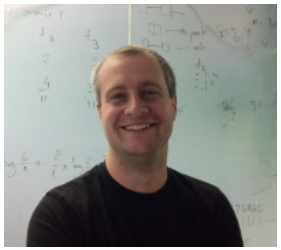
Bioinformatics, master and slave

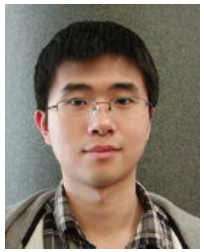
William Ritchie
Bioinformatics Lab,
Centenary Institute

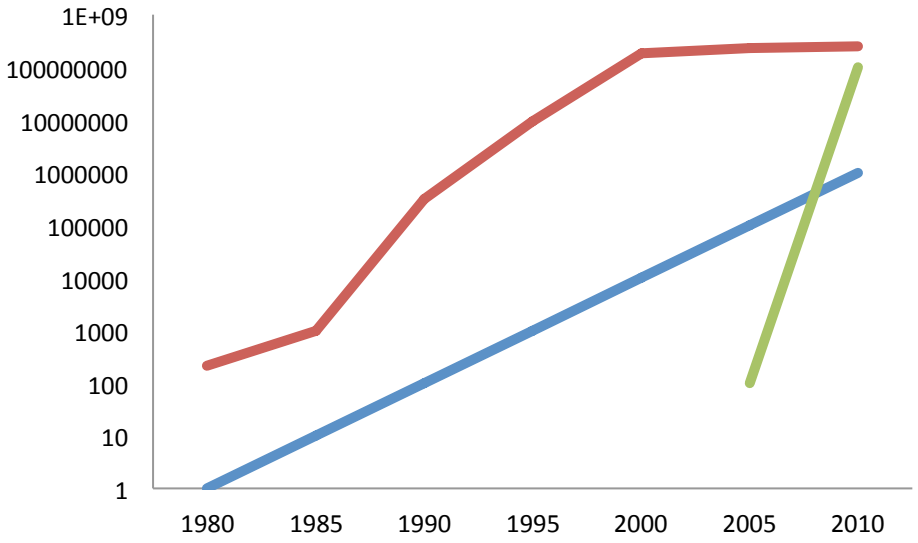
The University of Sydney

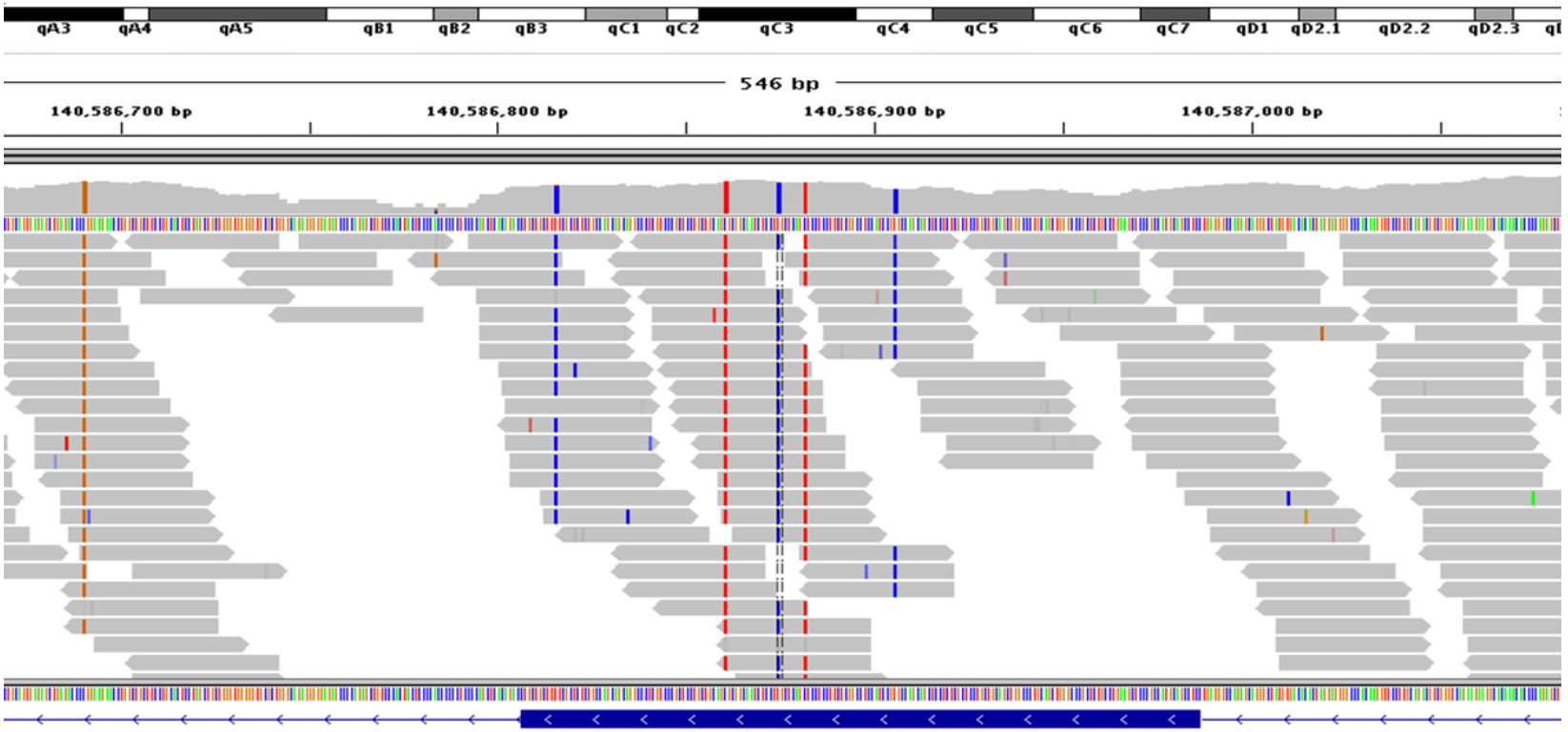
Centenary Institute research for life

Bioinformatics

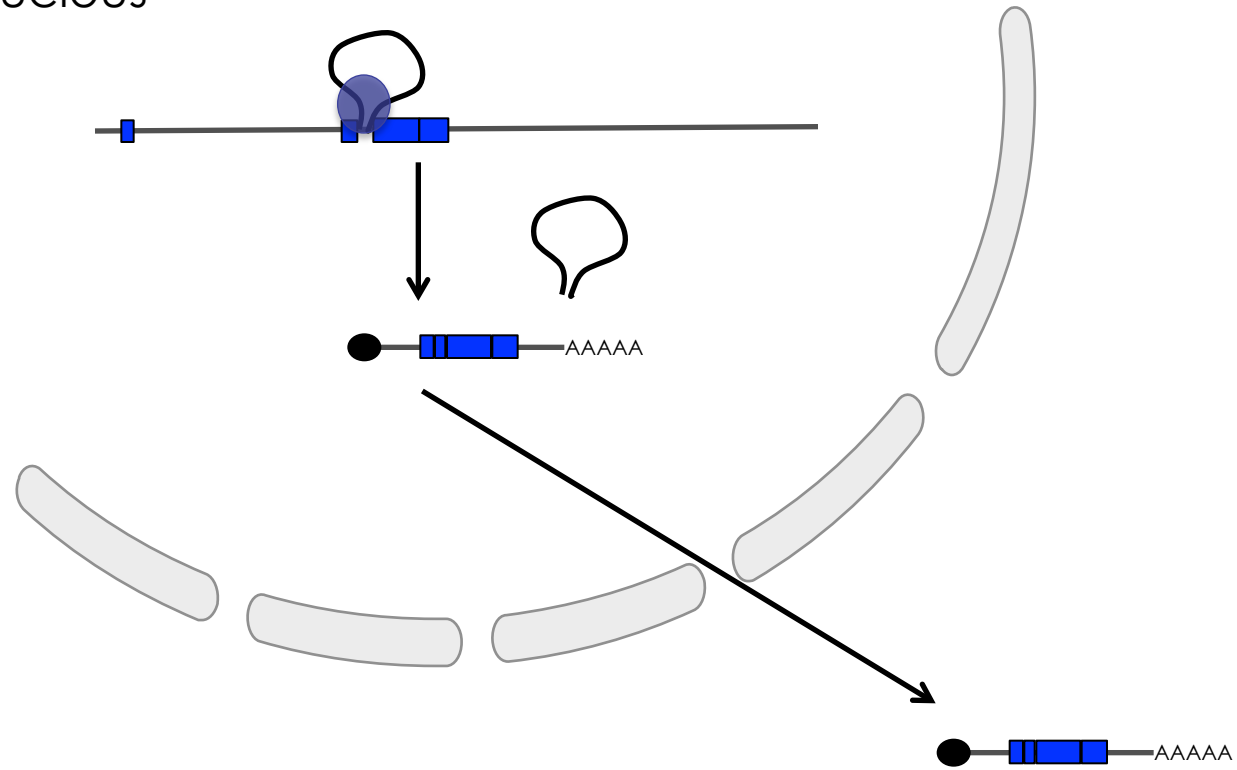





1E+09
100000000
10000000
1000000
100000
10000
1000
100
10
1
1980
1985
1990
1995
2000
2005
2010

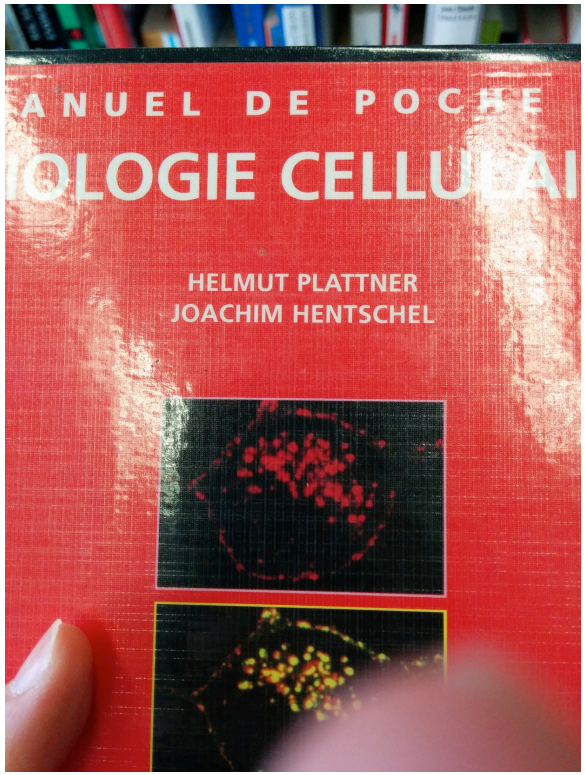

qA3
qA4
qA5
qB1
qB2
qB3
qC1
qC2
qC3
qC4
qC5
qC6
qC7
qD1
qD2.1
qD2.2
qD2.3
546 bp
140,586,700 bp
140,586,800 bp
140,586,900 bp
140,587,000 bp

nucleus



Splicing machinery

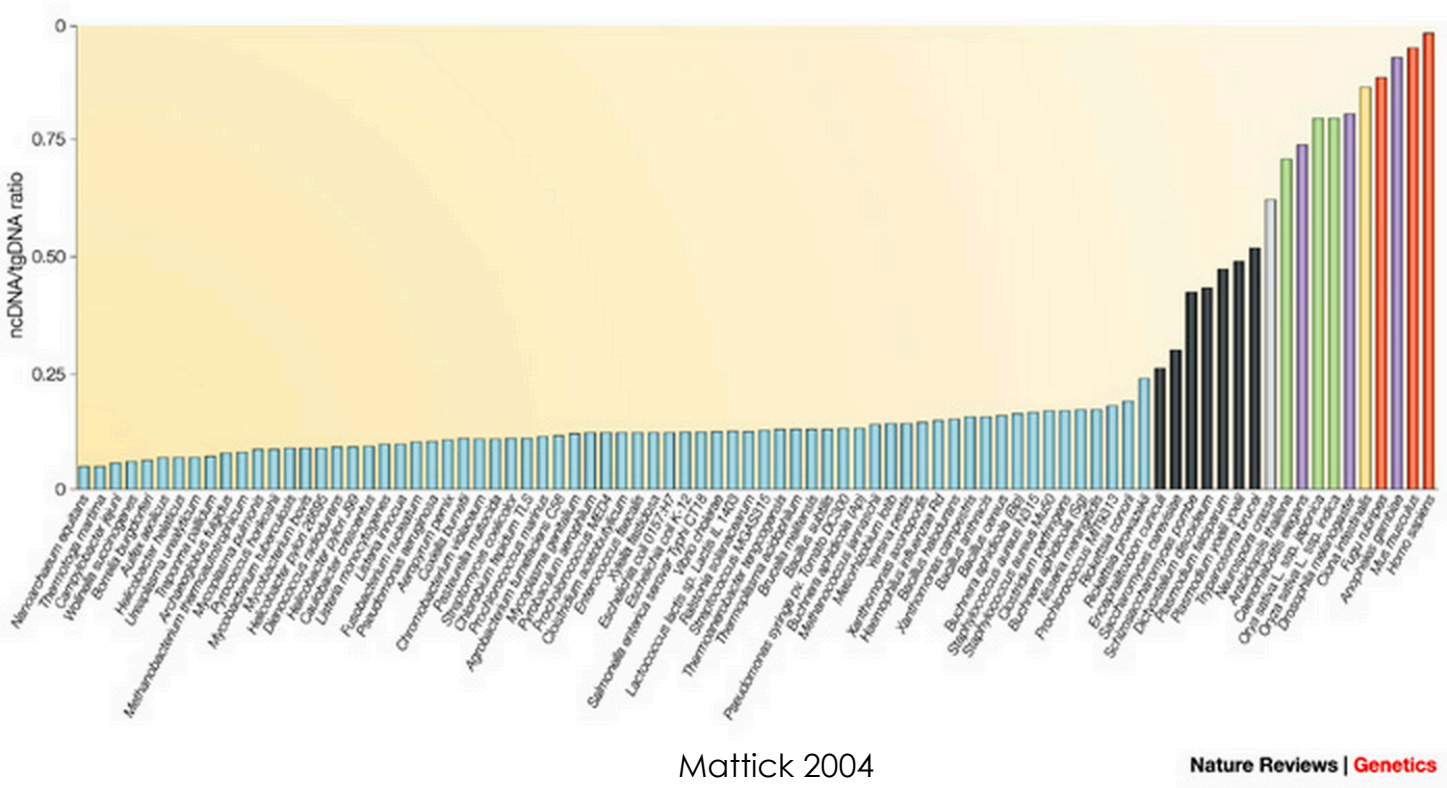
cytoplasm



séquence équivalente du brin ADN sens (porteur de l'information hé

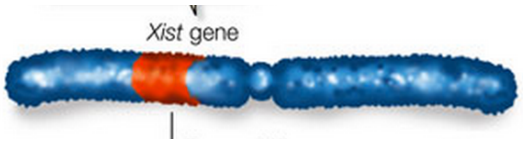
– L'ARN se présente d'abord sous forme d'un précurseur (**pré-ARN**
encore être mis en forme à l'intérieur du noyau (**épissage**, c'est-à-d
des morceaux ne contenant pas d'information, les **introns**).
– L'**ARNm** terminé (mature) quitte le noyau par les pores nucléaires

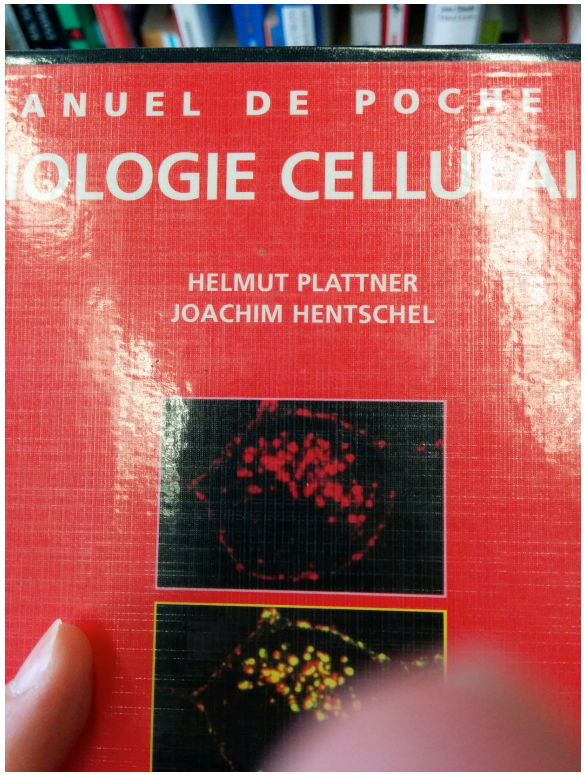
2% protein coding VS 80% transcription



Mattick 2004

Nature Reviews | Genetics

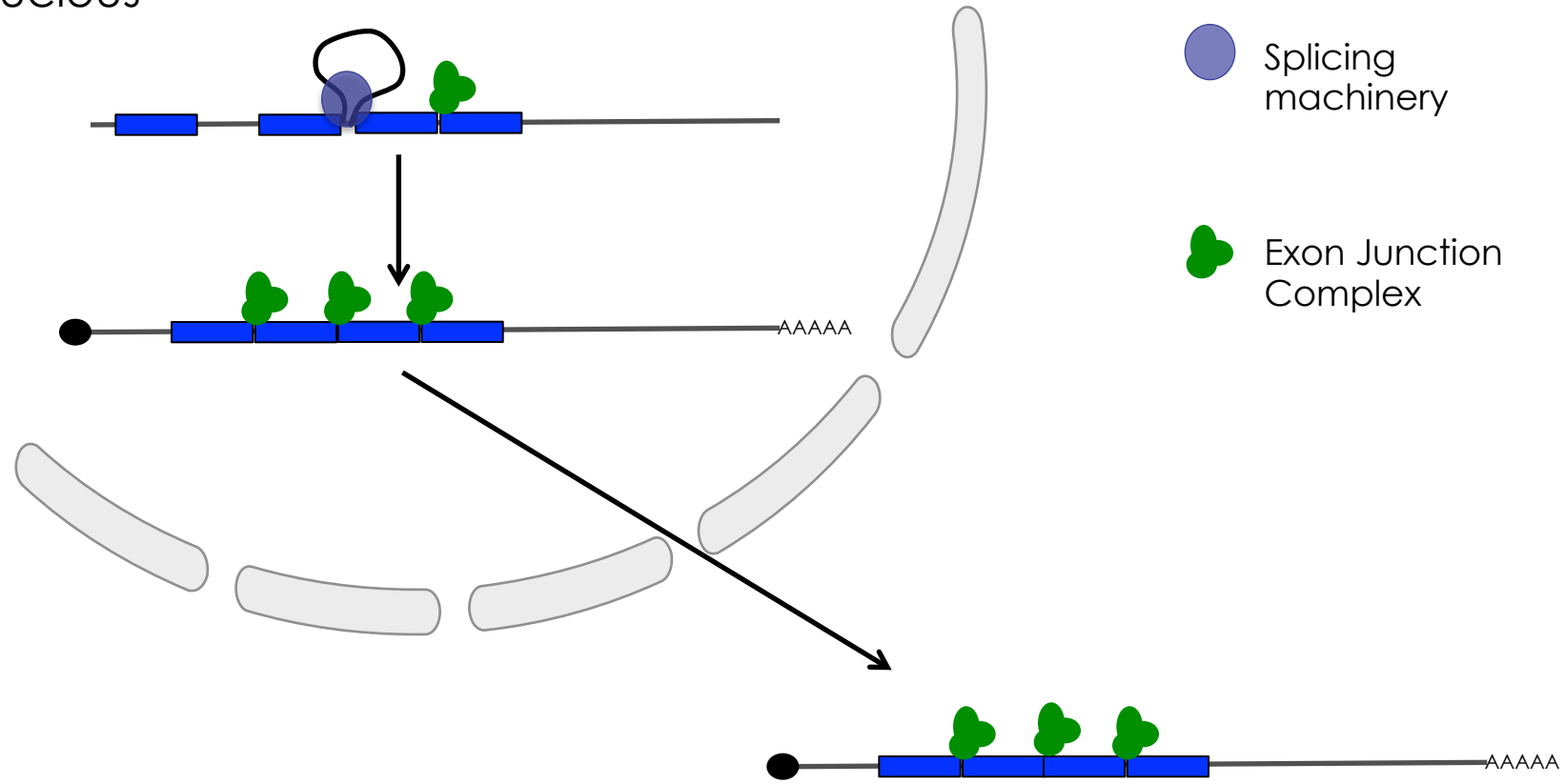




séquence équivalente du brin ADN sens (porteur de l'information hé

– L'ARN se présente d'abord sous forme d'un précurseur (**pré-ARN**
encore être mis en forme à l'intérieur du noyau (**épissage**, c'est-à-d
des morceaux ne contenant pas d'information, les **introns**).
– L'**ARNm** terminé (mature) quitte le noyau par les pores nucléaires

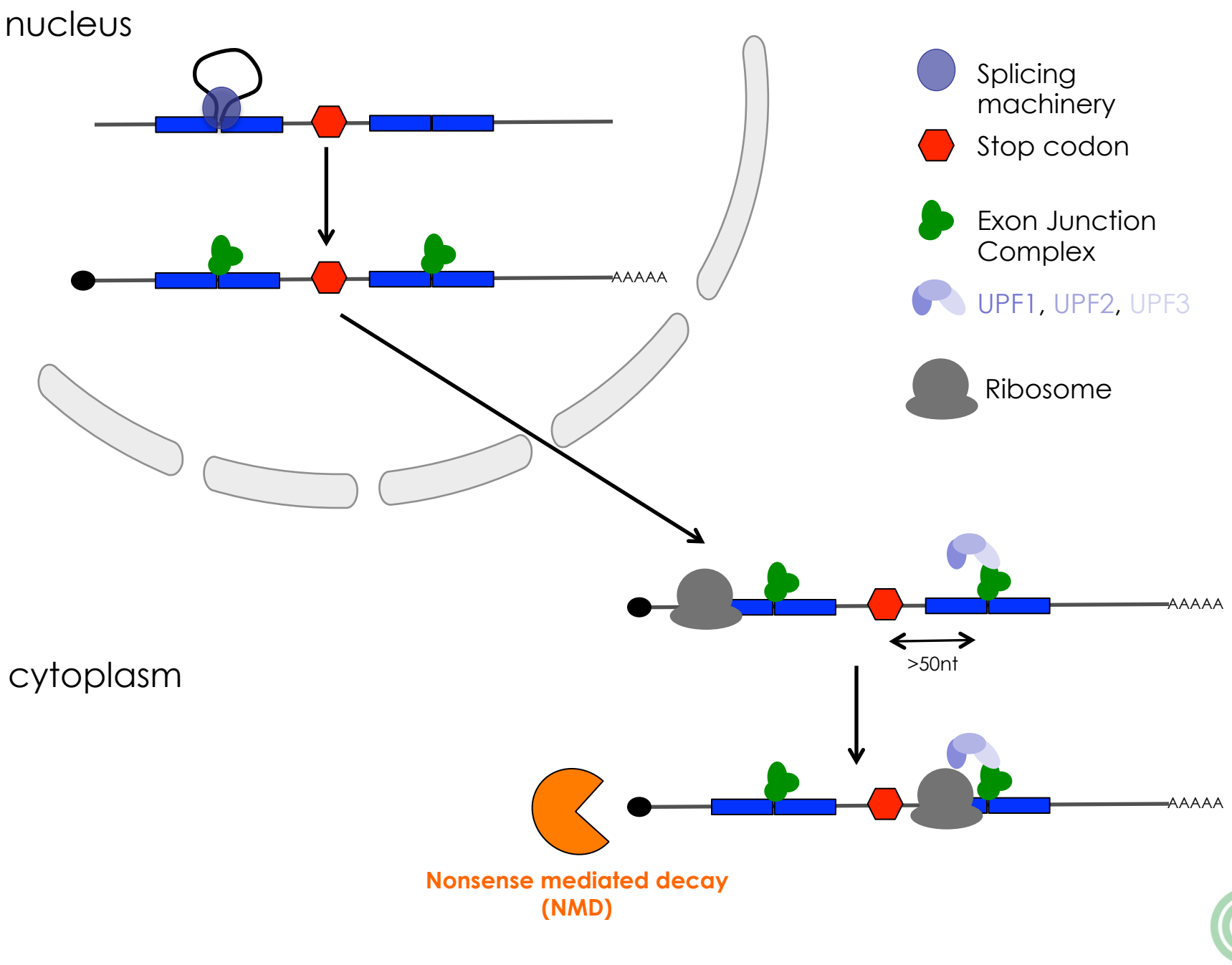
nucleus



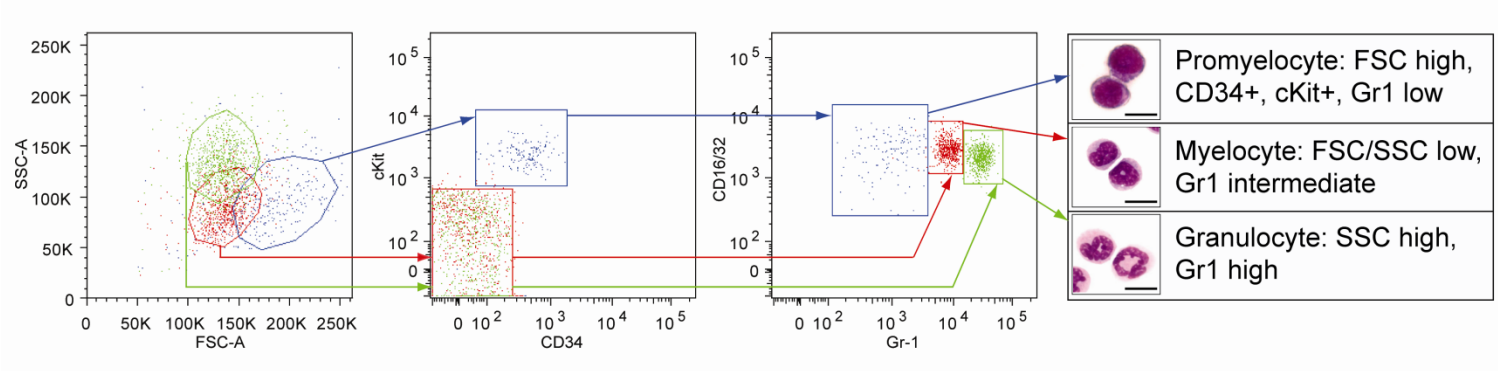
Splicing machinery

Exon Junction Complex

cytoplasm

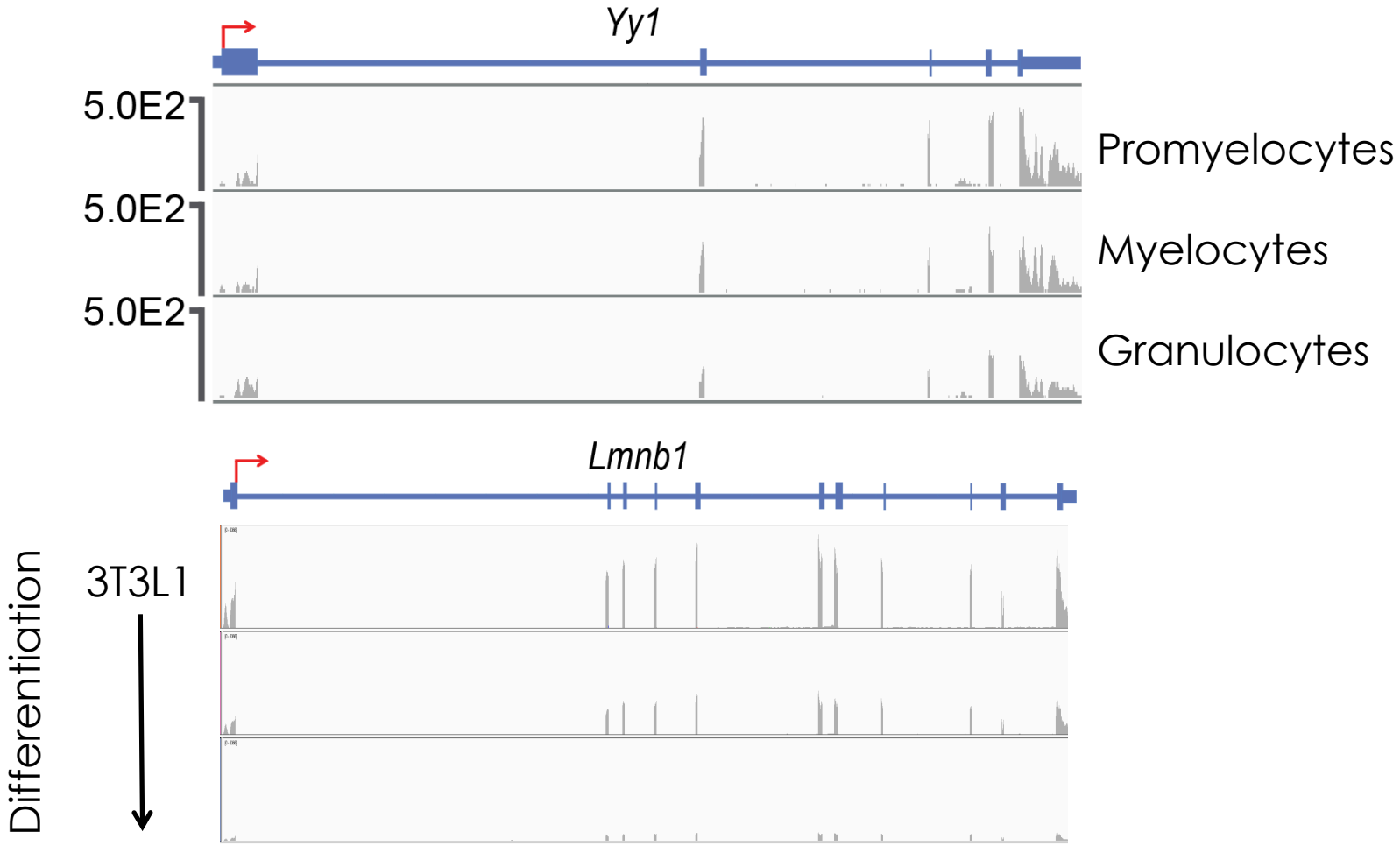

nucleus
Splicing machinery
Stop codon
Exon Junction Complex
UPF1, UPF2, UPF3
Ribosome
AAAAA
AAAAA
>50nt
cytoplasm
AAAAA
Nonsense mediated decay (NMD)

Granulopoiesis = good model

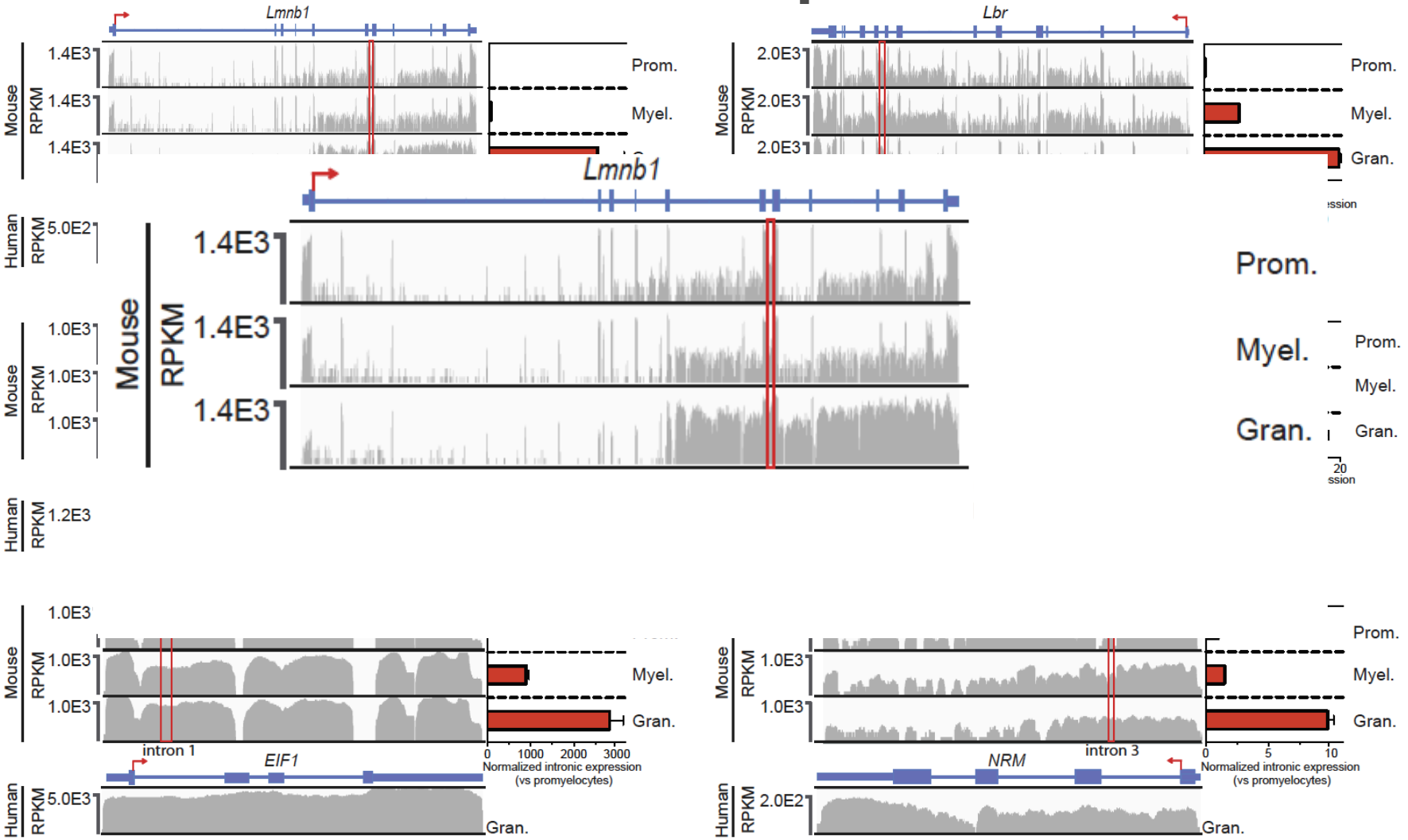


- **mRNA-seq**
- **Mass Spectrometry**

Reads in the intron?

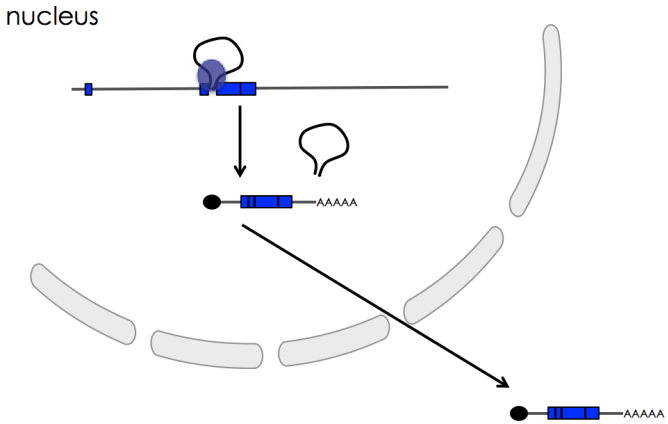


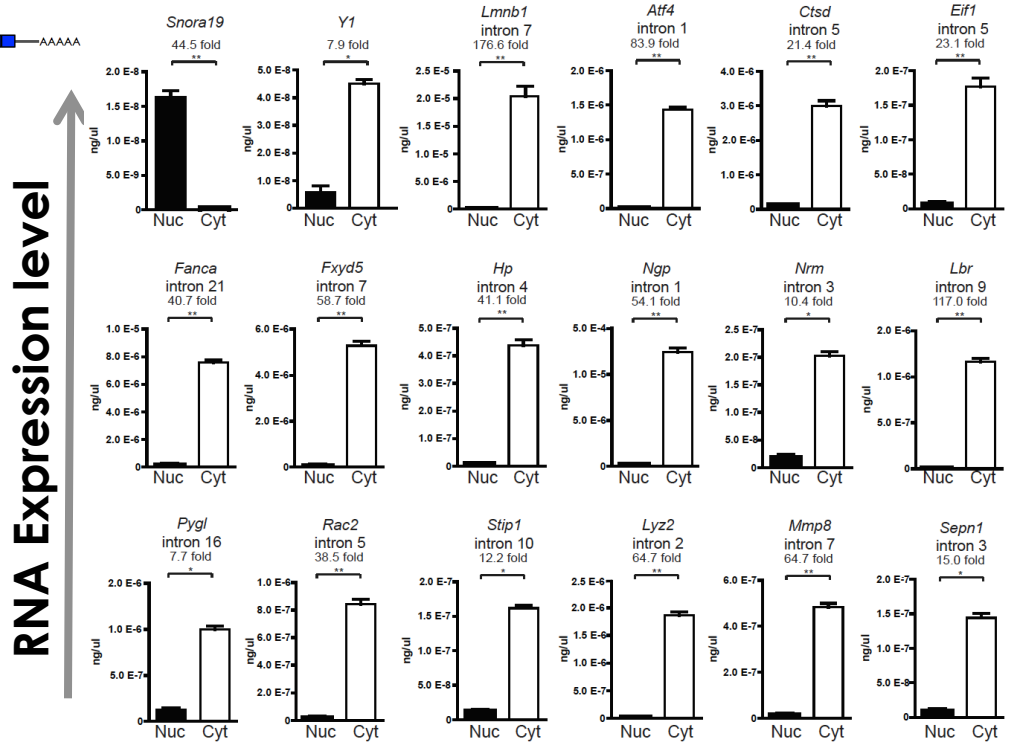
IR transcripts



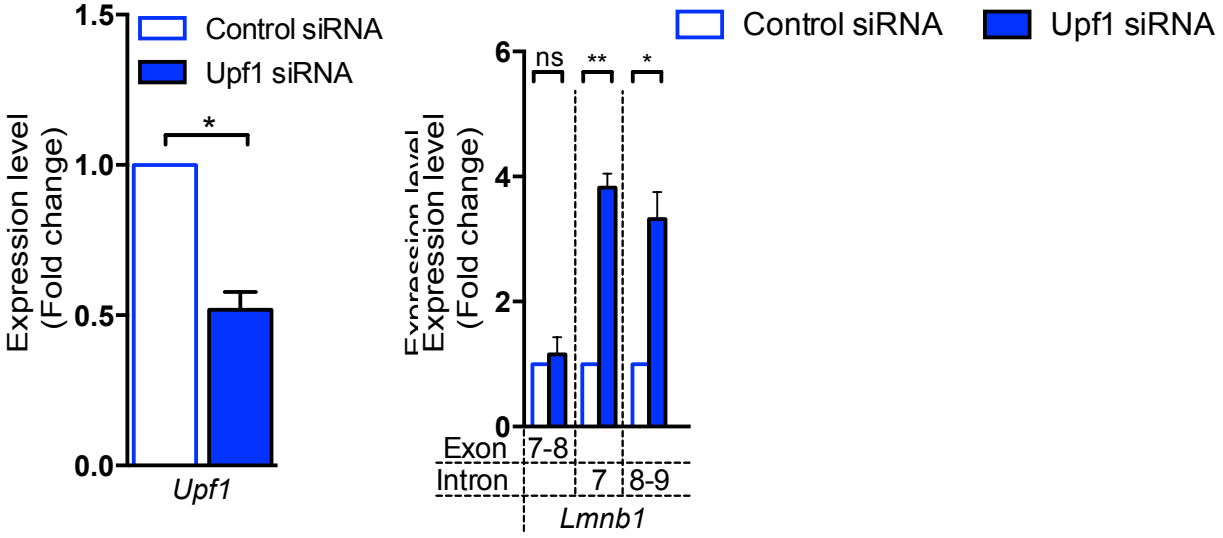
IR conserved in human & mouse ($p = 2.85 \times 10^{-22}$, hypergeometric test)

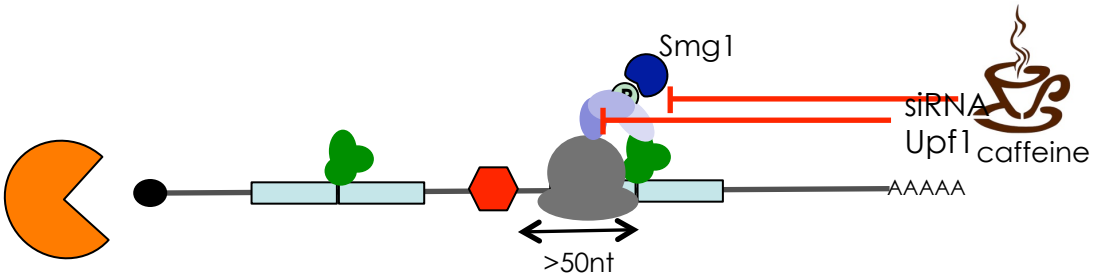
IR transcripts are not splicing intermediates



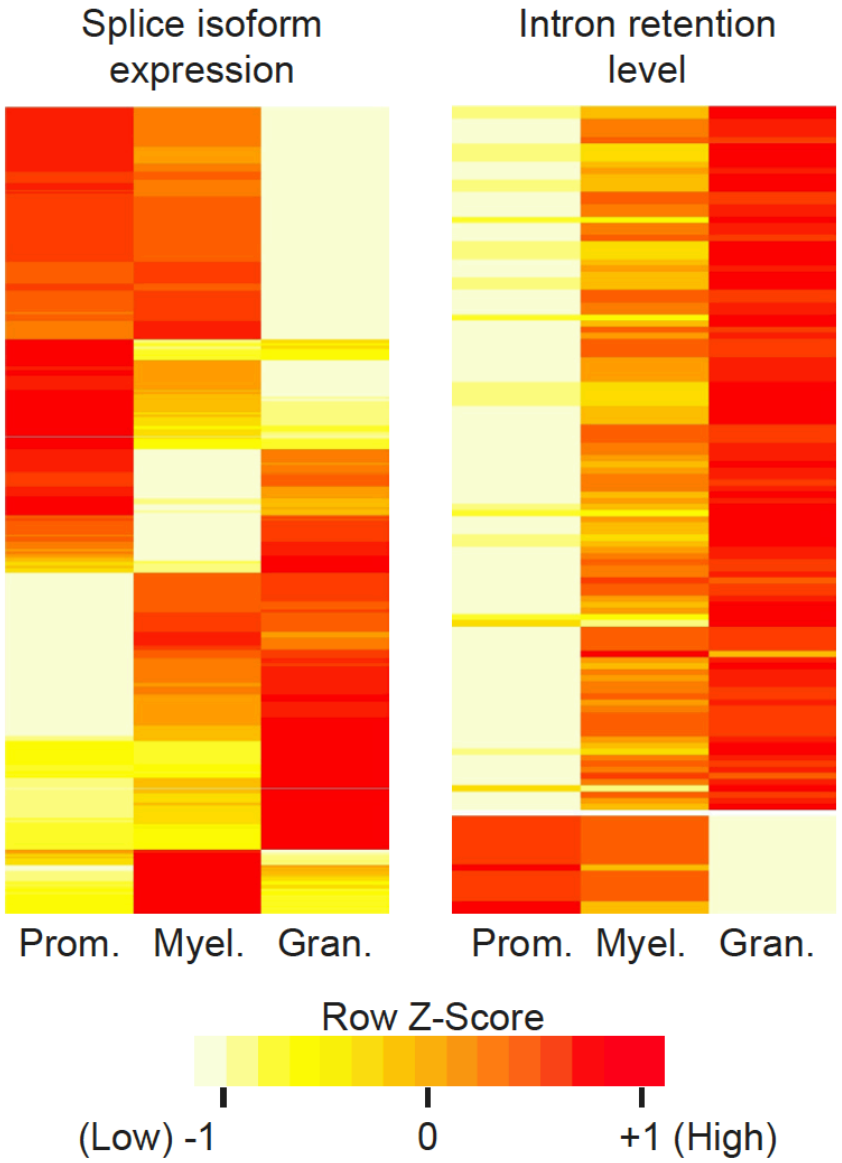


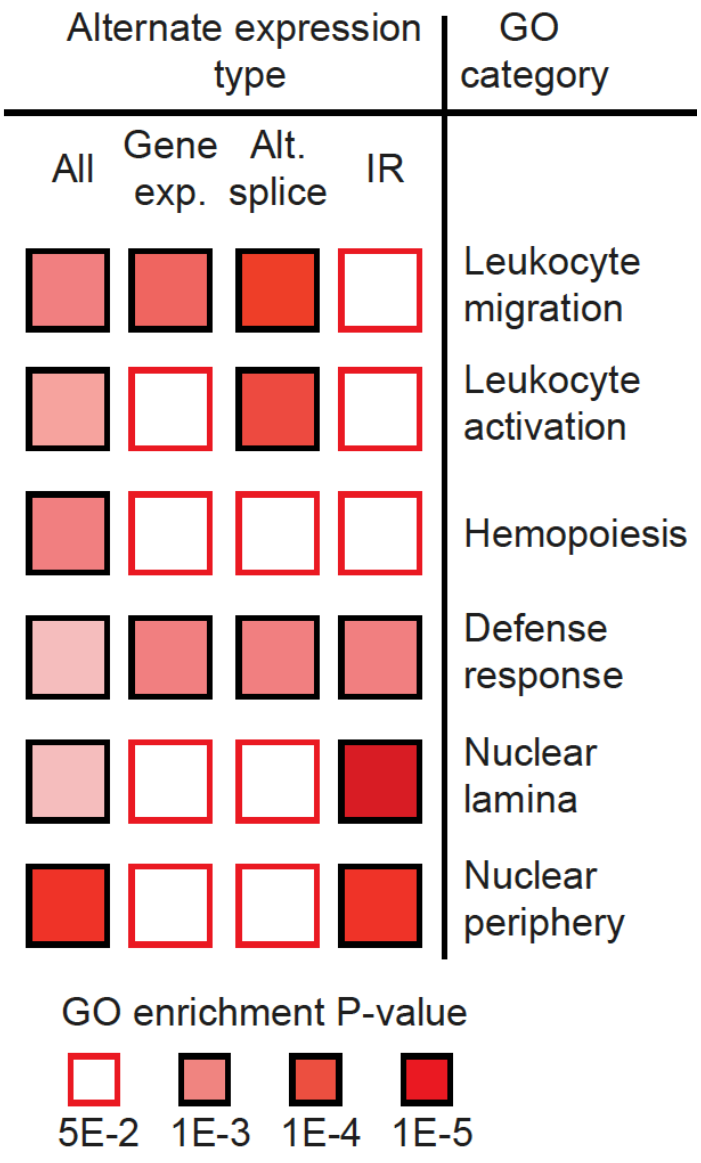
IR transcripts are degraded by NMD





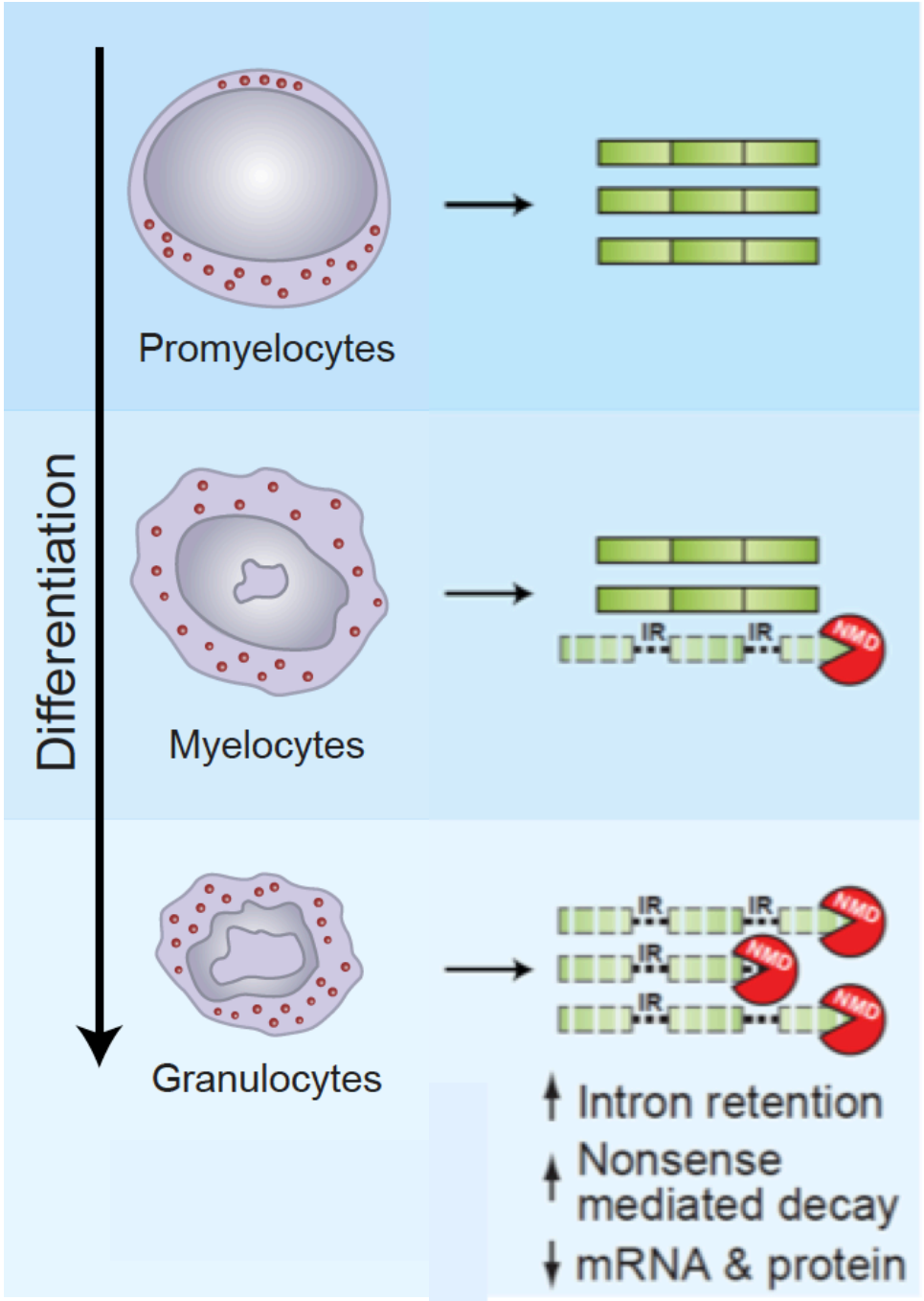
IR plays a role in granulopoiesis



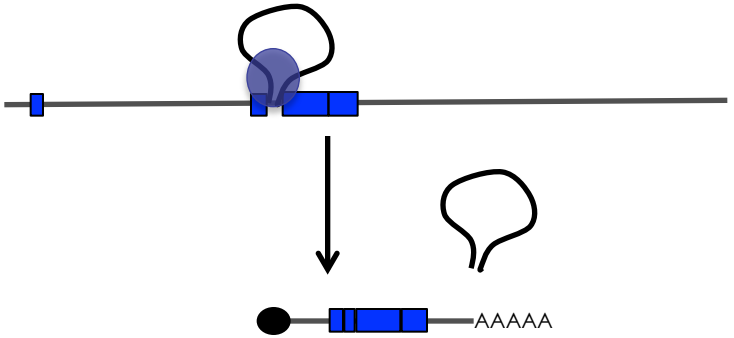


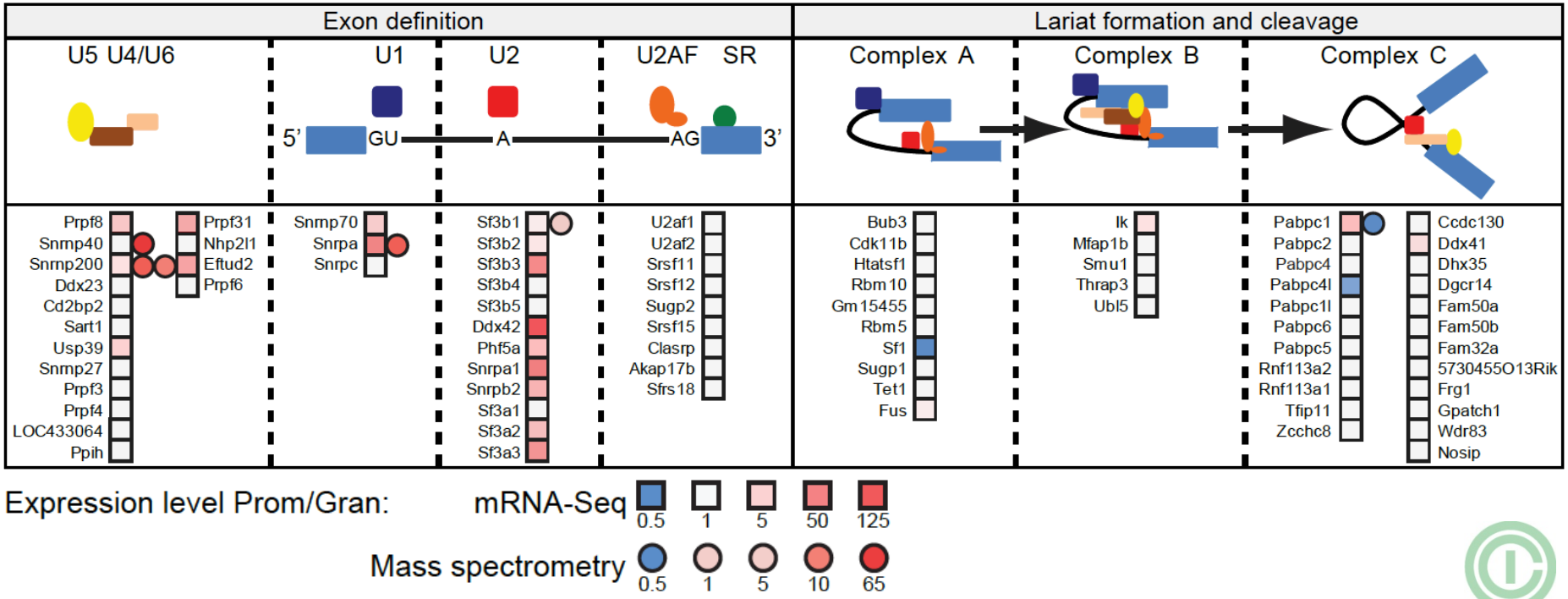
IR ↗ protein↘

Gene/Protein Symbols	Fold expression Gran/Prom*	Highest IR
Gnai2;Gnai-2	1.569	Gran
Tuba4;Tuba4a	0.998	Gran
Tubb2c;Tubb2c1;Tubb4	0.661	Gran
Gart;mCG_11450	0.574	Gran
Tubb5	0.497	Gran
D1Pas1-rs2;Ddx3	0.417	Gran
Lbr	0.403	Gran
Hnrnpa2b1;Hnrpa2b1	0.386	Gran
Grp78;Hspa5	0.217	Gran
Rpl7a	0.036	Gran
H3.3a;H3.3b;H3f3a;H3f3b	0.017	Gran
Lmnb1	0.014	Gran
Stip1	0.0008	Gran
Lmnb2	0.0007	Gran
H2afy	0.0002	Gran
Cd97	0.00004	Gran

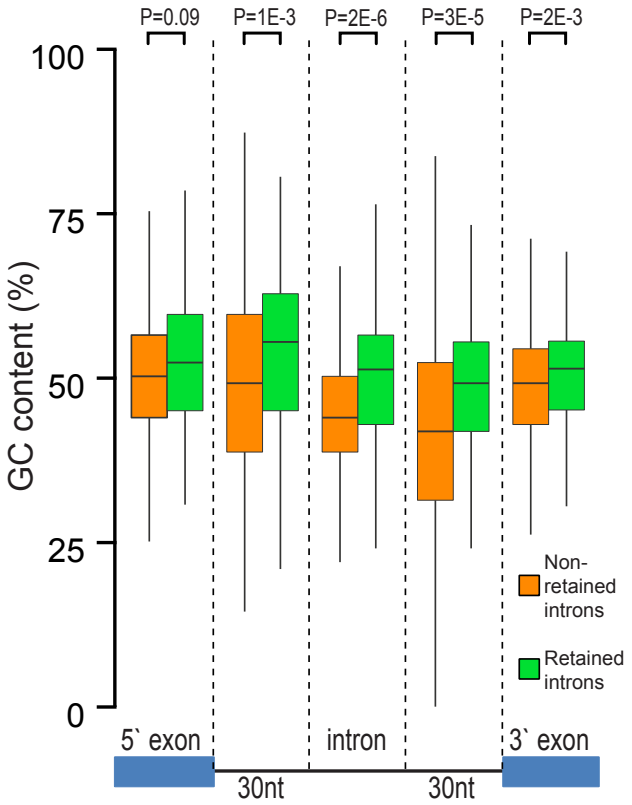

Differentiation
Promyelocytes
Myelocytes
Granulocytes
IR
IR
NMD
IR
IR
NMD
IR
NMD
Intron retention
Nonsense mediated decay
mRNA & protein

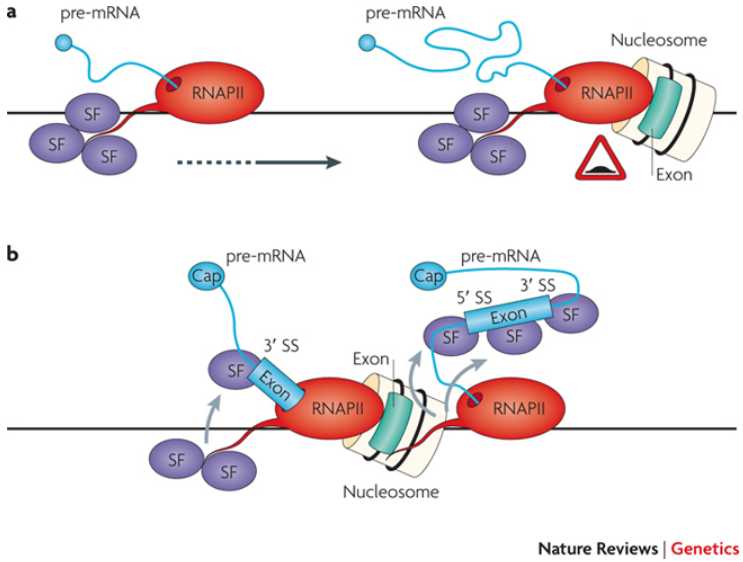
IR is associated with differential expression of splicing factors



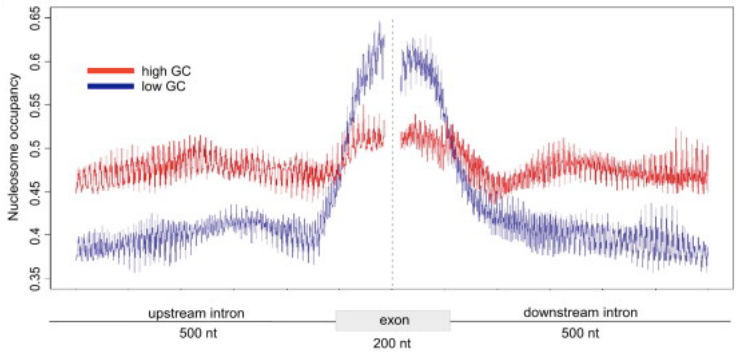


IR is associated with high intronic GC content





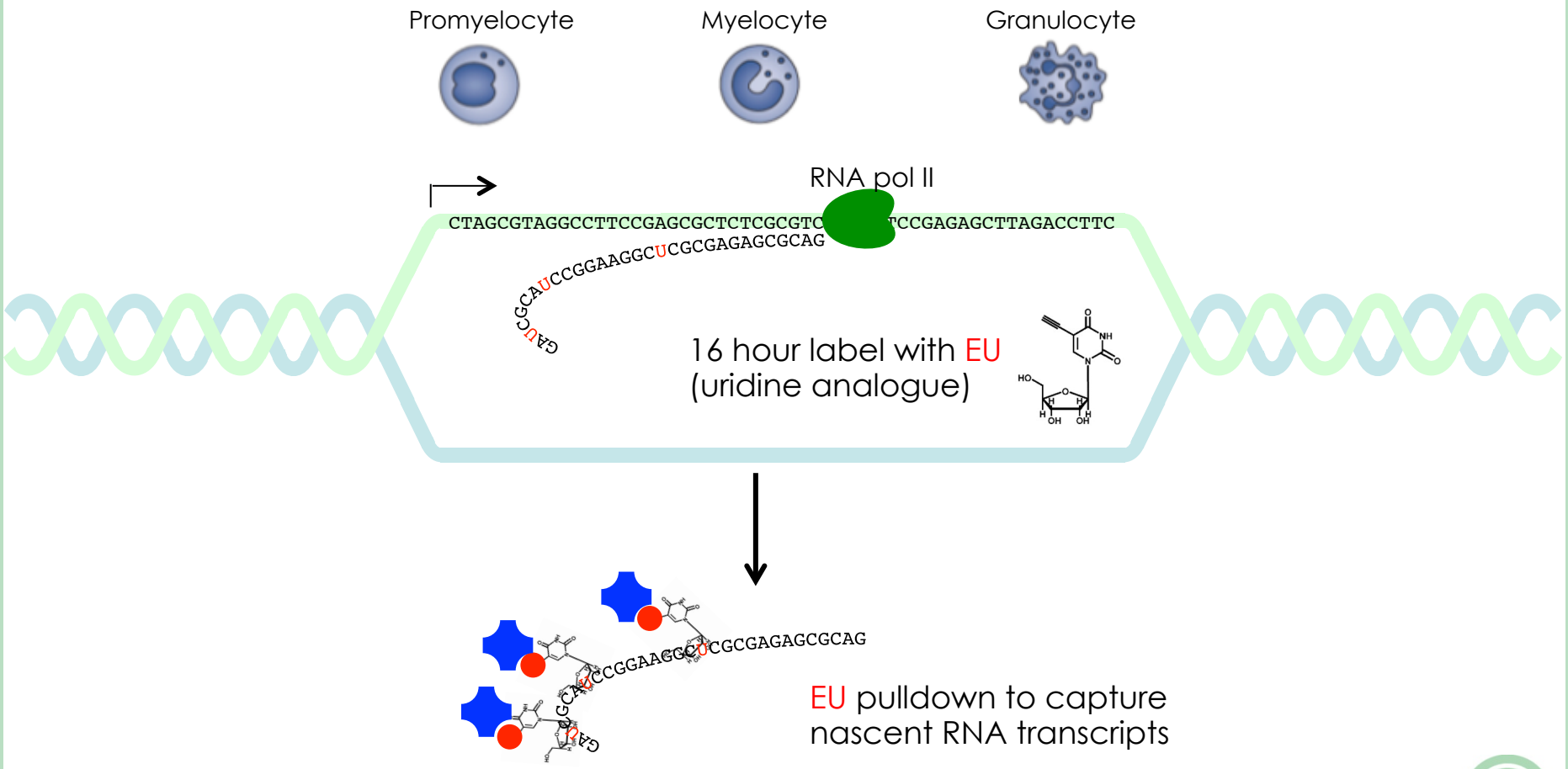
Keren et al., 2010



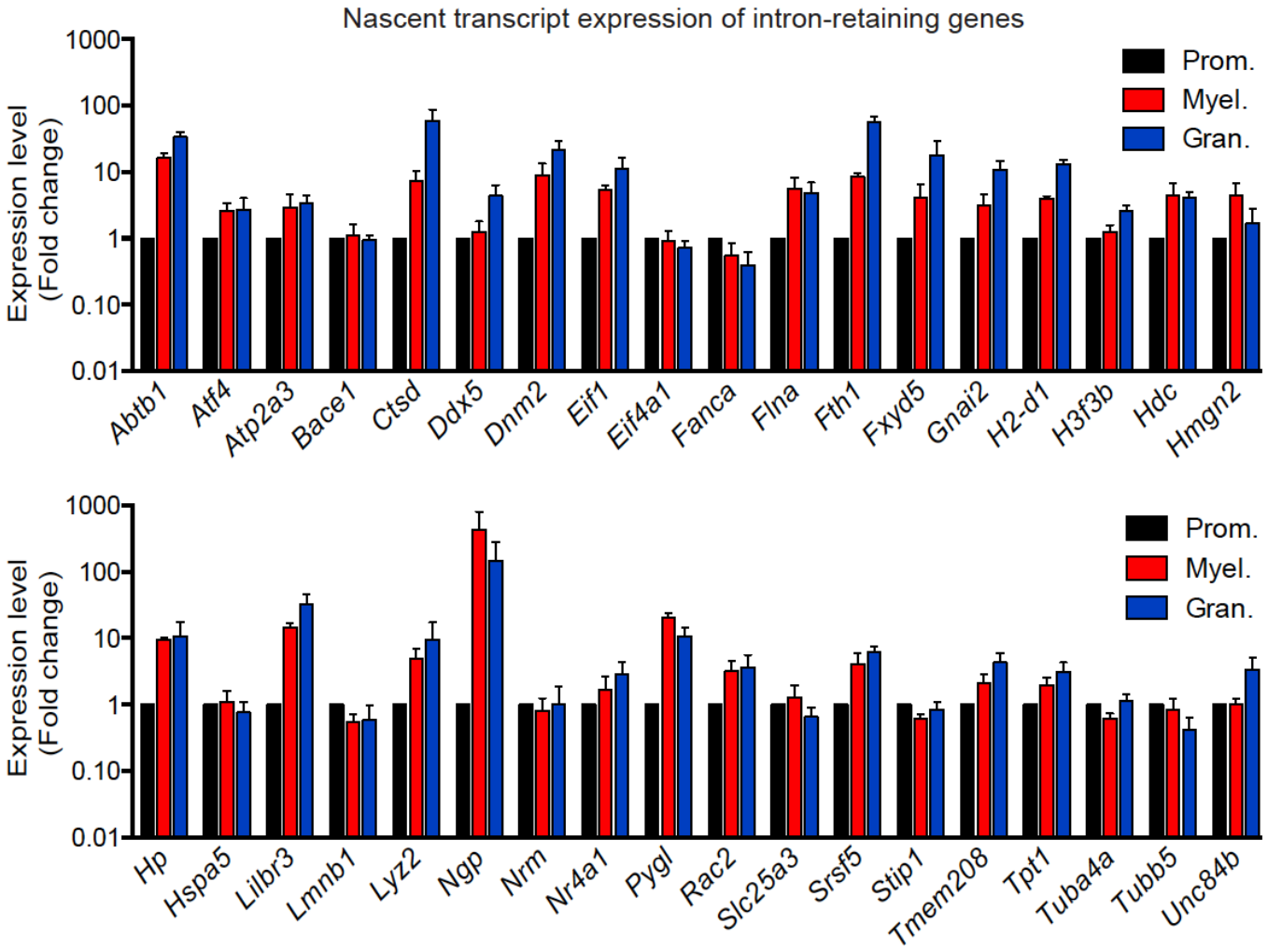
Amit et al., Cell Reports 2012

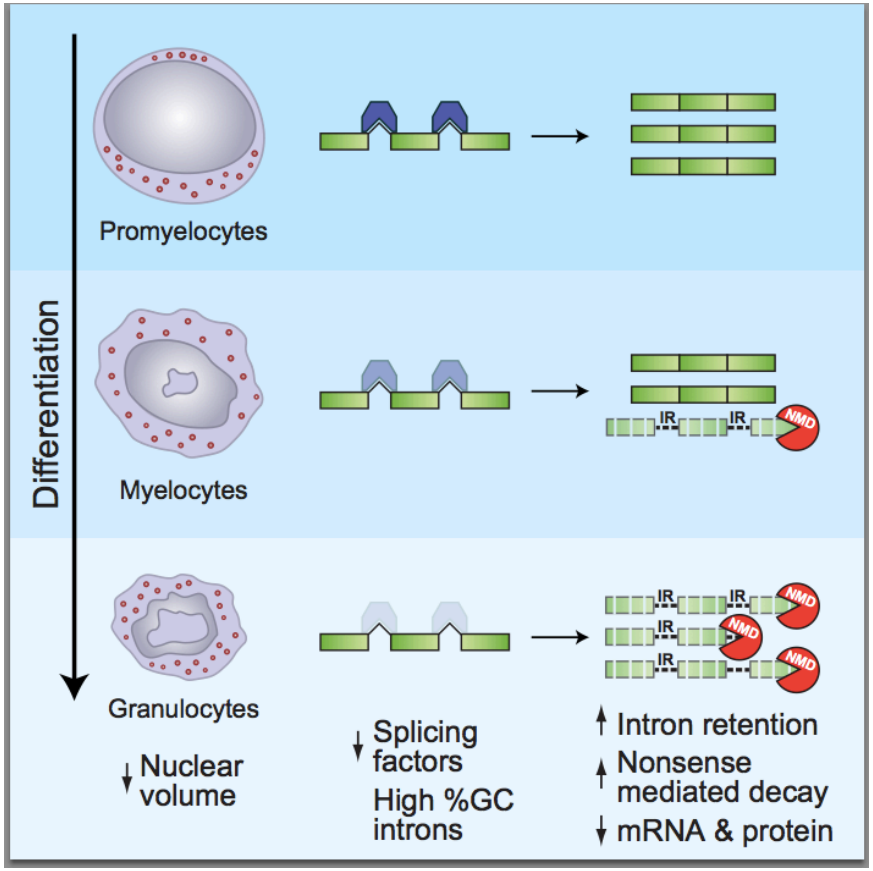
IR versus transcriptional regulation

- Does transcription play a role or cause IR?



IR & NMD regulate gene expression independently of transcriptional regulation



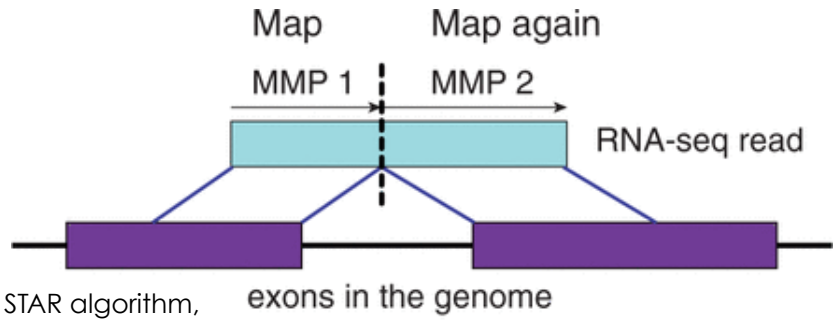


Cell

Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation

Justin J.-L. Wong,[1,4,8] William Ritchie,[1,2,4,8] Olivia A. Ebner,[5] Matthias Selbach,[5] Jason W.H. Wong,[6] Yizhou Huang,[6] Dadi Gao,[1,2,4] Natalia Pinello,[1,4] Maria Gonzalez,[1,4] Kinsha Baidya,[1,4] Annora Thoeng,[1,4] Teh-Liane Khoo,[1,4] Charles G. Bailey,[1,4] Jeff Holst,[1,3,4,9] and John E.J. Rasko[1,4,7,9,*]

Finding IR: why has it been ignored?



STAR algorithm,
Dobin et al, 2012

Stitching introduces artificial bias
Intron length
Splice Site "bonus"

Mappability
Repeats and low complexity regions

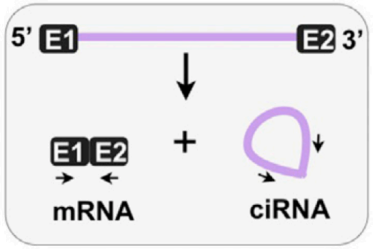
Intron-centric Biology

Neuron

Report

Cytoplasmic Intron Sequence-Retaining Transcripts Can Be Dendritically Targeted via ID Element Retrotransposons

Peter T. Buckley,[1,2,5] Miler T. Lee,[2,3,4,5] Jai-Yoon Sul,[1] Kevin Y. Miyashiro,[1] Thomas J. Bell,[1] Stephen A. Fisher,[2,3] Junhyong Kim,[2,3,4,6] and James Eberwine[1,2,4,6,*]



ciRNAs

Zhang et al., 2013

nature

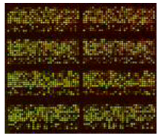
Natural RNA circles function as efficient microRNA sponges.

Hansen TB[1], Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J.

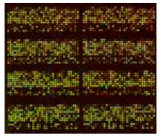
1
2
3

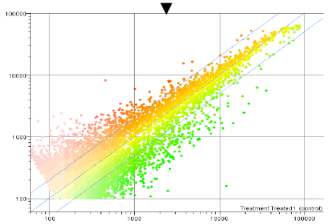
Functional analysis of miR-302c: mixed opinions

Melanoma metastasis GSE8401



Melanoma GSE8401





a

b

picTar predicted has-miR-302c targets (142 genes)

TargetScan predicted has-miR-302c targets (131 genes)

GO enrichment analysis

Genes downregulated in Melanoma metastasis (1311 genes)

GO enrichment analysis

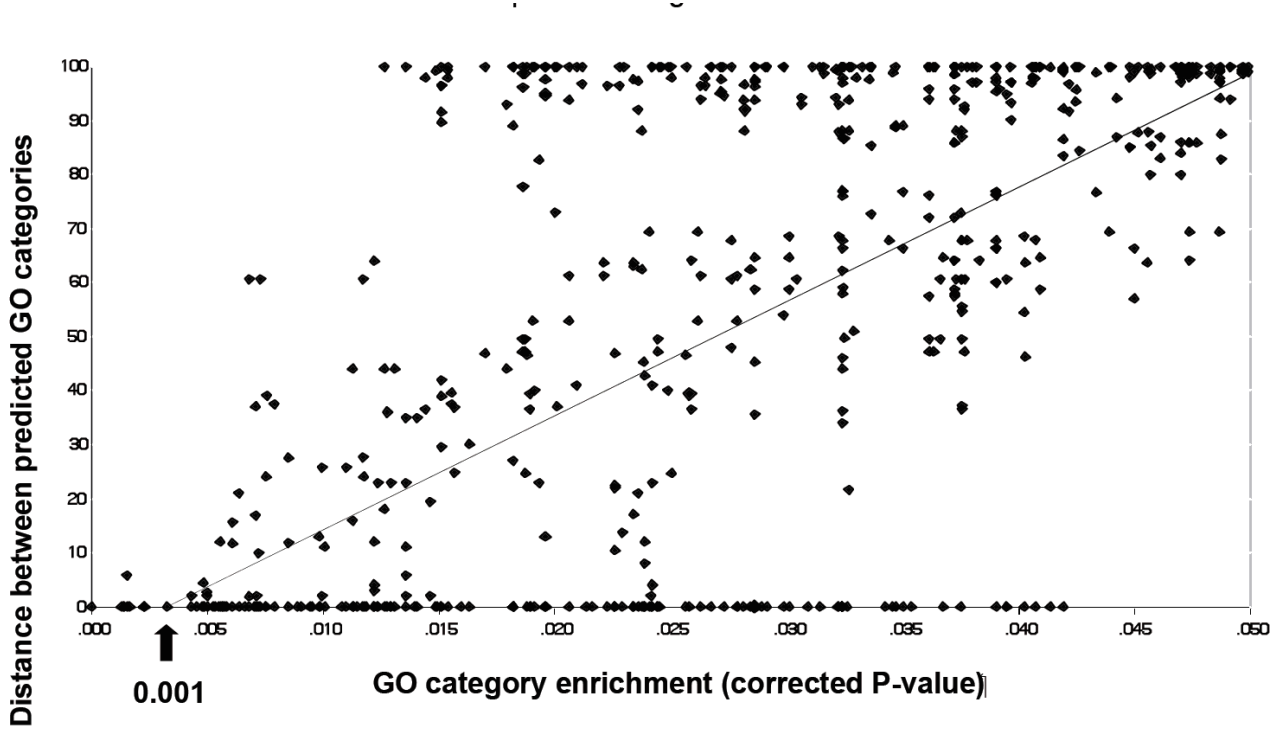
"blood vessel morphogenesis" GO:0048514, p<0.05

"transcription factor binding" GO:0008134, p<0.05

"microRNA 302c is involved in blood vessel morphogenesis in melanoma metastasis"

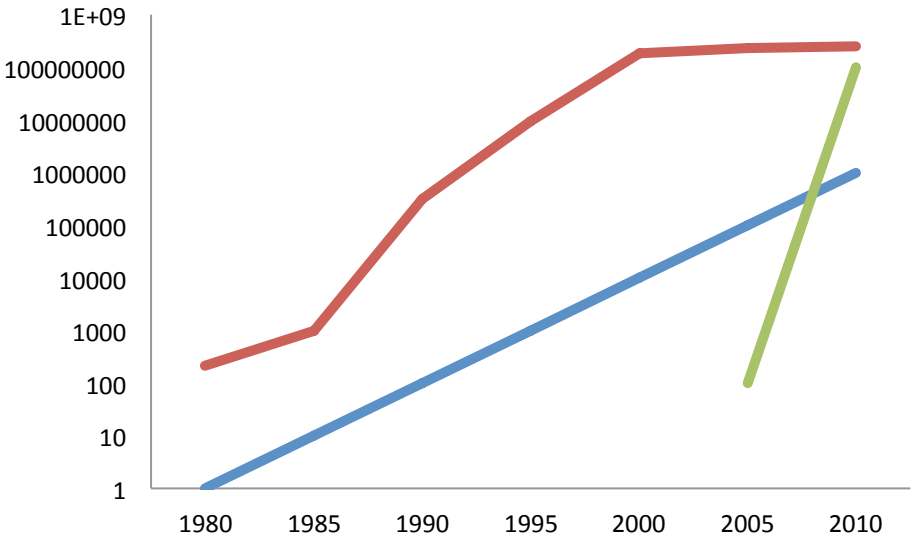
"microRNA 302c is involved in transcription factor binding in melanoma metastasis"

P=0.001 the golden number



Ritchie et al., Nature Methods 2010

The next 5 years



…10 years

RESEARCH Open Access

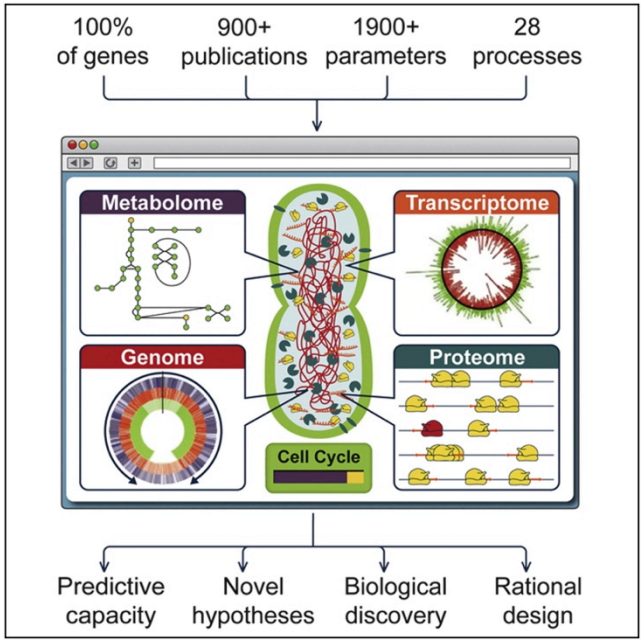
Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors

Steven JM Jones[1*], Janessa Laskin[2], Yvonne Y Li[1], Obi L Griffith[1], Jianghong An[1], Mikhail Bilenky[1], Yaron S Butterfield[1], Timothee Cezard[1], Eric Chuah[1], Richard Corbett[1], Anthony P Fejes[1], Malachi Griffith[1], John Yee[3], Montgomery Martin[2], Michael Mayo[1], Nataliya Melnyk[4], Ryan D Morin[1], Trevor J Pugh[1], Tesa Severson[1], Sohrab P Shah[4,5], Margaret Sutcliffe[2], Angela Tam[1], Jefferson Terry[4], Nina Thiessen[1], Thomas Thomson[2], Richard Varhol[1], Thomas Zeng[1], Yongjun Zhao[1], Richard A Moore[1], David G Huntsman[3], Inanc Birol[1], Martin Hirst[1], Robert A Holt[1], Marco A Marra[1]

nature International weekly journal of science

Genome complexity, robustness and genetic interactions in digital organisms

Richard E. Lenski[1], Charles Ofria[2], Travis C. Collier[4] & Christoph Adami[3]



Karr et al., 2012

Mycoplasma genitalium

Acknowledgements

Gene and Stem Cell Therapy Lab

John Rasko
Justin Wong
Chuck Bailey
Natalia Pinello
Jeff Holst
Annora Thoeng
Maria Gonzalez
Kinsha Baidya
Liane Khoo

Flow Cytometry and Microscopy

Adrian Smith
Steven Allen
Suat Dervish
Shihong Yang
Rob Salomon
Frank Kao
Kristina Jahn
Ben Roediger

Bioinformatics Program

William Ritchie
Dadi Gao
Rob Middleton

Max-Delbrück Centre for Molecular Medicine

Matthias Selbach
Olivia Ebner

University of New South Wales

Jason Wong
Yizhou Huang

Geneworks

Rob King

- **Funding:**