

# A Functional Model of LRPPRC in LSFC

Chaoyang Deng<sup>1</sup>, John D. Rioux<sup>2</sup>

**Keywords:** microarray data analysis, gene ontology, COX deficiency, Leigh Syndrome

## 1 Introduction.

LRPPRC is an RNA and single-stranded DNA-binding protein [1]. By integrative genomics LRPPRC has been identified as the causal gene for Leigh syndrome, French-Canadian type (LSFC), a variant of cytochrome c oxidase deficiency found in Quebec [2]. An A354V mutation in exon 9 and/or an 8-bp deletion in exon 35 (results in C1277STOP) of LRPPRC were found in LSFC patients. However, the biological function of LRPPRC remains largely unresolved, and how the mutations of LRPPRC cause COX deficiency and develop LSFC is still unknown.

Microarray data analysis is a powerful approach to extract biological pathways and to find differently expressed gene sets related to a specific disease or condition. With gene ontology information, it will even enable us to construct a functional model for a totally unknown gene.

## 2 Methods and Results.

The NGO program was developed based on Pearson correlation efficiency algorithm and Gene Ontology information. Using a public Affymetrix Microarray dataset, we performed Neighborhood GO analysis of LRPPRC and found three pathways, including translation factors and electron transport chain, which are significantly enriched in LRPPRC neighborhood. Many other interesting gene sets in categories of molecular function, biological process and cellular component (table 1) are also highly correlated with LRPPRC expression, especially gene set related to RNA splicing.

We also searched LRPPRC protein domain using PSORT and Pfam tools and identified the following putative domains: (1) A mitochondrial targeting sequence (ARLYAI); (2) 10 previous reported PPR domains; (3) a MIF4G domain encompassing the point mutation site, which bind to eIF4A, eIF3, RNA and DNA; and (4) a SECA-DEAD domain at the site of truncation mutation. LRPPRC Protein sequence blast in GenBank also identified several homologous proteins from other genome with functions of mRNA stabilization, proper splicing and translation initiation.

Based on above study, as well as knowledge of human mitochondrial genome and case reports of other forms of Leigh Syndrome caused by mutations in tRNA, we finally constructed a functional Model for LRPPRC in LSFC (Figure 1). In our model, we predict that LRPPRC may recognize and bind to the junction site between COX mRNA and tRNA, help correct splicing to produce mature COX mRNAs, interact with mitochondrial translation initiation factors and ribosome by docking at the short 5'UTR, stabilizing the COX mRNA by binding to the 3'UTR, and even assist new born COX peptides folding and assembly. We believe that this integrative approach will not only be an important tool for our research of the LRPPRC gene but will also be a useful approach that will be applicable to other genes that have poorly characterized function.

---

<sup>1</sup> Broad Institute of MIT and Harvard, One Kendall Square, Cambridge, MA, USA. E-mail: cdeng@broad.mit.edu

<sup>2</sup> Broad Institute of MIT and Harvard, One Kendall Square, Cambridge, MA, USA. E-mail: rioux@broad.mit.edu

### 3 Figures and tables.

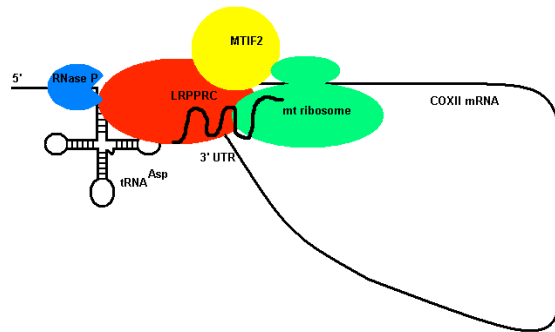


Figure 1: The predict function model of LRPPRC in LSFC

Table 1: Neighborhood GO analysis of LRPPRC.

Category	Terms	100	250	500	12482	Significance
pathway (159)	Translation_Factors	1	4	8	48	***
	Cell_Cycle	1	8	14	76	*****
	Electron_Transport_Chain	1	3	11	83	***
cellular (483)	mitochondrion	7	18	35	313	*****
	nucleus	40	90	144	1837	*****
	cytoplasm	11	26	40	472	***
	complex	15	38	75	680	*****
	membrane	3	12	34	2498	*****
	function	26	42	68	338	*****
(4211)	RNA binding	9	14	19	58	*****
	splicing	3	4	14	81	*****
	translation	4	9	16	78	*****
	chaperone	3	9	11	87	***
process (1532)	regulation of translation	3	5	8	50	***
	assembly	7	18	21	182	***
	RNA splicing	13	18	30	83	*****
	RNA processing	15	22	29	110	*****
	protein biosynthesis	1	8	19	162	***
	protein folding	5	9	12	97	***

Note: For each neighborhood of 100, 250 or 500 genes, the number of genes corresponding to a given GO term is tabulated. The significance was calculated by chi-squared test, and those marked by three stars has P value <  $10^{-5}$ , five stars has P value <  $10^{-10}$  and seven stars has P value <  $10^{-20}$ . The numbers under each category are the total number of terms in that category.

### References

- [1] Mili S, Pinol-Roma S. 2003. LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs. *Mol Cell Biol.* 23(14):4972-82.
- [2] Mootha VK et al. 2003. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci USA.* 100(2):605-10.