A Bayesian model for single-cell gene expression analysis with MERFISH data

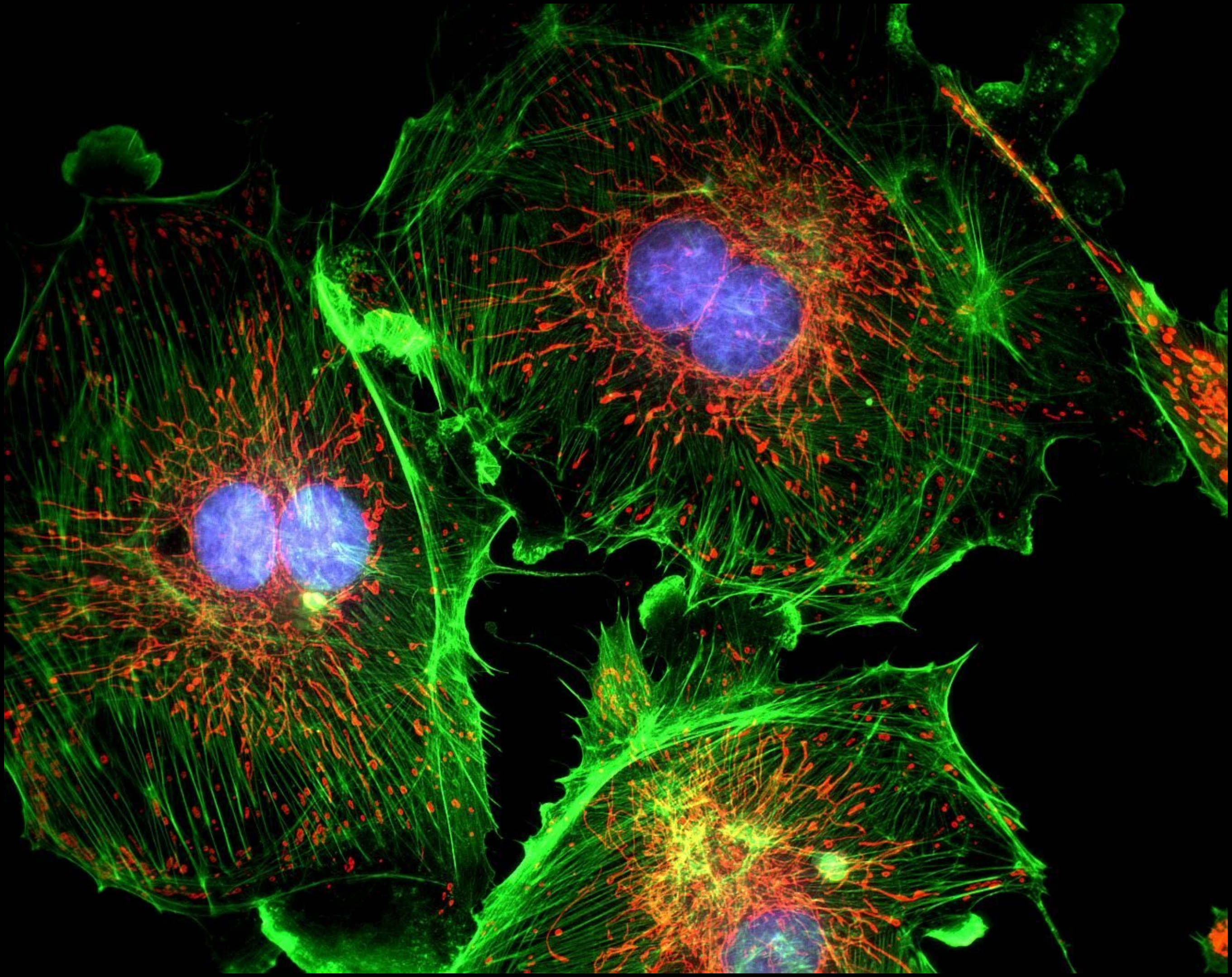
Johannes Köster

2016

CWI
Centrum Wiskunde & Informatica

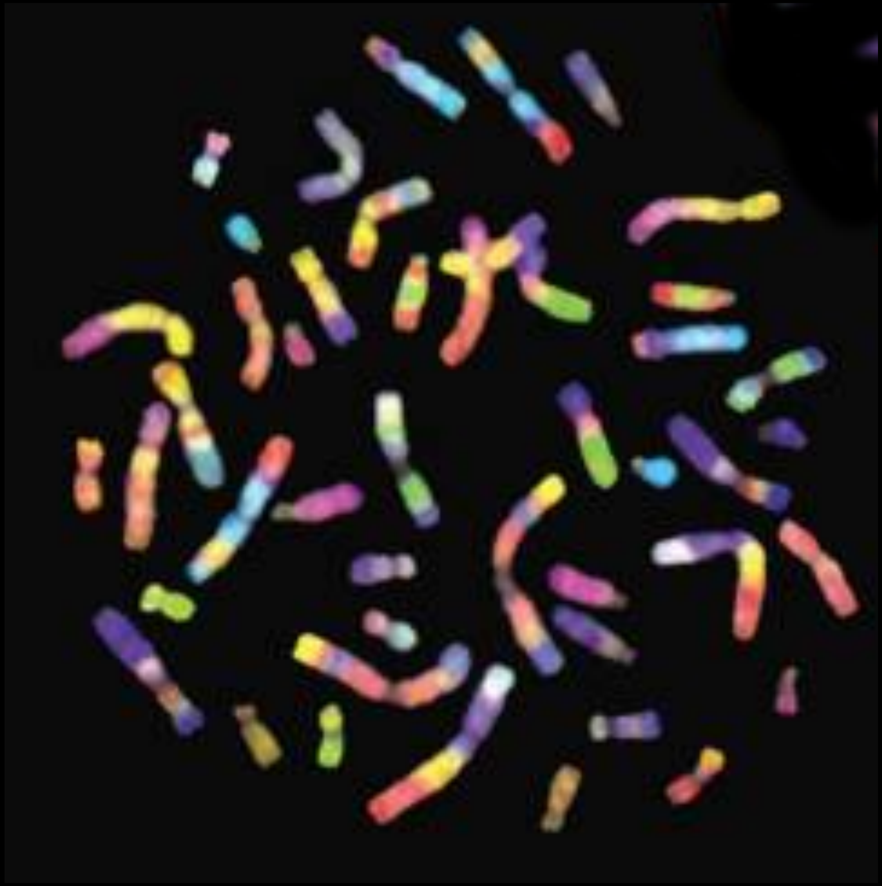
VE RI TAS
HARVARD
MEDICAL SCHOOL

Cells



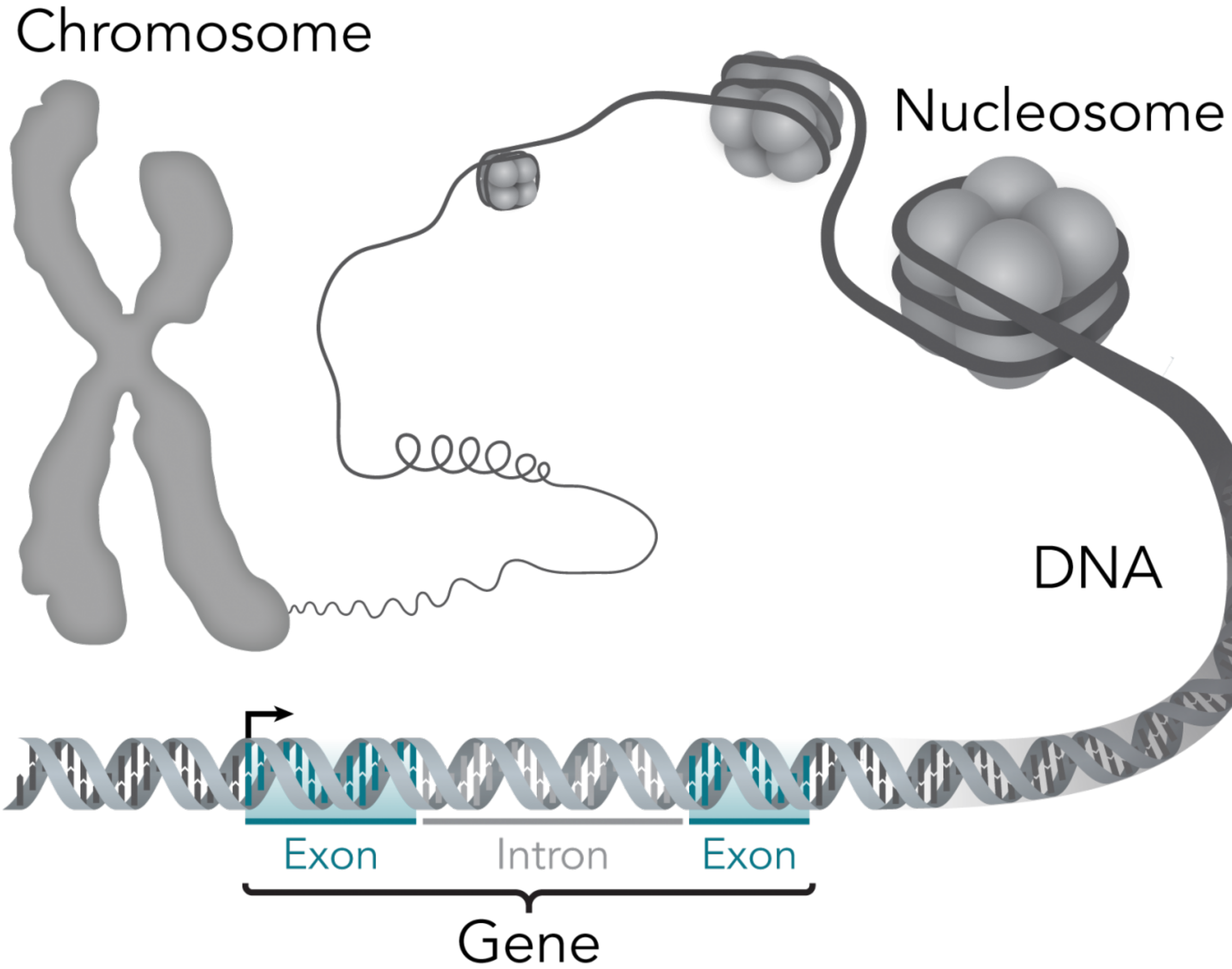
www.copdri.com

Chromosomes



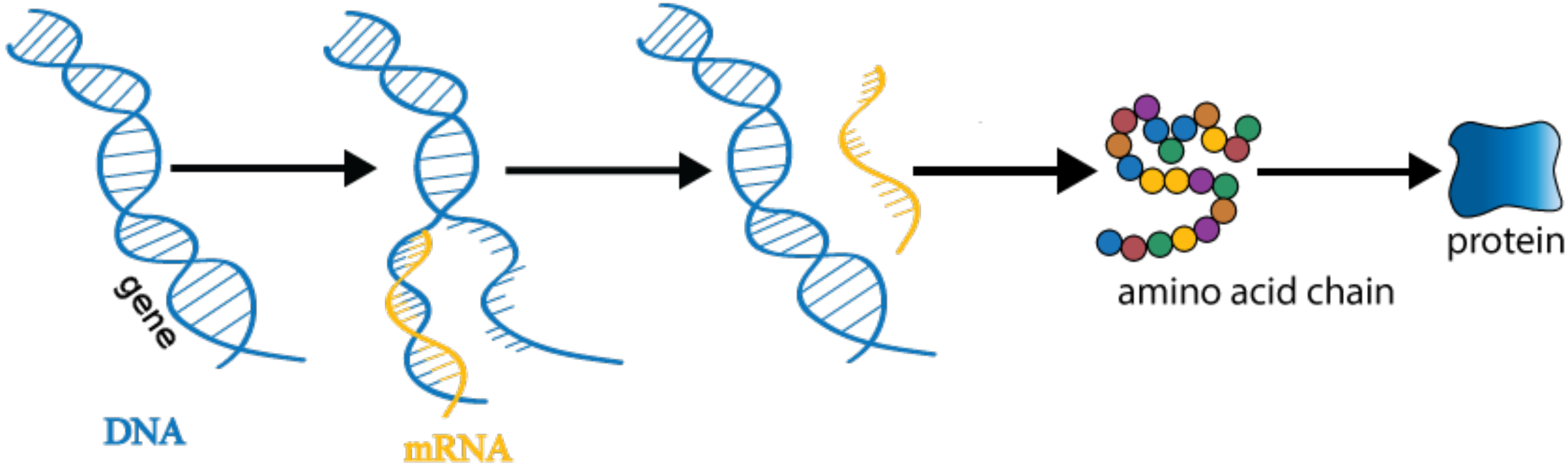
www.austincc.edu

Genes



www.medicalxpress.com

From genes to proteins



Gene activity/expression:

the more RNA, the more protein

biosocialmethods.isr.umich.edu

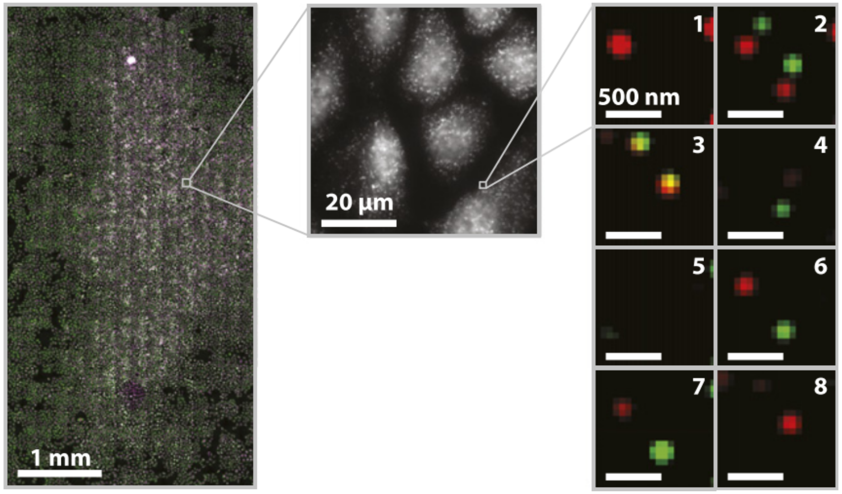
FISH

Fluorescence in-situ hybridization:

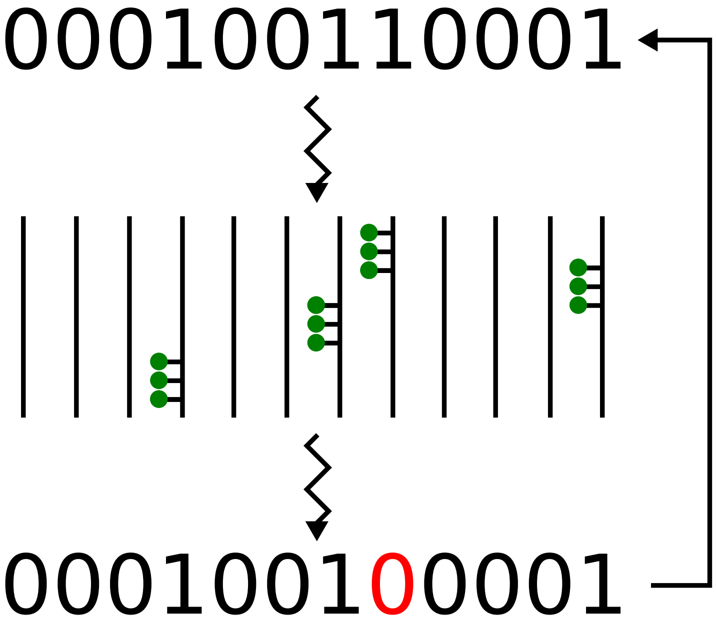
- label RNA with fluorescent probes
- see RNA molecules in single cells

Problem:

needs 1 color per gene



MERFISH



design
for each
RNA species

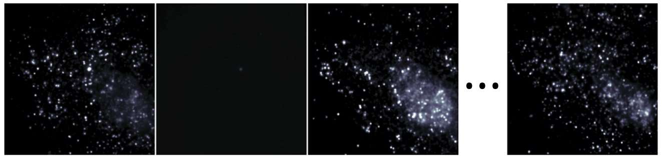
000100110001

hybridization
of fluorescent
probes

calling
by assignment
and correction
of 1-bit errors

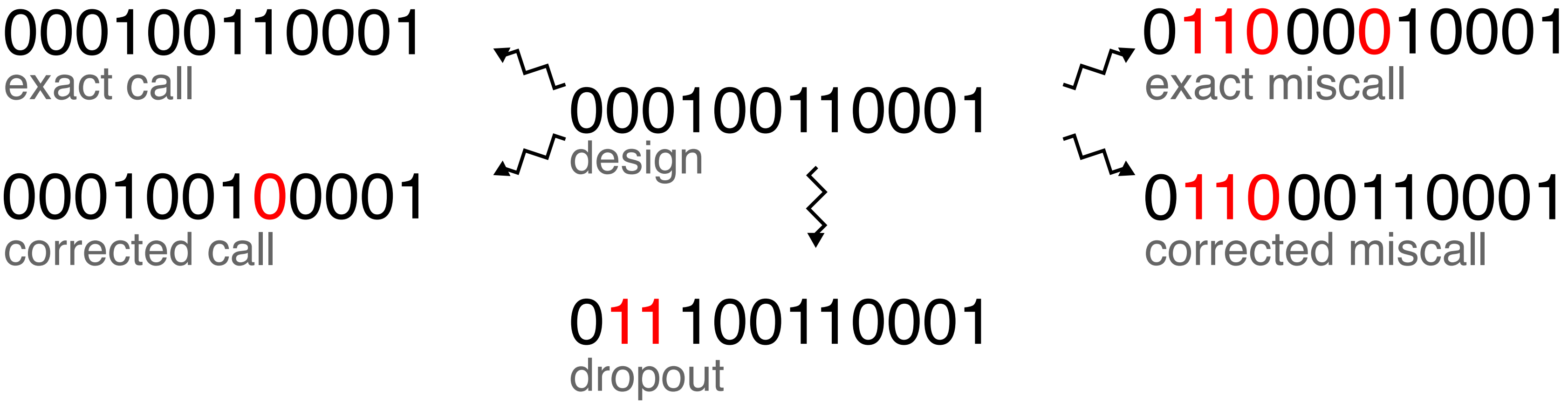
readout
for each
RNA molecule

000100100001



Chen et al. 2015

MERFISH



000100110001
exact call

000100100001
corrected call

000100110001
design

011100110001
dropout

011000010001
exact miscall

011000110001
corrected miscall

Problem with raw counts:

up to 20% loss and misidentification

Research question

Can we recover the true counts from the raw, biased MERFISH counts?

Approach

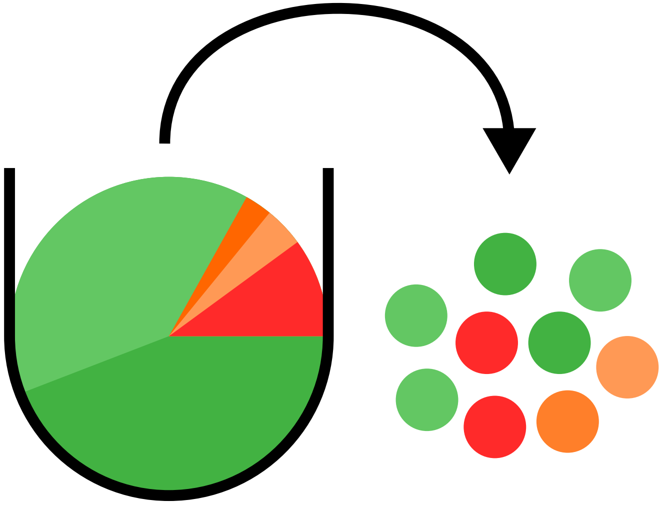
Event probabilities

Known error probabilities:

1→0 error: 1% - 10%

0→1 error: 0.5% - 4%

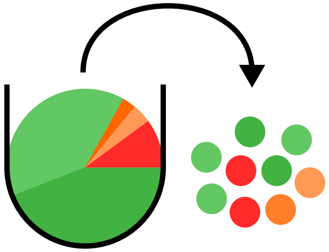
Urn model for scenario probabilities

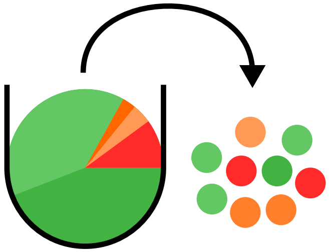


Scenario:

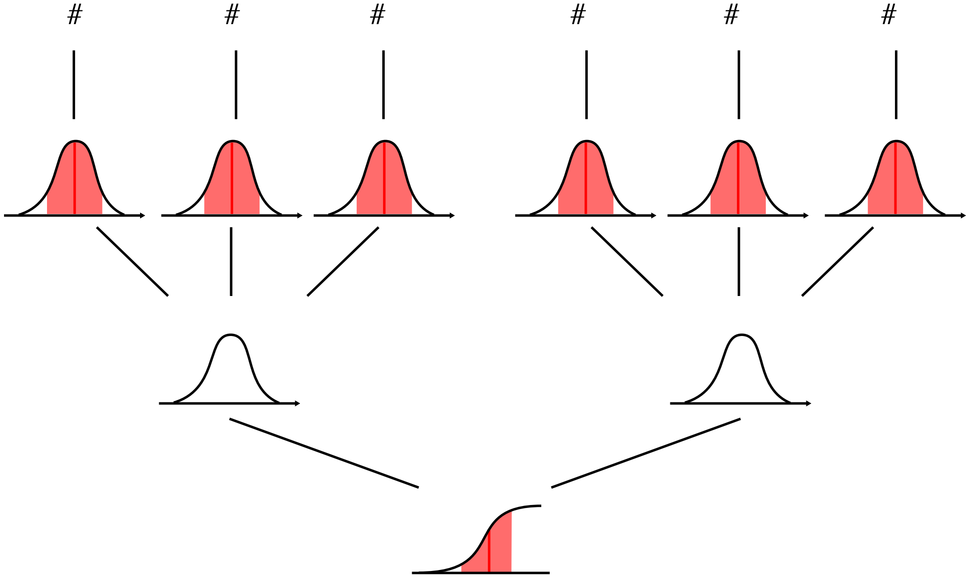
one **combination of events** that explains observed counts given true counts

Likelihood of observed counts given truth

$\Pr(\# \mid X=x) = \Pr($  $) +$

$\Pr($  $) + \ldots$

Bayesian model for differential gene activity



Input/Output

observed counts

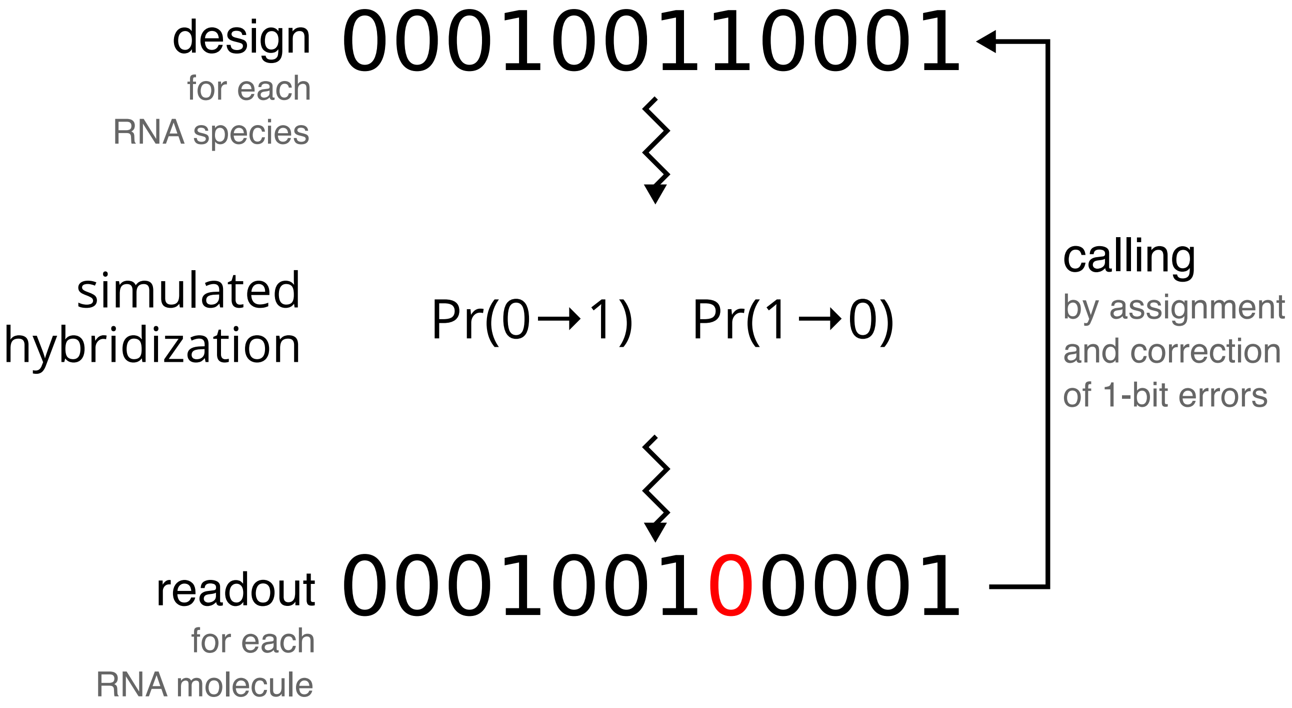
posterior probability for expression

posterior probability for mean and variance in groups of cells

posterior probability for fold change or coefficient of variation

Results

Simulation



design
for each
RNA species

000100110001

simulated
hybridization

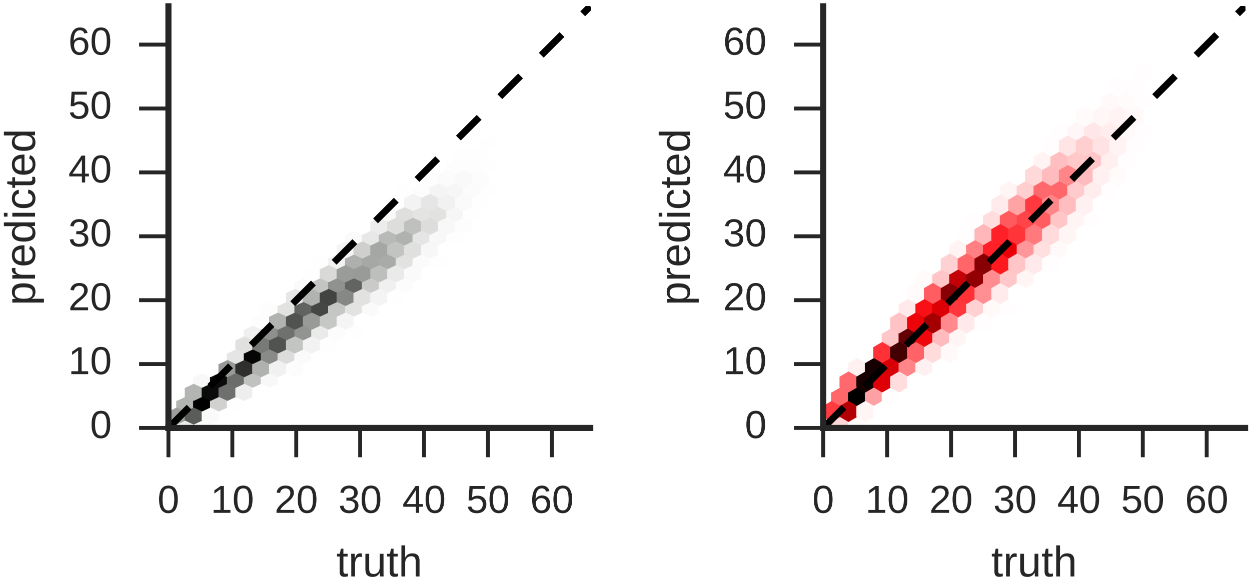
Pr(0→1) Pr(1→0)

readout
for each
RNA molecule

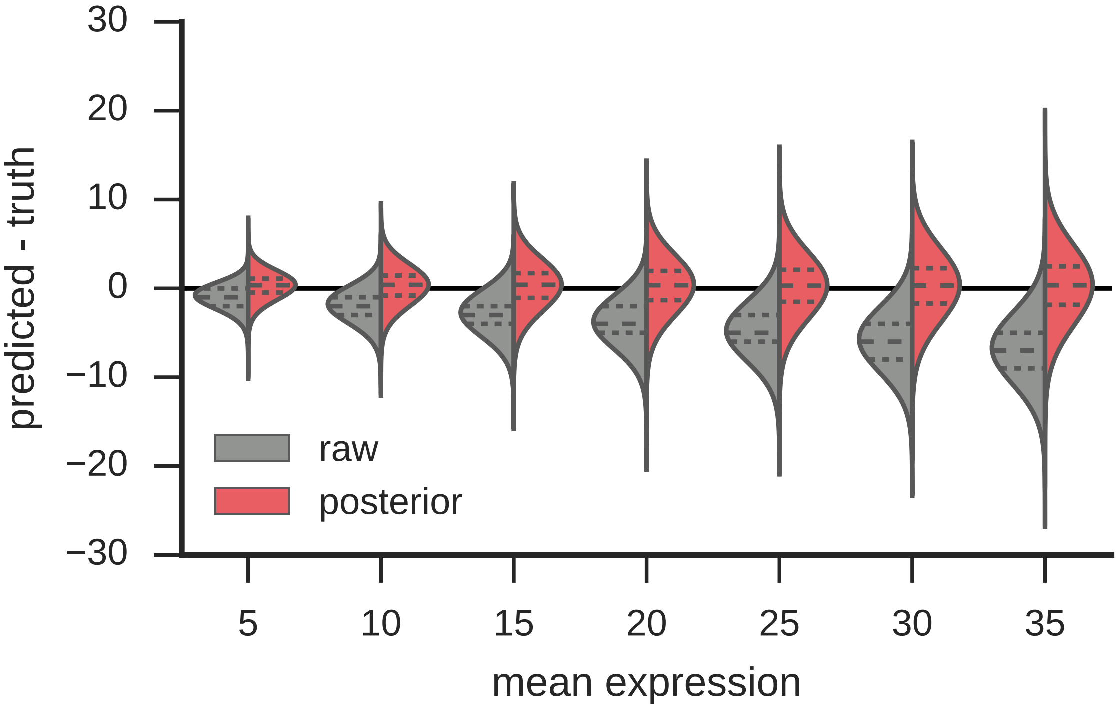
000100100001

calling
by assignment
and correction
of 1-bit errors

Bayesian model recovers biased counts



Bayesian model recovers biased counts



Conclusion

Conclusion

- MERFISH is a technology to visualize and count RNA molecules in single cells.
- Raw counts are **biased**.

The presented Bayesian model provides:

- **accurate estimates** of the true RNA counts
- a way to **access differential gene activity** between cells while maintaining **uncertainty information**
- ability to measure **more genes in parallel**

Acknowledgements

Shirley Liu

Myles Brown

Bo Li

Peng Jiang

Eric Severson

DANA-FARBER CANCER INSTITUTE

HARVARD SCHOOL OF PUBLIC HEALTH

HARVARD MEDICAL SCHOOL