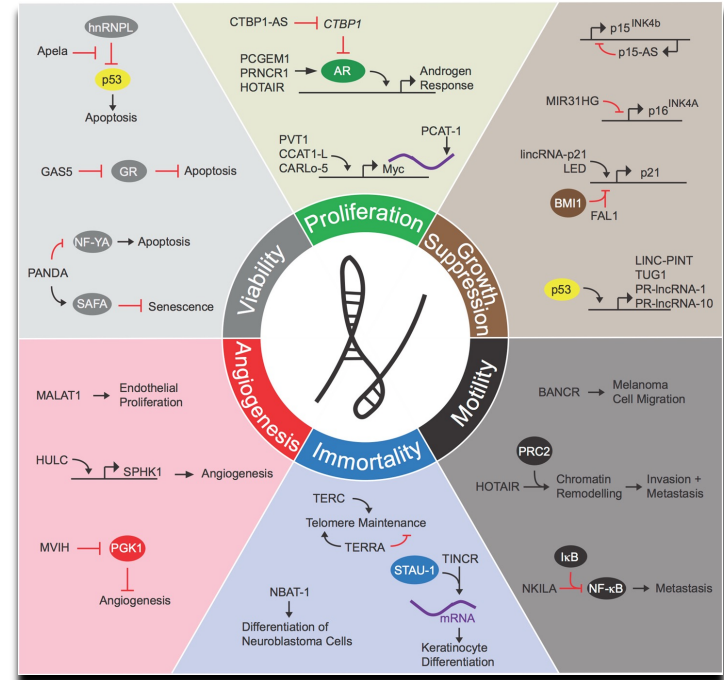
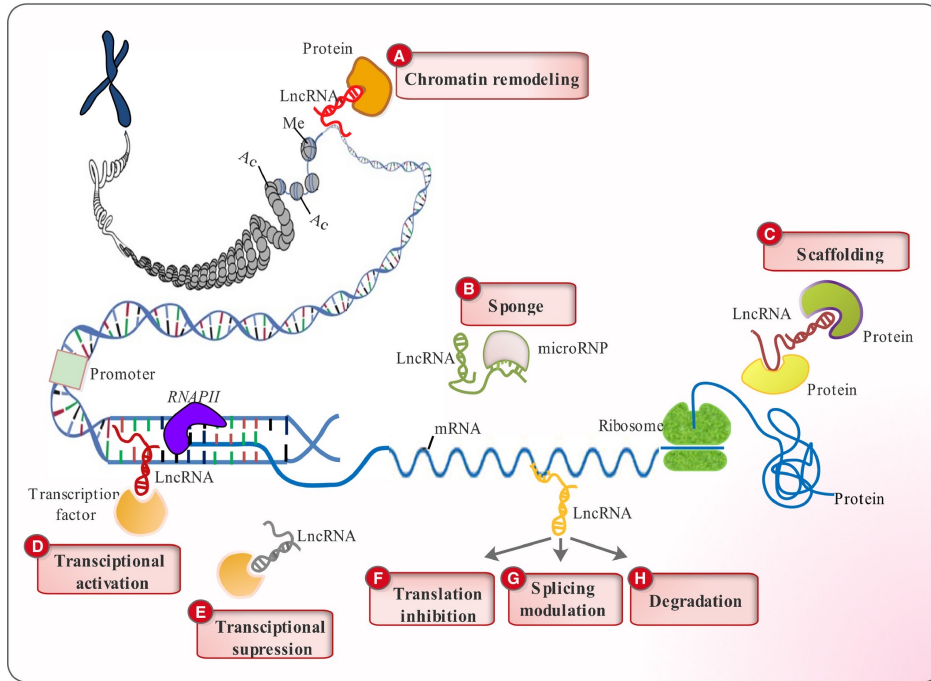
## Most human/mouse genes are non-coding



Mouse Genome Database (Gencode)  
showing different classes of mouse genes.

# Selected lncRNAs show diverse regulatory functions

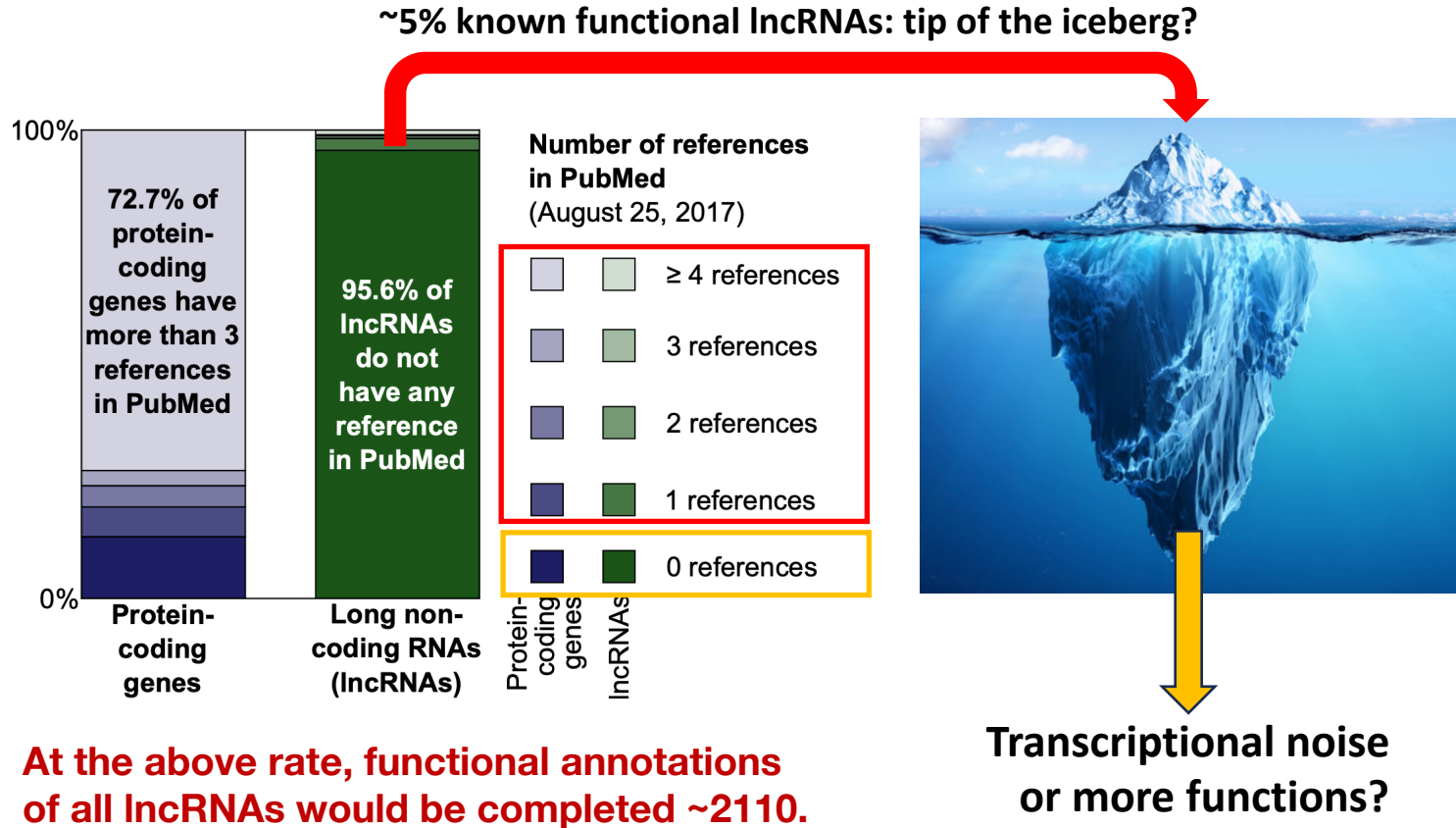
lncRNAs can function in a variety of molecular process including regulation in cancer.



Salehi S. *et al.*, *Journal of Cellular and Molecular Medicine* 21, 3120, (2017)



# lncRNAs are NOT studied enough...

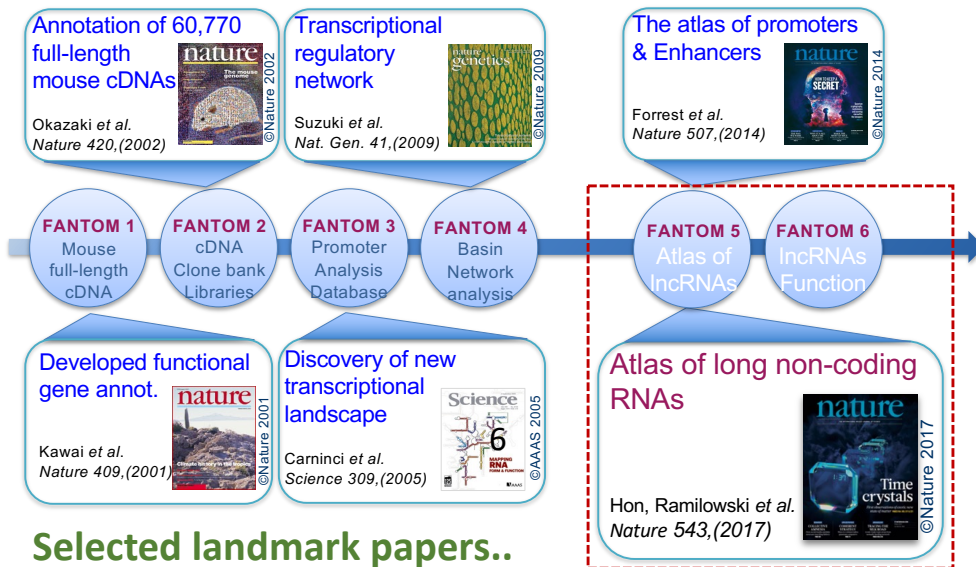


# FANTOM: Functional Annotation of Mammalian Genomes

Worldwide Genomic Consortium led by RIKEN Yokohama since 2000



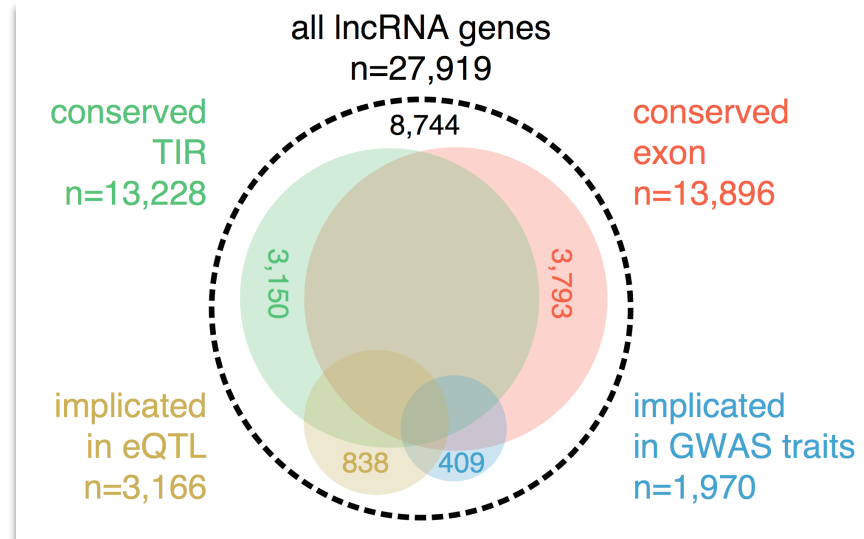
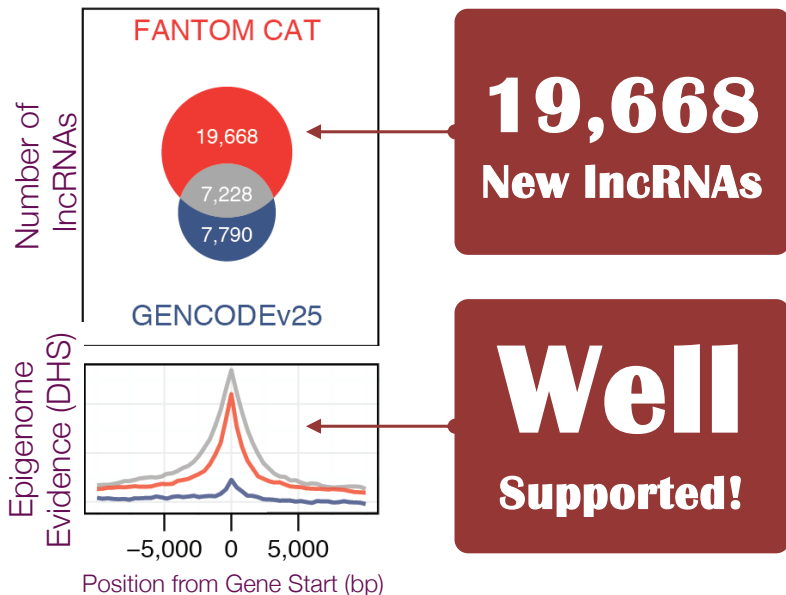
FANTOM Consortium 2017 Summer Meeting:  
*Cracking mysteries of human lncRNAs*



# FANTOM 5: Computational Atlas of lncRNA Functions



FANTOM CAT (CAGE Associate Transcriptome) of 27,919 human lncRNAs:  
~10,000 CAGE/RNA-seq expression profiles + Epigenome Data + Annotations



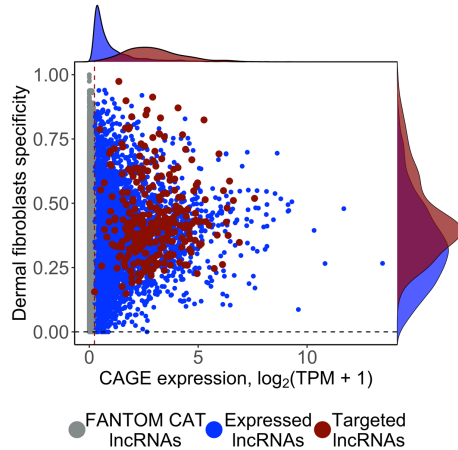
Hon, Ramilowski, et al. *Nature* 543, 2017

67% of all human lncRNAs (19,175/27,919) show various functionally relevant traits.

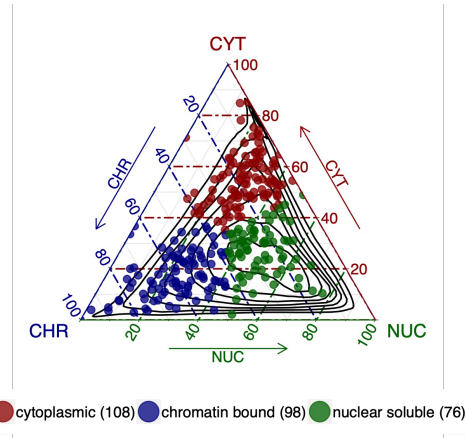
# FANTOM 6: Functional Screenings of lncRNAs

## 1. Unbiased selection of 600+ lncRNAs:

- Induced Pluripotent Stem Cells (~300 lncRNAs)
- Human Dermal Fibroblasts (~300 lncRNAs)
- other cell types (more limited)



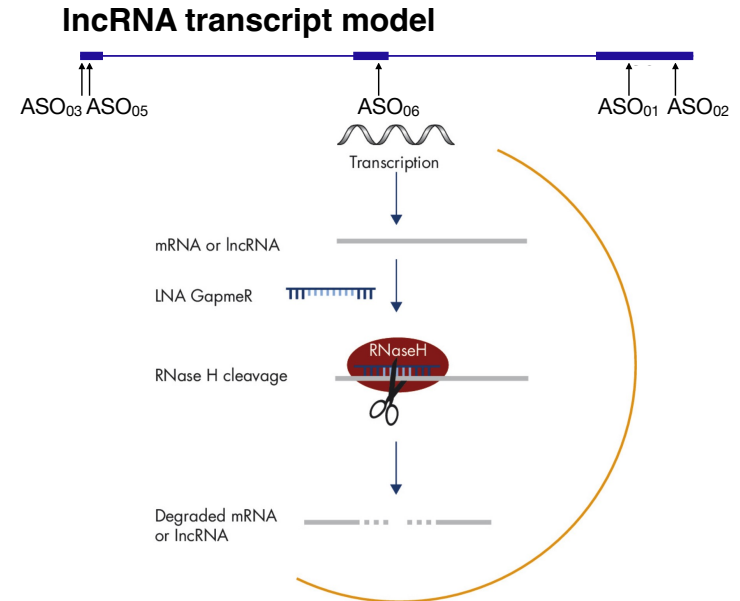
**VARIABLE EXPRESSION LEVELS**  
(CAGE EXPRESSION DATA)



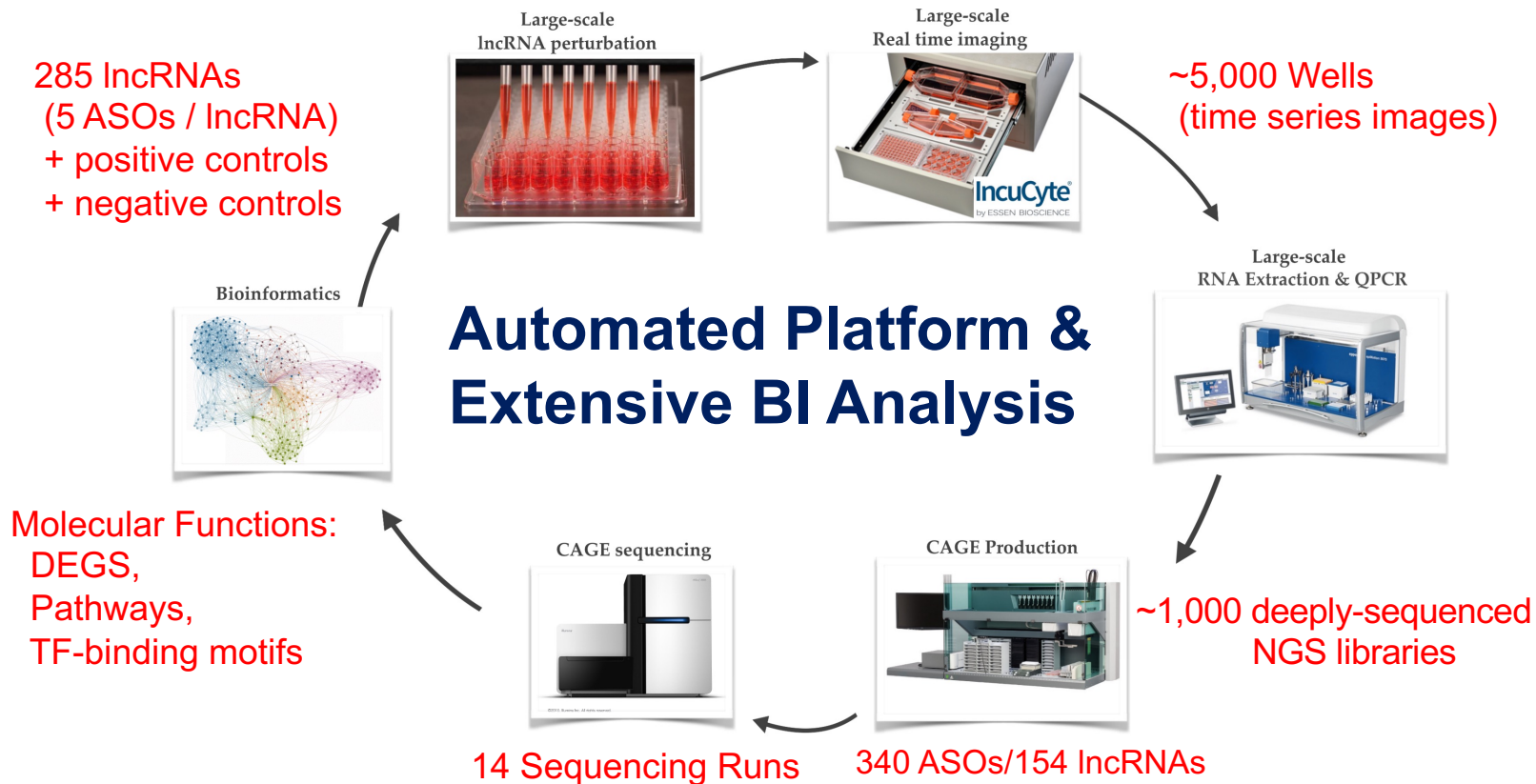
**62% NUCLEAR & 38% CYTOPLASMIC**  
(CELL RNA FRACTIONATION DATA)

## 2. Antisense Oligo (ASO) knockdown mechanism:

- ASO is ~12-18 nt long DNA sequence
- designed to bind to a selected RNA transcripts
- more

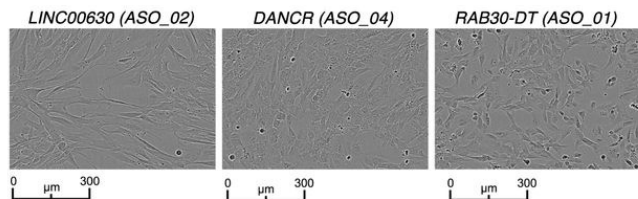


# FANTOM 6: Human Dermal Fibroblast study overview

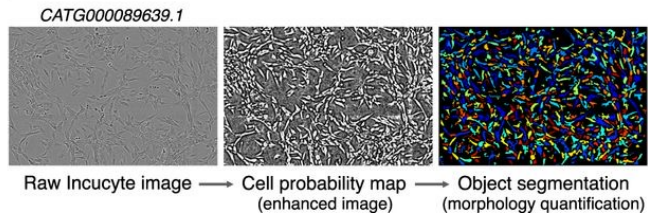


# ~30% lncRNAs Regulate Cell Growth & Morphology

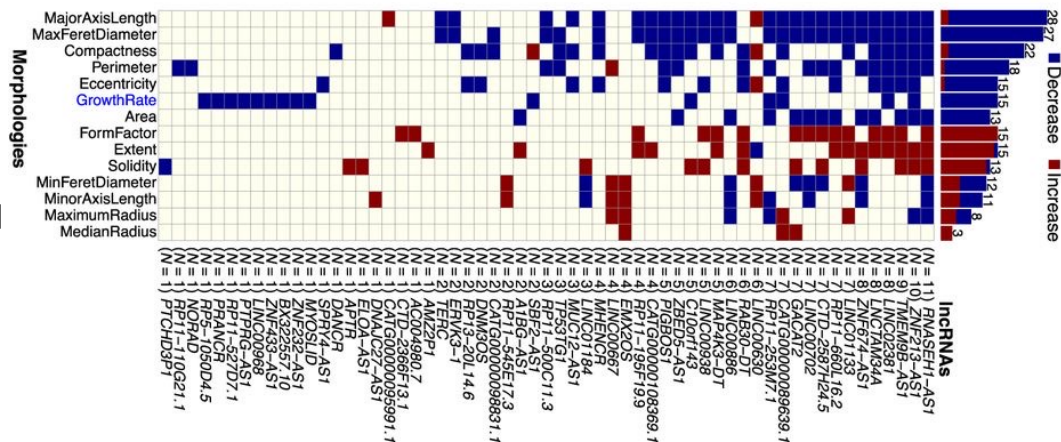
## 1. Distinct cell morphologies changes (example: three selected lncRNAs)



## 2. Morphology imaging processing using AI (example: one novel lncRNA target)

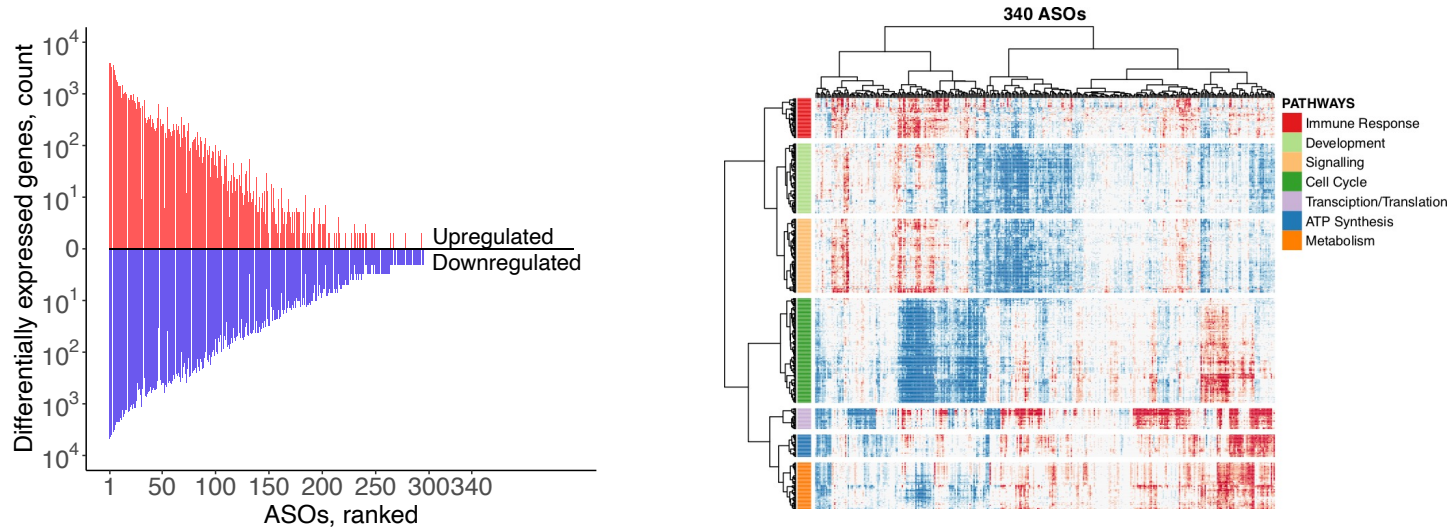


## 3. lncRNAs can regulate multiple morphologies (all lncRNA in HDF data)



# IncRNAs show diverse molecular signatures of functions

*Molecular functions of IncRNAs, were inferred by comparing transcriptome after each ASO KD targeting one IncRNA with the transcriptome of matching negative controls.*

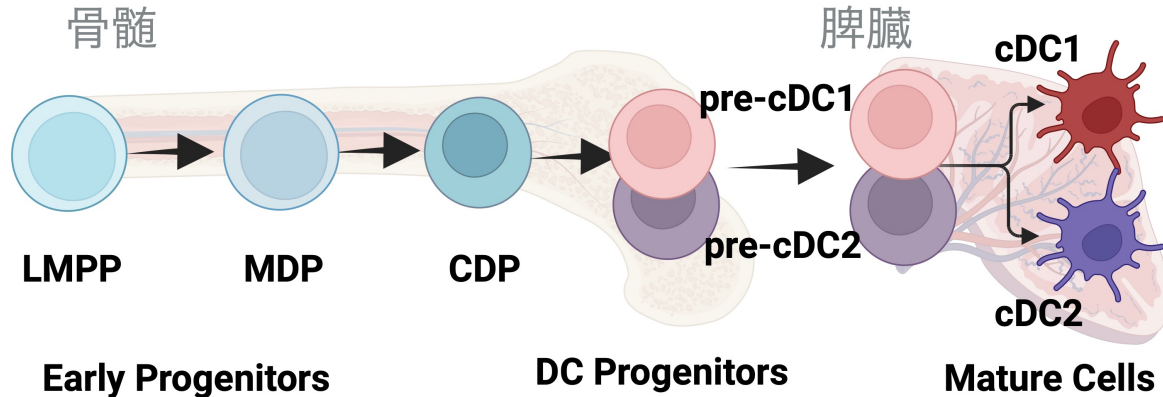


**IncRNA ASO KDs showed a wide ranged of DE genes (left) & dysregulated pathways (right). Overall ~20% IncRNAs show biological function signatures.**

# Dendritic Cells: Function & Differentiation

**DCs:** are professional Antigen Presenting Cells priming regulatory and cytotoxic T-cells to orchestrate a variety of adaptive immune responses upon infection and in cancers.

DCs differentiate mainly in bone marrow and mature in spleen in an IRF8-dependent manner.



1. **LMPP:** Lymphoid-Myeloid Primed Progenitors
2. **MDP:** Monocyte Dendritic Cell Progenitors

3. **CDP:** Common Dendritic Cell Progenitors
4. **pre-cDC1/pre-cDC2:** pre-Dendritic Cells
5. **cDC1/cDC2:** classical Dendritic Cells

How can we find lncRNA expression in our data?

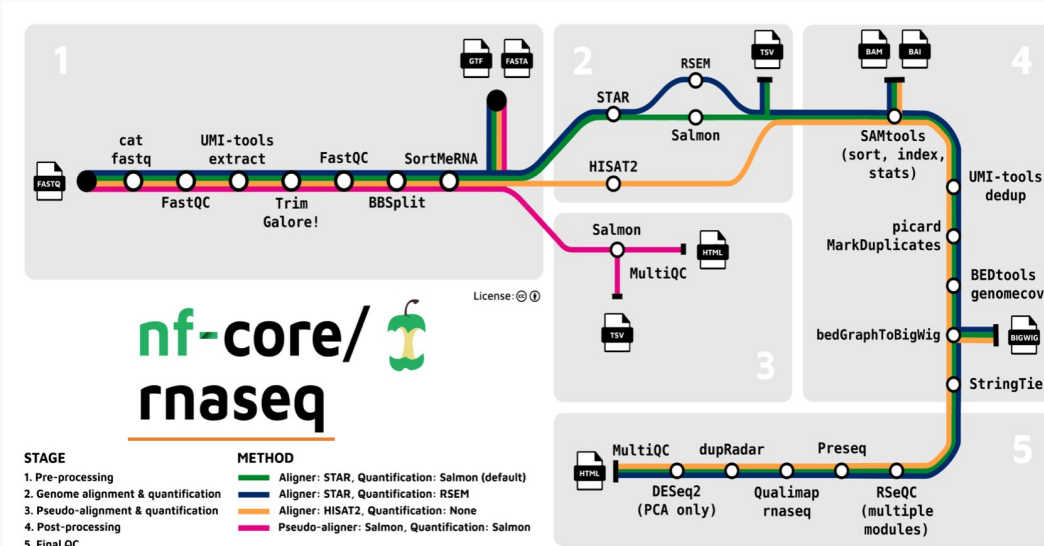


# RNA-seq Data Analysis Pipelines

Growing & diverse NGS data is NOT only a challenge for the hardware, but also requires BI pipelines to process and analyze the data reliably and reproducibly using the best standards in the field.

**nf-core** is a collection of 86 pipelines developed first at the Genome Institute of Sweden to process essentially any NGS data on Amazon Cloud or on own server. Anybody can develop a nf-core pipeline if it complies with the standards.

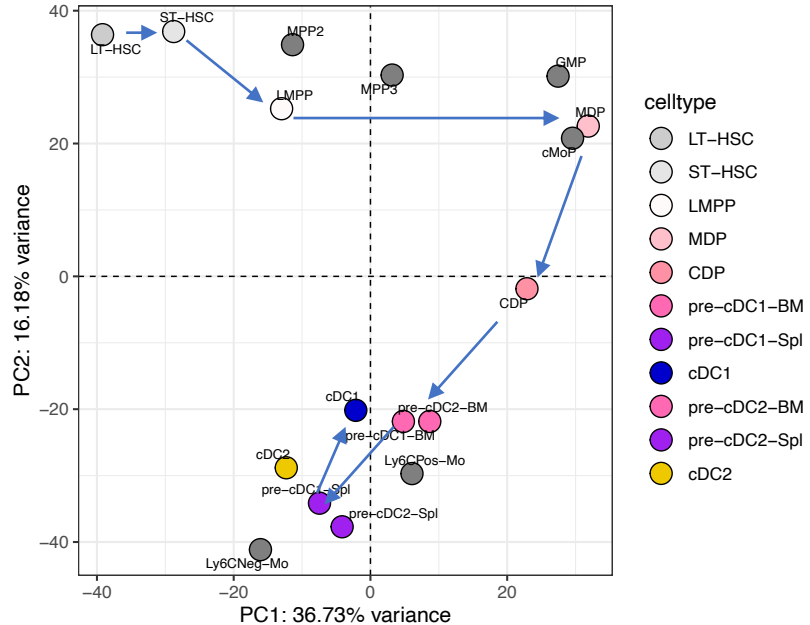
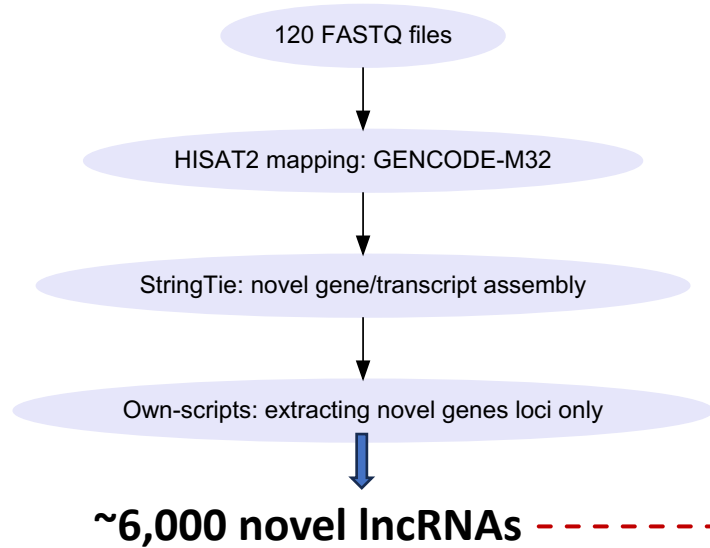
## Pipeline summary



Server: 2x2TB RAM / 240CPUs

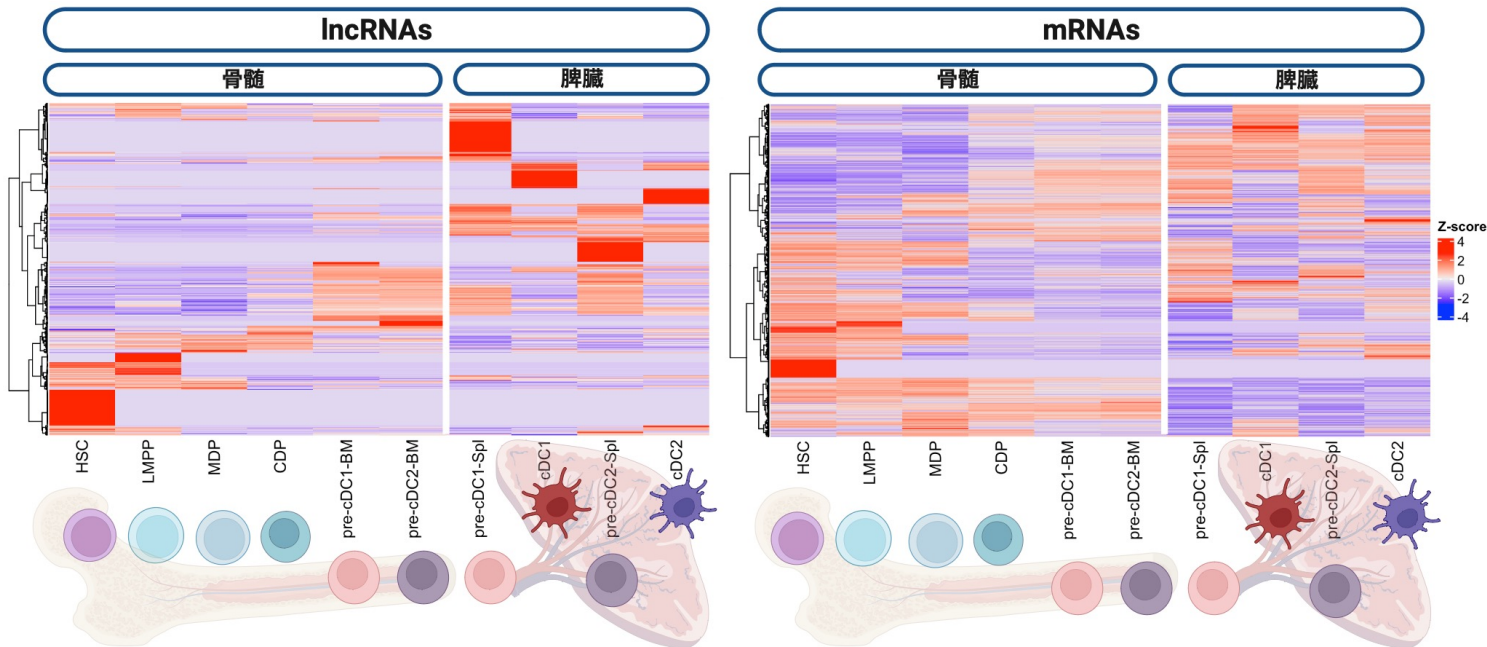
# We found thousands of lncRNA specific to DC differentiation

**de-novo gene assembly:** applying nf-core & own bioinformatics pipelines to RNA-seq in-vivo data we found ~6,000 novel lncRNAs in cDC differentiation time-course.



# lncRNAs are much more cell type specific than mRNAs

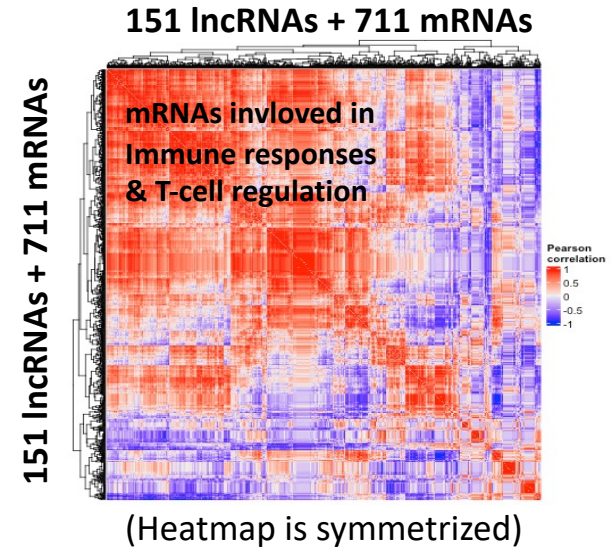
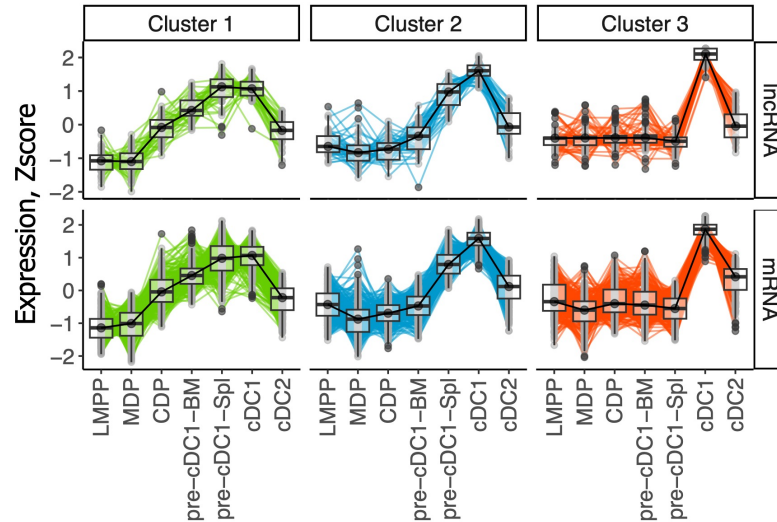
When plotting scaled across cell types expression (Z-score) of lncRNA (left) & mRNAs (right) across differentiation data we notice more restricted patterns of lncRNA expression.



Let's focus on DC1 specific mRNAs and lncRNAs.

# cDC1-specific mRNAs & lncRNAs form a co-regulatory network

Differential expression analysis and clustering identified 151 lncRNAs and 711 mRNAs specific to cDC1. We next looked at their expression profiles across cDC1 differentiation.

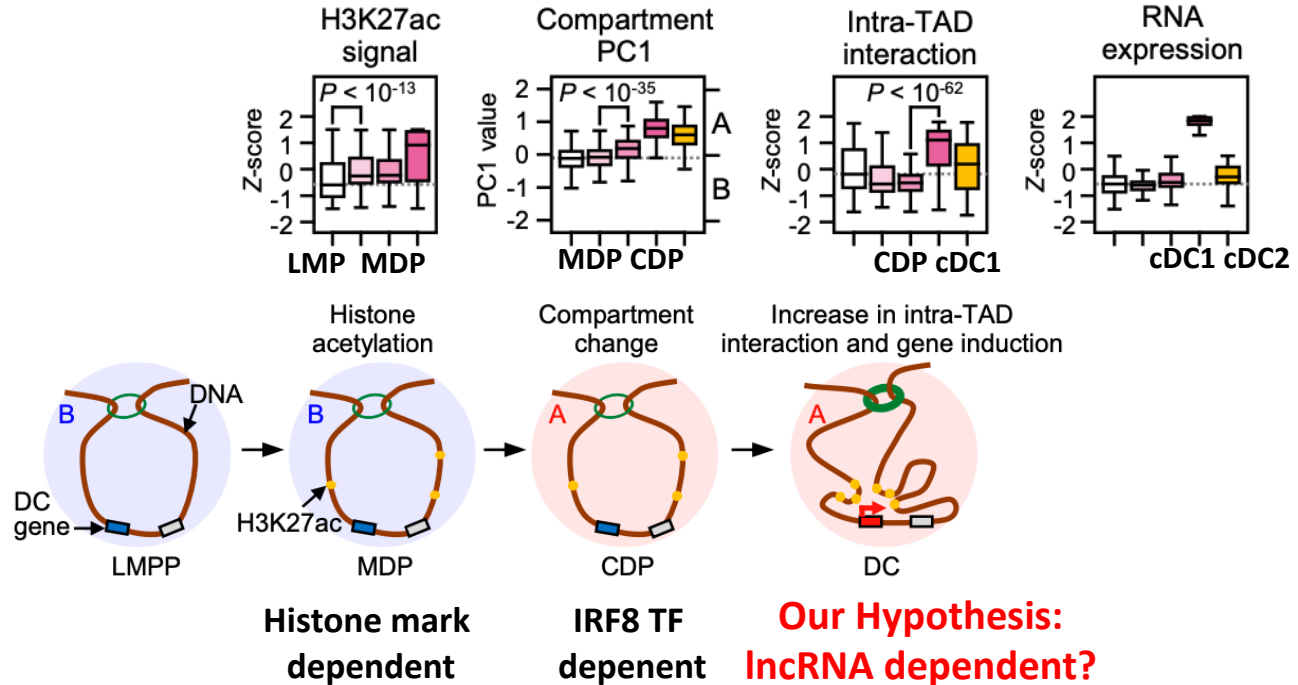


Many cDC1-specific lncRNA & mRNAs form a highly-correlated expression network in cDC1 differentiation. Usually, such *quilt by association* suggests that genes have commonly regulated.

# Making cDC1-specific protein landscape starts in early progenitors

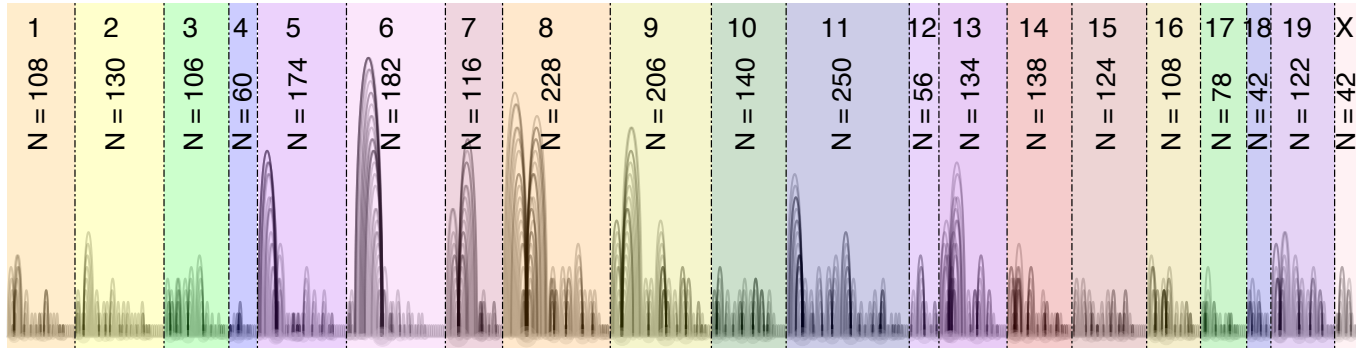
Dendritic cells develop & establish their functional protein landscape following:

(1) chromatin activation, (2-3) chromatin structure establishment, (4) expression of specific mRNAs.



# e-lncRNAs are widely transcribed from enhancer regions

We made predictions of potential enhancers regulating cDC1-specific mRNAs.  
We found that Many enhancers transcribe e-lncRNAs.



Contact map of cDC1 enhancer regions predicted to regulate cDC1-specific mRNAs and transcribing e-lncRNAs.

**Although it is premature to conclude that e-lncRNA transcription has regulatory functions in cDC1 differentiation, yet it is worth further investigating:**  
e-lncRNA DNA, RNA & protein binding motifs, structures, cellular localizations, etc..

# Summary

1. Collectively, lnc-RNAs are widely expressed in a variety of cell types, **yet many are restricted to a few cell types making their studies hard.**
2. Evidence for widespread functions of lncRNAs is growing, **yet far majority of lncRNAs remain functionally uncharacterized.**
3. lnc-RNA form co-expressed networks with mRNAs in cDC1 differentiation, **yet we still do not know what exactly their roles are**
4. Many enhancers regions express e-lncRNAs – **need to be further investigated**