Determining Transcriptional Regulators of Peroxisome Biogenesis via *in Silico* Analysis

Edwin Paz[1], Clarence Mah[2], Michelle Ragsac[2], Heather Chen[2], Zachary Chiang[2], David Tran[2], Terry Gaasterland[3]

[1]Albert La Spada Laboratory, Sanford Consortium for Regenerative Medicine; [2]Bioengineering Department, University of California, San Diego;
[3]Laboratory of Computational Genomics, Scripps Institute of Oceanography

Abstract

Recent studies have shown that certain PEX genes are among those associated with peroxisome biogenesis disorders and peroxisomal defects. It is currently unknown whether any of these genes are coordinately regulated by the same transcription factors. We have utilized several different *in silico* approaches to analyze the upstream regulatory regions of these genes for common binding site sequence motifs in addition to enriched transcriptional regulators.

Introduction

Peroxisomes are highly dynamic organelles that are involved in important catabolic processes such as β-oxidation of fatty acids, reduction of reactive oxygen species, and biosynthesis of ether phospholipids. As indispensable players in human health, it is important to further characterize unknown aspects of peroxisome biogenesis.

Peroxisomal disorders represent a class of medical conditions that share dysfunction of peroxisomes. Dysregulation of peroxisomal processes are associated with a group of brain developmental disorders including Zellweger syndrome, neonatal adrenoleuko-dystrophy, and infantile Refsum disease. Underlying these disorders are mutations in PEX genes resulting in faulty peroxin production and function.

We hypothesize that important contributors to neurodegenerative diseases are dysregulated transcriptional networks affecting peroxisome biology. Using a bioinformatics approach, we have identified potential transcription factors (TF) and their consensus binding motifs that have a high probability of coordinately regulating peroxisome biogenesis and function. Identifying novel transcriptional regulators of peroxisome biogenesis may help uncover novel targets for therapeutic intervention.

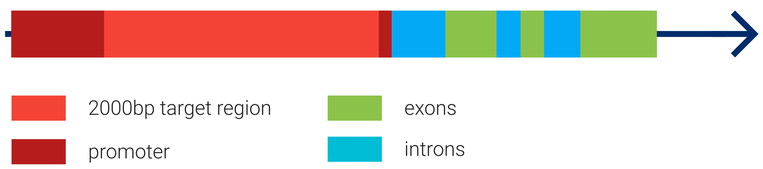

Figure 1

1 Our analysis focuses on analyzing the 2000bp upstream promoter region of target genes.
2 The oPOSSUM analysis yielded potential overrepresented motifs, characterized as having a fisher score >1 SD more than the mean.
3 Examples of *de novo* motifs identified in the PBD related genes.

Materials & Methods

We performed three different analyses to search for regulatory motifs in the promoter regions of two gene subsets: peroxisome biogenesis disorder related genes (PBD) and peroxisome defect related genes (PD). The target region was the 2000bp upstream region of each gene.

PBD Genes		PD Genes	
PEX1	PEX12	ABCD1	HSD17b4
PEX2	PEX13	ACOX1	PHYH
PEX3	PEX14	AGPS	TRIM37
PEX5	PEX16	AGXT	
PEX6	PEX19	AMACR	
PEX7	PEX26	CAT	
PEX10		GNPAT	

HOMER
TRANSFAC
oPOSSUM

oPOSSUM 3.0, a web-based tool for detection of overrepresented TF binding sites, was used to query the JASPAR database at the target regions of the PBD and PD genes. HOMER, a *de novo* motif discovery software suite, was used to exhaustively look for overrepresented motifs that are then matched to known motifs in the JASPAR database. TRANSFAC, a commercial transcription factor binding profile database for eukaryotic TFs, was also queried for overrepresented motifs with a custom script: overrepresented TFs were determined by comparing their frequency in the target region of our gene sets versus a background set of the target region of all genes in the TRANSFAC database. Finally, a z-score was calculated for enrichment to screen for TFs that were least expected by random chance.

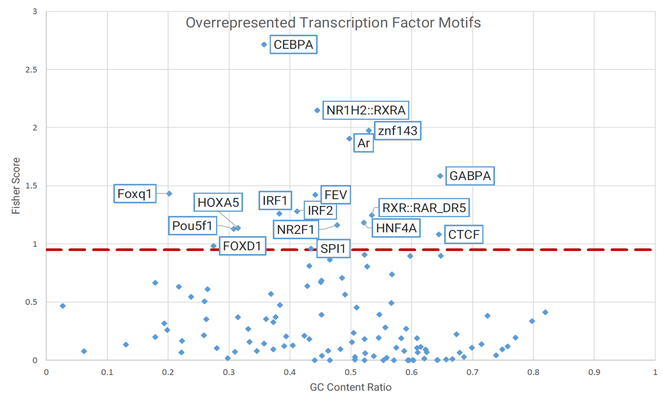

Figure 2

Results & Discussion

A preliminary set of known TF candidates were selected based on our statistical analyses. Figure 2 shows significant TF motifs comparing Fisher scores against %GC content using all TFs in JAPSAR as background and PBD genes as our target set. Statistically significant motifs exhibited a Fisher Score of at least one standard deviation above the mean. The TF motifs depicted in Figure 2 are above the threshold described. GC content also provides a way to measure any bias present in our data set; the wide distribution shows that our target set has a similar nucleotide composition as our background set, suggesting that our analysis is not biased towards a subset of TF motifs.

Motif	log pValue	%Targets	%BG	STD	BG STD
GAGGTCTAGAAA	-2.128e+01	53.85%	1.66%	399.0bp	676.5bp
CTTCCGTCAG	-2.123e+01	61.54%	2.91%	502.5bp	659.4bp
AGGATAACGACC	-1.939e+01	53.85%	2.19%	504.3bp	685.0bp

Figure 3

Conclusion

Our computational analyses produced a small list of likely transcription factor candidates associated with PDBs and PDs. These studies in conjunction with experimental methods may help guide *in vitro* experimentation. Knockdown of these candidates will allow us to see downstream effects on PBD and PD pathways, specifically if candidates impact peroxisomal function independently or in coordination with other transcription factor complexes. We also demonstrate the ability to investigate *de novo* motifs produced by HOMER as possible TF binding candidates. While we only analyzed genes directly related to PBDs and PDs, further TF analyses may investigate other genes connected to other peroxisomal functions.

Acknowledgments

Special thanks to Professor Terry Gaasterland for the privilege to access the San Diego Supercomputer as well as her bioinformatics insight and guidance on how to approach this project. Special thanks to Dr. Albert La Spada who is principal investigator for this project. Thank you to Dr. Edwin Paz for the opportunity to contribute to the project and as a mentor.