



# **A Knowledge-based Approach for Genome-wide Genotyping Analysis of Parkinson's Disease**

**A NextBio™ Whitepaper By:**

**Mostafa Ronaghi, Ph.D.  
Senior Research Associate  
Stanford Genome Technology Center**

## INTRODUCTION

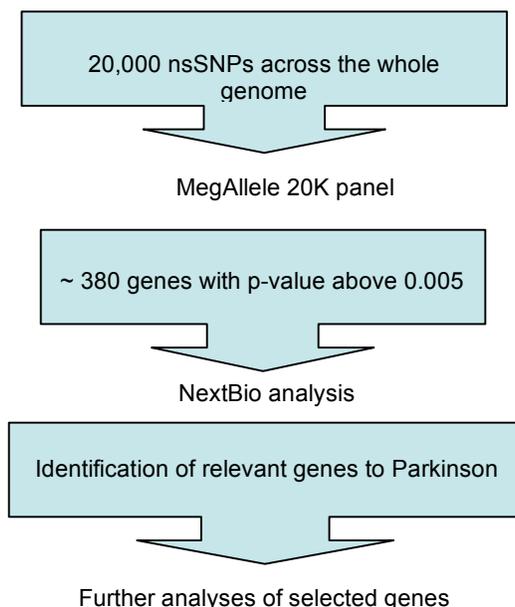
In this paper, we utilize a novel knowledge platform from NextBio to gain new insights into Parkinson's disease (PD). We were able to explore candidate non-synonymous SNPs using NextBio's reference library of neurological expression studies, regulatory motifs, gene ontology functional groups and pathways to identify candidate genes for further study.

## BACKGROUND

To study PD, different approaches have been used to identify important genes including PARK1-11. The protein products of these genes are involved in different pathways of neurodegeneration, making elucidation of the underlying disease mechanism difficult. The emergence of high-throughput gene expression and genotyping technologies enabled a systematic approach to the study of PD in human, as well as in model organisms. With the accumulation of high-throughput data, a single-gene approach in the study of complex diseases is giving way to global, genome-wide approach that captures the complexity of associated phenotypes.

## STUDY RESULTS

In this report we explore results of a high-throughput genotyping study of PD patients using NextBio (Fig. 1). In the approach presented here, we explore non-synonymous SNPs with statistically significant association with PD within the context of gene expression signatures of phenotypes related to PD.



**Figure 1.** Analysis workflow diagram of the Parkinson's disease genotyping study using Affymetrix GeneChip® arrays and NextBio.

### Genome-wide SNP genotyping

To explore the mechanism behind development of PD, we performed a genome-wide genotyping study on 400 patient samples. We chose a commercial panel (the MegAllele system marketed by Affymetrix) to analyze 20,000 polymorphic non-synonymous SNPs (nsSNPs). Our hypothesis in choosing the nsSNP-specific sets for genome-wide coverage, were to test the functional variants directly, as opposed to other approaches that indirectly test many variants genome-wide. The rationale behind this approach was that the causal SNPs for disorders are highly penetrant and often lead to severe phenotypes. We could therefore directly associate these SNPs to a disease rather than dealing with more complex indirect associations. Although we have a comprehensive list of nsSNPs in the chosen panel covering ~13,000 genes in human, the complete list could be as large as ~60,000 nsSNPs. Several genes do not contain nsSNP and others appear to be rare. This panel will therefore exclude some of the major genes which are highly conserved throughout evolution and may play an important role in different diseases. MegAllele software was used to generate a

list of SNPs with significant p-values. We set a p-value of 0.005 to generate a list consisting of 300 genes. Just one gene (p-value 0.0043) from that list had previously been reported in PD-related studies.

### Meta-analysis of genotyping data with NextBio

There are a number of approaches for analyzing large-scale data from gene expression and genotyping platforms. Biological relevance of a given set of genes is typically evaluated using pathway and gene ontology information, as well as publications-based regulatory network analysis. Here, we present an example of a knowledge-based approach for analyzing results of Parkinson’s disease genotyping study. It involves meta-analysis of a given set of genes (containing non-synonymous SNPs from MegAllele) across other large-scale studies related to PD – a process in which we identify important connections across independent sets of related data. In addition, a regulatory motif analysis of PD-associated genes serves as the backbone for generating and exploring a hypothesis presented here. NextBio currently contains thousands of study results across diverse therapeutic areas. All of the studies were systematically analyzed and corresponding signatures (results) captured within the platform. NextBio content includes data from human, mouse and rat, as well as the capability for a seamless cross-species analysis. Once our data was imported into NextBio it was ready for exploration. To evaluate the significance of specific genes in our set we focused on a number of studies on neurological disorders related to Parkinson’s disease. We then asked the following questions.

*Are there any known PD-associated genes in the gene list generated in this study?*

We searched NextBio’s reference library of neurological studies and found that park2, a gene correlated to late-onset Parkinson, has shown a significant p-value of 0.0043 in our genotyping study. Although this is a relatively modest p-value, it serves as a good orthogonal dataset to provide validation of our study.

*Can we draw any important information through the study of regulatory motifs upstream of genes identified in our PD genotyping study?*

For this specific purpose we imported a set of regulatory motifs available at the Broad Institute Molecular Signature Database into NextBio. Although this subset focuses primarily on highly conserved motifs we were able to find several motifs upstream of many genes with significant PD association p-values in our genotyping study. About 100 genes contained one of the four conserved motifs (Fig. 2). There are four transcription factors that are known to bind to these sites – thus serving as potential regulators of genes that contain them. Particularly interesting was the regulatory motif that binds MEF2A transcription factor. The MEF2A-binding motif indicates a significant association with genes derived from our study (p-value = 0.0148), potentially binding to and regulating upstream sequences of 30 genes from our dataset.

Matching Results							
Details	RankScore	P-Value	Q-Value	S Biogroup Name	Common Genes	Species	Common Features
		0.0148	1	YTATTTNR_MEF2	30	Homo sapiens	40
		0.0544	1	CTTTGA_LEF1	52	Homo sapiens	62
		0.0613	1	TAANNYSGCG_Unknown motif	4	Homo sapiens	7
		0.1477	1	RTAAACA_FREAC2	38	Homo sapiens	45

**Figure 2.** Regulatory motifs identified for the gene list generated by genome wide genotyping.

*Is the expression of MEF2A significantly regulated in other studies within NextBio related to Parkinson’s disease?*

MEF2A has not previously been directly associated with Parkinson’s disease. However, it is upregulated in the brain and has been involved in other neurological disorders. Searching through NextBio, we found that MEF2A expression is modulated in PD related studies (Fig. 3). Most notably, the expression of MEF2A is downregulated in the brain of alpha-SYN-overexpressing transgenic mice (-1.5 fold), as well as in rotenone-treated neuroblastoma cells derived from human (PD model system) by -1.4 fold. Expression modulation of MEF2A in PD-related phenotypes provided us with a hint about the importance of this gene.

A separate important piece of evidence came from the study on mouse brain tissue specificity. MEF2A was shown to be specifically up-regulated in the striatum relative to other mouse brain tissues. As we know, Parkinson’s disease results from loss of Dopaminergic innervations to the striatum (and other basal ganglia). In addition, we found that MEF2A is highly upregulated in specific subsets of cingulate cortex neurons in one study. Cingulate cortex’s Lewy body formations

have been associated in a number of studies with dementia accompanying development of Parkinson's disease. These orthogonal datasets provide important underlying evidence for the potential involvement of MEF2A in Parkinson's disease.

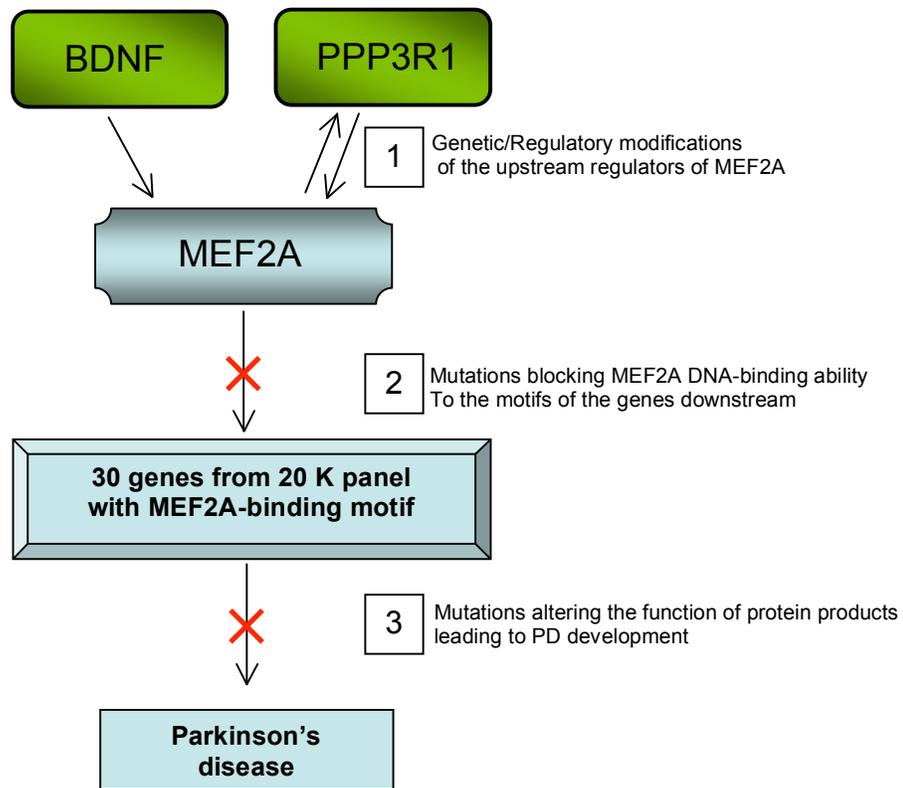
Details	RankScore	Study	Gene Symbol	P-Value	Fold Change	Species
		MouseForebrainAtlas - Up-regulated expression	MEF2A	7.73E-5	2.002	Mus musculus
— Biosets						
Details	RankScore	S Bioset Name	Gene Symbol	P-Value	Fold Change	Species
		Cingulate cortex - GABA G42 neurons_UP-regulated	MEF2A	7.73E-5	2.002	Mus musculus
		Cingulate cortex - Glu CT6 neurons_UP-regulated	MEF2A	0.00496	1.919	Mus musculus
		BrainMapAtlas_Mm - Up_Tissue-specific expression	MEF2A	1.98E-5	1.588	Mus musculus
— Biosets						
Details	RankScore	S Bioset Name	Gene Symbol	P-Value	Fold Change	Species
		Striatum_Up - Tissue-specific expression	MEF2A	1.98E-5	1.588	Mus musculus
		Parkinson's disease - protective effects of Hsp70	MEF2A	0.00931	-1.512	Mus musculus
— Biosets						
Details	RankScore	S Bioset Name	Gene Symbol	P-Value	Fold Change	Species
		PD prevention by HSP70_DOWN - aSYN transgenic vs. wildtype mice	MEF2A	0.00931	-1.512	Mus musculus
		PD prevention by HSP70_CHGN - aSYN transgenic vs. wildtype mice	MEF2A	0.00931	-1.512	Mus musculus
		Parkinson's disease model - Rotenone treated neuroblastoma cells	MEF2A	0.027	-1.409	Homo sapiens

**Figure 3.** Expression of MEF2A across mouse and human studies investigating Parkinson's disease model systems, as well as tissue-specificity in the brain. Blue sections represent the expanded details of each study (above). For example, the "MouseForebrainAtlas" study contains a number of signatures that can be viewed within the blue "details" section.

Surprisingly, we found in the literature that MEF2A is regulated by BDNF and PPP3R1 (Flavell et al. 2006). BDNF stimulates neuronal growth and protects nigral dopamine neurons and has previously been associated with Parkinson disease (Karamohamed et al. 2005). PPP3R1, on the other hand, is among the genes with significant p-value in our genotyping study.

**CONCLUSION**

This data led us to several interesting hypotheses related to PD and MEF2A. We hypothesize that MEF2A may play a significant role in Parkinson's disease via multiple distinct mechanisms (Fig. 4).



**Figure 4.** MEF2A is regulated by BDNF and PPP3R1 and regulates 30 genes that also have significant association with PD in our genome-wide genotyping study.

If indeed MEF2A gene is an important regulator of pathways leading to PD development then the actual disease-causing aberrations could involve genes upstream of MEF2A (e.g. BDNF and PPP3R1), the MEF2A itself (sequence changes altering its ability to bind and regulate genes downstream) or the genetic alterations in the downstream targets of MEF2A regulation (which includes a set of 30 genes with the upstream regulatory motifs that bind MEF2A). All of these potential mechanisms require separate investigations. Currently, our plan is to sequence MEF2A in our population set. In addition, we plan to perform antibody assays to look at the phosphorylation state of this protein in brain tissues of Parkinson's disease patients.

## REFERENCES

- Flavell SW, Cowan CW, Kim TK, Greer PL, Lin Y, Paradis S, Griffith EC, Hu LS, Chen C, Greenberg ME. **Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number.** (2006) *Science*. 311:1008-1012.
- Karamohamed S, Latourelle JC, Racette BA, Perlmutter JS, Wooten GF, Lew M, Klein C, Shill H, Golbe LI, Mark MH, Guttman M, Nicholson G, Wilk JB, Saint-Hilaire M, DeStefano AL, Prakash R, Tobin S, Williamson J, Suchowersky O, Labell N, Growdon BN, Singer C, Watts R, Goldwurm S, Pezzoli G, Baker KB, Giroux ML, Pramstaller PP, Burn DJ, Chinnery P, Sherman S, Vieregge P, Litvan I, Gusella JF, Myers RH, Parsian. (2005) **BDNF genetic variants are associated with onset age of familial Parkinson disease: GenePD Study.** *Neurology* 13:1823-1825.

For more information please contact:



NextBio  
3 Results Way  
Cupertino, CA 95014  
Phone: 408 200 2042  
Fax: 408 777 8091  
info@nextbio.com  
www.nextbio.com

#### **About NextBio**

NextBio accelerates biological and clinical discovery by overcoming many of the challenges created by an overwhelming abundance of disparate, high-throughput experimental data. Founded in 2004 by a team of industry leaders, NextBio addresses a fundamental need within R&D organizations to effectively leverage vast quantities of internal and public information to advance research-critical discoveries. NextBio's knowledge-based discovery platform enables all researchers to tap into the power of information combined from different assays, platforms and research silos to draw new scientific and clinical insights. Its value in improving research efficiency and effectiveness has already been proven at some of the world's leading research organizations.