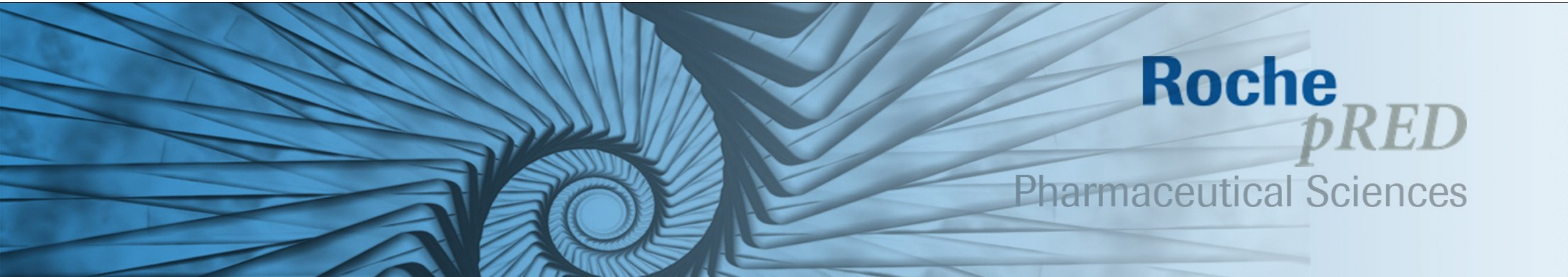
“Groundhog day” and what we can learn from that…

Fabian Birzele, Head of Bioinformatics & Exploratory Data Analysis,
Pharmaceutical Sciences



“Groundhog day” and what we can learn from that…

And what does that have to do with bioinformatics???

The Genome Sequence of the Eastern Woodchuck (*Marmota monax*) – A Preclinical Animal Model for Chronic Hepatitis B

Tyler S Alioto, Fernando Cruz, Jèssica Gómez-Garrido, Miriam Triyatni, Marta Gut, Leonor Frias, Anna Esteve-Codina, Stephan Menne, Anna Kiialainen, Nadine Kumpesa ...
Show more
Author Notes

G3 Genes|Genomes|Genetics, Volume 9, Issue 12, 1 December 2019, Pages 3943–3952,
https://doi.org/10.1534/g3.119.400413
Published: 01 December 2019 **Article history ▼**

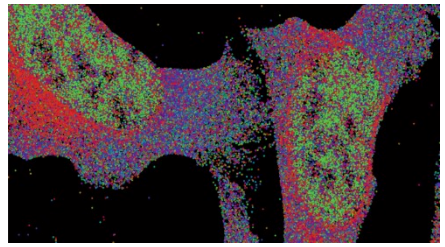
PDF Split View Cite Permissions Share ▼

Abstract

The Eastern woodchuck (*Marmota monax*) has been extensively used in research

… nothing… aside from the fact that I was involved in sequencing the Woodchuck genome

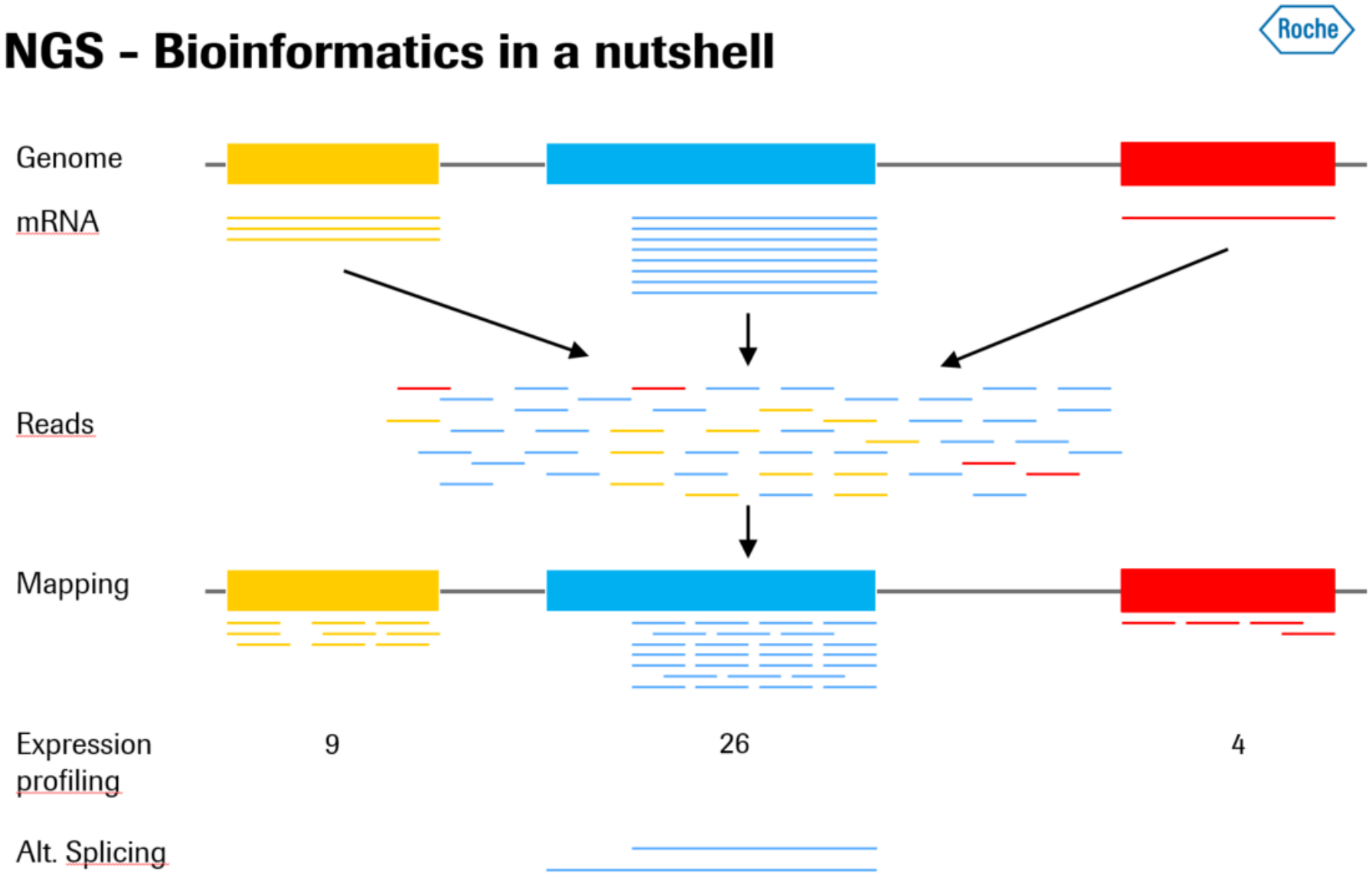
“Groundhog day” and what we can learn from that…



- Since I started to study bioinformatics in 2000, the world has seen an unprecedented rise of high throughput data generation technologies.
- Affymetrix gene expression **arrays**, followed by next generation **sequencing**, and **better** next generation **sequencing** machines, followed by **single cell sequencing** and now followed by **multi-modal, spatial resolution on protein and RNA level**.
- In this talk I would like to look back a few years and show you some «lessons learned» of this ever repeating cycle of a **new technology emerging and establishing this technology in drug research**.

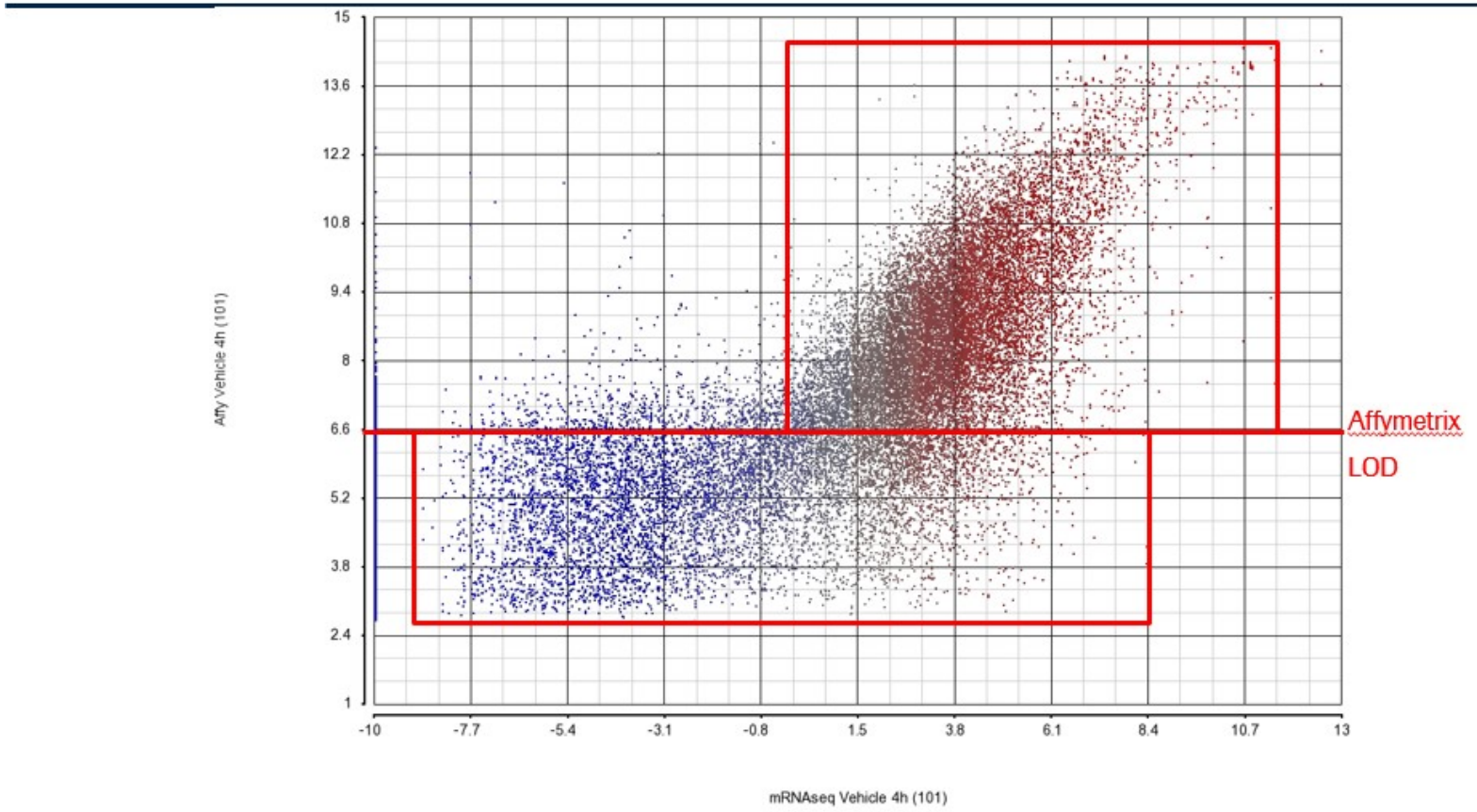
Lesson 1 – convince yourself

Convince yourself – implement a pipeline



Convince yourself – Compare to other technologies

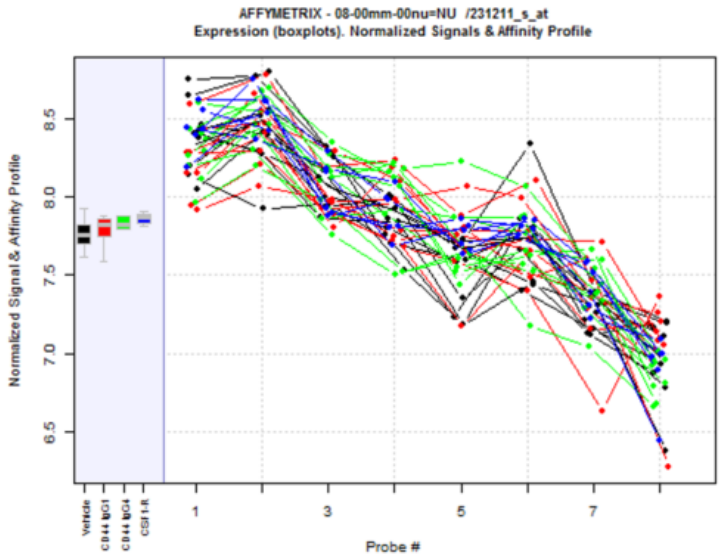
Correlation of signal intensities

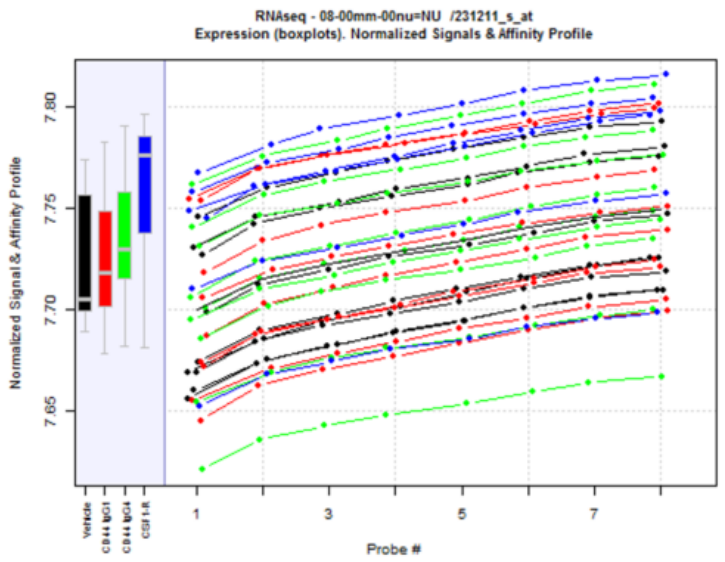


- In general good correlation between Affymetrix and mRNAseq above Affymetrix detection limit (~6.6)
- Overall, full dynamic range of mRNAseq is higher compared to Affymetrix

Convince yourself – Compare to other technologies

Affymetrix and mRNAseq “on the probe level”





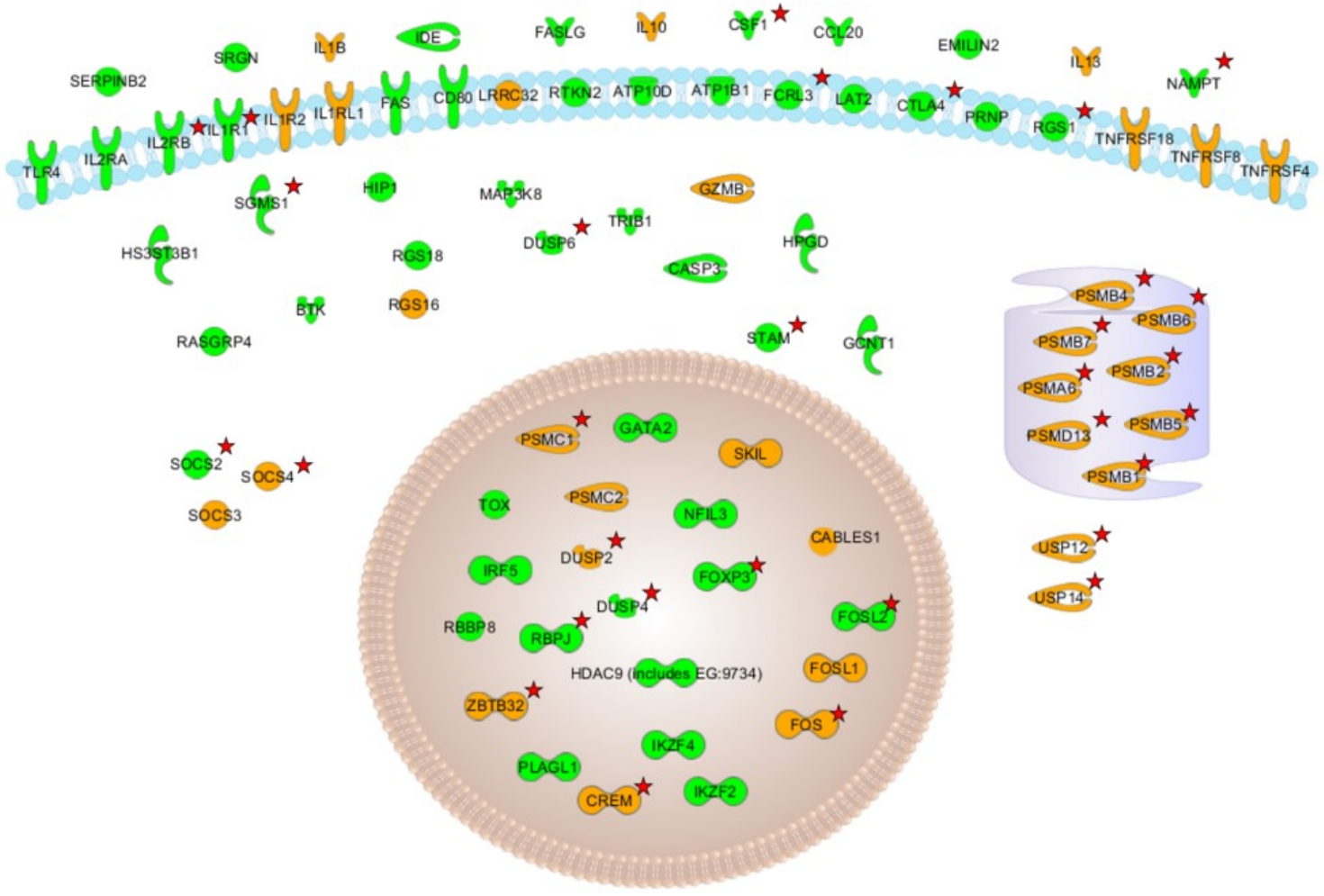
- Intensities for Affymetrix probes of the same gene can vary greatly (in this example 4-fold). Affinity depends on probe position nontrivially
- In comparison, mRNAseq signals perfectly follow the “Signal ~ Affinity * Expression” model (note that the Y-axis range in the right plot is only 7.65-7.8 (on log2 scale)
- **But:** overall signal does only change slightly with both methodologies

Anton Belousov, TRS-BEDA

Lesson 2 – Convince biologists

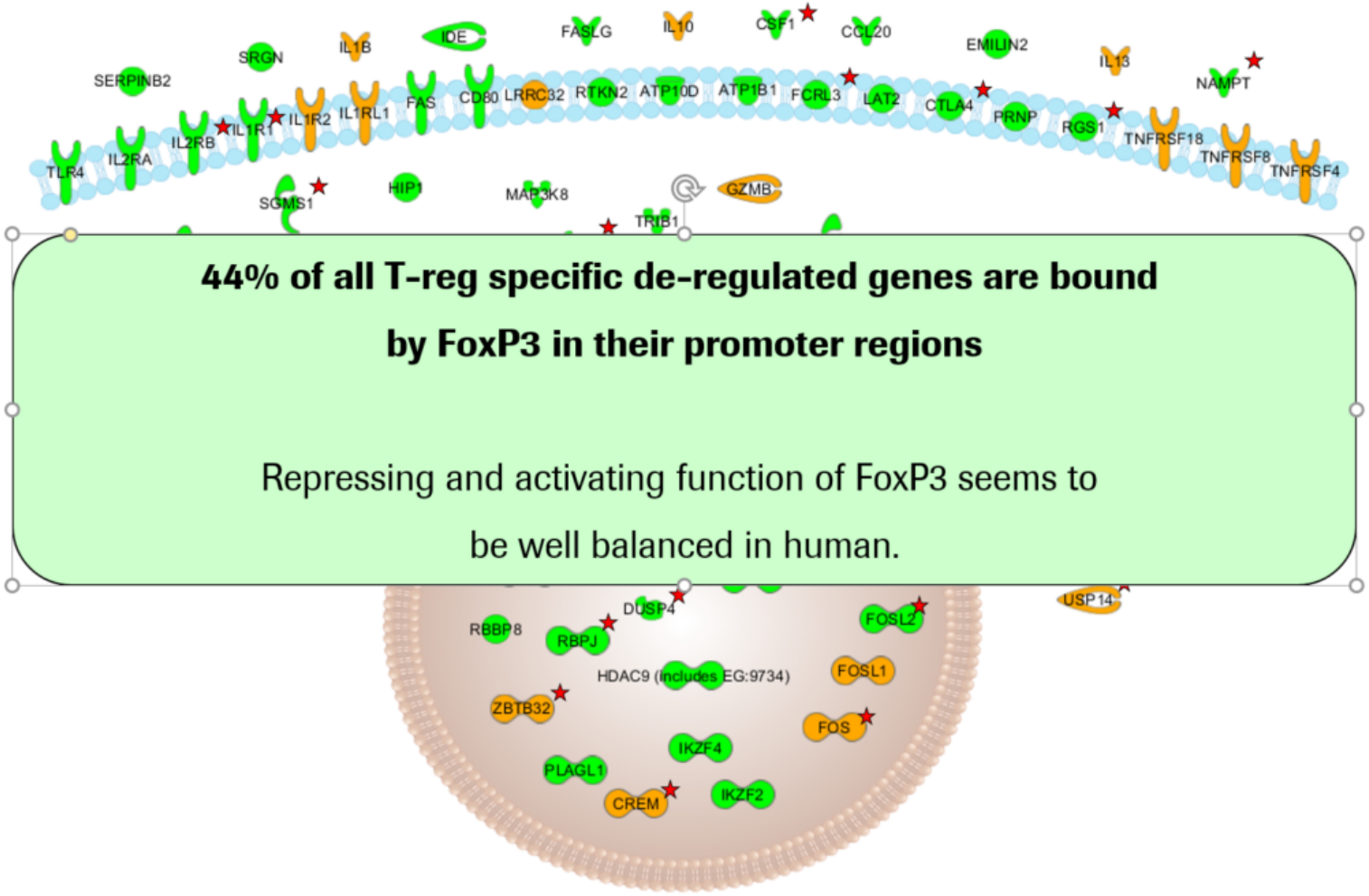
Convince biologists – Understand biology

Combination of mRNAseq and ChIPseq



Convince biologists – Understand biology

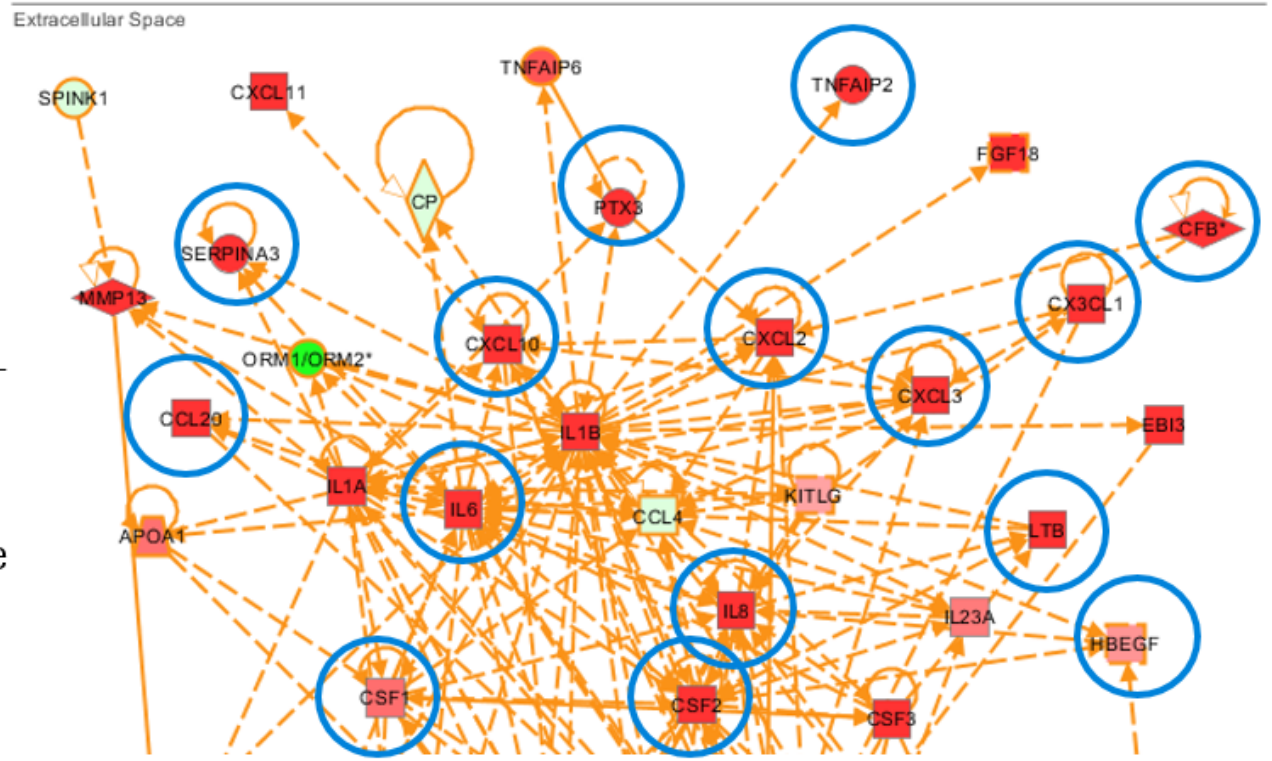
Combination of mRNAseq and ChIPseq



Convince biologists – Show the advantage...

Understanding biology

- Both platforms identify similar sets of de-regulated genes (overlap at a fixed threshold between 42-67%)
- Overrepresentation and signature analyses lead to similar results with better p-Values for mRNAseq
- Genes identified in addition in mRNAseq fit well into the biological context

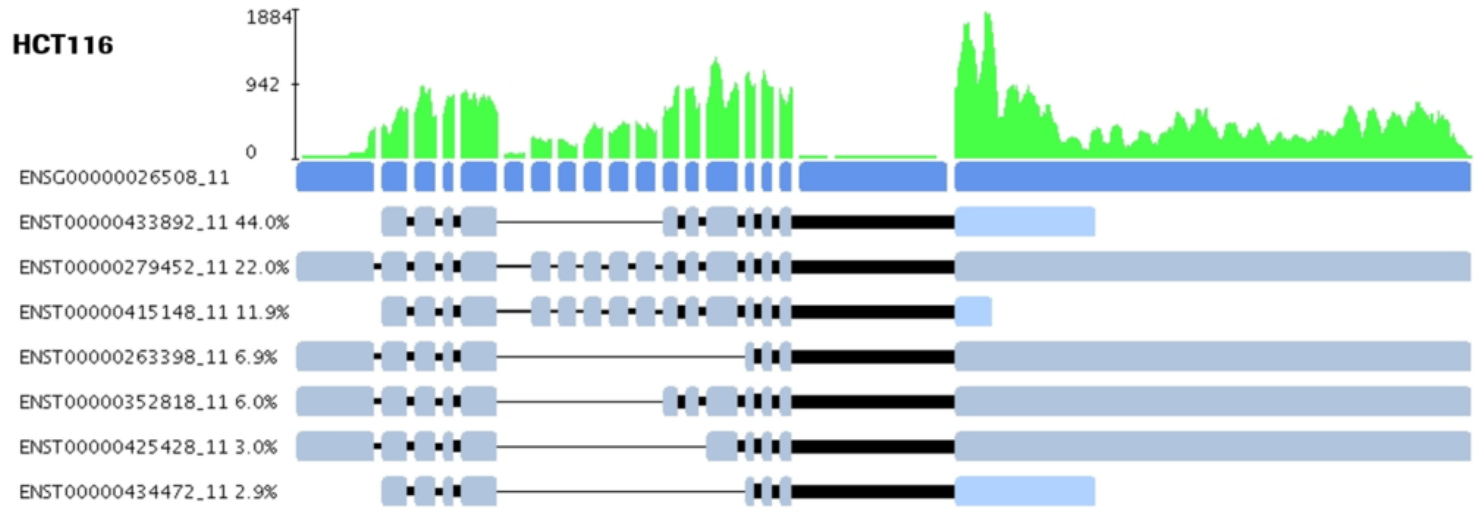


Human extracellular genes deregulated at 4h after treatment, red: up under treatment, green: down under treatment

Lesson 3 – Fully exploit the opportunities you are given

Fully exploit the opportunities you are given – Splicing

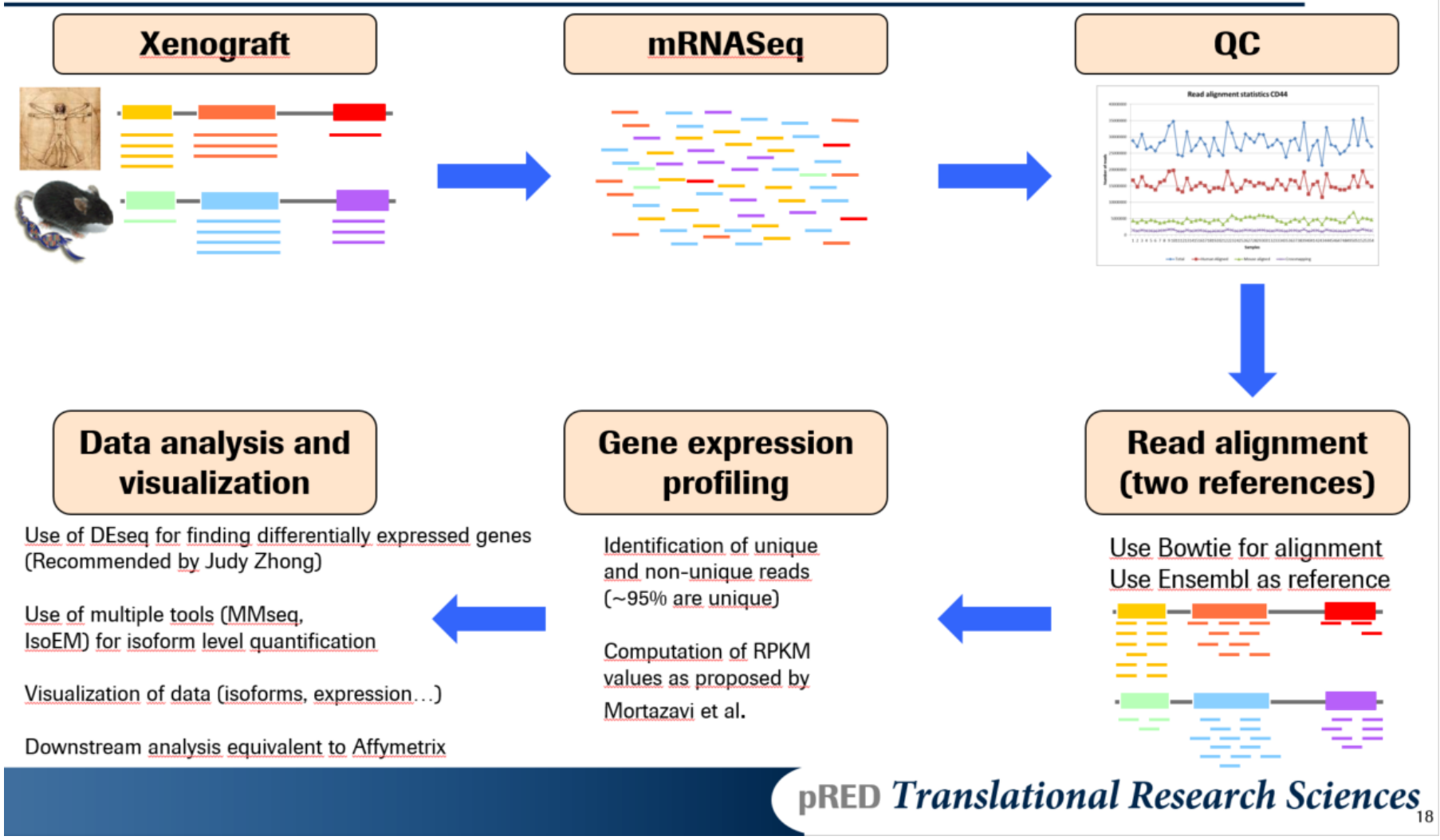
Isoform level gene expression profiling



- Major isoforms (e.g. of CD44) can be identified
- CD44 is expressed in a mixture of mainly two isoforms which differ in their N-terminal sequence and the 3' UTR.
- Other cell lines have very different expression patterns for CD44 isoforms
- Overall, no strong changes in isoform expression are identified upon anti-CD44 treatment.

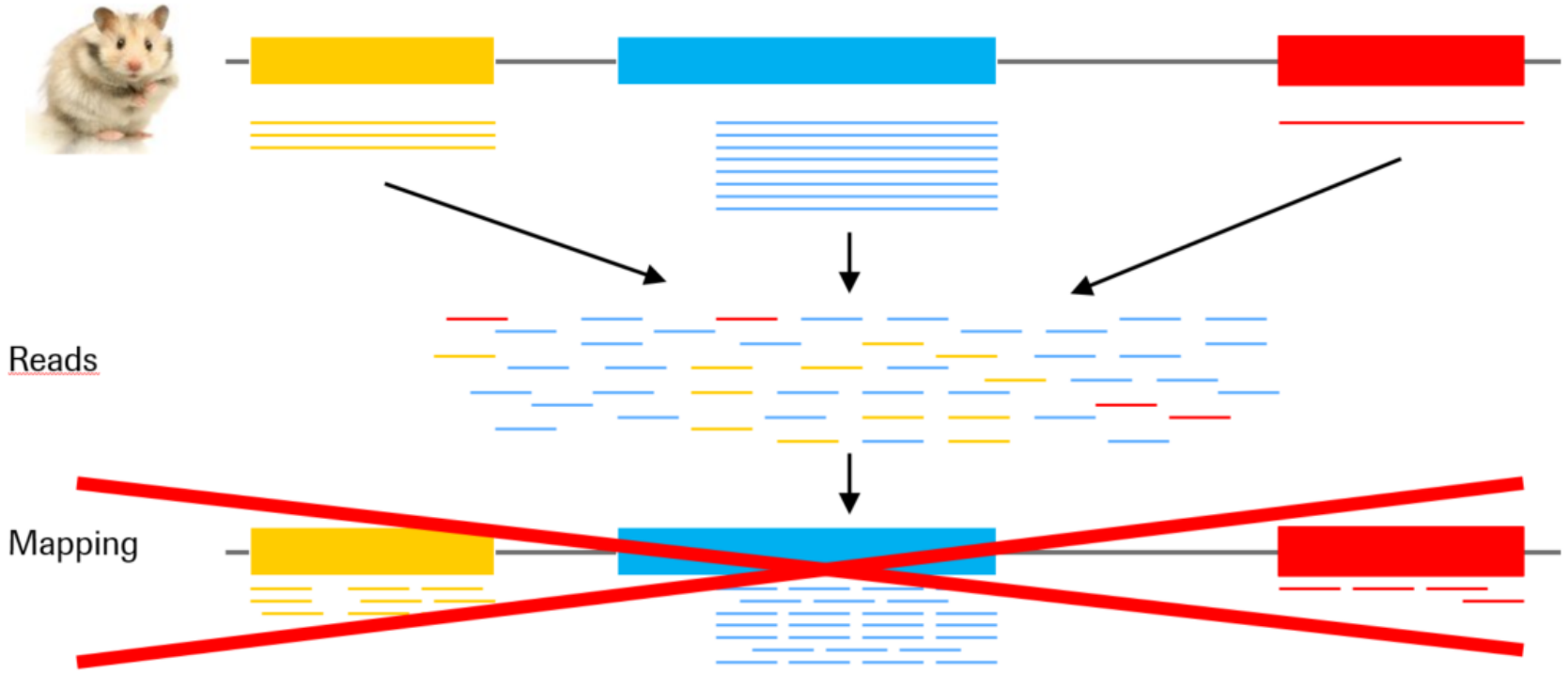
Fully exploit the opportunities you are given – Xenografts

mRNAseq in Xenograft models



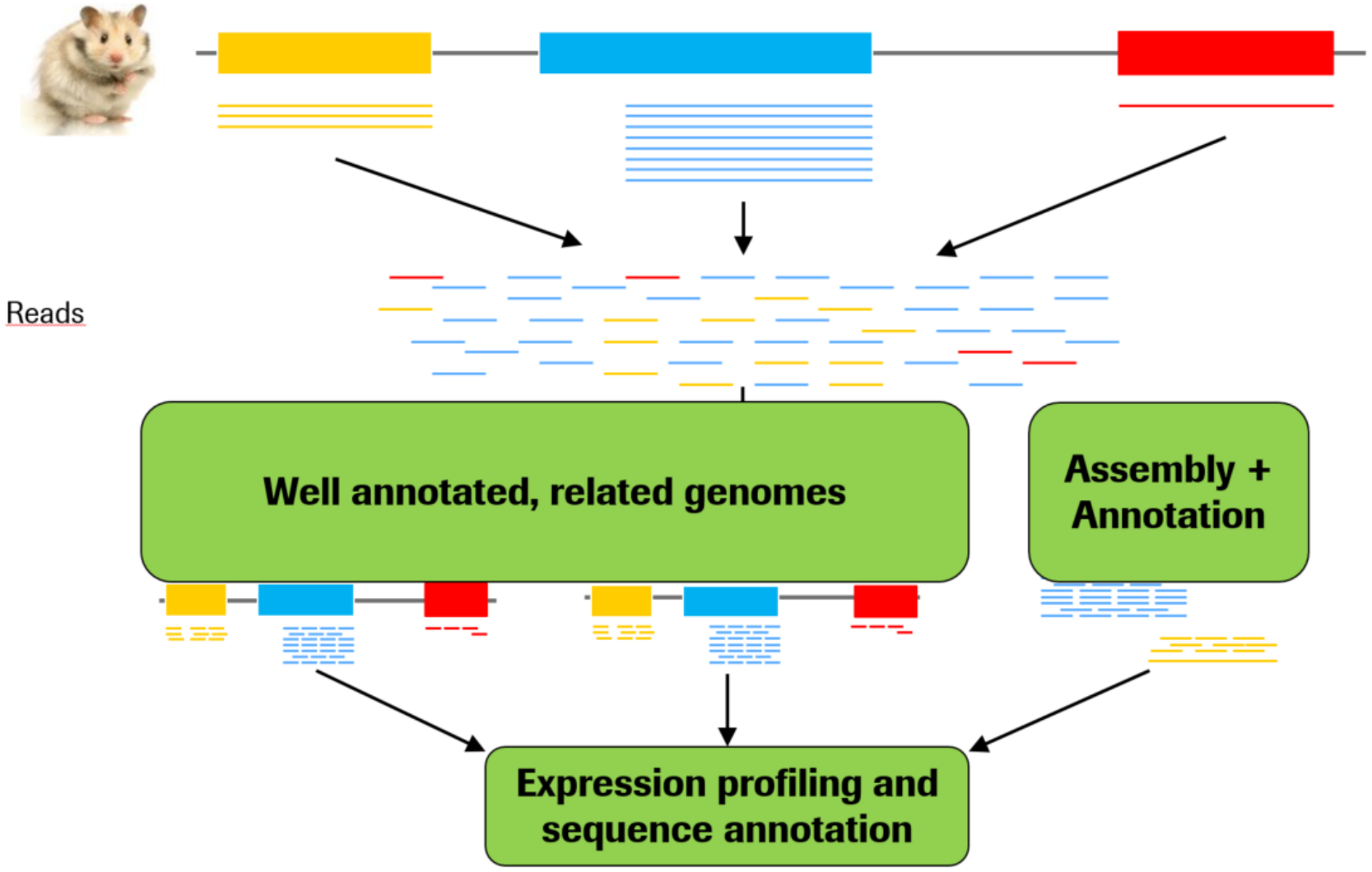
Fully exploit the opportunities you are given – CHO

NGS – Bioinformatics without a reference genome



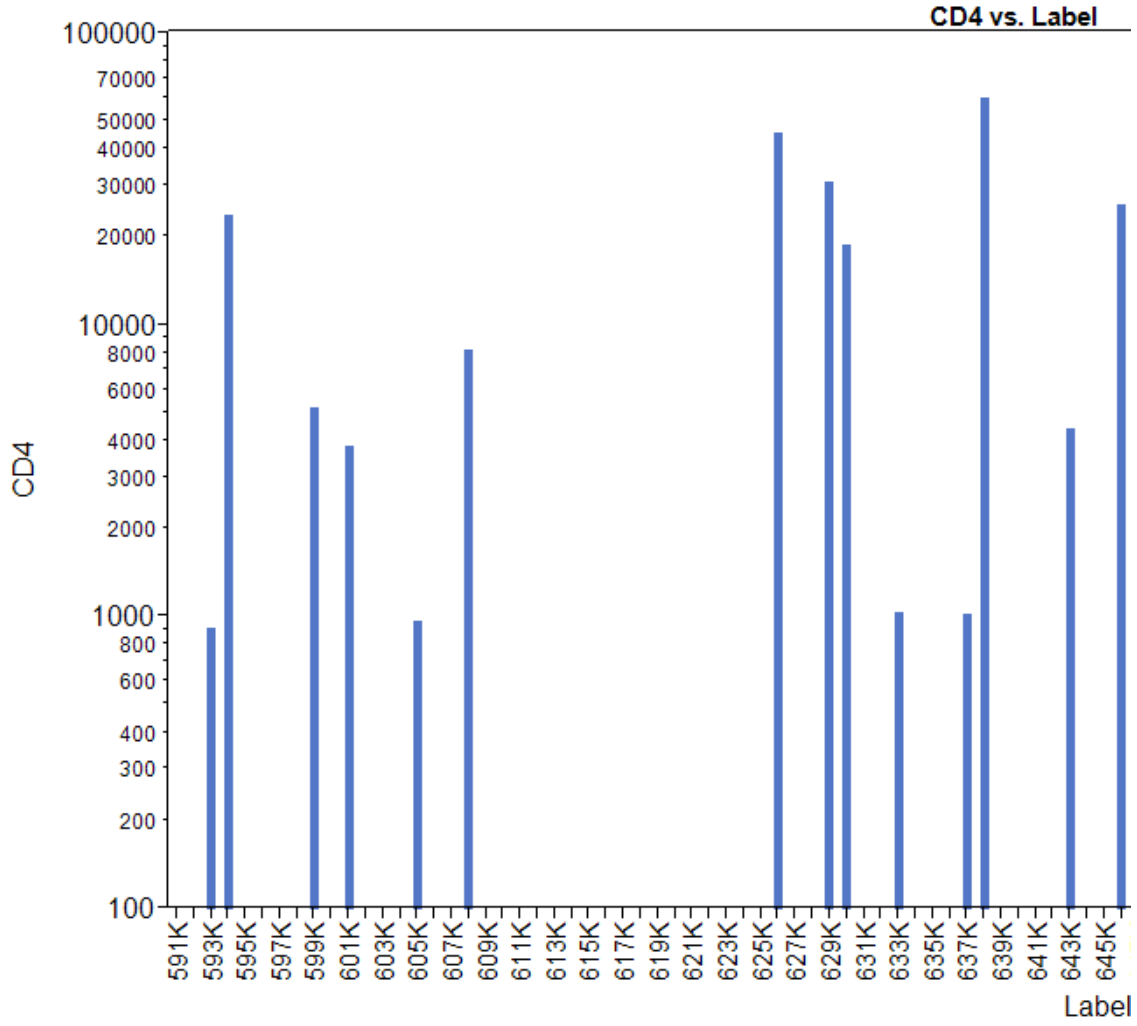
Fully exploit the opportunities you are given – CHO

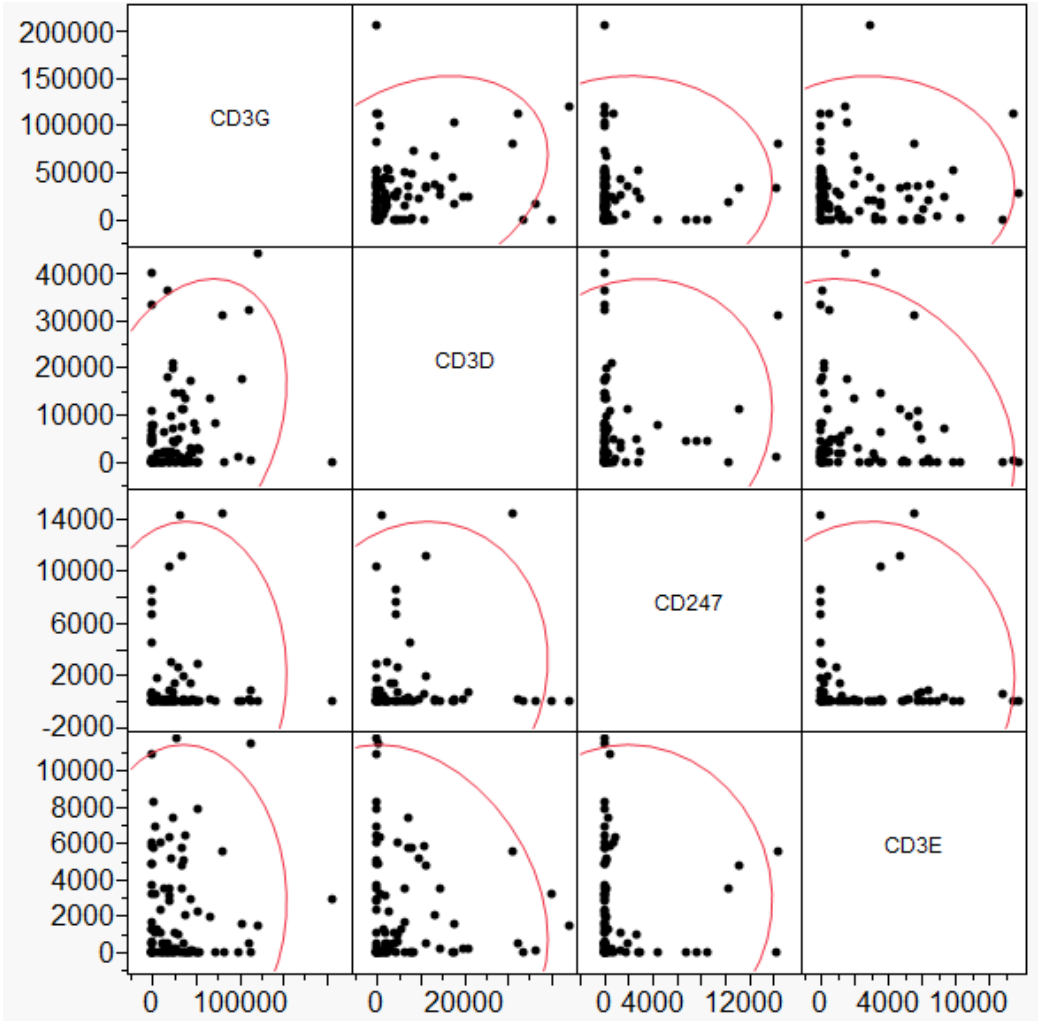
NGS – Bioinformatics without a reference genome



Lesson 4 – Trash it if it is not good enough yet

Trash it if it is not good enough yet – Early Single Cell data



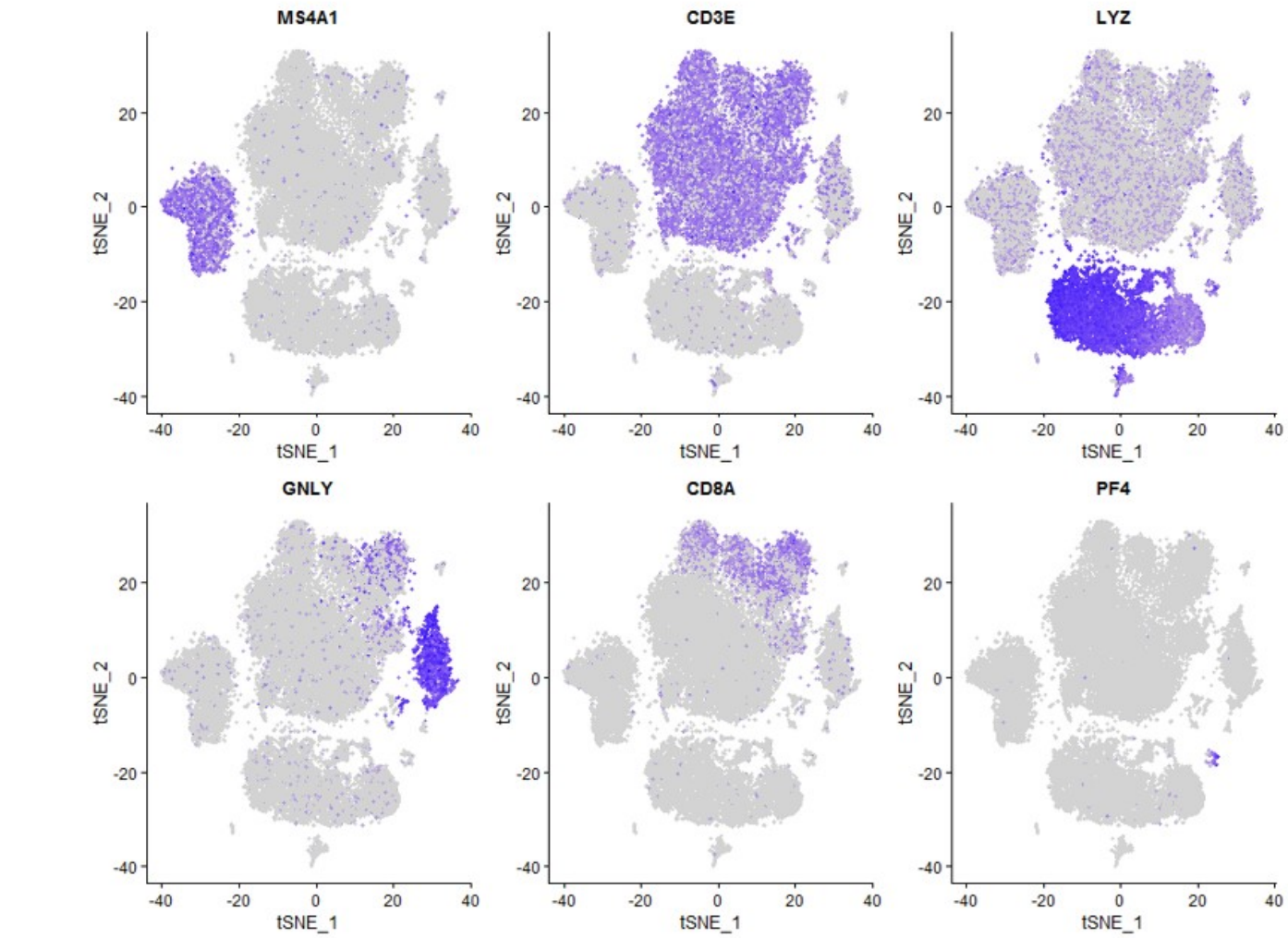


2015 – early single cell sequencing data. We were not convinced of stability, correlation of mRNA and protein…

… And be fast if it finally works – One year later…

End of 2016 – arrival of 10X

Technology is ripe and now we need to be fast!



Conclusions

Lessons learned…

- **Convince yourself – deeply understand the technology**
  - For that you need to get the technology under control (**data, pipelines**…), understand the properties of the data (**variation, statistical properties**…) and finally compare to what is already there (**how do they compare, what is similar, what is different**…)
- **Convince biologists – it is not enough that you find it cool**
  - Once you are sure that you have the data under control, work with biologists to convince them of the usefulness of the new tool. For that, **you need to understand their questions and biology**.
- **Fully exploit the opportunities you are given**
  - Once you and your collaboration partners are convinced – use all the opportunities that the new technology gives you. **Be creative, be fast, bring together QUESTION and TECHNOLOGY**
- **Trash it if it is not good enough yet**
  - Sometimes technologies are fancy and cool but not ripe enough yet. In that case, **do not waste your time but take a step back and observe**…

“Groundhog day” and what we can learn from that…

There will always be a new technology!!!

… And the mechanism to make them productive are always the same….

Doing now what patients need next