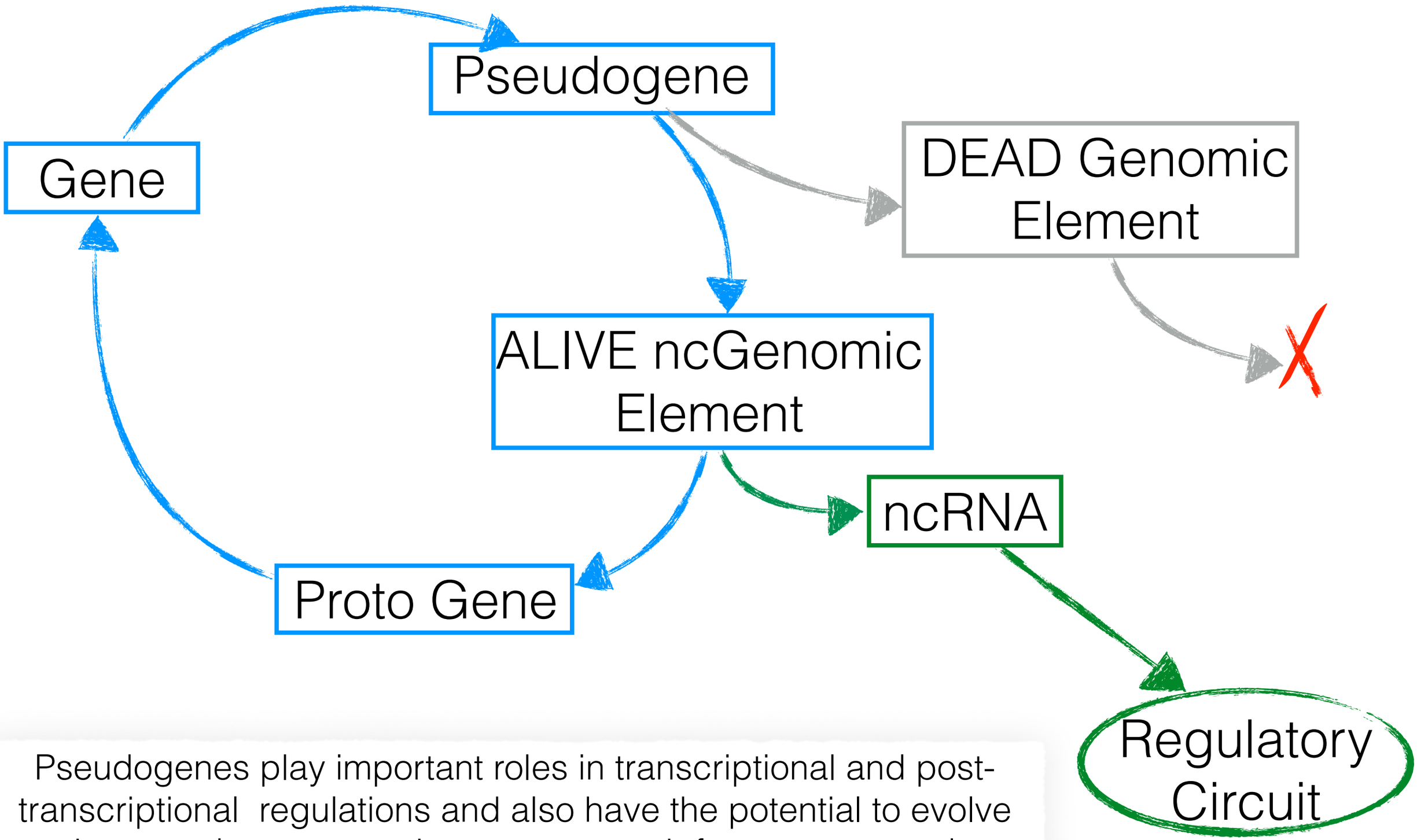
PSEUDOGENES
~ CANCER'S BIG CATCH ~

Cristina Sisu

Journal Club 30th July 2015

Pseudogene circle of life



Pseudogenes play important roles in transcriptional and post-transcriptional regulations and also have the potential to evolve into novel genes, serving as a reservoir for gene renewal

REVIEW

Pseudogenes: Pseudo-functional or key regulators in health and disease?

RYAN CHARLES PINK, KATE WICKS, DANIEL PAUL CALEY, EMMA KATHLEEN PUNCH, LAURA JACOBS, and DAVID RAUL FRANCISCO CARTER

School of Life Sciences, Oxford Brookes University, Headington, Oxford, OX3 0BP, United Kingdom

REVIEW

MOLECULAR BIOLOGY

Pseudogenes: Newly Discovered Players in Human Cancer

Laura Poliseno*

Review

Pseudogene in cancer: real functions and promising signature

Lu Xiao-Jie,[1] Gao Ai-Mei,[2] Ji Li-Juan,[3] Xu Jiang[3]

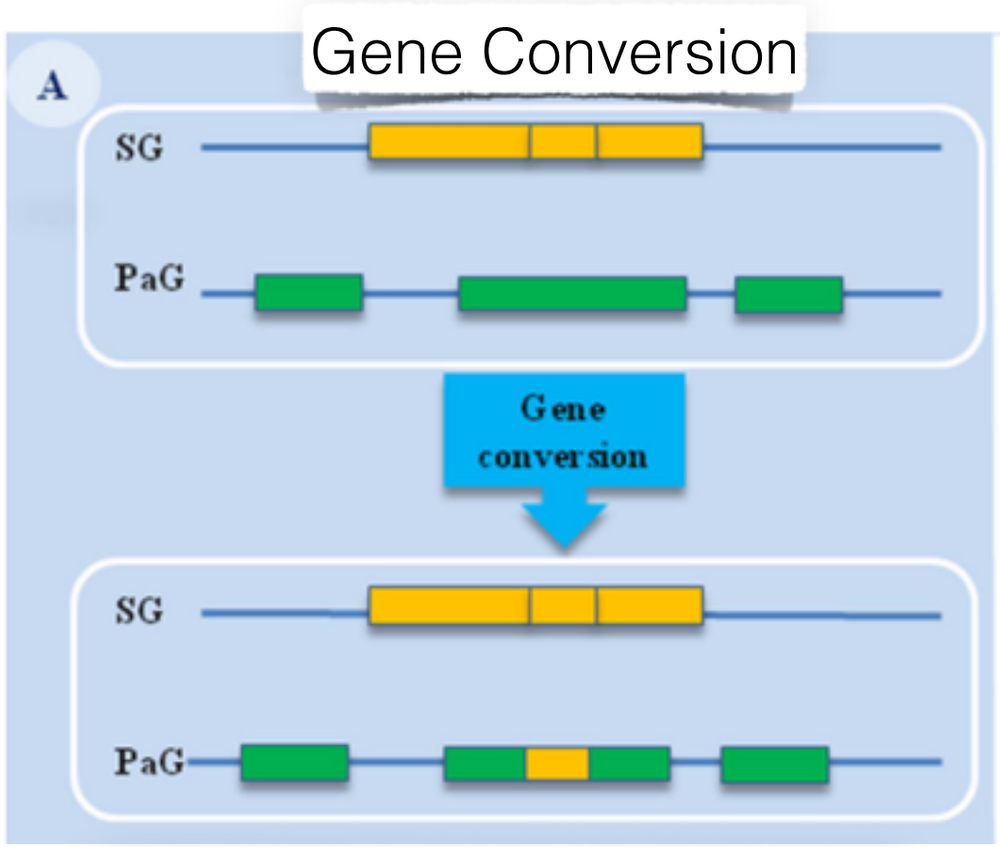
Pseudogenes **regulate & interfere** with the expression & activity of functional protein coding genes

- Through their pseudogene DNA sequence
- Through their RNA product
- Through their pseudo-protein/polypeptide product

Pseudogenes **regulate & interfere** with the expression & activity of functional protein coding genes

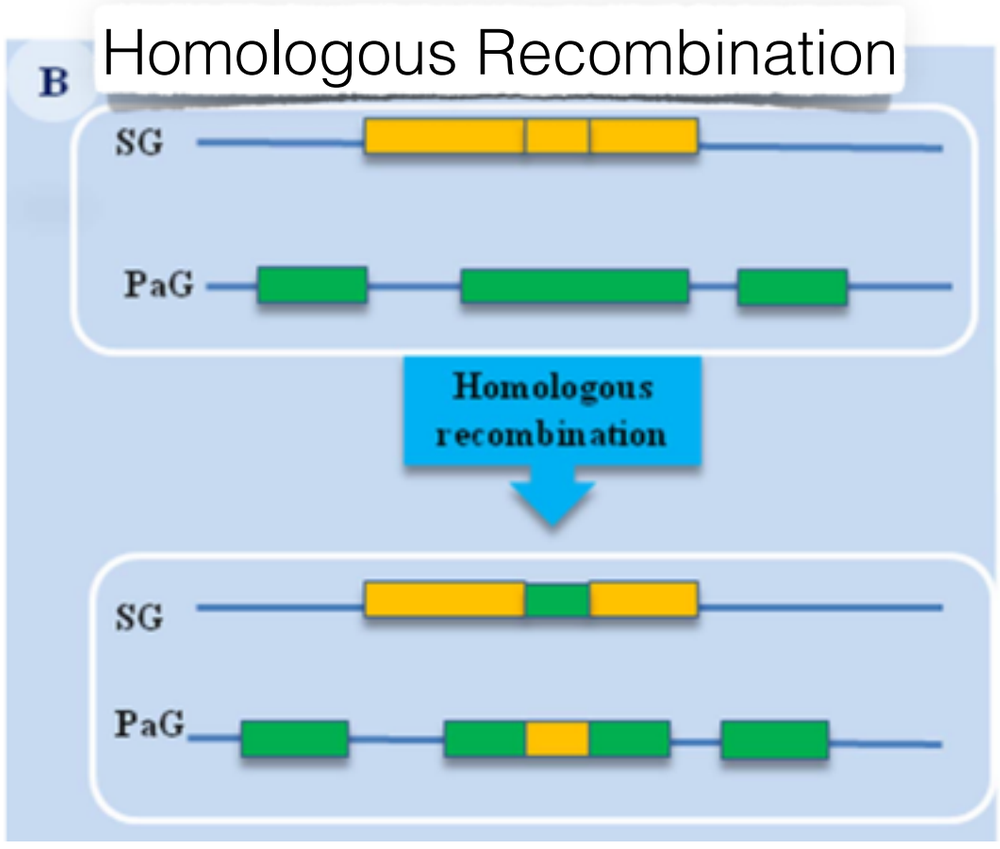
- Through their pseudogene DNA sequence
- Through their RNA product
- Through their pseudo-protein/polypeptide product

Functions of pseudogene DNA



The pseudogene DNA sequence replaces a sequence in the parent such that the two sequences become identical.

e.g: CYP2A6 gene is converted by its pseudogene to a new variant CYP2A6*1B that influences the smoking-induced lung cancer risk.

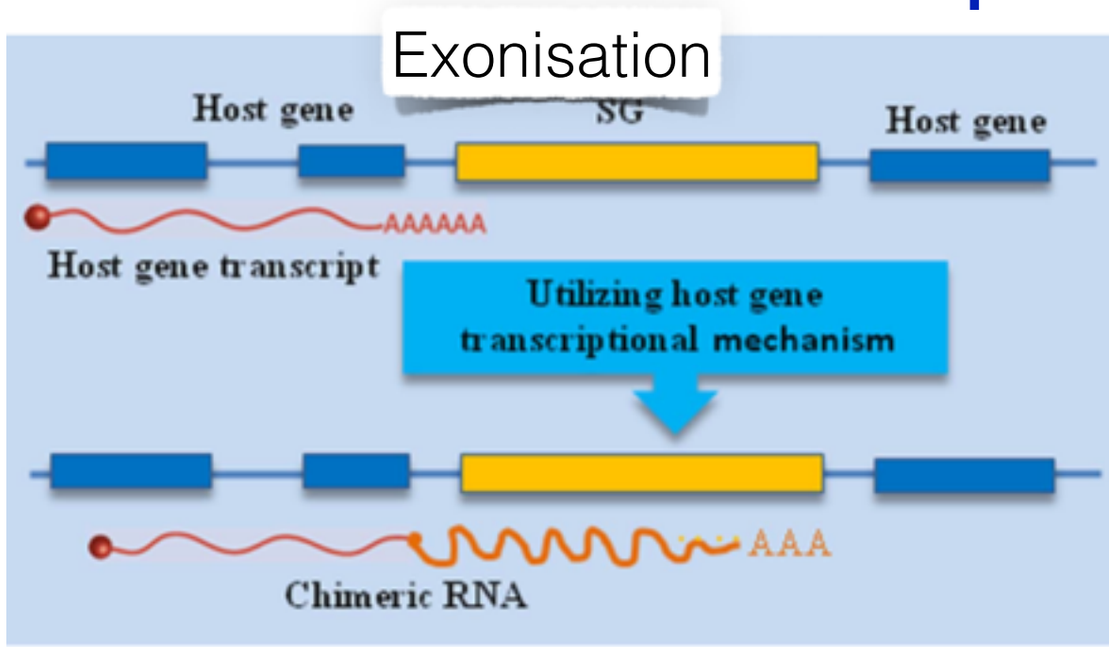


Exchange of DNA between the parent and the pseudogene sequence.

e.g: homologous recombination between BRCA1 gene and psiBRCA1 pseudogene resulted in a 37kb deletion that pseudogenizes the original BRCA1 gene (by removing the promoter and start codon sequence)

—> new mechanism for oncosuppressor gene inactivation.

Functions of pseudogene DNA



The pseudogene uses the transcriptional mechanism of the host gene.

e.g: somatically acquired pseudogenes during cancer development

nature COMMUNICATIONS

ARTICLE

Received 6 Feb 2014 | Accepted 13 Mar 2014 | Published 9 Apr 2014

DOI: 10.1038/ncomms4644 OPEN

Processed pseudogenes acquired somatically during cancer development

Susanna L. Cooke[1], Adam Shlien[1], John Marshall[1], Christodoulos P. Pipinikas[2], Inigo Martincorena[1], Jose M.C. Tubio[1], Yilong Li[1], Andrew Menzies[1], Laura Mudie[1], Manasa Ramakrishna[1], Lucy Yates[1], Helen Davies[1], Niccolo Bolli[1,3], Graham R. Bignell[1], Patrick S. Tarpey[1], Sam Behjati[1,3], Serena Nik-Zainal[1], Elli Papaemmanuil[1], Vitor H. Teixeira[2], Keiran Raine[1], Sarah O'Meara[1], Maryam S. Dodoran[1], Jon W. Teague[1], Adam P. Butler[1], Christine Iacobuzio-Donahue[4], Thomas Santarius[5], Richard G. Grundy[6], David Malkin[7], Mel Greaves[8], Nikhil Munshi[9], Adrienne M. Flanagan[2,10], David Bowtell[11], ICGC Breast Cancer Group[1,17], Sancha Martin[1], Denis Larsimont[12], Jorge S. Reis-Filho[13], Alex Boussioutas[11,14], Jack A. Taylor[15], D. Neil Hayes[16], Sam M. Janes[2], P. Andrew Futreal[1], Michael R. Stratton[1], Ultan McDermott[1,5], Peter J. Campbell[1,3,5]

Background

- Cancer development is based on DNA mutations
- Processed pseudogenes
  - are the product of the LINE-mediated retrotransposition
  - influence the evolution though gene duplication, novel exons, gene fusions, antisense transcript productions

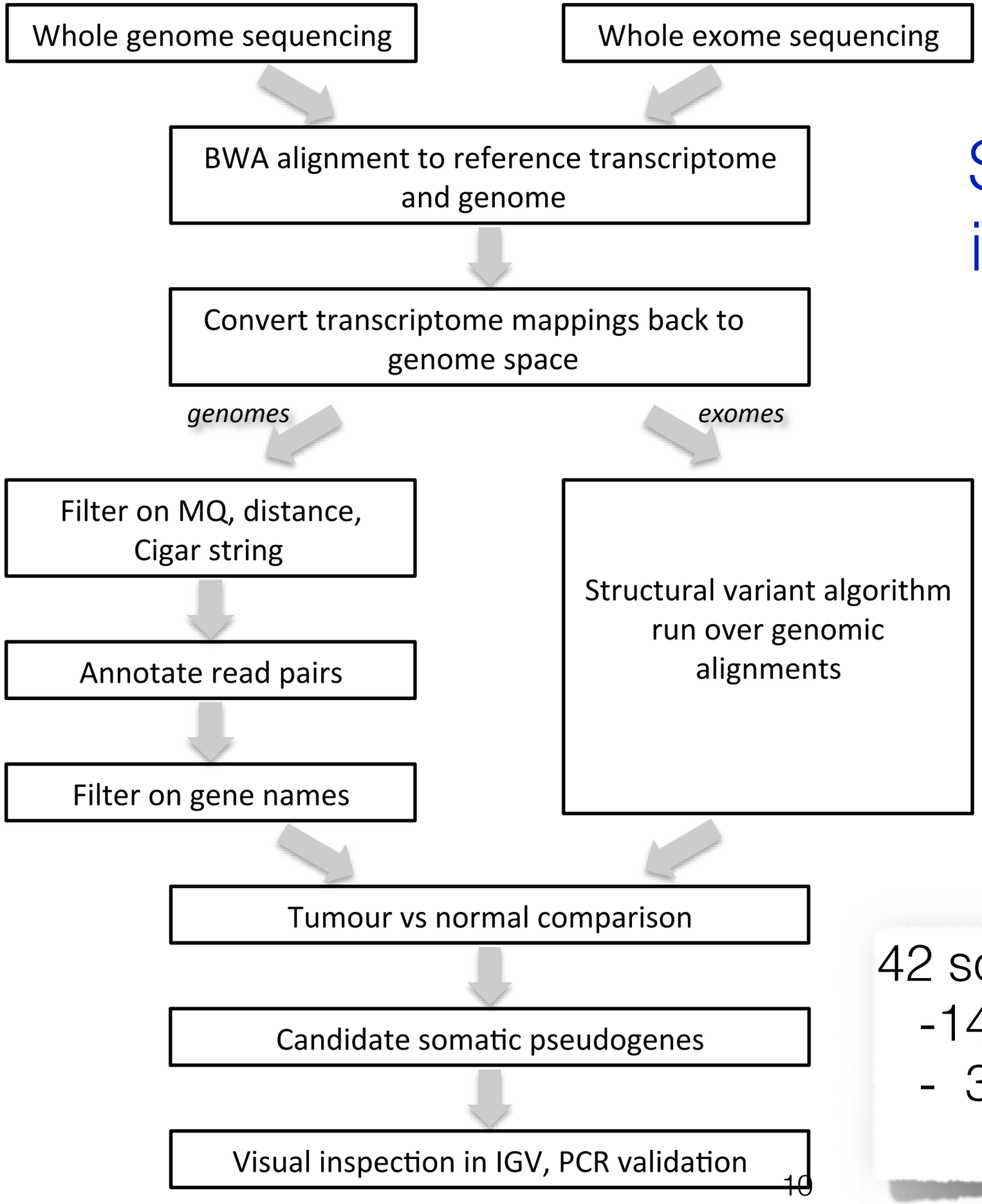
Aim

- Develop a bioinformatics method to detect somatically acquired processed pseudogenes in cancer samples through targeted exome and genome-wide studies

Data

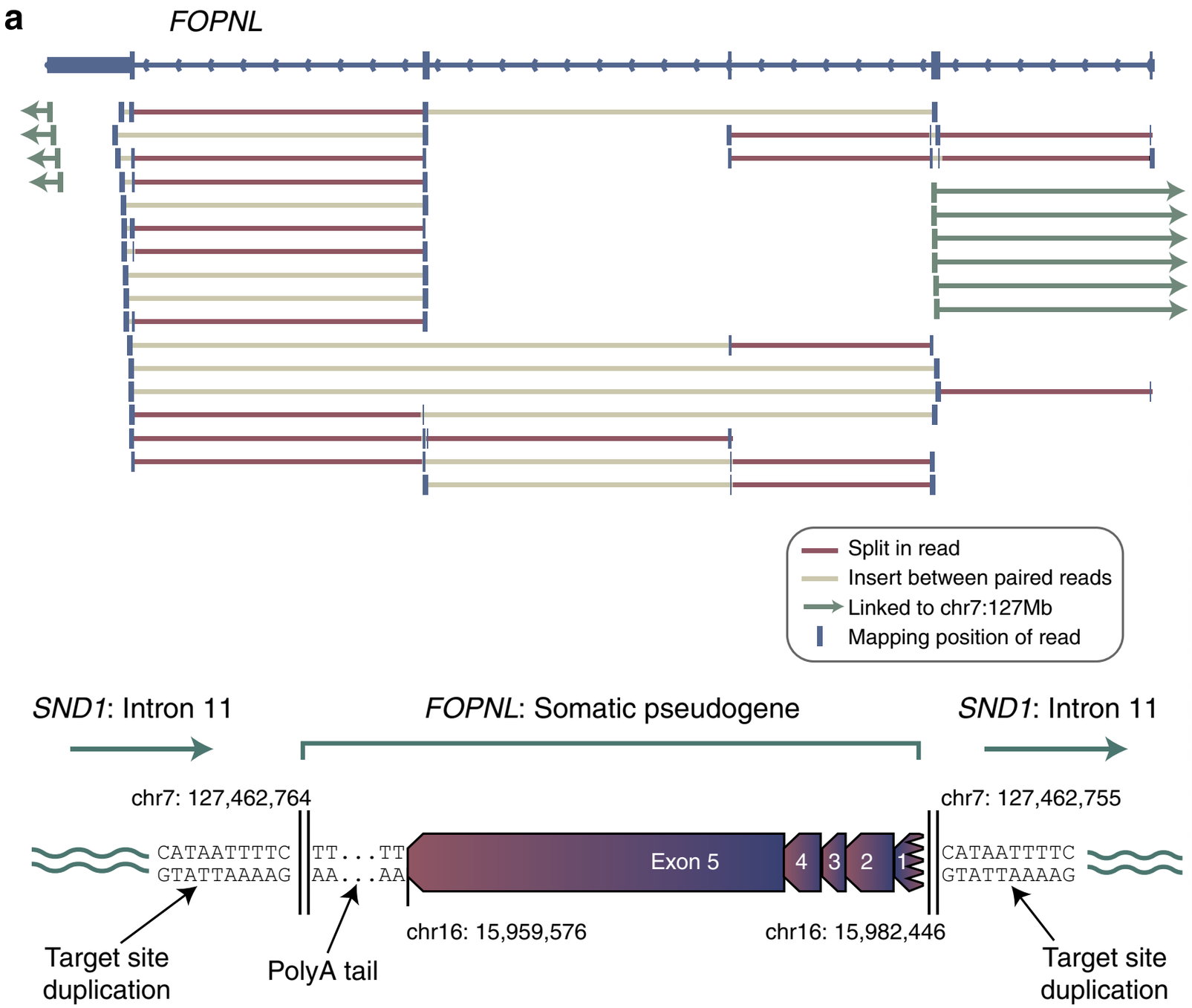
- 660 cancer samples: 629 primary & 31 cell lines
- 18 tumour types
- low coverage (2-5x) pair-end, high-coverage (30-40x) pari-end shotgun

Somatic pseudogene identification pipeline



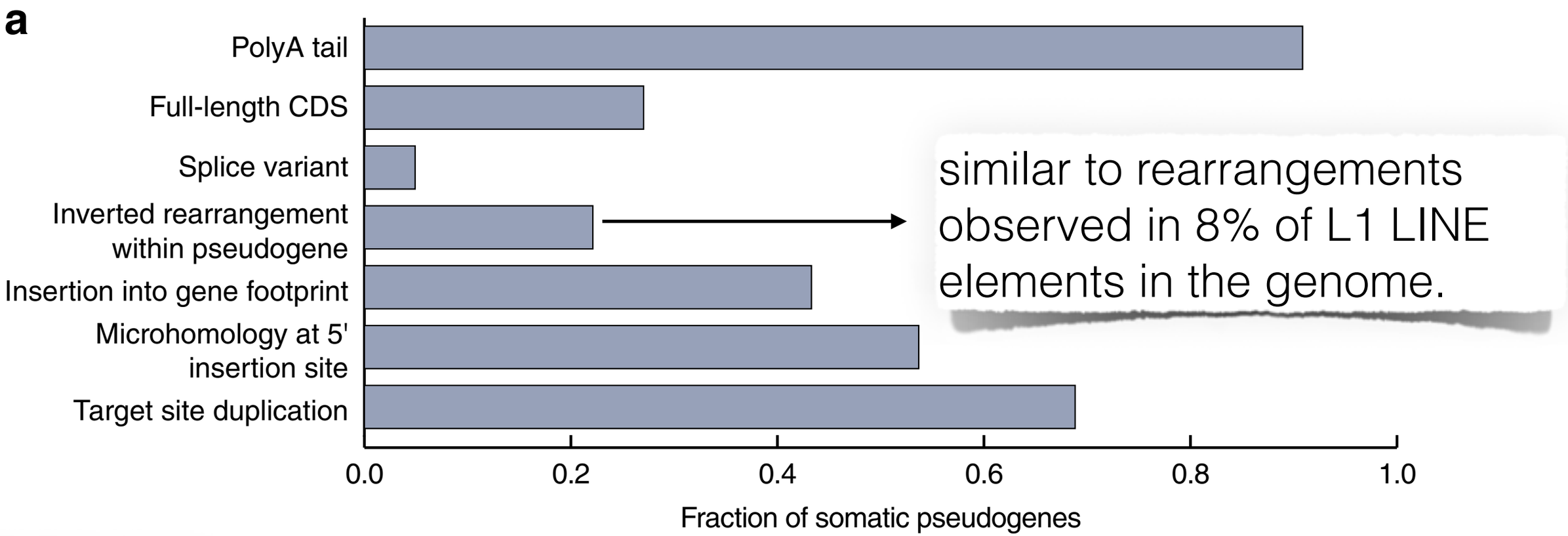
42 somatic pseudogenes:
- -14 out of 629 primary cases
- - 3 out 31 cell lines

e.g. Lung cancer — FOPNL

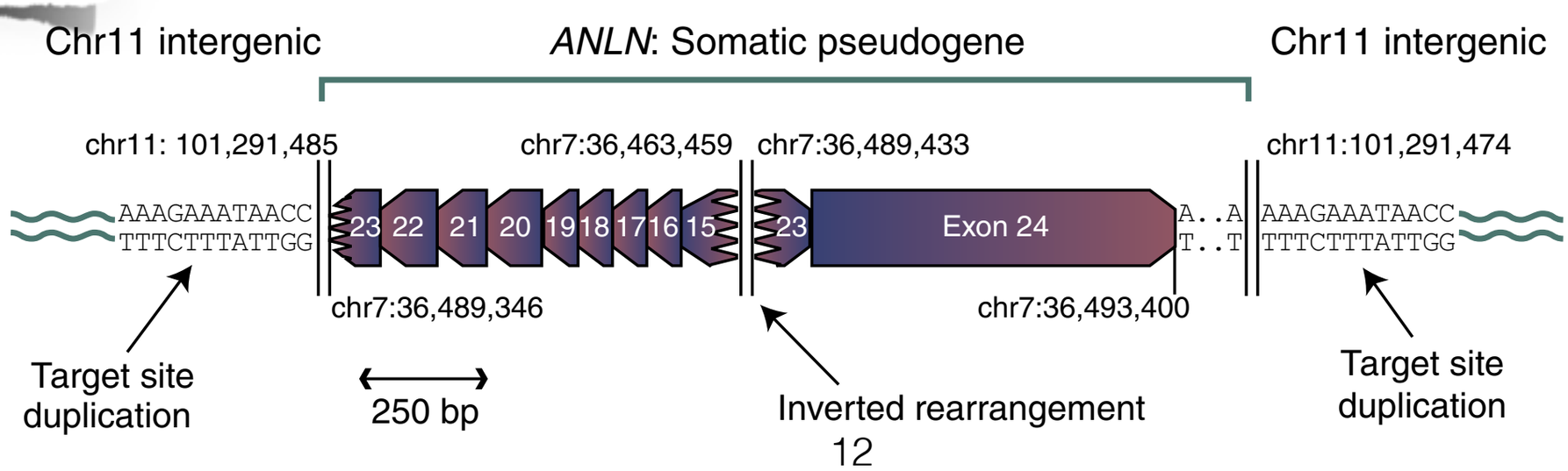


- all 5 FOPNL exons are inserted in SND1's 11th intron in oposite orientation
- 5'UTR & full 3'UTR sequence conserved
- The pseudogene is not seen in the germline DNA

Properties of somatic pseudogenes

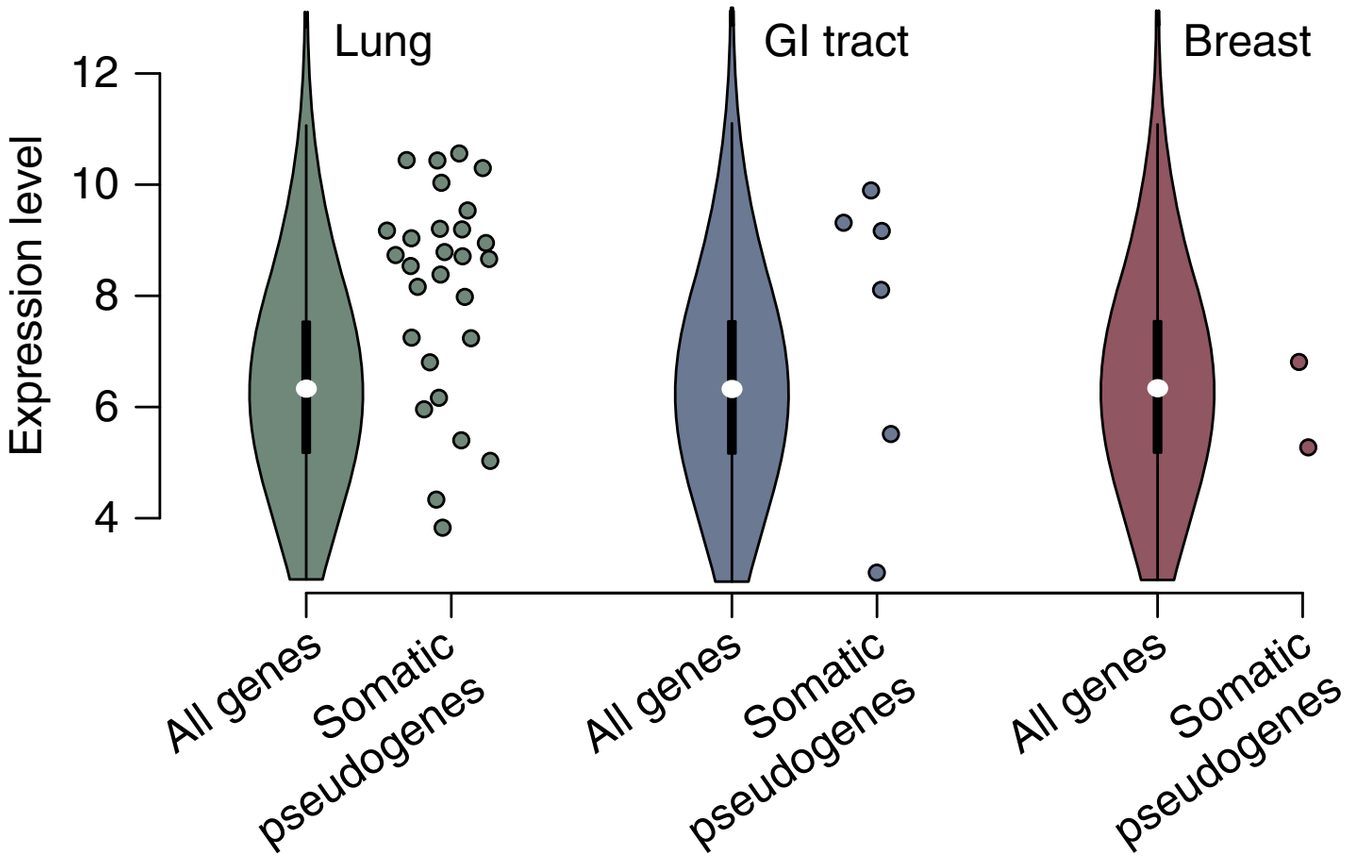


Example



Tissue specific patterns of somatic pseudogenes in cancer

- Somatic pseudogene were most frequent in non-small cell lung cancer (19%) & colorectal cancer (18%)
    * high rate of somatic retrotransposition of the LINE elements



- Highly expressed transcripts are more likely to be templates for somatic pseudogenes

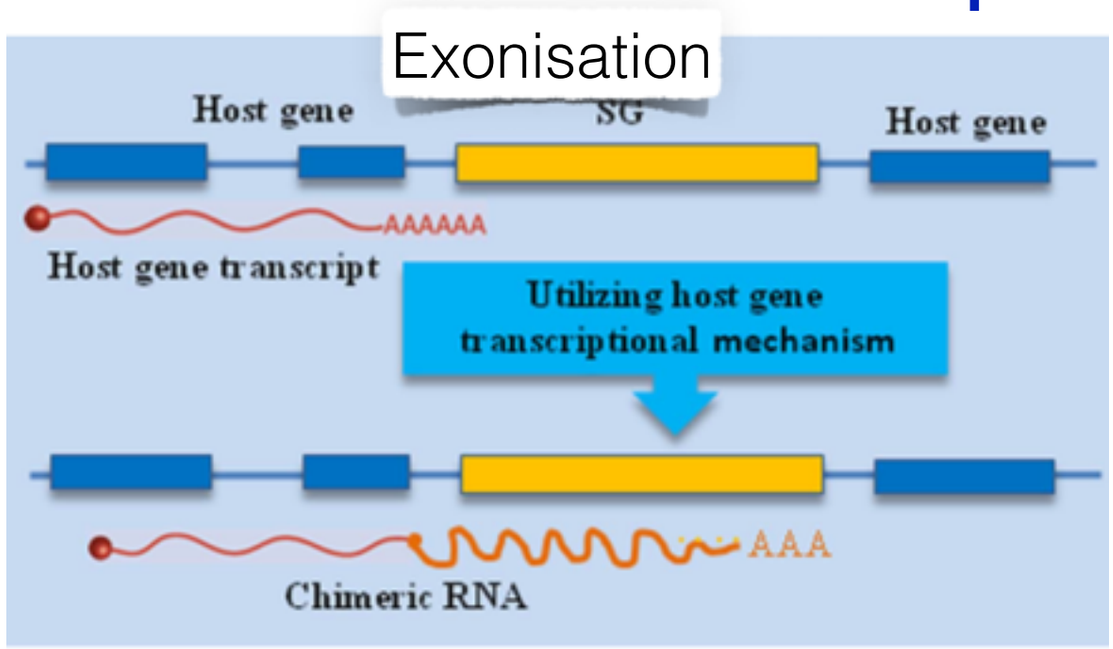
Why do somatic pseudogenes matter?

- if inserted in introns, UTRs, or exons the pseudogene is most likely to be expressed

- the somatic pseudogene insertion can stop the expression of a target gene at the insertion site

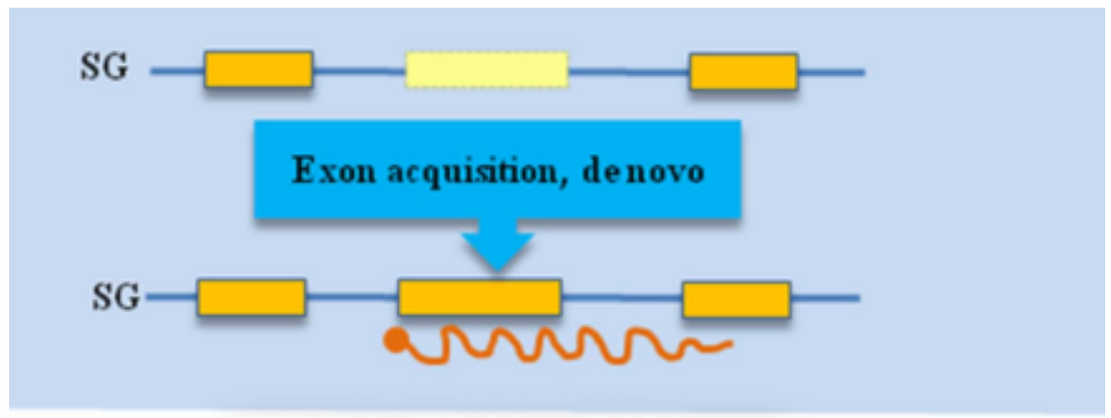
- no evidence was found if the pseudogene was expressed when inserted in intergeneric region

Functions of pseudogene DNA



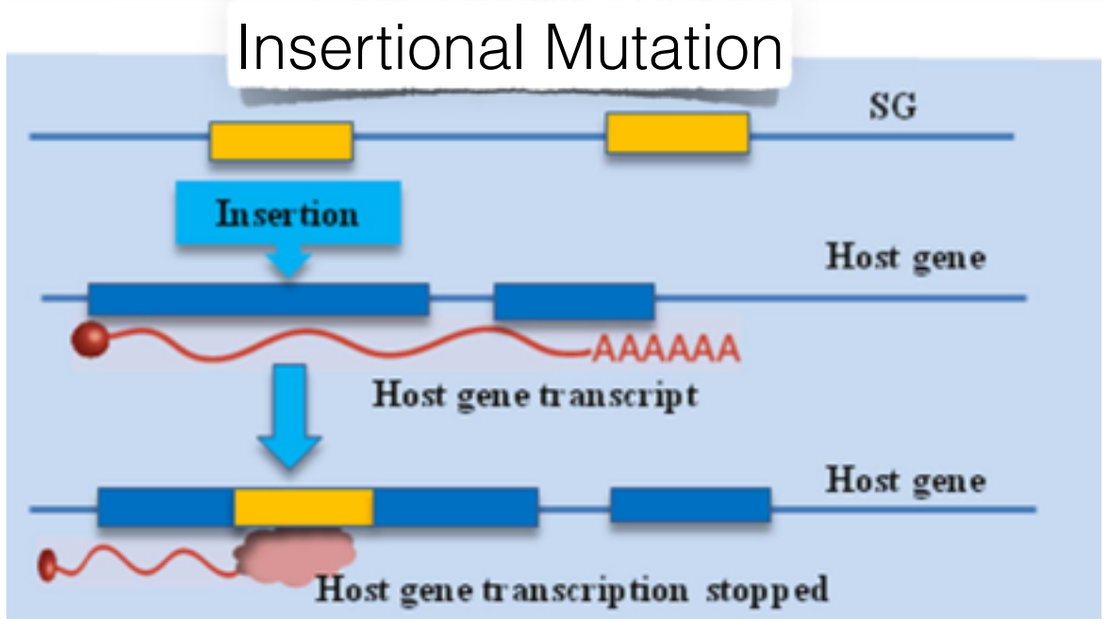
The pseudogene use the transcriptional mechanism of the host gene

e.g: somatically acquired pseudogenes during cancer development



The pseudogene can obtain exons de novo.

e.g: KLK3 pseudogene has 2 extra exons compared to the parent. Both are regulated by androgen but show different expression patterns

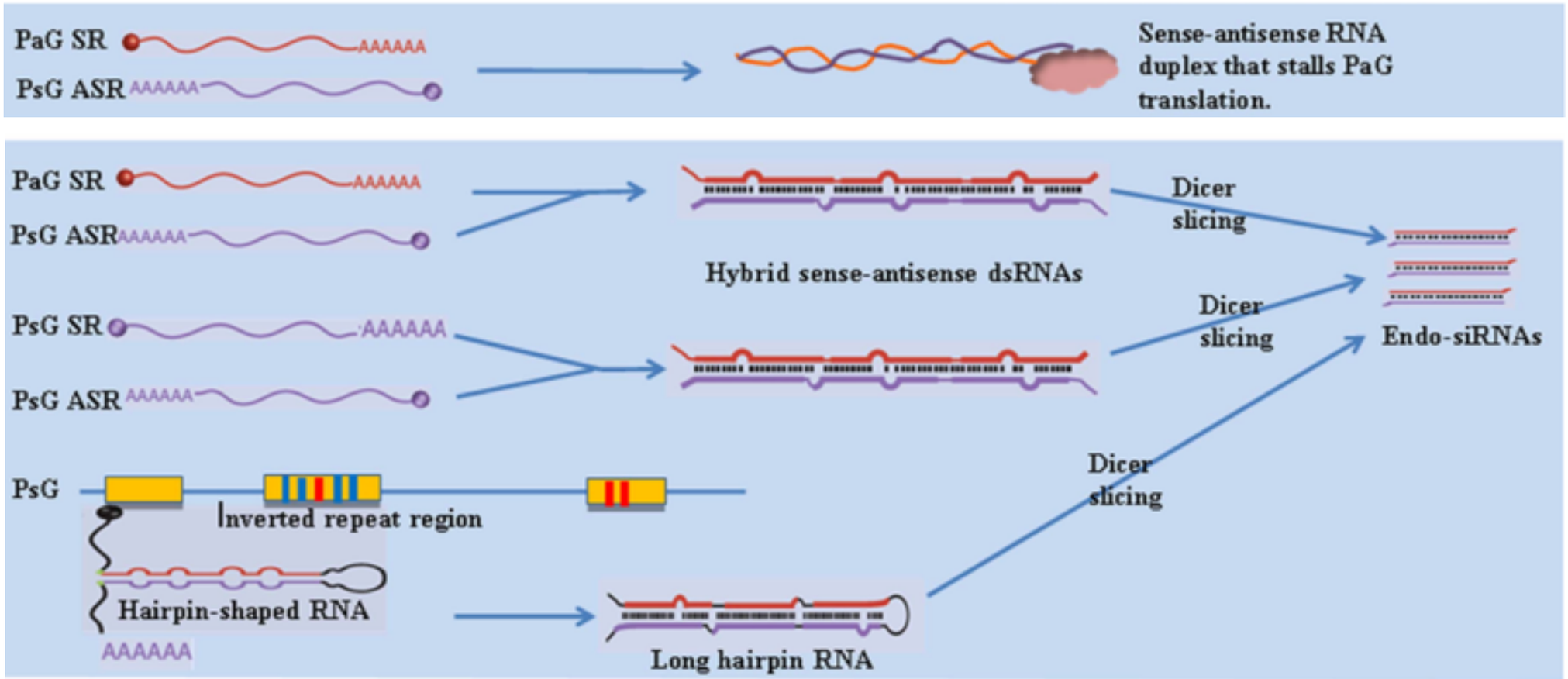


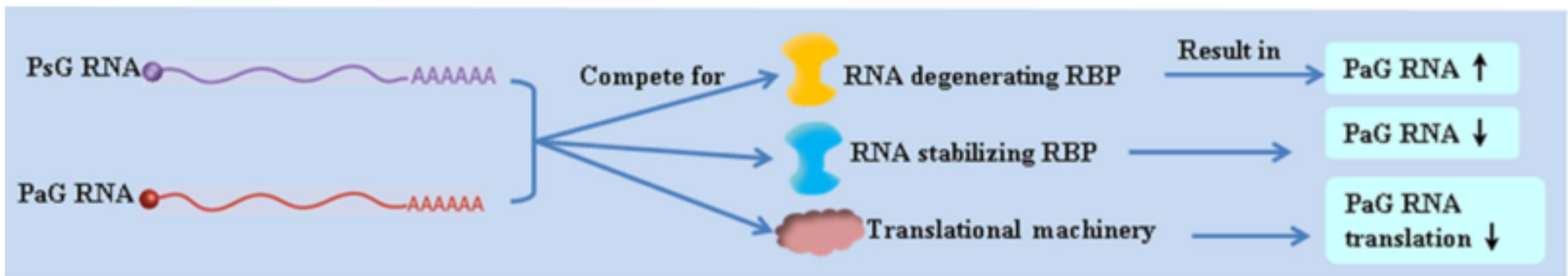
The pseudogene inserts itself into the promoter/exons of the host gene and stops its expression

Pseudogenes **regulate & interfere** with the expression & activity of functional protein coding genes

- Through their pseudogene DNA sequence
- Through their RNA product
- Through their pseudo-protein/polypeptide product

Functions of pseudogenic RNA: to regulate the expression of parent genes

- 

- 

Cell

Resource

Expressed Pseudogenes in the Transcriptional Landscape of Human Cancers

Shanker Kalyana-Sundaram,[1,2,6,7] Chandan Kumar-Sinha,[1,2,7] Sunita Shankar,[1,2] Dan R. Robinson,[1,2] Yi-Mi Wu,[1,2] Xuhong Cao,[1,3] Irfan A. Asangani,[1,2] Vishal Kothari,[1] John R. Prensner,[1,2] Robert J. Lonigro,[1,2] Matthew K. Iyer,[1] Terrence Barrette,[1,2] Achiraman Shanmugam,[6] Saravana M. Dhanasekaran,[1,2] Nallasivam Palanisamy,[1,2] and Arul M. Chinnaiyan[1,2,3,4,5,*]

293 samples in 13 cancer and normal tissue types

2082 pseudogene transcripts:

* 154 highly tissue specific
* 218 expressed only in cancer

TCGA RNA-seq data

9925 expressed pseudogenes in 2808 cancer samples

547 breast cancer pseudogenes

ARTICLE

Received 12 Feb 2014 | Accepted 13 Jun 2014 | Published 7 Jul 2014

DOI: 10.1038/ncomms4963

The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes

Leng Han[1,*], Yuan Yuan[1,2,*], Siyuan Zheng[1], Yang Yang[1,3], Jun Li[1], Mary E. Edgerton[4], Lixia Diao[1], Yanxun Xu[1], Roeland G.W. Verhaak[1] & Han Liang[1,2]

RESEARCH ARTICLE Open Access

Pseudogenes transcribed in breast invasive carcinoma show subtype-specific expression and ceRNA potential

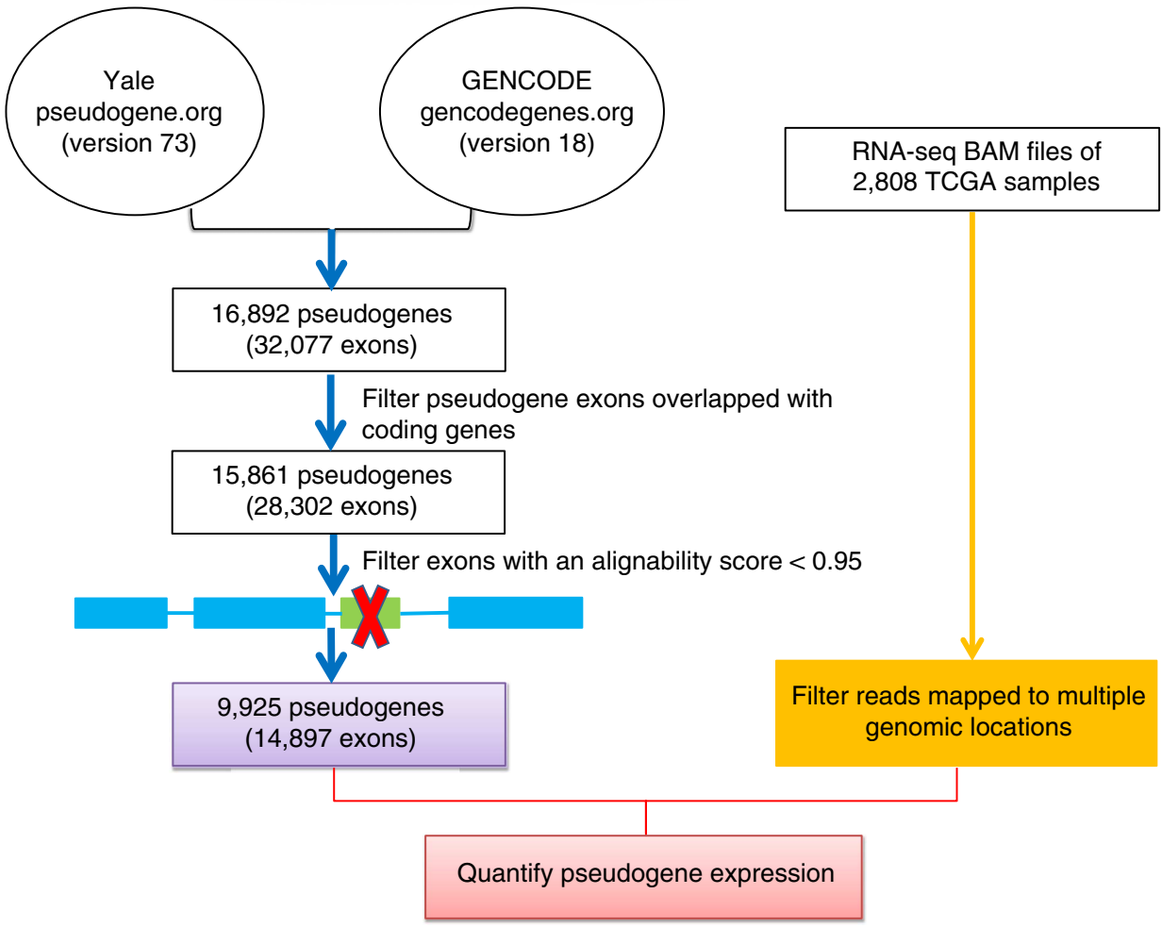
Joshua D Welch[1,2], Jeanette Baran-Gale[1,3], Charles M Perou[1,3,4], Praveen Sethupathy[1,3,4*] and Jan F Prins[1,2*]

TCGA RNA-seq data

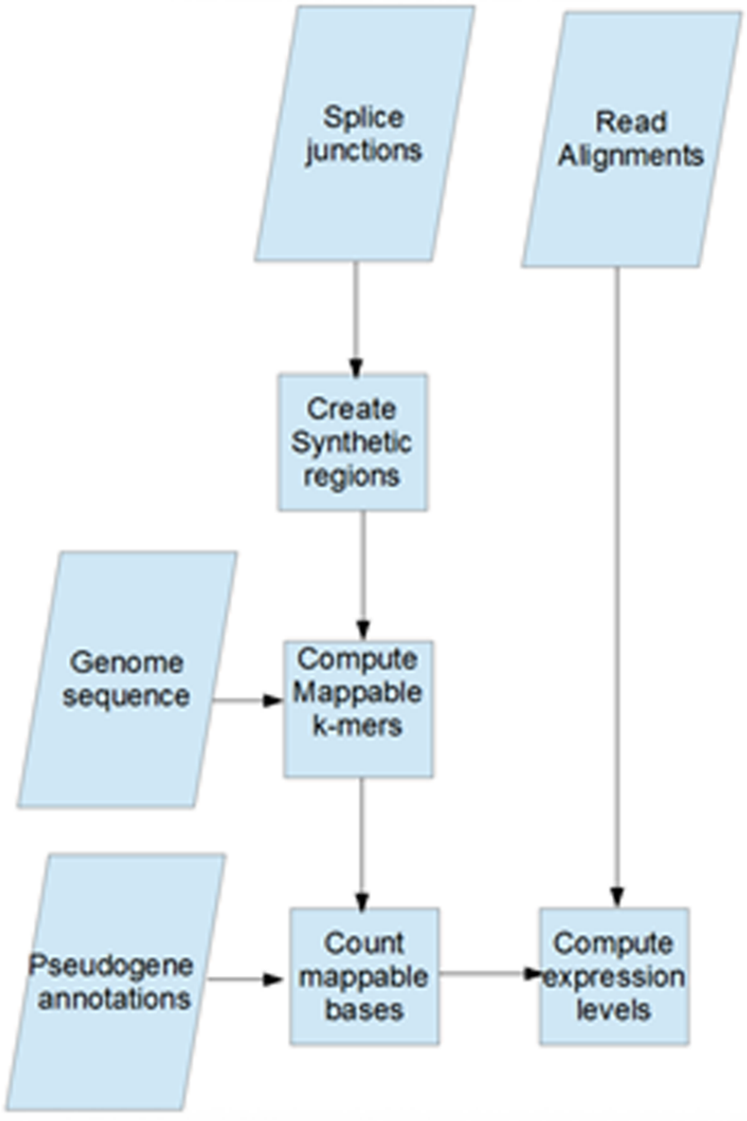
psiDR data

Pseudogene expression pipelines

Pan-Cancer

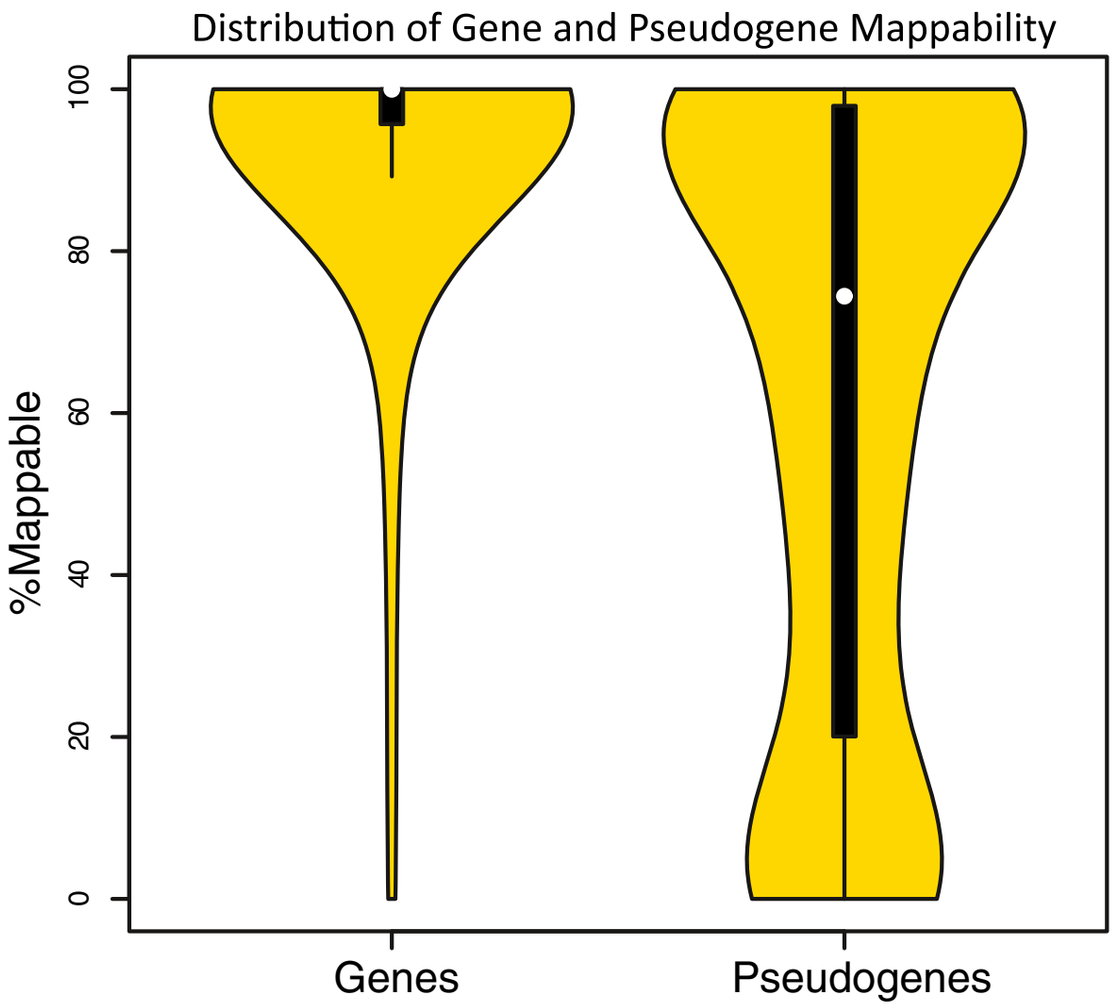


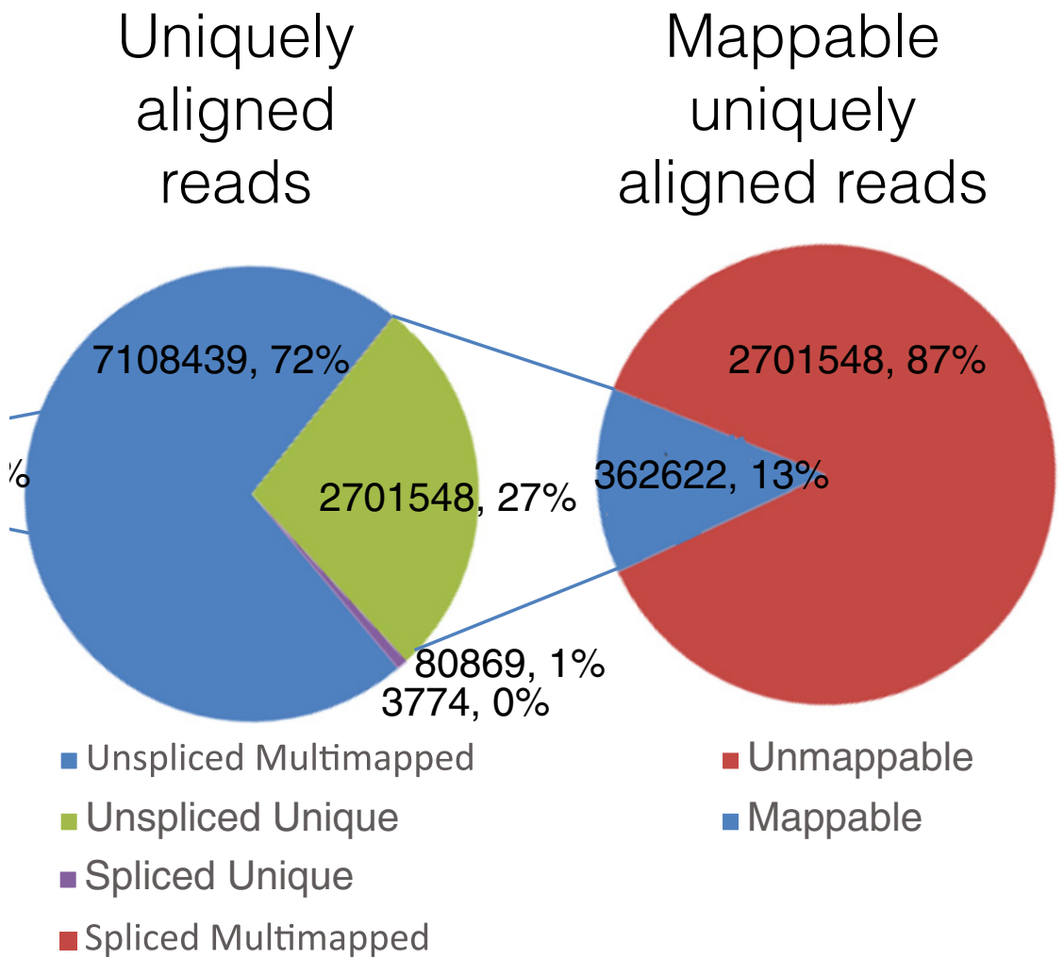
Breast Cancer



- pseudogene expression level
- coverage depth of RNA-seq
- mismatch distribution patterns between pseudogene and parental gene

Quality control on using genome mappability maps to quantify pseudogene expression

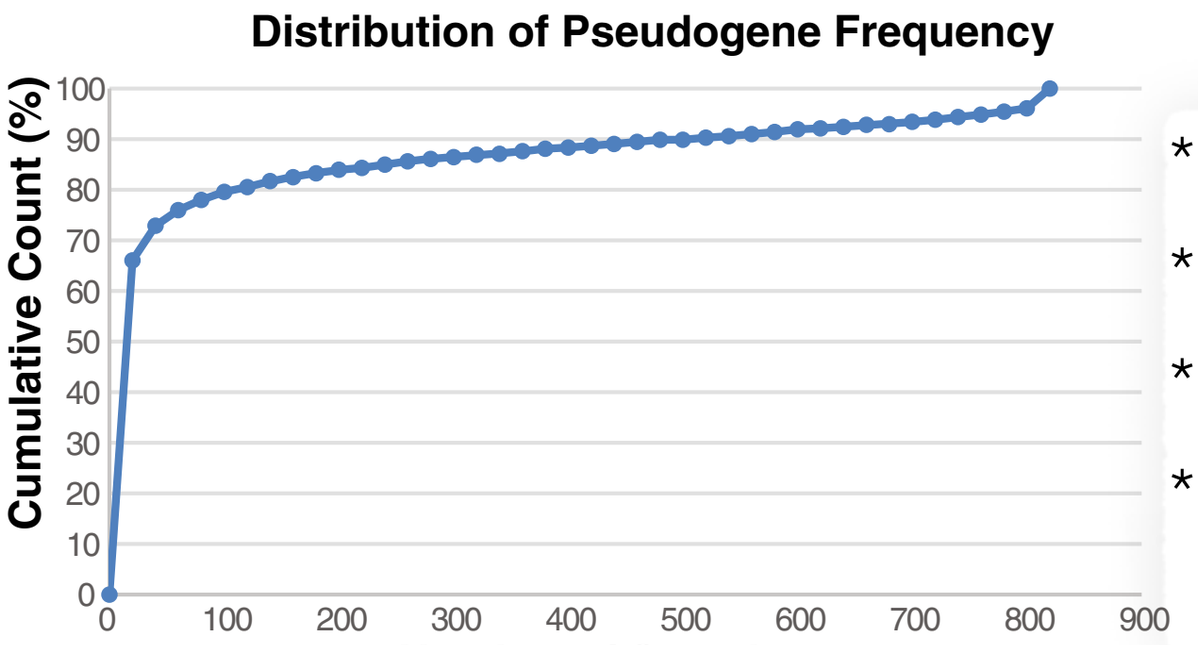




only 80% of pseudogene sequence length is usually mappable

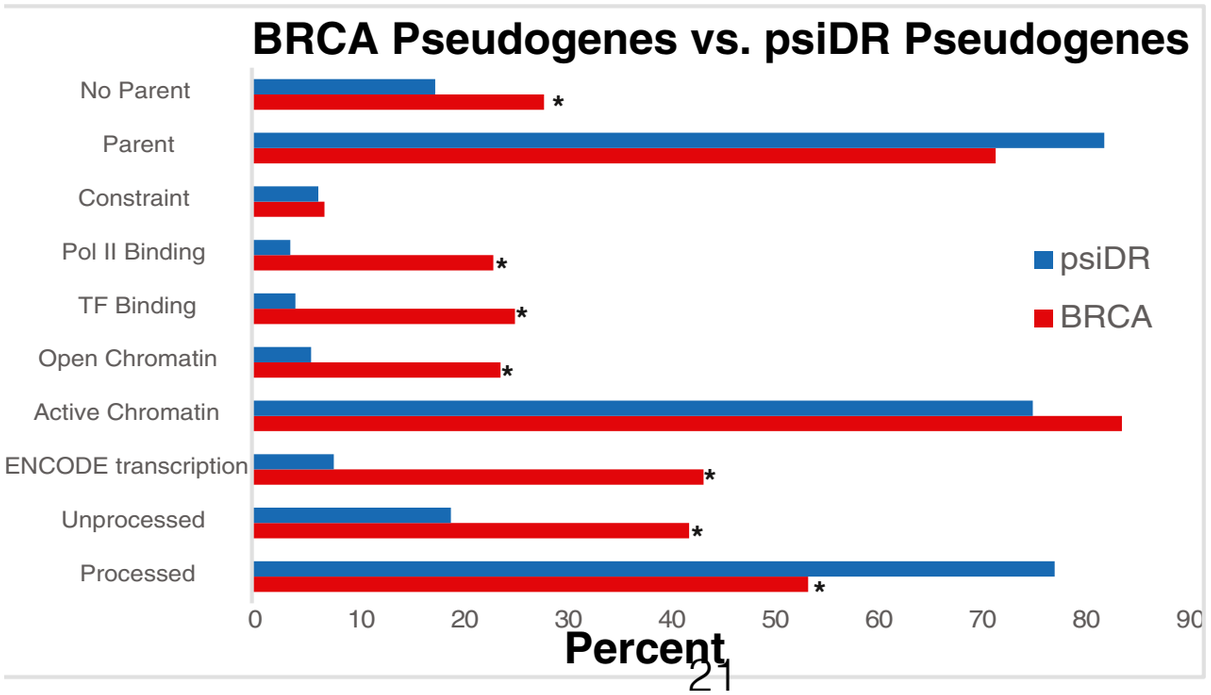
90% pseudogene are mappable while 10% are not

Transcribed pseudogenes

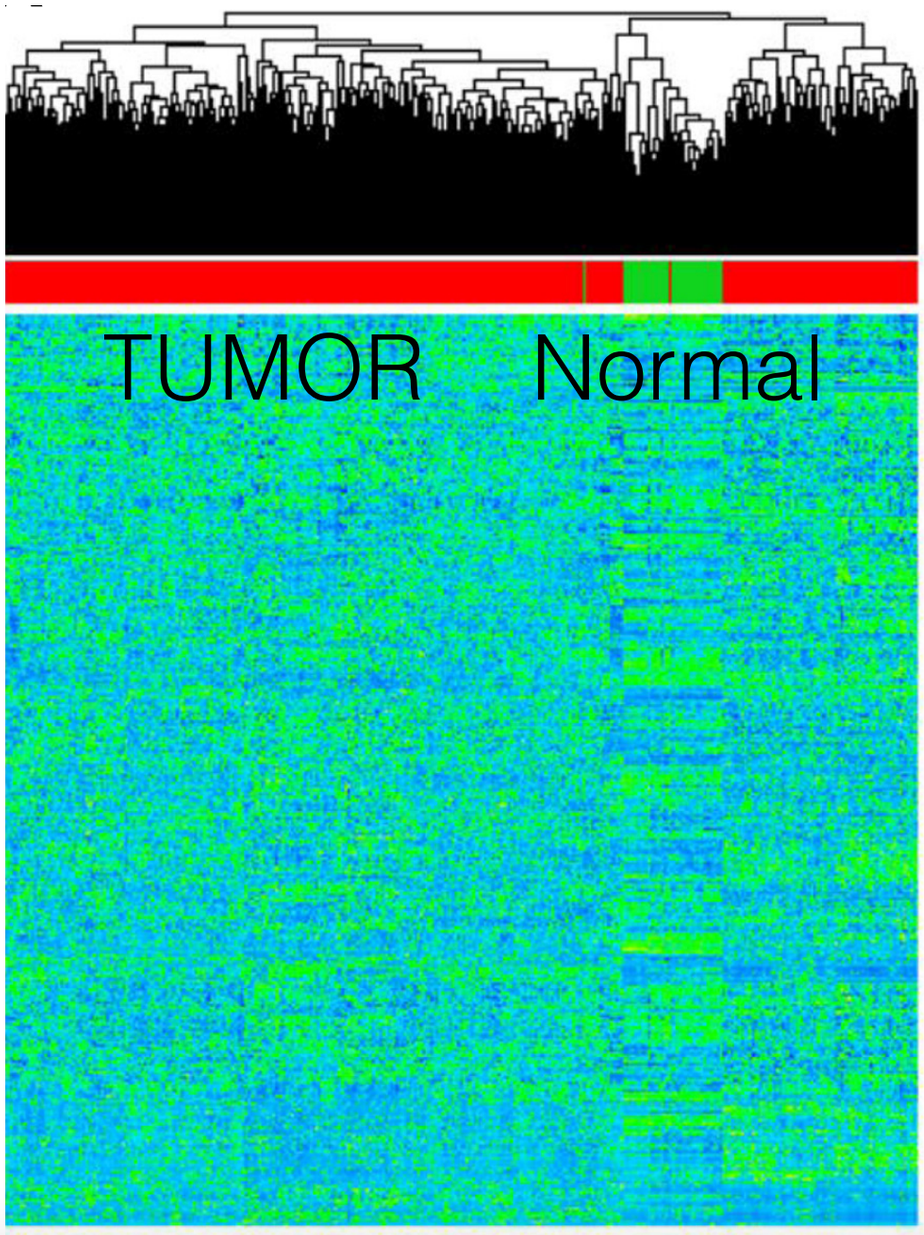


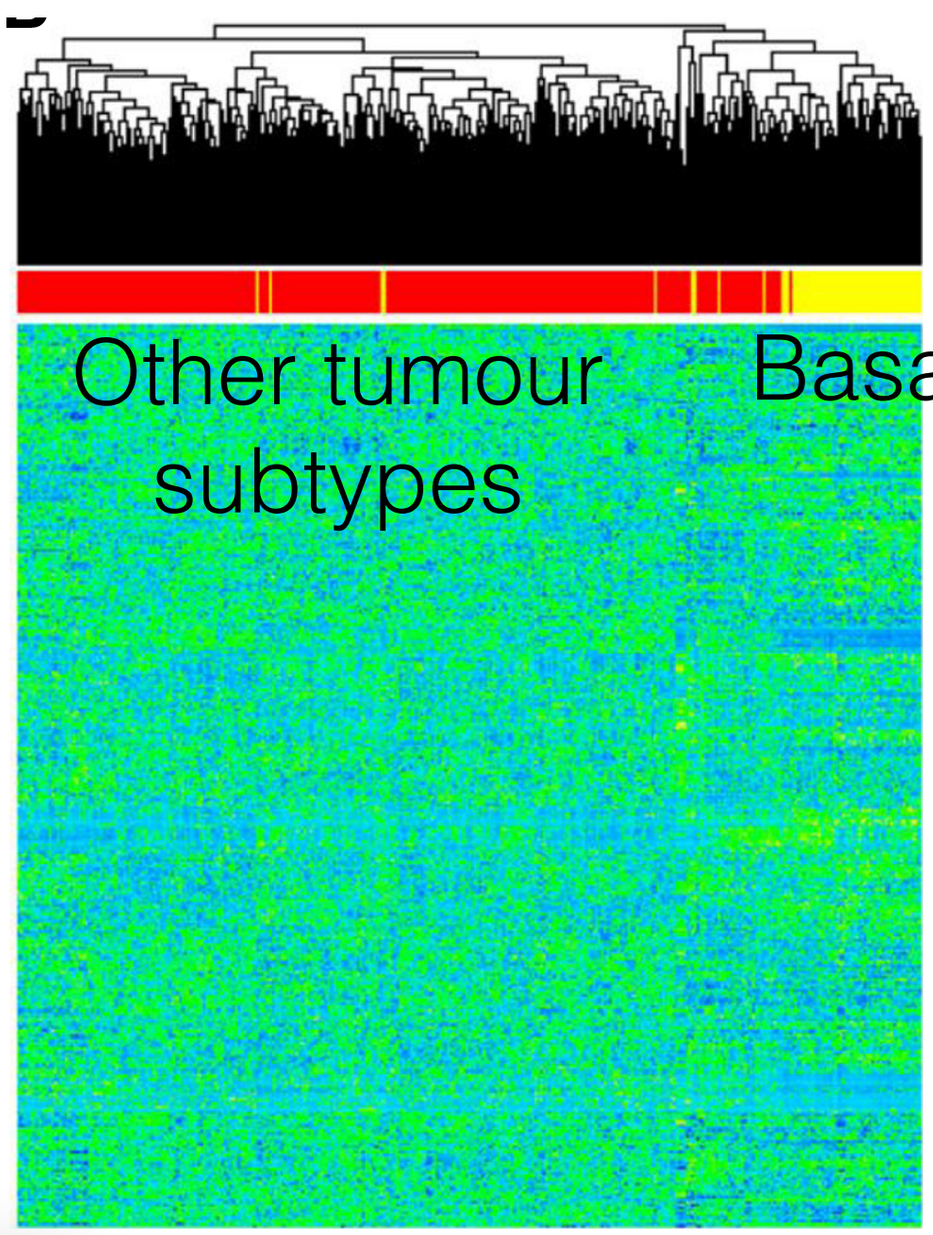
* Gencode 17
* 2012 transcribed pseudogenes
* 94 transcribed in 90% of samples
* **440** transcribed in at least 10% samples

—> 287 in psiDR



Using pseudogene transcription to differentiate between cancer subtypes





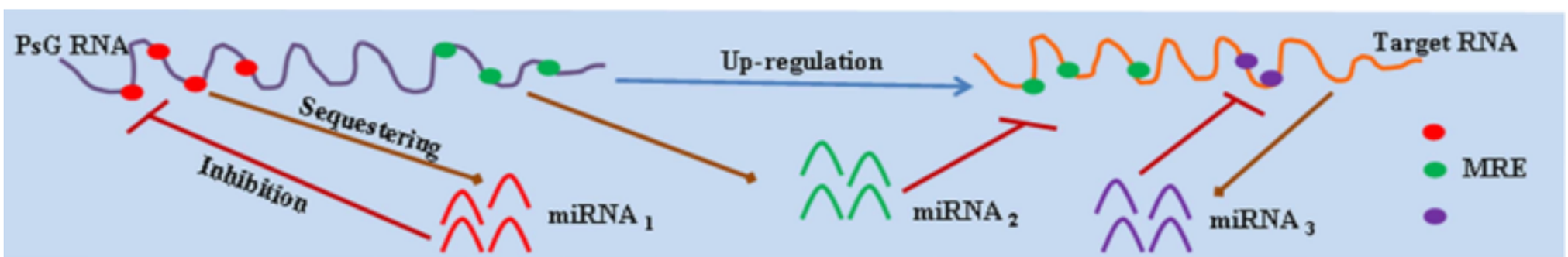
2 different tissues!
differentiation might be due to tissue specificity

e.g. CASP4

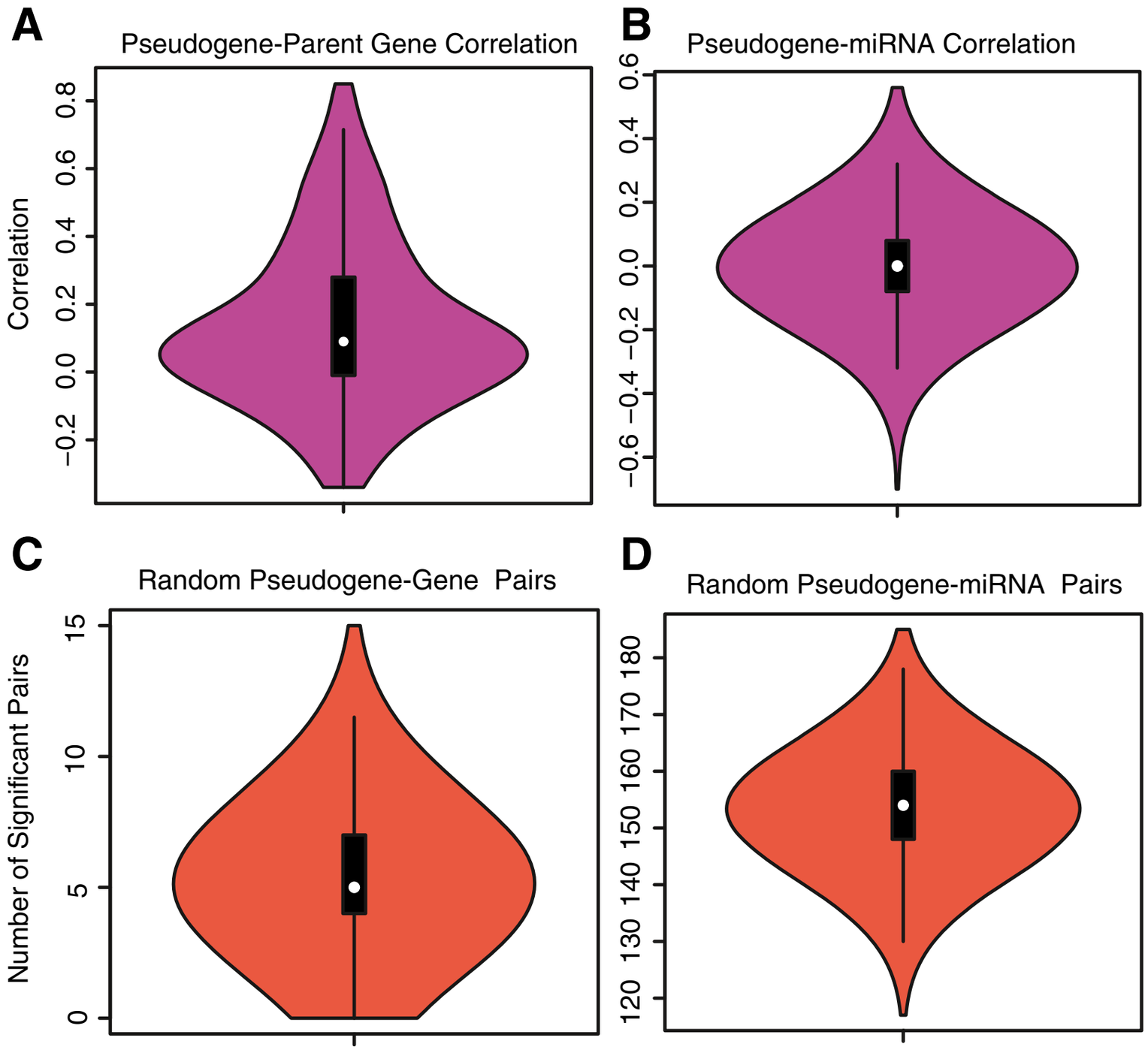

chr11
p15.4 p15.2 p14.3 p13 p12 p11.2 p11.11 q12.2 q13.2 q13.4 q14.1 q14.2 q21 q22.2 q23.1 q23.3 q24.2 q25
18 kb
4,772 kb 104,774 kb 104,776 kb 104,778 kb 104,780 kb 104,782 kb 104,784 kb 104,786 kb 104,788 kb 104,790 k
Transcriptome Mappability
[0 - 1.00]
Read Coverage
[0 - 405]
Gencode Genes (v17)
ENSG00000235505.3
ENST00000531091.1
ENST00000528437.1
ENST00000532510.1
ENST00000534659.1
ENST00000527617.1
ENST00000527637.1
ENST00000531371.1
ENST00000532441.1
ENST00000527425.1
Expression level (RPKUM)
1.6 2 2.4
Basal
Her2
LumA
LumB

Functions of pseudogenic RNA

- competing endogeneous RNA
- competition for the common pool of miRNAs thus regulation the parent's expression as competitive endogenous RNA (ceRNA) —> in particular in cancer:
  - e.g. PTEN & KRAS



Pseudogene-parent gene and pseudogene-miRNA pairwise correlations



Transcribed pseudogenes with ceRNA potential

- Integrate miRNA target prediction with pseudogene, gene, and miRNA expression levels
- 17 examples of pseudogenes with strong ceRNA potential
- **GBP1** pseudogene —> its parent gene is mediator of the anti-proliferative effect of inflammatory cytokines in endothelial cells
  - significant anti correlation with hsa-mir-199a-2-5p and significant expression correlation with parent

- **SUZ12P1** —> its parent gene is a polycomb group protein and part of the PRC2/EED-EZH2 complex; an important epigenetic regulator that performs histone methylation

Summary

- Mappability maps are useful in obtaining reliable results when quantifying pseudogene transcription, however the method will always be dependent of the data quality

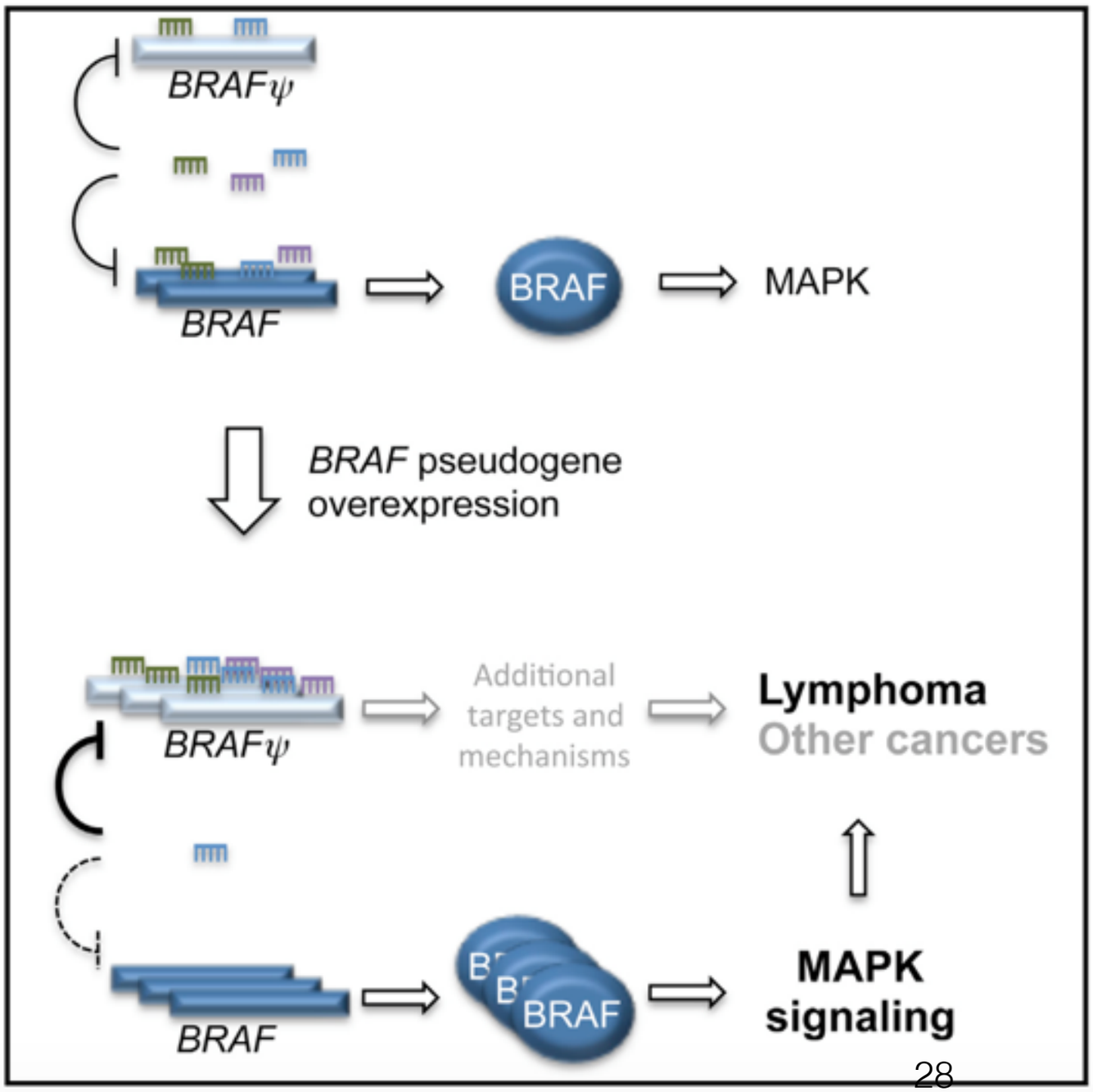
- Pseudogene transcription is a good marker for distinguishing between cancer types

- Integrating pseudogene, gene and miRNA expression highlights ceRNA potential as pseudoRNA function in cancer

Cell

The BRAF Pseudogene Functions as a Competitive Endogenous RNA and Induces Lymphoma In Vivo

Graphical Abstract



Authors

Florian A. Karreth, Markus Reschke, ..., Roberto Chiarle, Pier Paolo Pandolfi

Correspondence

ppandolf@bidmc.harvard.edu

In Brief

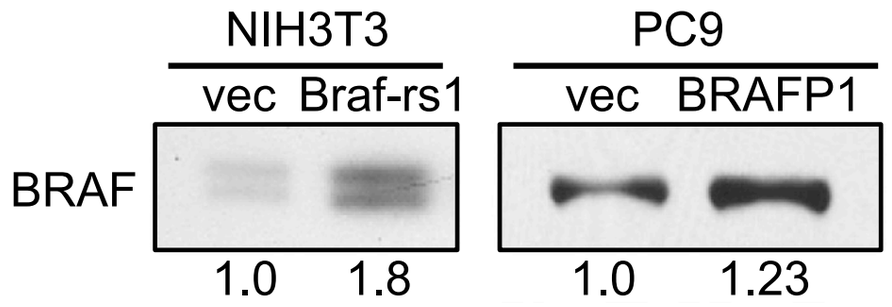
The in vivo evidence for the regulatory activity of pseudogenes has been lacking, and their role in disease progression has been correlative. This study now shows that transgenic expression of the BRAF pseudogene induces a malignancy in mice resembling human diffuse large B cell lymphoma, establishing its oncogenic function.

BRAF pseudogene

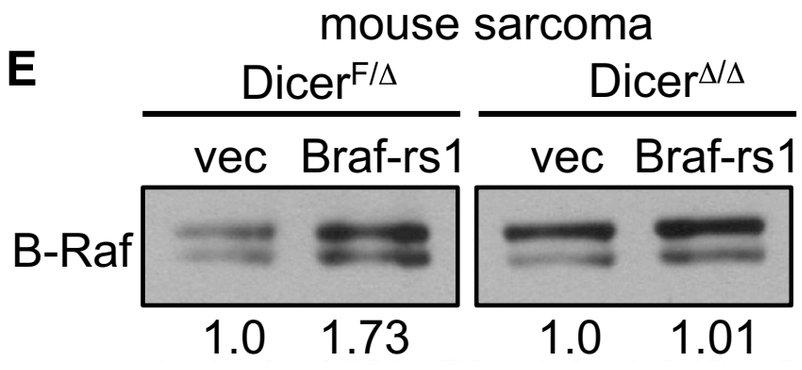
- overexpressed in various tumour types
- regulates the expression parent gene through sequestration of shared miRNAs and BRAFP1
- mediated elevation of BRAF may promote MAPK signalling and tumorgenesis.
- murine Braf-s1 and B-Raf are targeted by **54 and 114** miRNA families, **53** of which they have in **common**
- human BRAFP1 and BRAF are targeted by **60 and 48** miRNA families, **40** of which are **common** to both

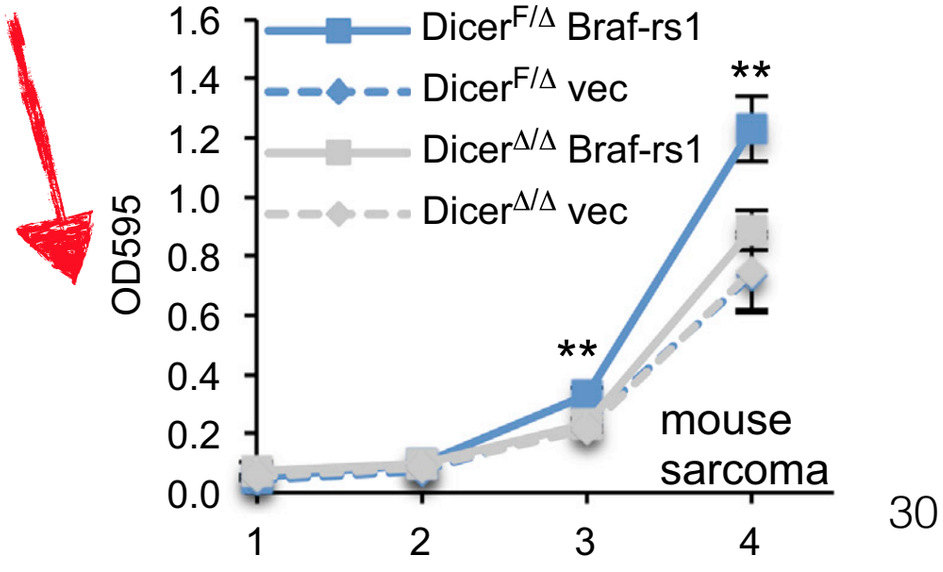
Regulatory role for BRAF pseudogene

- expression of BRAF pseudogene in human and mice cancer cells, elevate the BRAF protein phosphorylation



- ectopic expression of BRAF pseudogene increased the expression of its parent gene and elevated the proliferation of DICER1-proficient mouse cells

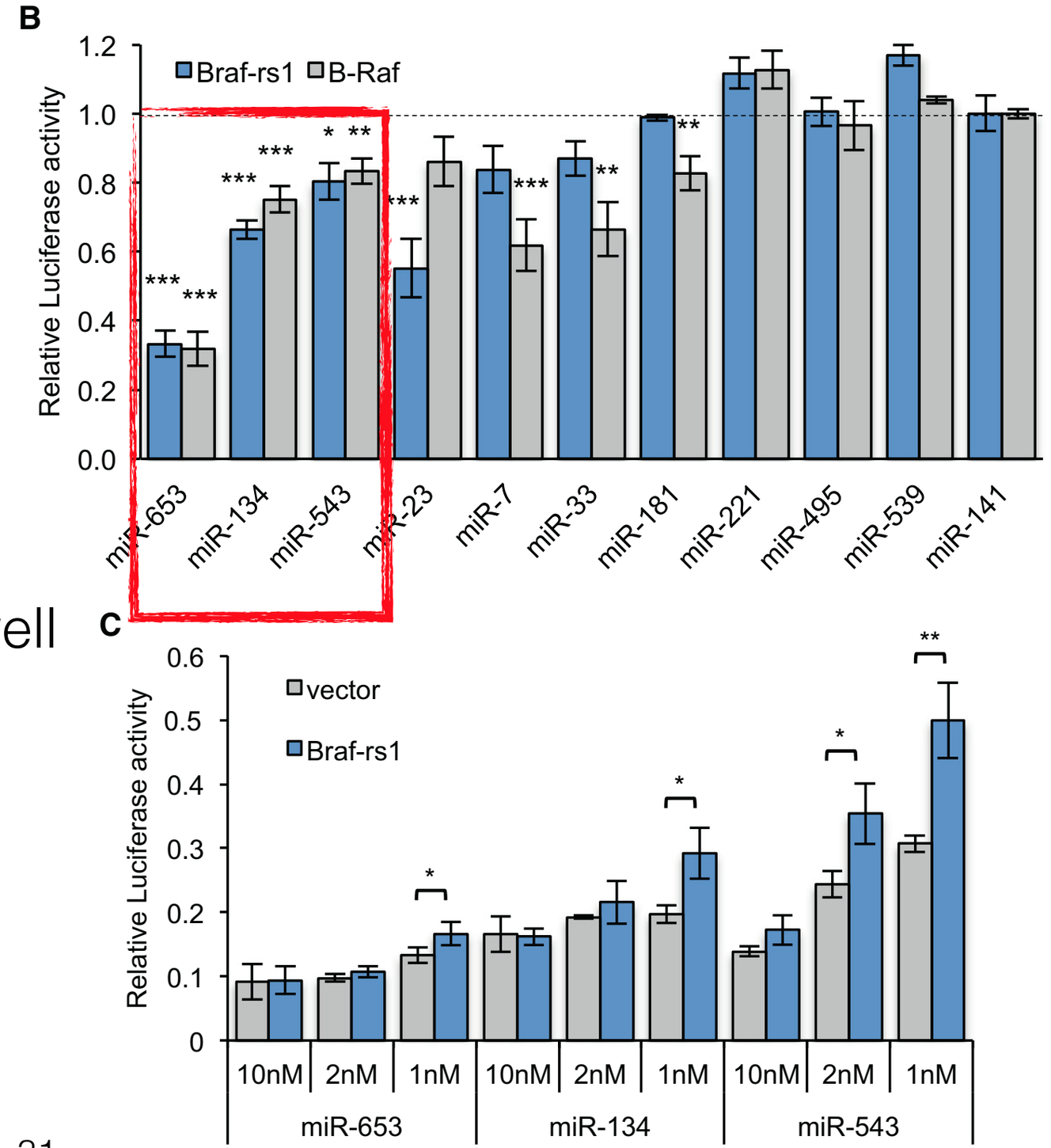




BRAF pseudogene induced effects are depended on BRAF and Dicer1

BRAF pseudogene as ceRNA

- Luciferase reported assay
- 10 mouse miRNA significantly repressed the BRAF gene luciferase reporters in both mouse and human
- variation in concentration as well as the potency of the miRNA affect the ceRNA crosstalk
  —> most effective at low concentrations

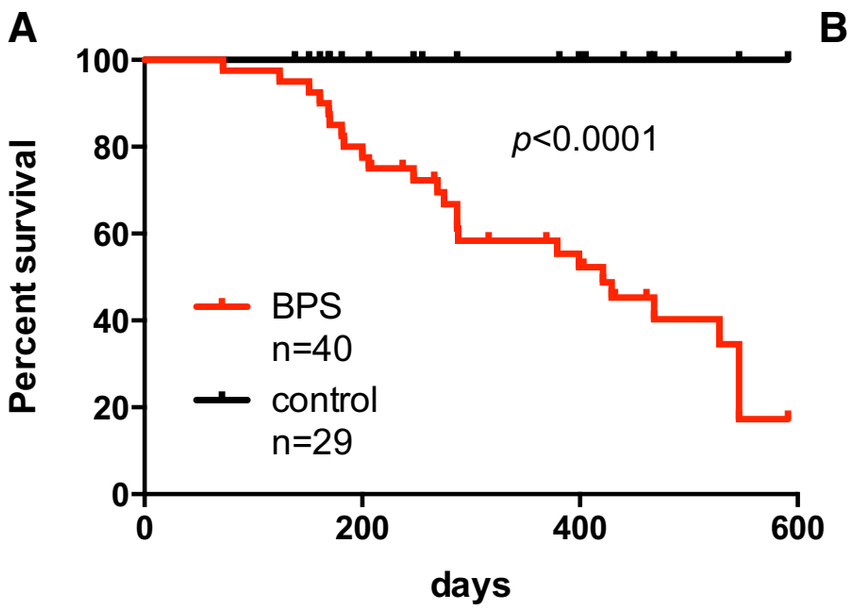


What happens in vivo?

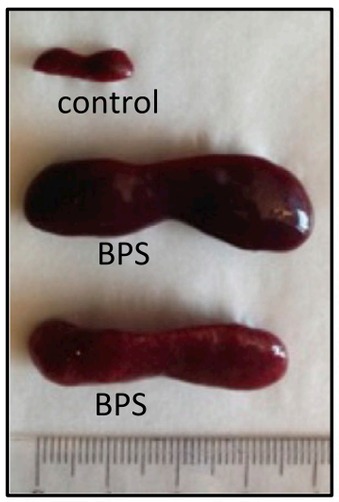
- generation of transgenic allele mice containing the BRAF-rs1 (mouse pseudogene) under the control of a doxyciline tet-response element

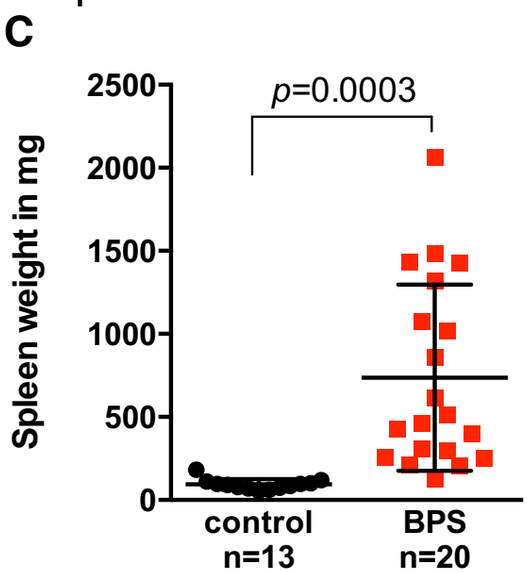
- expression of BRAF-rs1 resulted in a lymphoid malignancy

Survival rate

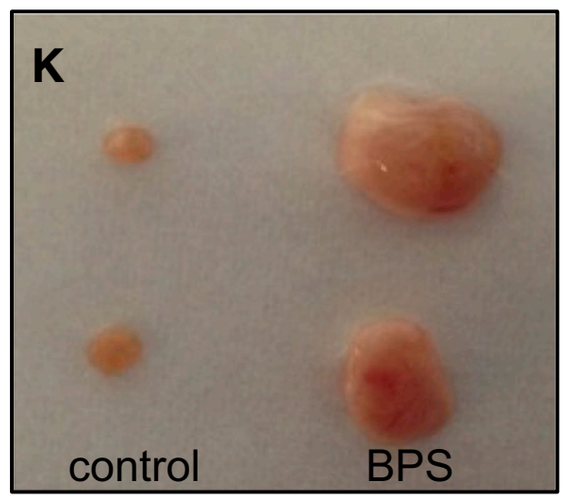


Size and weight of mouse spleen

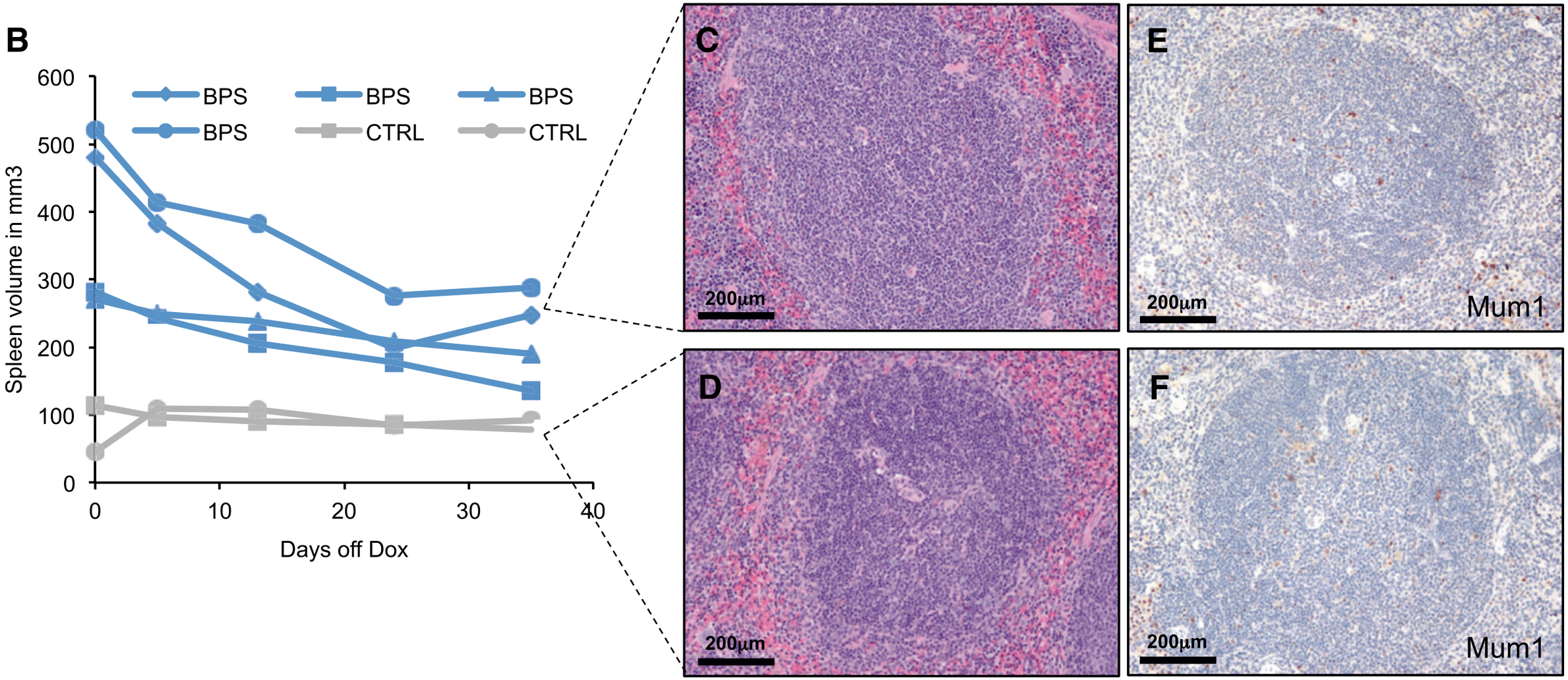




Lymph node size



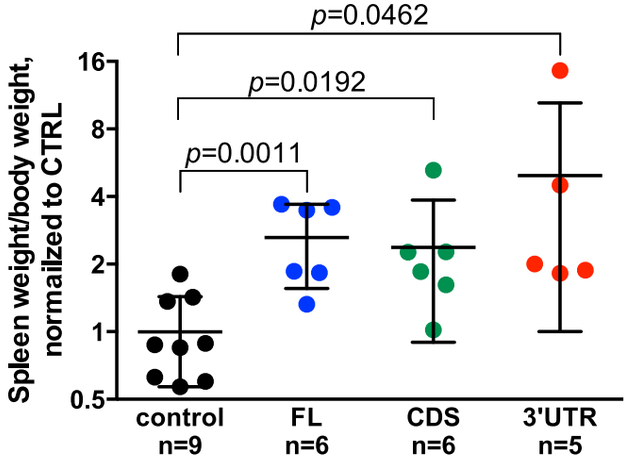
BRAF pseudogene is required for the development and maintenance of malignancy



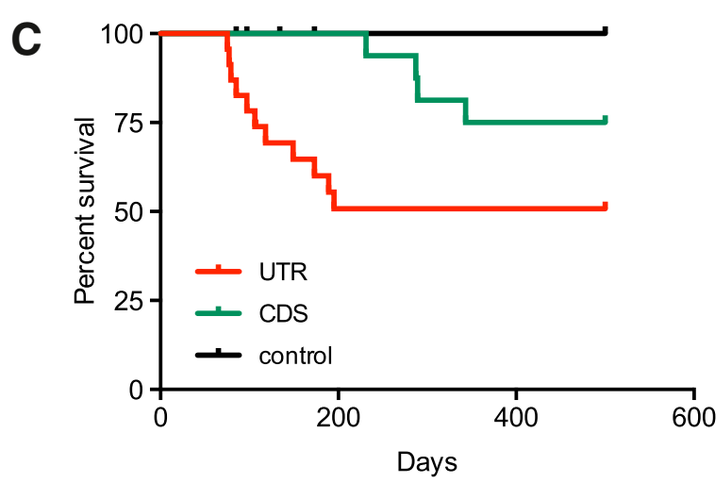
Which part of the pseudogene is actual culprit? CDS, UTR or the full length?

- Hypothesis: since BRAF-rs1 is able to decoy miRNAs shorter fragments may be able to crosstalk with the parent gene using the shared miRNA pool.

Spleen size



Survival rate



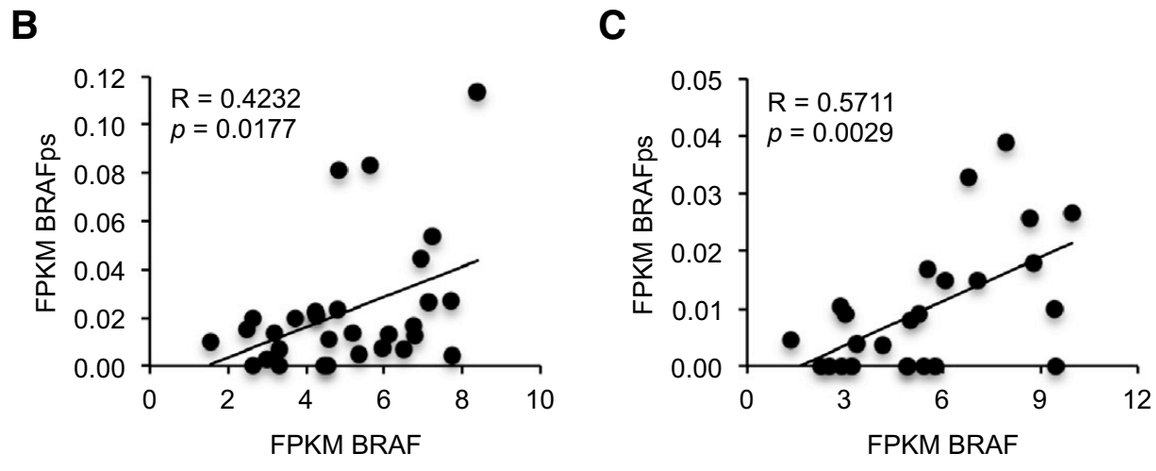
D

	Penetrance	Median survival	Onset (range)
FL	83%	421	72 (72-546)
3'UTR	49%	undefined	77 (77-195)
CDS	25%	undefined	231 (231-343)

- both CDS and UTR constructs have shown phenotypic changes
- only the BRAF-3'UTR had a significant effect on the parent gene expression and proliferation
- no significant effect from BRAF-CDS

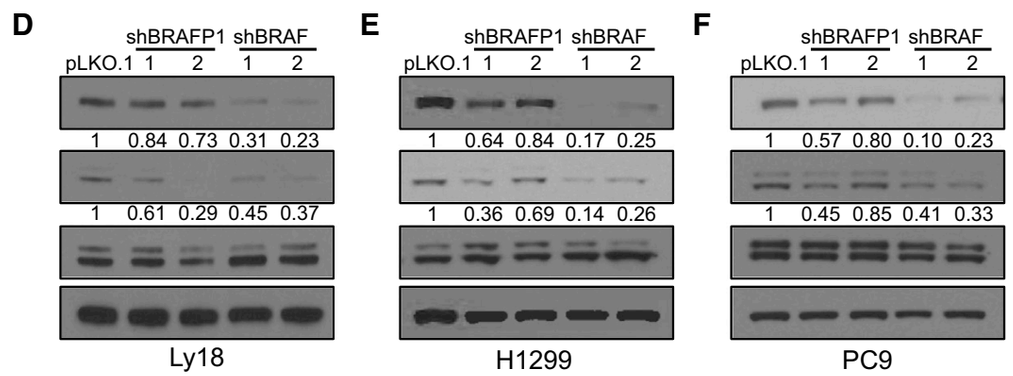
BRAFP1 in human cancer

- experiments in human cell lines indicate the BRAFP1 may operate as ceRNA to regulate its parent expression

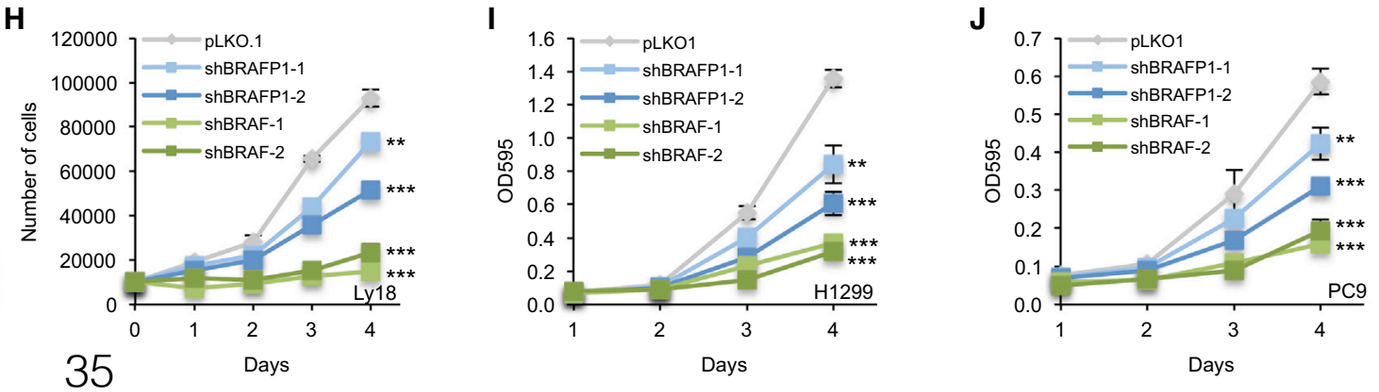


- knockdown of BRAFP1 in cancer cells reduced the expression of BRAF

- silencing BRAFP1 reduced the levels of mRNA levels in some but NOT all cell lines —> cell line specificity



oncogenic properties for BRAFP1



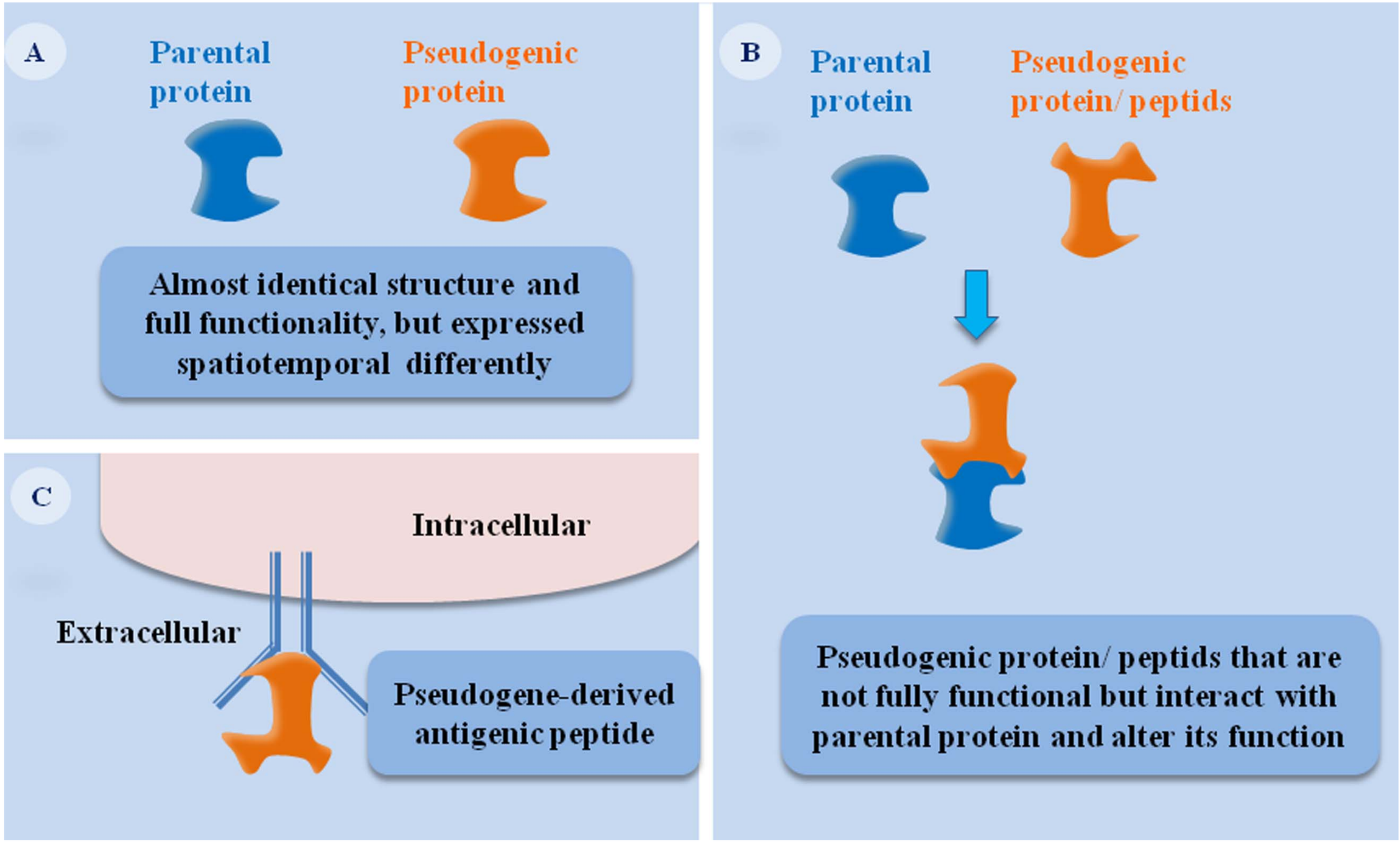
Are BRAF pseudogenes the same in mouse and human?

- NO
- evolved independently
- are not in syntenic regions
- however, both 3'UTR show high sequence similarity to parents
- both mediate the expression of their parent gene by acting as miRNA sponges

Pseudogenes **regulate & interfere** with the expression & activity of functional protein coding genes

- Through their pseudogene DNA sequence
- Through their RNA product
- Through their pseudo-protein/polypeptide product

Functions of pseudo-protein



Cancer pseudogenes…

- … can be used as biomarkers to differentiate between different types of cancer & also can be used as a prognostic

- … account for (possibly) the key difference between aggressive and mild forms of cancer

- … can regulate the expression of their parent genes but also of neighbouring protein coding genes

- “The term “pseudo” implies sequence variance compared to the parental gene, not indicating pseudo function […] many pseudogenes perform real and indispensable functions in physical and pathological processes.”

Literature

Main

- Xiao-Jie et al. J Med Genet (2015) 52, 17-24
- Welch et al. BMC Genomics (2015) 16:113
- Cooke et al. Nat Comm (2014) 5: 3644
- Karreth et al. Cell (2015) 161, 319-332

Extra

- Kalyana-Sundaram et al. Cell (2012) 149, 1662-1634
- Han et al. Nat Comm (2014) 5:3963