SVs & disease associated lncRNAs

Yan Zhang
Gerstein Lab
4/29/2015

Motivation

- It has been reported that deletion/CNV of lncRNA can be associated with a lethal lung development diseases.

Genome Res. 2013 Jan;23(1):23-33. doi: 10.1101/gr.141887.112. Epub 2012 Oct 3.

Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder.

Szafranski P[1], Dharmadhikari AV, Brosens E, Gurha P, Kolodziejska KE, Zhishuo O, Dittwald P, Majewski T, Mohan KN, Chen B, Person RE, Tibboel D, de Klein A, Pinner J, Chopra M, Malcolm G, Peters G, Arbuckle S, Guiang SF 3rd, Hustead VA, Jessurun J, Hirsch R, Witte DP, Maystadt I, Sebire N, Fisher R, Langston C, Sen P, Stankiewicz P.

- We look at functional impact of SVs (including CNVs) on known disease associated lncRNAs.

Three datasets

- **SVs**: 1000G phase 3 SV set.

- **"Conserved" lncRNAs**: A high-quality strict set of human lncRNAs (5413 transcripts) from Nitsche et al. 2015.

RNA. 2015 May;21(5):801-12. doi: 10.1261/rna.046342.114. Epub 2015 Mar 23.

Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved.

Nitsche A[1], Rose D[2], Fasold M[3], Reiche K[4], Stadler PF[5].

A little bit of detail:

GENCODE v14 (GRCh37) + a series of filters

- Remove transcripts that overlap with protein-coding sequences or pseudogenes in sense or antisense by at least one of GENCODE, ENSEMBL, UCSC, or RefSeq.
- Remove transcripts with putative coding regions.
- Remove unspliced entries
- Other cutoffs of PhyloCSF, possible ORF length, etc.

Three datasets

- **Disease associated lncRNAs**: The latest experimentally supported lncRNA-disease association data from LncRNADisease database (as of 4/27/2015).

Nucleic Acids Res. 2013 Jan;41(Database issue):D983-6. doi: 10.1093/nar/gks1099. Epub 2012 Nov 21.

LncRNADisease: a database for long-non-coding RNA-associated diseases.

Chen G[1], Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q.

Database summary from Chen et al. 2013:

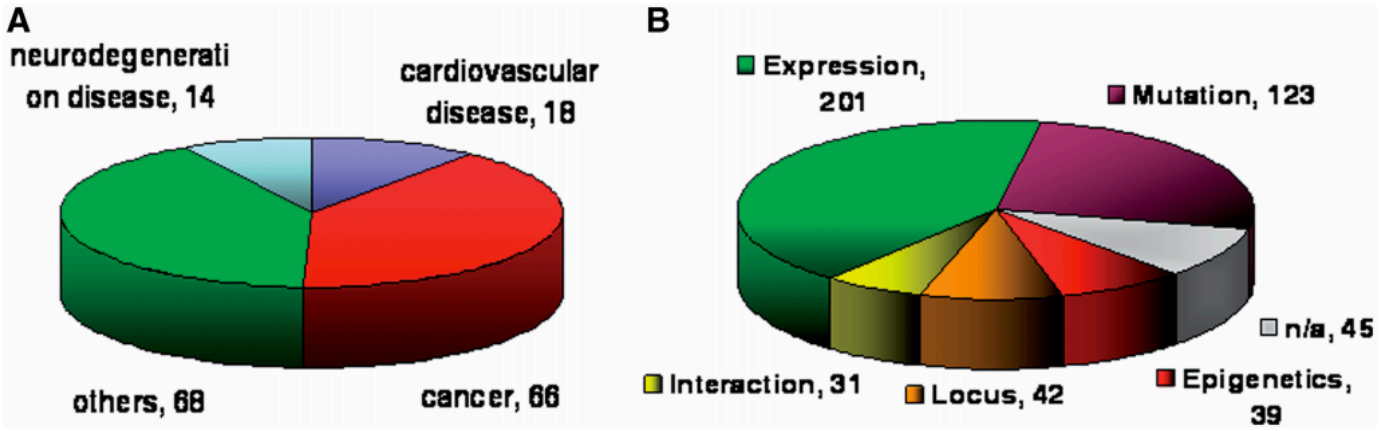
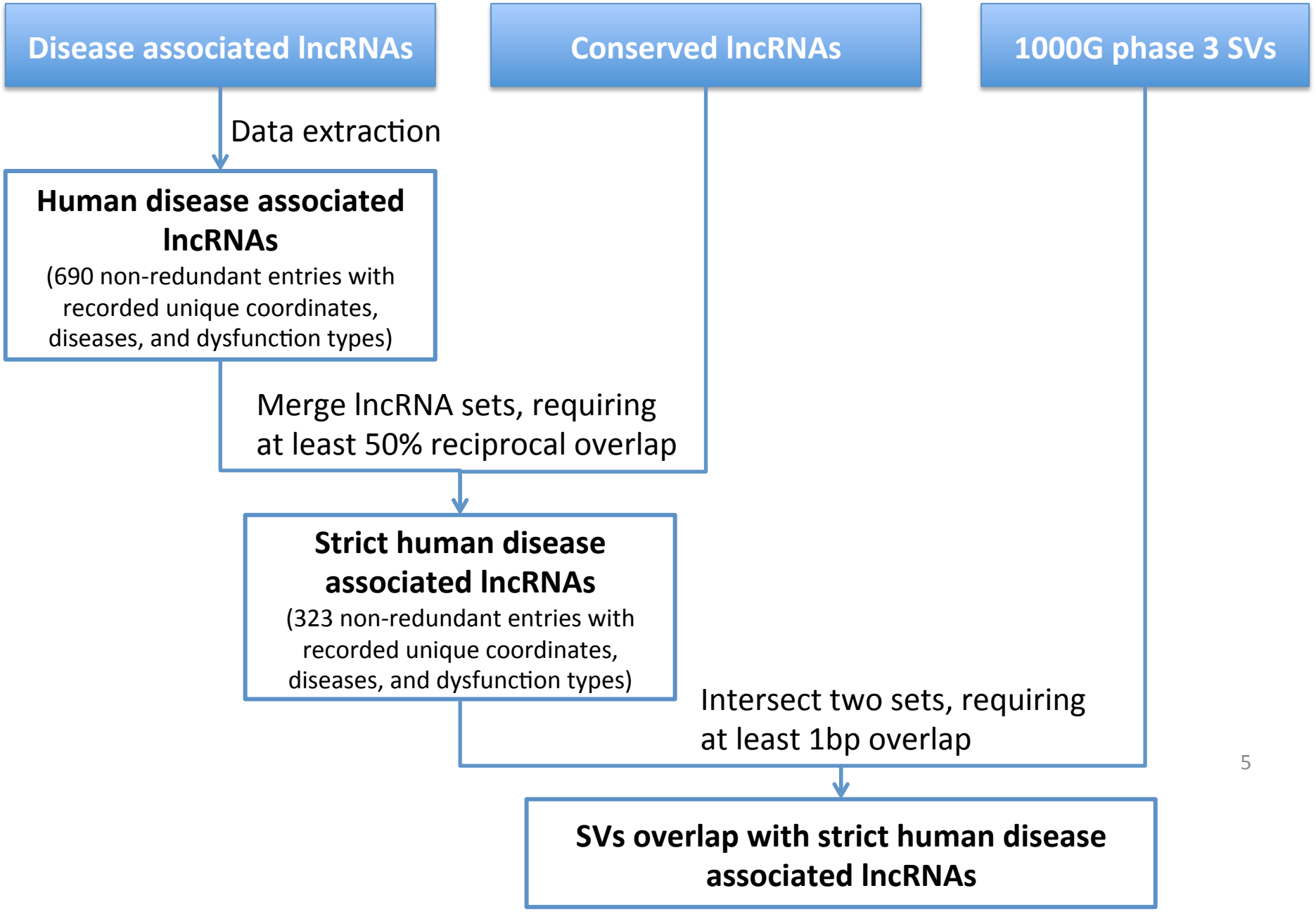


Figure 1. Statistics and distributions of diseases (**A**) and dysfunction types (**B**) of lncRNAs in the LncRNADisease database.

Analysis


Disease associated lncRNAs
Conserved lncRNAs
1000G phase 3 SVs
Data extraction
Human disease associated lncRNAs
(690 non-redundant entries with recorded unique coordinates, diseases, and dysfunction types)
Merge lncRNA sets, requiring at least 50% reciprocal overlap
Strict human disease associated lncRNAs
(323 non-redundant entries with recorded unique coordinates, diseases, and dysfunction types)
Intersect two sets, requiring at least 1bp overlap
SVs overlap with strict human disease associated lncRNAs

Result summary

- 44 unique SVs overlap with strict human disease associated lncRNAs.

DEL	DUP	mCNV	ALU	LINE1
30	4	1	7	2

- Example 1: The SV with the most (7) lncRNA entries

SV Information						lncRNA information						
Chr	Start (0-based)	End	Type	Frequency	ID	Chr	Start (0-based)	End	Strand	Symbol	Associated disease	Dysfunction type
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	AIDS	expression
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	amyotrophic lateral sclerosis	regulation
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	frontotemporal lobar degeneration	Interaction
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	Huntington's disease	expression
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	Intrauterine Growth Restriction	expression
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	TDP-43-associated pathological state	expression
chr11	65182225	65192548	DEL	0.0002	UW_VH_9761	chr11	65190268	65192232	+	NEAT-1	oral squamous cell carcinoma	expression

Result summary

- 135 unique disease associated lncRNA entries overlap with SVs.
- Example 2: The lncRNA overlap with the most SVs

SV Information						lncRNA information						
Chr	Start (0-based)	End	Type	Frequency	ID	Chr	Start (0-based)	End	Strand	Symbol	Associated disease	Dysfunction type
chr2	8170890	8182766	DEL	0.000599	BI_GS_CNV_2_8170891_8182766	chr2	8147900	8418214	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8210077	8210517	DEL	0.0002	DEL_pindel_2551	chr2	8147900	8418214	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8265735	8267776	DEL	0.000399	BI_GS_DEL1_B3_P0259_12	chr2	8147900	8418214	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8359006	8360475	DEL	0.0002	UW_VH_14482	chr2	8147900	8418214	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8383265	8383514	ALU	0.0002	ALU_umary_ALU_988	chr2	8147900	8418214	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8391683	8393675	DEL	0.000599	BI_GS_DEL1_B5_P0259_533	chr2	8147900	8418214	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8170890	8182766	DEL	0.000599	BI_GS_CNV_2_8170891_8182766	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8210077	8210517	DEL	0.0002	DEL_pindel_2551	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8265735	8267776	DEL	0.000399	BI_GS_DEL1_B3_P0259_12	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8359006	8360475	DEL	0.0002	UW_VH_14482	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8383265	8383514	ALU	0.0002	ALU_umary_ALU_988	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8391683	8393675	DEL	0.000599	BI_GS_DEL1_B5_P0259_533	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation
chr2	8426073	8426304	ALU	0.000399	ALU_umary_ALU_989	chr2	8147900	8464760	-	LINC00299	Intellectual and developmental disability	mutation