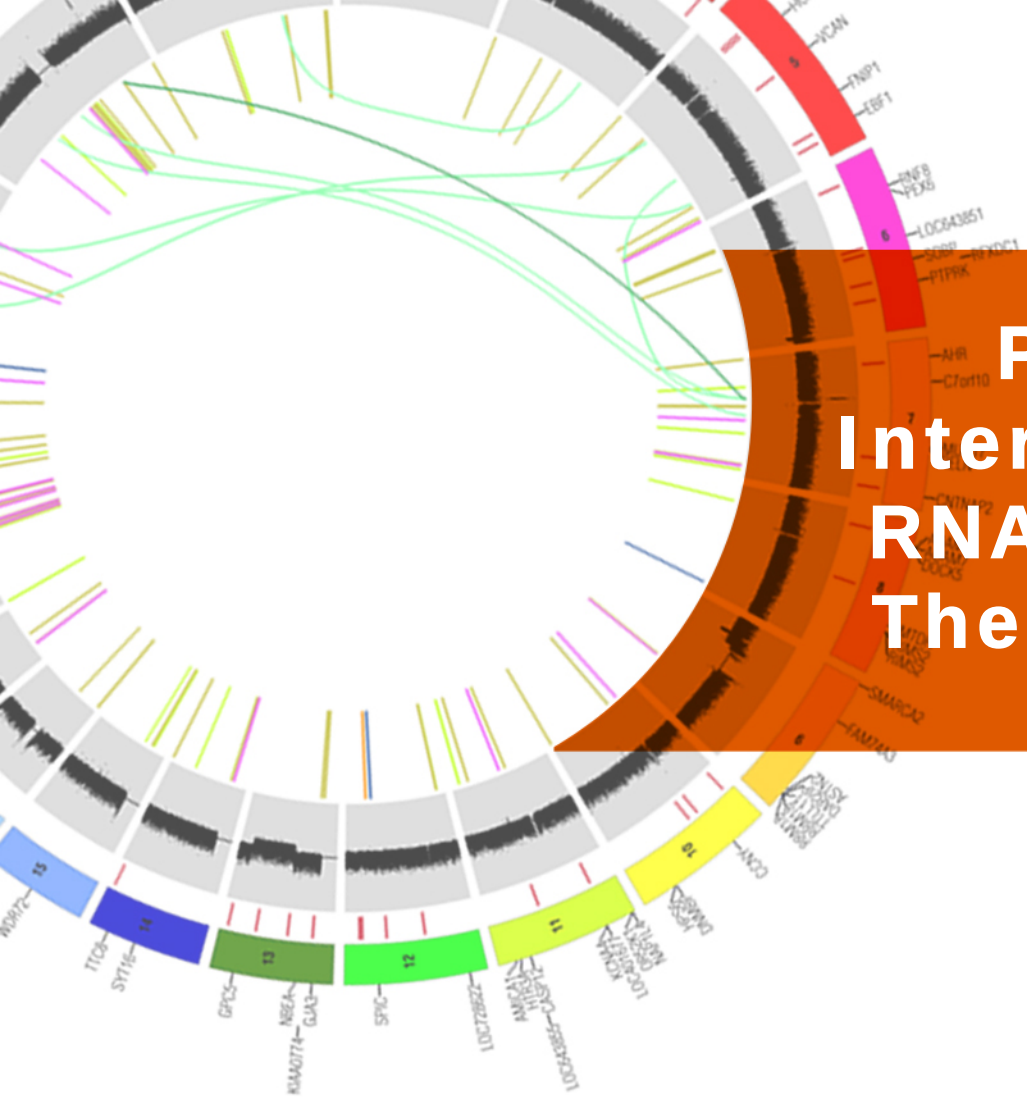


The Cancer Genome Atlas

Profiling Long Intergenic NonCoding RNA Interactions In The Cancer Genome

Samir B. Amin

Dr. Lynda Chin Lab

May 12, 2014

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Baylor College of Medicine

Primary Focus

- To catalogue expression levels long intergenic non-coding RNA (lincRNA) in TCGA cancer types.
  - To extend initial profiling efforts by MSKCC group and MD Anderson group.[1,2]
- To facilitate integrative analysis in understanding emerging gene regulatory role of lincrnas in progression of cancer.
  - Whether or not such pervasively transcribed non-coding RNAs are of regulatory importance or merely a transcriptional noise.

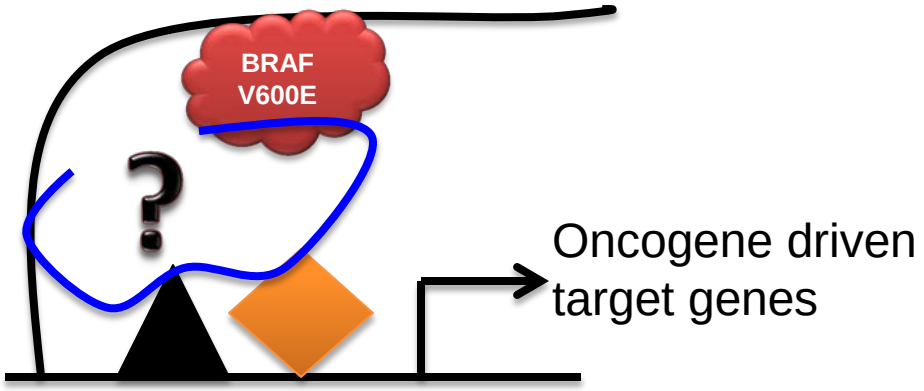
1. Akrami R. PLoS One 2013
2. Han L, Yuan L. Nat Commun. 2014

lncRNA…

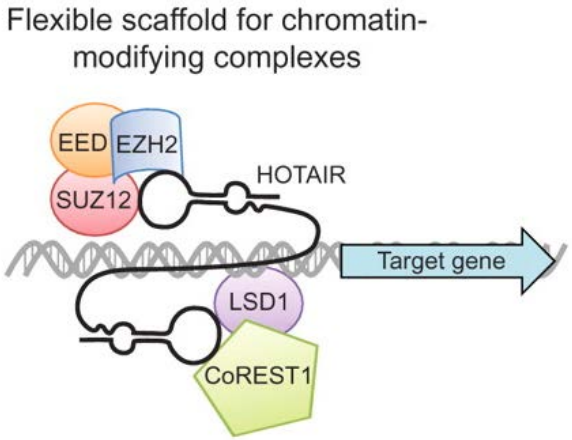
- > 200 bp & no coding potential.
- Poly-A tail present for majority of lncrnas.
- Epigenetics marks consistent with that of a transcribed gene, i.e., H3K4me3 at promoter, H3K36me3 throughout gene body.
- ≈ 10,000 putative transcripts, 3,000 show highly conserved ‘patches’ within coding region.[#]

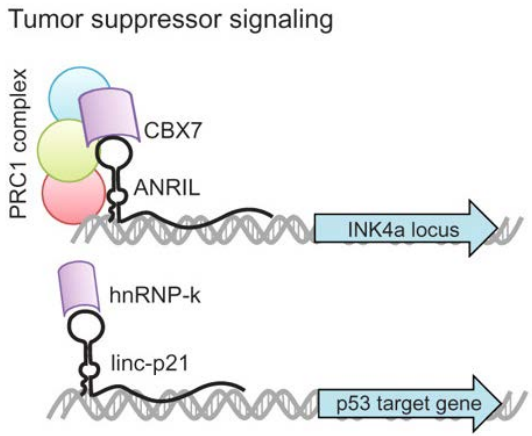
Prensner J, Chinnaiyan A. Cancer Discovery 1 (2011)
#: Guttman M. Nature 458 (2009), Ponjavic J. Genome Res. 17 (2007)

lncRNA interactions can facilitate oncogene driven downstream gene regulation.



lncRNA can act as a **molecular scaffold** to mediate RNA-Protein interactions at regulatory regions of oncogene targeted genes.



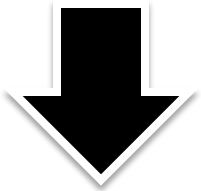


Prensner J, Chinnaiyan A. Cancer Discovery 1:391 (2011)

How does lncRNA interact?

- Is there a sequence-specific motif in lncRNA structure to form a functional scaffold?

  i.e., Transposable elements, microrna recognition elements (MRE), G4-DNA.



- Is there an enrichment of lncRNA harboring such motif to drive downstream gene regulation of a driver gene?

Outline – Analyses

- To quantify lincrna expression in TCGA tumor types.
- To correlate lincrna expression with gene expression, mutation and methylation subtypes.
- To identify enrichment of sequence-specific lincRNA-DNA interactions at regulatory domains of cancer genes.

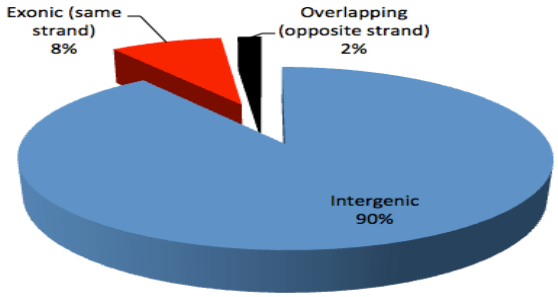
lncrna quantification

lncRNA annotations

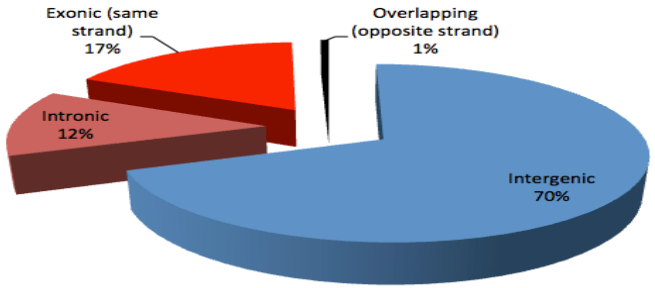
	Broad	ENCODE	UMICH
Publication	Cabili et al., 2011	Derrien et al., 2012	Prensner et al., 2011
Transcript reconstruction	RNA-seq	cDNA	RNA-seq
Length of lincRNA	> 200 nt	> 200 nt	> 200 nt
Structure of lncRNA	Multi-exon	Multi-exon	Multi-exon
Total transcripts	8,263	14, 880 (6000 + genic lncRNAs)	99 PCATs

Quantification Methods

(Intra) genic & overlapping transcripts with mRNAs give false estimation of lncRNA abundance in *non*-strand-specific RNA-seq, and being excluded from estimation.



Broad

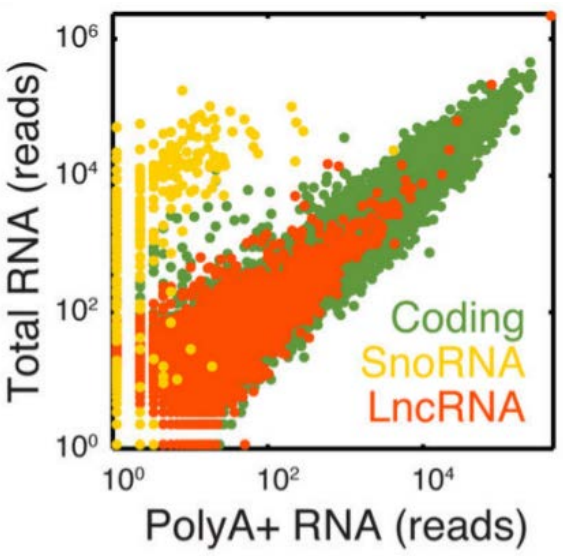


ENCODE

- Cufflinks
- HTSeq for PRAD and OV[$]

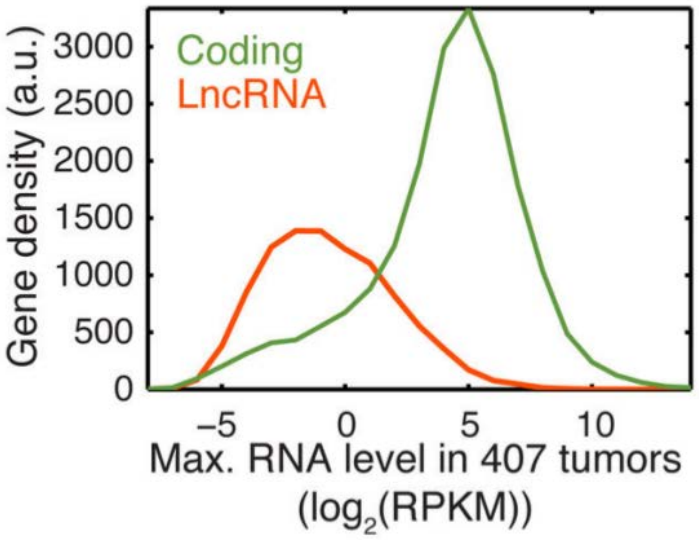
$ By Erik Larsson

Majority of lincrnas are PolyA enriched & show very low expression in comparison to mRNA expression

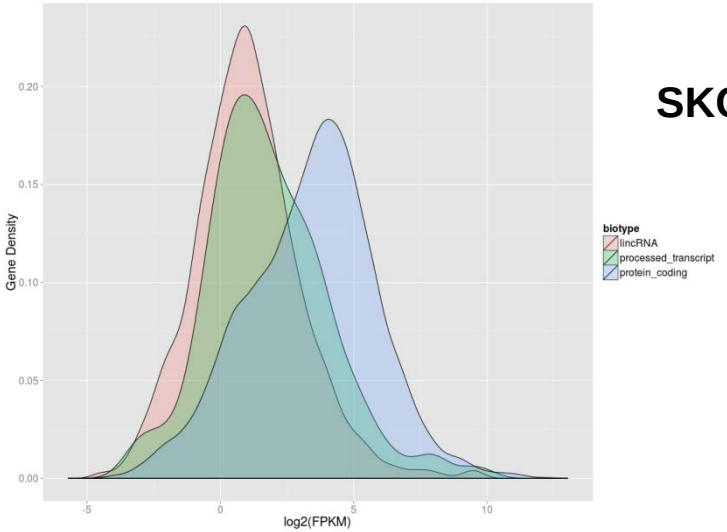


16 tissues

Akrami R. PLoS ONE 2013.
via Erik Larsson

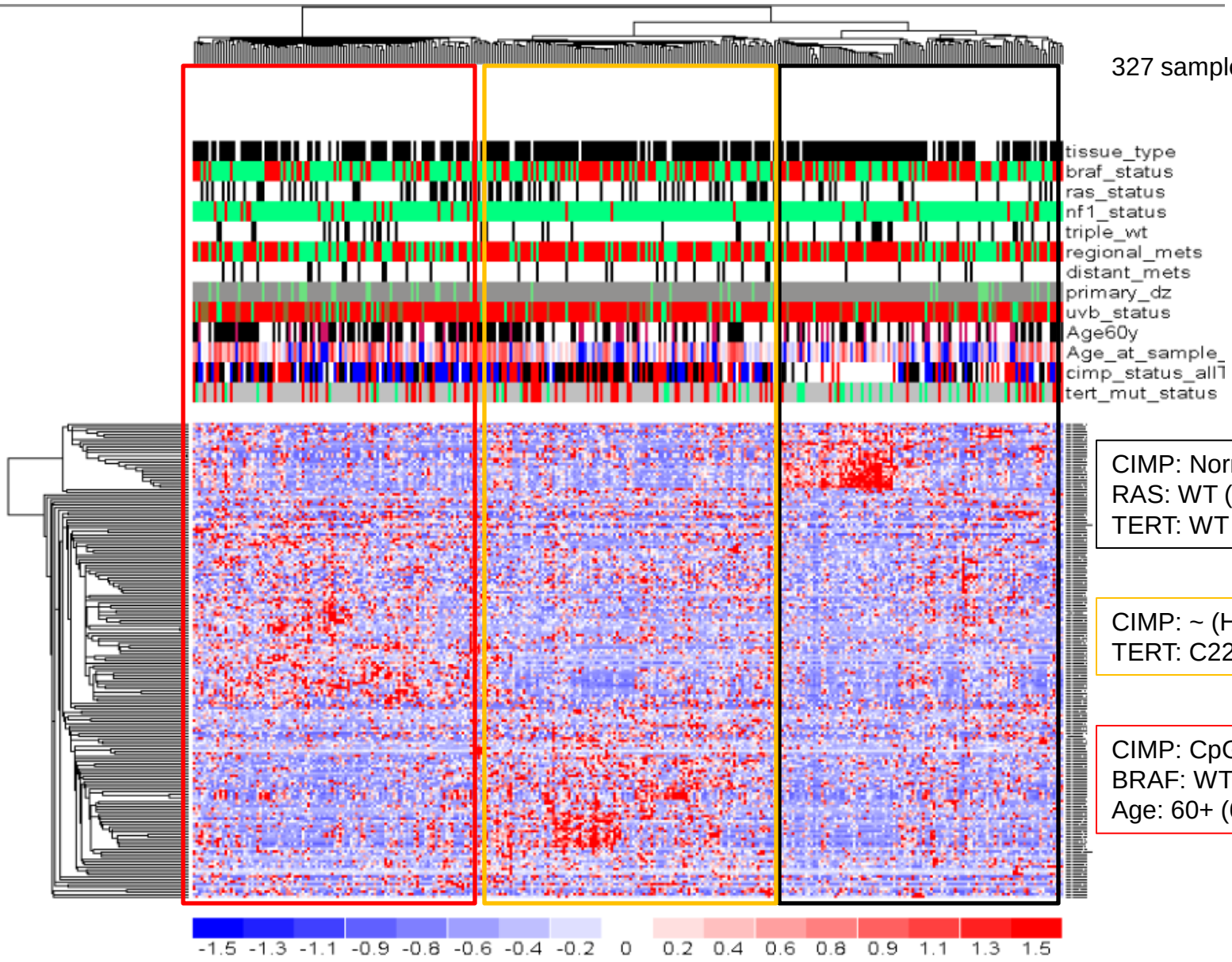


OV (n=407)



SKCM (n=355)

Unsupervised analysis show differential lincrna expression based on CIMP subtype and mutation signatures in SKCM



327 samples, 189 lincrnas

CIMP: Normal-like (52/109 | 70);
RAS: WT (91/109 | 236);
TERT: WT (22/109 | 39);

CIMP: ~ (Hypo + Hyper) methy;
TERT: C228T (13/78 | 25);

CIMP: CpG methy (45/112 | 81);
BRAF: WT (77/112 | 179);
Age: 60+ (65/112 | 138);

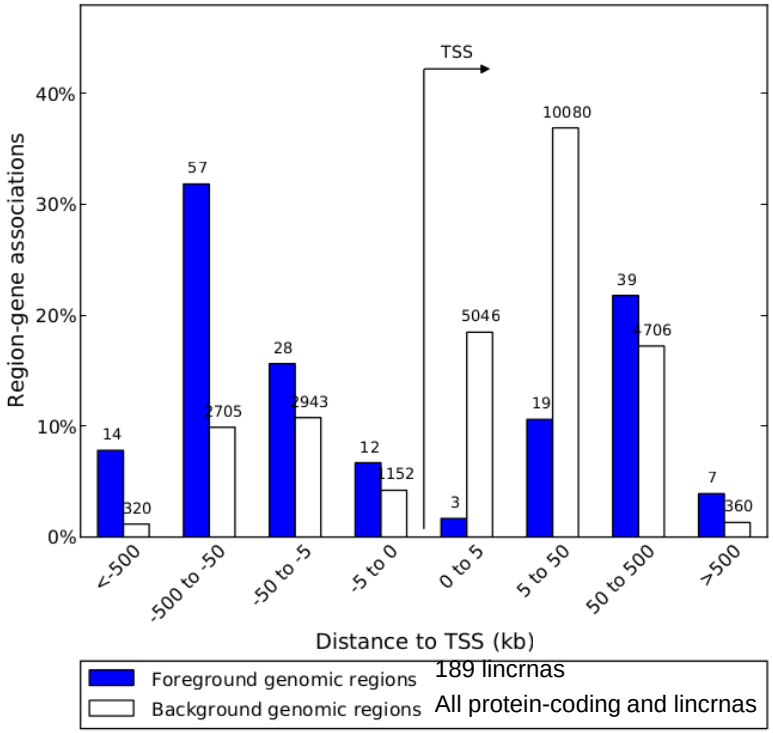
Experimental evidence indicating potential functional role of lincrnas in cancer

lincrna	Location	Properties
TRBV11-2	7q34	Upregulated in triple WT (BRAF, RAS, NF1) cluster; **Regulator of ERK signaling pathway.**
CT49 or TAG1	5p15	Upregulated in TERT C228T cluster; Expressed in melanoma, retinoblastoma, pineocytoma, mesothelioma. **Role in inactivation of Rb signaling pathway.**
MEG3	14q32	Uniformly low expression in melanoma; GISTIC2 wide deletion peak in the same region. **Known tumor suppressor lncrna.**

Differentially expressed lincrnas show proximity to several oncogene regulated genes

- MSigDB Cancer neighborhood analysis show significant enrichment (FDR q:0.02-0.03) of gene annotations regulated by HOXA9 (9/75), ERCC4 (5/27), FOXO1 (8/67) and CDH5 (5/34).

- lincrnas are enriched more in distal regulatory sites as compare to proximal regulatory sites, possibly indicating enhancer like activity.

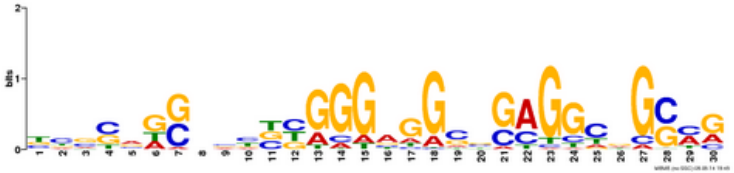
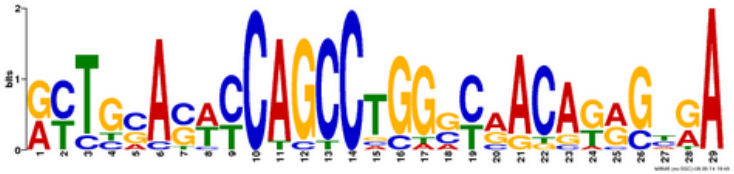
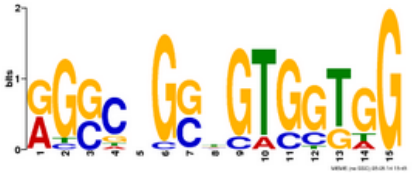


Relative distance of 189 lincrna regions from their putatively regulated genes.

Enrichment q value < 0.04

Identification of sequence-specific lincrna interactions in regulatory domains of cancer genes

Motif discovery analysis of variably expressed lincrna (n=189) indicates potential role of lincrnas in transcription regulation and cancer growth signaling pathways

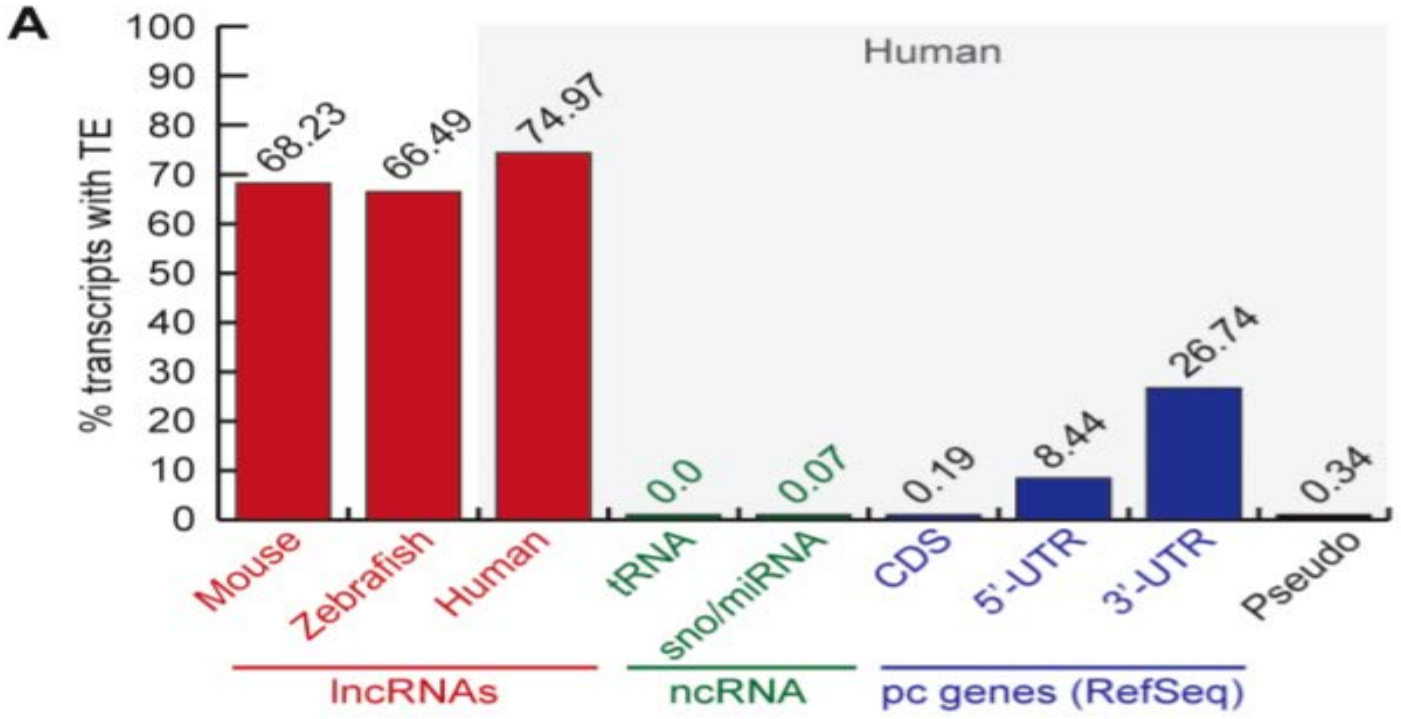
	Total Predictions	Top 5 Predictions
	373	CC transcription factor complex BP anterior/posterior pattern formation BP small GTPase mediated signal transduction BP negative regulation of signal transduction BP embryonic limb morphogenesis
	38	CC mitochondrion BP oxidation reduction CC nucleoplasm part MF catalytic activity BP gland development
	290	BP negative regulation of signal transduction MF transcription repressor activity BP Wnt receptor signaling pathway, calcium modulating pathway CC cytoplasmic part BP histone modification

Analysis shows known TF motifs for EGR1 (q: 0.04), ZEB1 (q: 0.03), SP2 (q: 0.03), Tcf3 (q: 0.1).

Alu elements are preferentially enriched in lincRNA exonic regions

- ~23% of lincRNA transcripts have at least 1 *Alu* sequence in their coding region.

Transposable elements enrichment in lincrna exonic regions

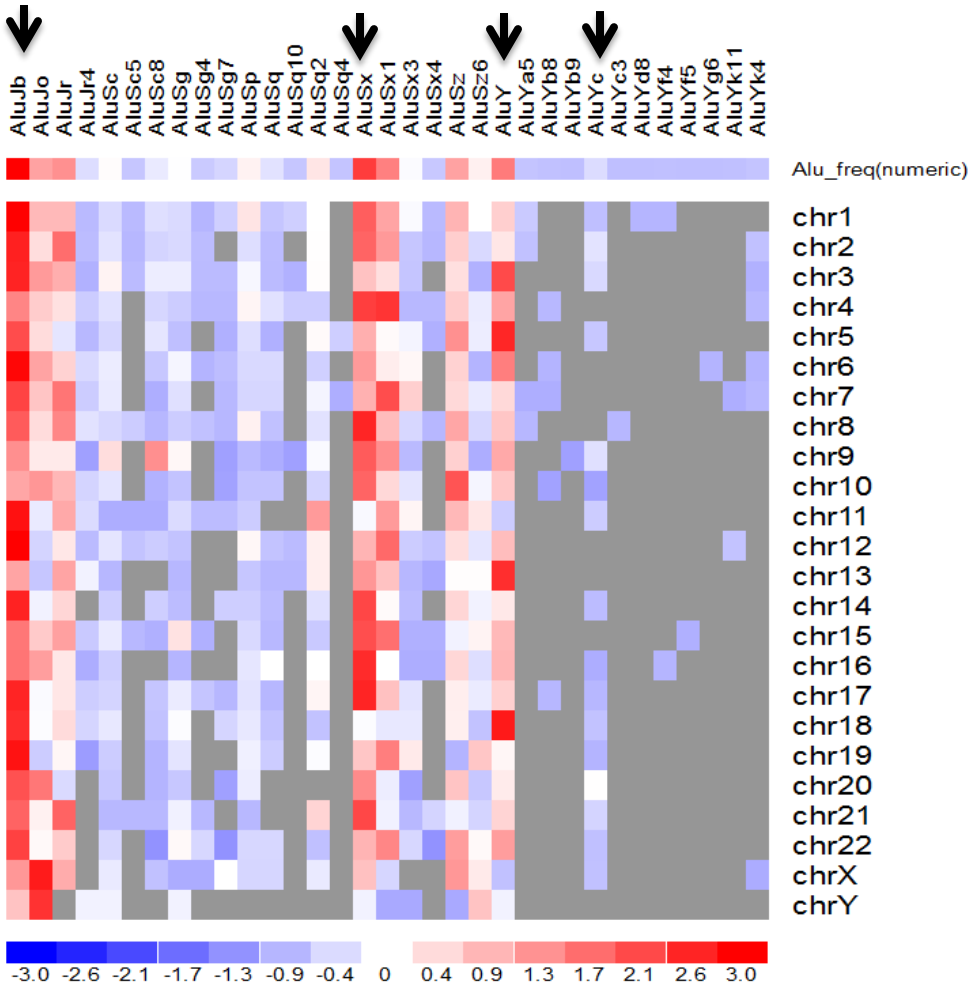


Kapusta A. PLoS Genet. 2013

Alu elements are preferentially enriched in lincRNA exonic regions

- Preferential hits from AluJb, AluSx and AluYc subfamilies which corresponds to most recent expansion of *Alu* elements with primate evolution.

Frequency of subfamilies of Alu elements within exonic regions of lincrnas, sorted by chromosomal location.



Summary

- Expression profiling of lincrna in SKCM and PRAD.
- Differential lincrna expression based on CIMP subtype and mutation signatures in SKCM.
- lincrna exonic regions are enriched in *Alu* elements and possibly play a role in sequence specific lincrna interactions in regulatory domains of cancer genes.

Ongoing Tasks

- Outline functional relevance, if any of differentially expressed lincrnas by…
  - Co-expression network analysis using **predicted lincrna – mrna** and **lincrna – microrna** partners sharing common sequence specific motif.
  - Copy number alterations involving differentially expressed lincrna regions.
- Making lincrna data and analyses accessible via Synapse portal (syn2426680) & Firehose pipeline.

Acknowledgement

Lynda Chin

Andy Futreal
Han Liang
Ching Lau

Terrence Wu
Kadir Akdemir
Sahil Seth
Chris Bristow

Roeland Verhaak

Kunal Rai

Firehose Team

TCGA SKCM and PRAD AWG Team

Erik Larsson, University of Gothenburg, Sweden

Ekta Khurana, Gerstein Lab


Andrew Rae | Nature 496 (2013)