

Data & approach for mouse pgene transcription

- **Data**
 - RNA-seq data on 30 mouse tissues, each with 2 bio-replicates;
- **Annotation**
 - 22,706 protein coding genes, from ensembl 65
 - 15,887 pseudogenes, from pseudopipe
- **Transcription of protein coding genes**
 - RPKM of each gene, done by CRG
- **Transcription of pseudogenes**
 - Remove low mappability regions (< 1) from pseudogene exons;
 - Filter out remaining regions less than 100 bp;
 - Calculate the RPKM values of the resulting regions, using uniquely mapped reads;
 - Set the threshold corresponding to human pgene threshold based on quantile normalization of human and mouse protein coding expression profiles

Results

- **Human**

- Threshold: rpk_m = 2
- Transcribed pseudogenes: 1441 (out of 12,007, 12%)

- **Mouse**

- Threshold: rpk_m = 3.28
- Transcribed pseudogenes: 878 (out of 15,887, 5.5%)

Result on Orthologous Pseudogenes

	Human	Mouse
Total Pseudogenes	12,007	15,887
In Synteny	687	610
In Synteny + orthologous parents + same biotypes	129 (127 processed, 2 duplicated)	129 (127 processed, 2 duplicated)

GO Enrichment of Ortholog Parents

Gene Ontology term	Cluster frequency	Genome frequency	Corrected P-value	FDR	False Positives	Genes annotated to the term
electron transport chain	7 of 61 genes, 11.5%	183 of 45554 genes, 0.4%	2.42e-06	0.00%	0.00	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , UQCR11 , TXN , MT-ND3
respiratory electron transport chain	6 of 61 genes, 9.8%	133 of 45554 genes, 0.3%	1.19e-05	0.00%	0.00	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , UQCR11 , MT-ND3
mitochondrial ATP synthesis coupled electron transport	5 of 61 genes, 8.2%	69 of 45554 genes, 0.2%	1.69e-05	0.00%	0.00	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , MT-ND3
ATP synthesis coupled electron transport	5 of 61 genes, 8.2%	74 of 45554 genes, 0.2%	2.42e-05	0.00%	0.00	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , MT-ND3
oxidative phosphorylation	5 of 61 genes, 8.2%	90 of 45554 genes, 0.2%	6.49e-05	0.00%	0.00	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , MT-ND3
cellular respiration	6 of 61 genes, 9.8%	214 of 45554 genes, 0.5%	0.00019	0.00%	0.00	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , UQCR11 , MT-ND3
cellular metabolic process	40 of 61 genes, 65.6%	16263 of 45554 genes, 35.7%	0.00088	0.57%	0.04	VHL , PCGF1 , LSM5 , HPRT1 , NUS1 , UBE2C , UQCR11 , RAB9A , CRLF3 , RPS25 , NPM3 , WDR45L , CIR1 , PSMB1 , NOLC1 , KAT2B , MND1 , MRPS5 , UQCQRH , RPL36AL , TOMM22 , GPBP1L1 , TXN , DPPA3 , PDLIM1 , NDUFA4 , EIF4EBP1 , DEPDC1 , MT-CO3 , NDUFS6 , AKR1D1 , GPS2 , SMG5 , MYCBP , LSM1 , FUNDC1 , ITCS , DPPA2 , MT-ND3 , H2AFZ
oxidation-reduction process	8 of 61 genes, 13.1%	721 of 45554 genes, 1.6%	0.00236	1.00%	0.08	UQCQRH , MT-CO3 , AKR1D1 , NDUFS6 , TXN , NDUFA4 , UQCR11 , MT-ND3
energy derivation by oxidation of organic compounds	6 of 61 genes, 9.8%	413 of 45554 genes, 0.9%	0.00864	1.11%	0.10	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , UQCR11 , MT-ND3
cellular component biogenesis	14 of 61 genes, 23.0%	2881 of 45554 genes, 6.3%	0.00938	1.00%	0.10	VHL , UQCQRH , RPS25 , NDUFA4 , NPM3 , WDR45L , HPRT1 , NOLC1 , UBE2C , ISG20L2 , POMP , EIF4EBP1 , RSU1 , H2AFZ
generation of precursor metabolites and energy	7 of 61 genes, 11.5%	631 of 45554 genes, 1.4%	0.00956	0.91%	0.10	NDUFA4 , UQCQRH , MT-CO3 , NDUFS6 , UQCR11 , TXN , MT-ND3

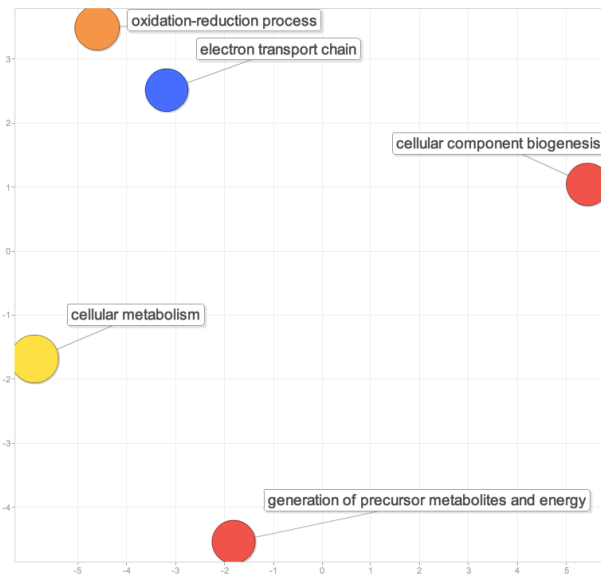


Figure from Revigo

Transcribed Orthologs

- **Human**
 - 19 of 129 orthologs are transcribed (14.7%)
- **Mouse**
 - 7 of 129 orthologs are transcribed (5.4%)
- One ortholog is transcribed in both organisms:
 - ENST00000453380.1 in human, chrX:139173901-139174720
 - chrX:57801214-57802021 in mouse