RNA Repeat-Mediated Transcription Dosage Networks

Small RNA Silencing: Little Guides, Big Biology (A6)
January 24-28, 2016
Keystone Resort – Keystone, Colorado USA

Melanie Adams MD, NuclearRNANetworks.com®

Abstract

Allele dosage variability is key to ongoing evolution of viable differences between cells and individuals. Presence of numerous conserved retrotransposed repeat sequences (particularly conserved AluRepeat SINEs in introns of human genes) suggests that their control contributes to dosage fine-tuning as well as seeding of heterochomatin.

Interestingly, because retrotransposed elements (rTE) replicate through an RNA intermediate (copy and paste), conservation of retrotransposed sequences of parent and progeny loci provides a mechanism by which transcription from one gene can directly modify the transcription rate of another.

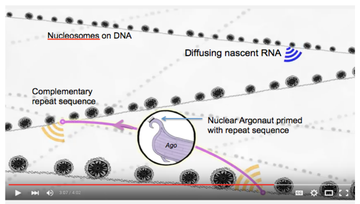


Fig 1. Nuclear RNA Networks Pt 1 - Nodes & Edges

In my model, stable nascent sequences diffuse in a gradient from the locus of a paused RNAPII. If nascent sequences diffusing from 2 loci are complementary, and recognized as endo-siRNA precursors, intranuclear Argonaut-guided histone modifying complexes 'connect' the transcription rate of one gene locus to that of the other. (Fig 1 and 2A. Animation at NuclearRNANetworks.com)

Functional nuclear RNA networks could form in this way, particularly as successful retrotransposition is more likely within the co-localized open chromatin coordinating a particular cell function.[1] (Fig 2B) While inhibition of transcription through a single repeat sequence might minimally delay progress of an initiated RNAPII, the elongation rate through a cluster of TE would be the sum of multiple rate control points governed by transcription of other loci. (Fig 2C)

Results Summary:

Experiment 1: Matches to RPL7L1* (n=45) and PPIA** (n=76) fall into two groups: Immune Response (IR) vs. non-IR genes, ***P* < 0.03**
Experiment 2: AluSp derived from IR genes are more likely to match IR genes (14/64) than the AluSp of non-IR genes (2/63), ***P* < 0.001**
Experiment 3: Random sequence (20nt, n=50). No matches.

* RPL7L1: 60S Ribosomal Protein L7-like, required for blastocyst formation
** PPIA: Peptidylprolyl Isomerase A (Cyclophillin A). Cis-trans isomerization, a rate-limiting step in protein folding.

Conclusions:

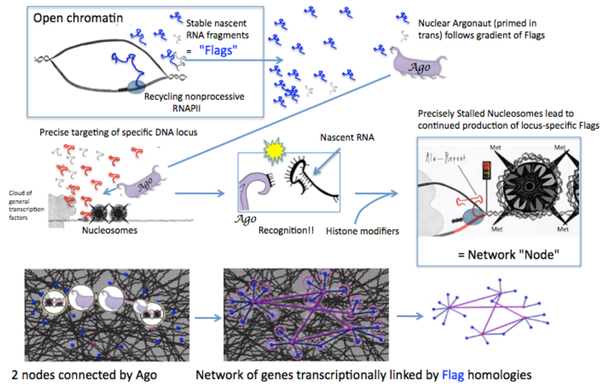
There is *in silico* evidence that retroTE stratify genes into Immune Response (IR) *vs.* non-IR, depending on the IR/non-IR origin of parental TE (*P* < 0.03). Specificity of the linkages between IR genes can be traced to Alu Repeat sequence: AluSp of high homology is present in greater frequency in these IR genes (*P* < 0.001).

Therefore, transcribed rTE fragments may be recognized by rTE-primed Argonaut-guided complexes that 'connect' these genes into a network, precisely linking the transcription rate of one locus to that of others. Within a single gene, the net effect of multiple small control points is to fine-tune RNAPII elongation rates and therefore, gene dosage. This mechanism provides the regulatory flexibility necessary for ongoing evolution of cells, individuals, and species within a background of genetic stability. In addition, AluSp piRNAs may provide a mechanism for stress-induced gene dosage effects to be transgenerational.

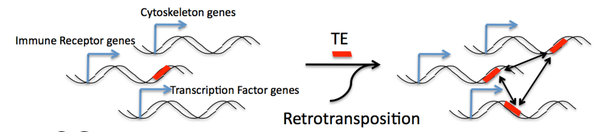
Ref 1. **Nuclear architecture dictates HIV-1 integration site selection.** Marini B, Kertesz-Farkas A, Ali H, Lucic B, Lisek K, Manganaro L, Pongor S, Luzzati R, Recchia A, Mavilio F, Giacca M, Lusic M. *Nature 521, 227-231 (14 May 2015) doi:10.1038/nature14226*

BACKGROUND

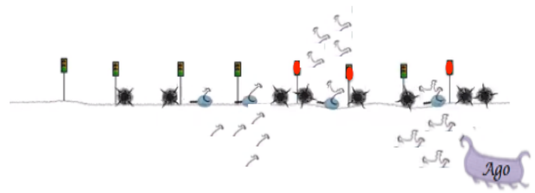
FIG 2A. Formation of a Nuclar RNA Network.



2B. Retrotransposition into co-localized open chro
Genes necessary for coordinated function become networked.



2C. Multiple rate control points within a single gene.



METHODS

1. **Identification of Pseudogene, Human Chained Self Alignments (HCSA) and high homology sequence**
   - *** NCBI BLAST-2013: Transcripts + top 15 intronic hits, E = 0.0, % homology >75%**
   - **• UCSC Genome Browser: Duplicates >1000, HCSA, BLAT- 2013 top 20 hits, homology >75**
2. **Groups compared with the chi square test**

This model predicts that retro-TEs with high homology will be present in greater frequency in functionally-grouped genes.

RESULTS

EXPERIMENT 1: Do matches to RPL7L1 (parent gene of pseudogene GR-retroRPL7L1) fall into different functional groups than match to PPIA (parent of housekeeping gene STIM-rPPIA)?

INDEX GENE:	**RPL7L1** (10,262 bp)	**PPIA** (6,438 bp)
high homolgy matches:	**Intragenic x 19**	**Intergenic x 40**
Pseudogenes, HCSA,*other	**CLEC16A**	**CD244**
	GR (GR)	**CMIP**
	MEFV	
	OXR1	
	PCCA (5')	
	rUK114	
	COBL	KIF9
	DOCK9	LRRC49
	POR, steroid synthesis	COX18
	WBP2NL: PR and ER fctn	A2ML1-AS
	Dpp2	CCR4-NOT
	SUFU	POLR3K
	WAS-AS 3'ncRNA	ATG10
	ZNF81	CUX1
	CHRNB3	GRID2
	NBEAL1	C2
	MRO	CTC-439O9.3
	TMEM168	CYP1A2
	AP3M1	FAM65B
	CCDC142	KIAA0355
	CPEB1	MEI1
	CPEB3	MSH5
	linc CTC-347C20.1	PCYT1A
	AC002310.13	PKD1L3
	RP11-15B17.1	PRKCA
		RAD51B

Functional Catagories
red= stress and/or immune response pathway
Pink = hormone receptor pathway
brown = apoptosis pathway
dark yellow = Ras signalling. activates MAP kinase cascade
dark red = serine/threonine protein kinase
blue = actin, llamnin or other filmentous protein
dark green = a/w tethering, vessicle trafficking
plum = Ca++ responsive signaling, buffering or transport proteins
purple = mitochondrial protein
Light blue = protein folding
teal = neurotranmitter receptor, neural specific protein
green = ubiquitinilation pathyway
gold = transcription factor
dark orange = growth factor
grey = protease inhibitor
black = other (including genes of mitosis or meiosis) or unknown

RESULTS Red = Inflammatory Response pathway

	RED	NON RED
RED GENES TOTAL (not including index)	6	2
NON RED GENES TOTAL	39	67
TOTAL GENES	45	69

The P value is 0.033014.

Pink = Hormone Receptor pathway	PINK	NON PINK
PINK GENES TOTAL	3	0
NON PINK GENES TOTAL	42	69
TOTAL GENES	45	69

The P value is .029739.

EXPERIMENT 3: Random sequence (20nt, n=50)
Results: No matches

EXPERIMENT 2: Do AluSps (250-350bp) derived from IR genes (n=7) match (top 10 hits) with genes in different functional groups than AluSsp derived from non-IR genes (n=7)?

Top 10 genes with high homolgy AluSp match

	AluSp from IR genes Match#		AluSp from non-IR genes Match#	
1	GR	1	-	STIM1: transmembrane protein that mediates Ca2+ influx
		2	RPL7L1	-
		3	SNORA51	IG
		4	CLEC16A	IG
		5	IG	MVB12A
		6	DSTN	IG
		7	EIF3AB11	IG
		8	CDON	IG
		9	RSG1	IG
		10	NTN1	IG
2	NFKB1	1	-	ARP:1 actin-related protein 1 homolog A
		2	GPC3	ZUFSP
		3	TTL14	ANKRD27
		4	CRTC3-AS1	CHRM5
		5	RP11-108SN6.3	FUT10 fucosyltransferase
		6	IG	SLC26A8
		7	APITD1-CORT	DIAPH2
		8	PPP4R1	IG
		9	MBNL2	IG
		10	IG	IG
3	IL1: interleukin 1 receptor, type I	1	-	RAB7A: vesicle traffic regulation in late endosomes
		2	FAM65B	IG
		3	LPXN	IG
		4	IG	IG
		5	IG	IG
		6	IG	AKHGEF3
		7	IG	MYO18A
		8	CERKL	COL4A4
		9	FAIM2	IG
		10	RCN1	IG
4	PSMB2 subunit of immunoproteasome	1	-	EMC7: tethering between the ER and mitochondrial outer membranes
		2	F11R (insertion)	IG
		3	NUAK2	IG
		4	IRAK2	LMX1B
		5	PPP3R1	IG
		6	IG	IG
		7	MATN2	IG
		8	PRKA	IG
		9	PCDH15	uncharacterized LOC256021
		10	IG	IG
5	MAP3K9	1	-	ALAS1: Succinyl-CoA + glycine = 5-aminolevulinate + CoA + CO(2)
		2	IG	SPATA5
		3	IG	IG
		4	IG	PGPEP1
		5	IG	PTGR1
		6	SMAD5	EIF2B5
		7	IG	ELMOD1
		8	DHX57	SETDB1
		9	IG	PAXBD
		10	IG	IG
6	CIITA	1	-	ATP2A2: ATPase, Ca++ transporting, cardiac muscle, slow twitch 2
		2	IG	IG
		3	IG	CTNND1
		4	SYCE2	IG
		5	IG	IG
		6	IG	FAF1
		7	RP11-544M22.13	SNTA1
		8	DFNA5	IG
		9	NUMB	IG
		10	MYO1B	ITGAL
7	TGFB2	1	-	C1orf43: chromosome 1 open reading frame 43
		2	CCNT1	FAF1
		3	SIRT1	IG
		4	DLGAP1	PADI3
		5	DIAPH1	IDE
		6	CSNK1G3	DOCK3
		7	IG	PLDS
		8	KIF5A	GPATCH2
		9	IG	CACNA1E
		10	IG	10.0 IG

IR genes	Non-IR genes
14	2
50	61
64	63

The P value is 0.001497