Genetic biomarkers for brain hemisphere differentiation in Parkinson's Disease

Mou'ath Hourani*, Alexandre Mendes†, Regina Berretta† and Pablo Moscato†

*Newcastle Bioinformatics Initiative (NBI), School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW, 2308, Australia
hourani@cs.newcastle.edu.au
†Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine (CIBM), The University of Newcastle, Callaghan, NSW, 2308, Australia
{Alexandre.Mendes, Regina.Berretta, Pablo.Moscato}@newcastle.edu.au

Abstract. This work presents a study on the genetic profile of the left and right hemispheres of the brain of a mouse model of Parkinson's disease (PD). The goal is to characterize, in a genetic basis, PD as a disease that affects these two brain regions in different ways. Using the same whole-genome microarray expression data introduced by Brown et al. (2002) [1], we could find significant differences in the expression of some key genes, well-known to be involved in the mechanisms of dopamine production control and PD. The problem of selecting such genes was modeled as the MIN (α,β)−FEATURE SET problem [2]; a similar approach to that employed previously to find biomarkers for different types of cancer using gene expression microarray data [3]. The Feature Selection method produced a series of genetic signatures for PD, with distinct expression profiles in the Parkinson's model and control mice experiments. In addition, a close examination of the genes composing those signatures shows that many of them belong to genetic pathways or have ontology annotations considered to be involved in the onset and development of PD. Such elements could provide new clues on which mechanisms are implicated in hemisphere differentiation in PD.

Keywords: Data Mining; Feature Selection; Microarray Data Analysis
PACS: 02.70.-c; 89.20.Ff

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disease characterized by four main symptoms: resting tremors, rigidity of the limbs, slowness of movement, and difficulty with balance and coordination. The disease is caused by a continuous loss of the dopaminergic neurons in the *substantia nigra* of the brain. The degeneration of these neurons reduces the amount of dopamine produced, which interferes in the functioning of the *basal ganglia* - a region of the brain involved in the control of muscle action. Although the exact underlying cause of PD in not yet known, most scientists believe that genetics and/or environmental factors play an important role. Family history is gradually being perceived to be a risk factor, with an estimated 15-25% of the Parkinson's patients reporting having a relative with the disease. This view was initially confirmed in 1996, when a candidate gene for some cases of PD was mapped to chromosome 4 [4]. Since then, several other Parkinson's-related genes have been found. Also, a series of genetic pathways, in particular those related to apoptosis and neurodegeneration have also been implicated to PD.

CP952, *Computational Models for Life Sciences—CMLS '07, 2007 International Symposium*
edited by T. D. Pham and X. Zhou

This work uses the same dataset introduced in Brown et al. (2002) [1] but goes one step further in terms of the analysis of results. While Brown et al. (2002) focuses in finding genes related to PD by comparing the brains of a PD model mouse against a control individual, in this work we aim at finding variations in gene expression between the left and right hemispheres of the mice brains. The dataset used comprises 7,035 genes and 80 experiments, corresponding to 40 voxels from a normal and 40 voxels from a Parkinson's affected rodent. It is available for download directly from the author's website[1]. The work of Brown provides extensive information on how the brain tissue samples were collected and how the microarray instance was generated. As these are not the central issues of this paper, we refer the reader directly to Brown's work.

The main method used to identify differentially expressed genes is based on a mathematical model for inference of gene expression patterns and NP-hard problem known as the MIN $(\alpha,\beta)-$FEATURE SET problem [3, 5]. This approach was previously used to identify genetic signatures for Alzheimer's disease [6], as well as for the molecular classification of cancer [3, 7]. Genetic signatures obtained using this method carry a mathematical guarantee of inter-class differentiation and intra-class similarity, which lacks in other traditional approaches based solely on statistics, such as p-value based selection. The MIN $(\alpha,\beta)-$FEATURE SET method is briefly described in the next section, but a more thorough discussion is found in reference [3].

MODELING THE GENE SUBSET SELECTION PROBLEM

To understand how we address the problem of finding genetic markers for the two hemispheres of the PD brain, we will describe the MIN $(\alpha,\beta)-$FS problem, which provides a combinatorial formalization of the problem of interest. The MIN $(\alpha,\beta)-$FS problem is a variation of the well-known $k-$FEATURE SET and it has been introduced with the aim of selecting robust feature sets of strong discriminative power and within class similarity [5]. The problem is described as:

- **Instance**: A set of m examples $X = \{x^{(1)},\ldots,\ x^{(m)}\}$, such that $\forall\ i = 1,\ldots,m$ $x^{(i)} = \{x_1^{(i)}, x_2^{(i)},\ldots,x_n^{(i)}, t^{(i)}\} \in \{\mathbb{N}\}^{n+1}$, and three integers $k > 0$, and $\alpha \geqslant 1, \beta \geqslant 0$

- **Question:** Does there exist an $(\alpha,\beta)-k-$Feature Set S, $S \subseteq \{1,\cdots,n\}$, with $|S| \leqslant k$ and such that:
  - for all pairs of examples $i \neq j$, if $t^{(i)} \neq t^{(j)}$ there exists $S'(i,j) \subseteq S$ such that $|S'| \geqslant \alpha$ and for all $l \in S'$ $x_l^{(i)} \neq x_l^{(j)}$?
  - for all pairs of examples $i \neq j$, if $t^{(i)} = t^{(j)}$ there exists $S'(i,j) \subseteq S$ such that $|S'| \geqslant \beta$ and for all $l \in S'$ $x_l^{(i)} = x_l^{(j)}$?

The problem is $NP-$hard as the $k-$FEATURE SET problem is a special case where $(\alpha,\beta) = (1,0)$ [8]. Furthermore, the MIN $(\alpha,\beta)-$FS problem is not likely to be fixed-

[1] http://labs.pharmacology.ucla.edu/smithlab/genome_multiplex/index.htm

parameter tractable for parameter k as Cotta and Moscato (2003) have proved that the k–FEATURE SET problem is $W[2]$-complete [2].

A fundamental distinction of the model used in this work, in relation to all previous applications, is that we take into account the fact that different regions of the brain are expected to have naturally distinct expression profiles due to normal tissue functional differentiation. Therefore, the aim of our approach becomes to find genes that are *differentially expressed in the same voxel in the PD and normal brains*. Under these circumstances, we will only work with the cases in which $\alpha \geqslant 1$ and $\beta = 0$. Our aim is then to find a set S' of k features (genes) such that for any pair of samples with different targets (PD/normal, same voxel), there are at least α genes in S' that support (i.e. have distinctive expression levels) this difference in all voxels. In our tests, the parameter α was adjusted to return a solution with around 36 genes, which is the number of genes reported in Brown's work.

The notion of distinctive expression levels is a critical issue. Some gene expression studies consider two expression levels as significantly distinct if there is at least a 2-fold difference between them. Such difference is high enough to reduce the influence of noise and other precision limitations in the cDNA microarray technology. Therefore, given any two samples, relative to the same voxel, one from the PD and the other from the normal brain, we only consider that a gene is discriminative for that pair if the expression levels for the two samples differ by at least a 2-fold ratio.

As said before, the MIN (α,β)–FEATURE SET Problem is NP-hard, but the use of a standard integer programming (IP) formulation, such as in references [3, 7], in conjunction with the IP solver ILog CPLEX 9.0[2], allows solving medium-sized instances to optimality in relatively short CPU times. To cite an exmple, the three optimal feature sets shown in Figure 2 were obtained in less than 5 minutes of CPU time each, using a Pentium IV 3.0 GHz computer with 1 Gb RAM.

PARKINSON'S DISEASE MICROARRAY DATA

The Parkinson's disease microarray data used in this work was introduced by Brown et al. (2002) [1]. They used a PD model created by the administration of toxic doses of methamphetamine to the C57BL/6J strain of mice[3]. These doses cause a destruction of the dopaminergic nerve, which is responsible for control movement initiation and coordination. The brains from the control and methamphetamine-treated mice were divided into 40 voxels each (ten volume slices taken horizontally, each divided into four voxels) and then analyzed, resulting in a 7,035-gene microarray with 80 samples.

Brown et al. (2002) reported 36 genes as differentially expressed in the brains of the PD and normal mice. The genes are listed in Figure 1 and will be used as benchmark for comparison against the ones selected using the MIN (α,β)–FS method.

---

[2] http://www.ilog.com/products/cplex

[3] http://jaxmice.jax.org/info/index.html

Gene symbol	Accession #	Name
Abca2	AA276158	ATP-binding cassette, sub-family A (ABC1), member 2
A1p2a2	AA222567	ATPase, Ca2+ transporting, cardiac muscle, slow twitch 2
C1r	AA261393	Complement component 1, r subcomponent
Lrpap1	AA253890	Low-density lipoprotein receptor-related protein associated protein 1
Papss2	AA244536	5'-Phosphoadenosine 5'-phosphosulfate synthase 2
Pglyrp	AA238752	Peplidoglycan recognition protein
Psme1	AA239485	Protease (prosome, macropsin) 28 subunit, alpha
Pura	AI894064	Purine-rich element-binding protein A
Rdh5	AA275664	Retinol dehydrogenase type 5
Rps5	AA240279	Ribosomal protein S5 (translation)
X66	AA249976	Xeroderma pigmentosum, complementation group C
Hdac5	AA017742	Histone deacetylase 5 (regulation of transcription)
Klf1	W97446	Erythroid Kruppel-like factor 1
Mata111	AA461637	Metastasis associated 1-like 1
Mxl1	AA472395	Ma-interacting protein 1 (antagonist of c-Myc transcription factor)
Nr2c2	AA501045	Nuclear receptor subfamily 2, group H, member 2
Sp1	AA212645	trans-Acting transcription factor 1
Cdc42	AA266975	Cell division cycle 42 homolog (ρ GTPase, cell morphology)
Crkas	AA240272	v-crk associated tyrosine kinase substrate
Mlapt/Mapl	AA028410	Microtubule-associated protein τ
Pkcq	W98195	Protein kinase C, theta (neurite outgrowth)
Stk2	AA268478	Serine/threonine kinase 2 (apoptosis, cytoskeletal remodeling)
Ecm1	AA237378	Extracellular matrix protein 1 (secretory glycoprotein)
Eln	AA239171	Elastin (extracellular matrix component)
Lamc2	W49392	Laminin, γ2 (extracellular matrix glycoprotein)
Grb2	AA183927	Growth factor receptor bound protein 2
Ppp2ca	AA245165	Protein phosphatase 2a, catalytic subunit, α isoform
S100a6	AA267952	Calcium-binding protein A6, or calcyclin
Parg	AA260570	Poly(ADP-ribose) glycohydrolase (apoptosis)
Siah1a	AA267965	Seven in absentia homolog 1A (cell cycle arrest)
Mor1	AA266087	Mitochondrial malate dehydrogenase (oxidative phosphorylation)
Mu1	AA250181	Methylmalonyl-coenzyme A mutase
Fxr2h	AA119248	Fragile X mental retardation gene, autosomal homolog 2
Qk	AA220551	Quaking (RNA binding protein required for myelin formation)
Ap1b1	AA221073	Adaptor protein complex AP-1, β1 subunit
Arf2	AA266938	ADP-ribosylation factor 2 (GTP-binding protein)

FIGURE 1. List of 36 genes reported in Brown et al. (2002) as differentially expressed in Parkinson's disease.

COMPUTATIONAL RESULTS

The application of the MIN (α,β)−FS method resulted in three genetic signatures for Parkinson's disease; one consisting of genes differentially expressed in the PD brain compared to the normal (Figure 2a), using all samples; a second signature considering only the samples from the left hemisphere of the PD and normal brains (Figure 2b); and the last signature considering only the right hemisphere samples (Figure 2c).

As the biomarkers will be compared to the 36-gene signature from Brown et al. (2002), we adjusted the parameters of the MIN (α,β)−FS approach to return optimal feature sets with the closest number of genes (probes). The parameters were (α,β) = (9, 0) for the whole brain signature; (α,β) = (10, 0) for the left hemisphere; and (α,β) = (11, 0) for the right hemisphere.

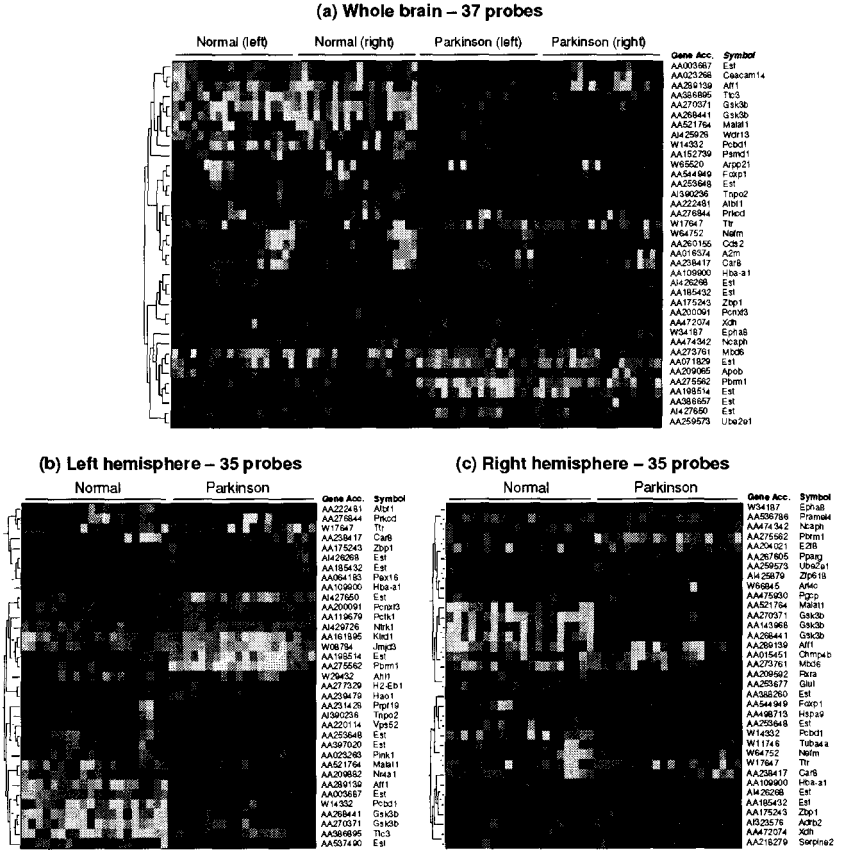

FIGURE 2. Three genetic signatures obtained using the MIN (α,β)−FS method using samples from the (a) whole brain; (b) left hemisphere only; (c) right hemisphere only. The parameters of the FS algorithm were adjusted to select the closest number of biomarkers to those reported in Brown et al. (2002) [1], in order to remove possible biases in the result analysis. Biological aspects of the findings are discussed and compared to those of reference [1].

Pathway analysis

An analysis of the pathways most represented in each of the signatures points to the MIN (α,β)−FS method is as a better biomarker selection method. The study was conducted using a web-based tool - the *GATHER* [9] from Duke University, USA. Next, we present the information retrieved using the pathway analysis tool and then we proceed to explain each of the findings. In Figure 3 we show the pathways most represented in the genetic signatures introduced in Brown et al. (2002) and in this work. Afterwards, we show current evidence of links between PD and the pathways present in Figure 3. Between parenthesis, we list the signatures in which the pathway is represented (**B** = Brown et al. (2002); **W** = MIN (α,β)−FS - whole brain; **L** = MIN (α,β)−FS -

(a) Brown et al. (2002)

Pathway	Gene symbol(s)	p-value
Focal adhesion	Bcar1 Cdc42 Grb2 Lamc2	0.004
Tight junction	Cdc42 Ppp2ca Prkcq	0.004
Retinol metabolism	Rdh5	0.005
Sulfur metabolism	Papss2	0.01
Glyoxylate and dicarboxylate metabolism	Mdh2	0.01
Reductive carboxylate cycle (CO2 fixation)	Mdh2	0.01

(b) MIN (α,β)−FS – Whole brain

Pathway	Gene symbol(s)	p-value
Alzheimer's disease	A2m Gsk3b	< 0.0001
Hedgehog signaling	Gsk3b	0.006
Complement and coagulation cascades	A2m	0.009
Purine metabolism	Xdh	0.01
Tight junction	Prkcd	0.01
Insulin signaling	Gsk3b	0.01

(c) MIN (α,β)−FS – Left hemisphere

Pathway	Gene symbol(s)	p-value
Glyoxylate and dicarboxylate metabolism	Hao1	0.002
Parkinson's disease	Pink1	0.004
Alzheimer's disease	Gsk3b	0.006
MAPK signaling	Nr4a1 Ntrk1	0.01
Hedgehog signaling	Gsk3b	0.01

(d) MIN (α,β)−FS – Right hemisphere

Pathway	Gene symbol(s)	p-value
Peptidoglycan biosynthesis	Glul	0.0002
Nitrogen metabolism	Glul	0.002
Glutamate metabolism	Glul	0.002
Alzheimer's disease	Gsk3b	0.003
Hedgehog signaling	Gsk3b	0.006
Purine metabolism	Xdh	0.01
Insulin signaling	Gsk3b	0.01

FIGURE 3. Pathways most represented in (a) the work of Brown et al. (2002) and (b-c-d) in the three genetic signatures depicted in Figure 2, respectively. The tables yield information on the pathway name; which biomarkers are present in the signature; and the statistical relevance of such finding, in terms of number of hits in relation to the size of the pathway – given by the p-value. Only pathways with p-value ≤ 0.01 are reported. It is noteworthy the high number of metabolism pathways and also the presence of Parkinson's and Alzheimer's disease pathways in the MIN (α,β)−FS signatures.

left hemisphere; **R** = MIN (α,β)−FS - right hemisphere).

- *Alzheimer's disease* (W, L, R) - It is widely accepted that neurodegenerative diseases in general have common characteristics. Recent studies further confirm that both Alzheimer's and Parkinson's diseases share similar genetic mechanisms that lead to alterations in physiological properties of the brain [10, 11].
- *Complement and coagulation cascades* (W) - No connection reported.
- *Focal adhesion* (B) - No connection reported.
- *Glutamate metabolism* (R) - The degeneration of dopaminergic neurons in the brain of PD patients has been linked to a malfunction of the complex interaction

between dopaminergic and metabotropic glutamate receptors (mGluRs) [12] , or to glutamate-induced toxicity [13].

- *Glyoxylate and dicarboxylate metabolism* (B, L) - No connection reported.
- *Hedgehog signaling* (W, L, R) - The Hedgehog pathway has a role in several developmental processes and in the maintenance of adult organs and cell types, including neuronal subtypes [14]. This was recently confirmed by a stem cell study for mesodiencephalic dopaminergic neuron replacement that confirmed the importance of Fgf8 (*fibroblast growth factor 8*) and Shh (*sonic hedgehog*) for the development of such cells [15].
- *Insulin signaling* (W, R) - The insulin signaling pathway has been linked to Parkinson's disease due the concentration of Igf1 (*insulin-like growth factor 1*) receptors in the *substantia nigra* region [16]. Also, a recent study has shown that human brain endothelial cells are specially vulnerable to hyperglycemic stress, resulting in apoptosis. Activation of insulin signaling thus protects the cell integrity by maintaining cellular reduction-oxidation balance [17].
- *MAPK signaling* (L) - This pathway has been directly involved in PD development by a study of a mutation in gene Lrrk2 (*leucine-rich repeat kinase 2*), which alters MAPK signaling cascades and triggers apoptosis, causing autosomal dominant Parkinson's disease [18].
- *Nitrogen metabolism* (R) - Nitrogen metabolism is linked to PD by the toxic oxidative stress response mechanism, which triggers dopaminergic neuron death [19, 20].
- *Parkinson's disease* (L) - The Parkinson's disease pathway is represented by Pink1 (*pten-induced kinase 1*), whose mutations cause autosomal recessive PD [21].
- *Peptidoglycan biosynthesis* (R) - No connection reported.
- *Purine metabolism* (W, R) - Purine metabolism is closely connected to dopamine metabolism through *adenosine*, an endogenous purine nucleoside that regulates it. Recent studies have established adenosine receptor-dopamine receptor interactions in PD and suggest adenosine as a target for PD therapy [22].
- *Reductive carboxylate cycle* (B) - No connection reported.
- *Retinol metabolism* (B) - Retinol metabolism is regulated by retinoic acid (RA) [23], and recently RA was found to be involved in the regulation of plasticity and regeneration in the adult brain, possibly playing a role in motor disorders such as PD [24].
- *Sulfur metabolism* (B) - Defects in sulphoxidation and sulphation of xenobiotics is a common hallmark of PD, indicating that endogenous sulphur metabolism is disturbed [25]. Xenobiotics effects on mitochondrial function have also been linked to neurotoxicity and apoptosis [26].
- *Tight junction* (B, W) - Alterations in tight junction and in the blood-brain barrier change the integrity of the cell membrane and permeability. Several studies describe their relation to brain degeneration and PD [10, 27].

From the 16 pathways cited in Figure 3, 11 of them can be linked to neurodegeneration or Parkinson's disease. Also, the appearance of Parkinson's and Alzheimer's diseases pathways in the genetic signatures introduced in this work is worth mention-

BROWN ET AL. (2002)

CSPG4 (chondroitin sulfate proteoglycan 4)
RAPGEF1 (Rap guanine nucl. exchange factor)
VAV1 (vav 1 oncogene)
E2F2 (E2F transcription factor 2)
WASF2 (WAS protein family, member 2)
WASF1 (WAS protein family, member 1)
WASL (Wiskott-Aldrich syndrome-like)
CD2AP (CD2-associated protein)
TNK2 (tyrosine kinase, non-receptor, 2)
ANXA2 (annexin A2)
HNRPC (heter. nuclear ribonucleoprotein C)
RTN4 (reticulon 4)
BCL6 (B-cell CLL/lymphoma 6)
S100B (S100 calcium binding protein, β (neural)
PTPN1 (tyr. phosphatase, non-receptor type 1)

WHOLE BRAIN

PAEP (progestagen-associated endometrial protein)
APOE (apolipoprotein E)
MUC1 (mucin 1, transmembrane)
HSPA5 (heat shock 70kDa protein 5)
SMAP (small acidic protein)
LOC220869 (dopamine responsive protein)
FRAT2 (freq. rearranged adv. T-cell lymphomas 2)
MTP (microsomal triglyceride transfer protein)
SNAI1 (snail homolog 1)
BTN1A1 (butyrophilin, subfamily 1, member A1)
ERP70 (protein disulfide isomerase related protein)
LRP2 (low density lipoprotein-related protein 2)
FRAT1 (freq. rearranged adv. T-cell lymphomas)
SGKL (serum/glucocorticoid regulated kinase-like)
LCAT (lecithin-cholesterol acyltransferase)
LIPC (lipase, hepatic)
RIPK4 (receptor-interac. serine-threonine kinase 4)
RBP4 (retinol binding protein 4, plasma)
ADAMTS1 (a disintegrin-like and metalloprotease)

LEFT HEMISPHERE

MUC1 (mucin 1, transmembrane)
YWHAG (tyrosine 3-monooxygenase/tryptophan)
PTPN1 (tyr. phosphatase, non-receptor type 1)
FRAT2 (freq. rearranged adv. T-cell lymphomas 2)
RUSC1 (RUN and SH3 domain containing 1)
RPS6KA6 (ribosomal protein S6 kinase, 90kDa)
SNAI1 (snail homolog 1)
FRAT1 (freq. rearranged adv. T-cell lymphomas)
SGKL (serum/glucocorticoid regulated kinase-like)
KLRC3 (killer cell lectin-like receptor)
RIPK4 (receptor-interac. serine-threonine kinase 4)
RBP4 (retinol binding protein 4, plasma)

RIGHT HEMISPHERE

BRD8 (bromodomain containing 8)
THRAP4 (thyroid horm. receptor assoc. prot. 4)
NCOA4 (nuclear receptor coactivator 4)
RNF8 (ring finger protein 8
ERBP (estrogen receptor binding protein)
PPARGC1A (peroxisome proliferative activ. recep.)
NR0B2 (nuclear rec. subfam. 0, group B, member 2)
EDF1 (endothelial differentiation-related factor 1)
GADD45G (growth arrest and DNA damage)
PPARBP (PPAR binding protein)
NCOA2 (nuclear receptor coactivator 2)
NRIP1 (nuclear receptor interacting protein 1)
FLJ22494 (hypothetical protein FLJ22494)
LOC220869 (dopamine responsive protein)
FRAT2 (freq. rearranged adv. T-cell lymphomas 2)
RARG (retinoic acid receptor, gamma)
SNAI1 (snail homolog 1)
BTN1A1 (butyrophilin, subfamily 1, member A1)
FRAT1 (freq. rearranged adv. T-cell lymphomas)
FABP1 (fatty acid binding protein 1, liver)
JMJD1C (jumonji domain containing 1C)
SGKL (serum/glucocorticoid regulated kinase-like)
NR4A2 (nuclear rec. subfam. 4, group A, member 2)
RBP4 (retinol binding protein 4, plasma)
TMPRSS3 (transmembrane protease, serine 3)
NFKB1 (nuclear factor of kappa light polypeptide)
NCOA1 (nuclear receptor coactivator 1)

FIGURE 4. List of protein binding candidates (protein-protein interaction) for all the biomarkers in each of the four signatures (with p-values < 0.01). The list was generated using the database GATHER, and among the most common ontologies for these proteins, we must cite *cell motility*, *regulation of neurogenesis*, *actin polymerization*, *lipid metabolism*, *regulation of axon extension* and *regulation of neuronal synaptic plasticity*.

ing, as Brown's biomarkers missed both. In terms of relevant biomarkers, we must give special emphasis to Pink1 (*pten-induced kinase 1*), found in the left hemisphere signature and that is a well-known PD gene that protects the neuron from mitochondrial oxidative stress [21]. Also worth mentioning is gene Gsk3b (*glycogen synthase kinase 3 beta*), present in the Alzheimer's pathway. Polymorphisms in Gsk3b were found to alter transcription and splicing and interact with *tao*-haplotypes to modify disease risk in Parkinson's [28].

Gene ontology analysis and protein-protein interaction

In this section we investigate the ontology annotations of the biomarkers present in the signatures and of the proteins most likely to bind with them. Using the biomarkers in each of the signatures – the one from Brown et al. (2002) and the three in Figure 2 – we determined the most common ontologies. Interestingly, there is a general absence of brain-related annotations; the majority is metabolism-related (protein, glycogen, sulphates, hormone, among others). With this result, we went one step further and using the GATHER protein binding database, we obtained the list of proteins most likely (p-value < 0.01) to bind with the genes in the four signatures (see Figure 4).

For the biomarkers in Brown et al. (2002), the majority of the binding proteins have brain-related ontology annotations such as *regulation of neurogenesis*, *actin polymerization*, *regulation of axon extension* and *regulation of neuronal synaptic plasticity*, whereas very few metabolism annotations are present. For the MIN (α,β)–FS signatures, in addition to brain-related ontologies, several other categories of interest are present. For instance, in the signature for the whole brain, metabolism (protein and lipid) and oxidative response ontologies are also present. For the left and right hemispheres, however, brain-related ontologies are almost non-existent, but there are plenty of metabolism (protein and phosphate), transcription and oxidative response annotations.

From these results, we can conclude that when looking at the whole brain at once, the proteins that better differentiate PD and normal brains are indeed related to brain functions, such as synapsis and neurogenesis. For the right and left hemispheres however, differentiation occurs via other mechanisms, related to metabolism, transcription and oxidative response. Supporting this conclusion, a recent study has put some light on asymmetries in PD, stating that metabolism and oxidative response variations between the two hemispheres of the brain could affect differently the dopaminergic neurons in the *substantia nigra*, resulting in asymmetric clinical effects [29]. However, the authors do not find an appropriate explanation for side preference in terms of symptoms and argue that the mechanisms behind side differentiation are still too complex to understand.

CONCLUSION

In this paper, we present three sets of biomarkers for Parkinson's disease (PD) using cDNA microarray data extracted from two C57BL/6J strain rodents. The first signature was obtained comparing the samples of the whole brain of the PD-affected against a control mouse. The second and third used the samples extracted from the left and right hemispheres of the brains, respectively. The problem of finding the signatures was modeled as a MIN (α,β)–FEATURE SELECTION problem. A pathway analysis conducted on the three signatures and on the biomarkers reported in Brown et al. (2002) indicate that the MIN (α,β)–FS approach retrieves biomarkers belonging to pathways more relevant to PD, including the Parkinson's and the Alzheimer's diseases pathways. Also, we made a study on the ontology of the biomarkers and of their most relevant binding proteins. The analysis shows that the biomarkers have mostly metabolism-related ontologies, whereas their binding proteins show more variation, with the presence not only of brain-related, but also metabolism, transcription and oxidative response ontologies.

REFERENCES

1. V. Brown, A. Ossadtchi, A. Khan, S. Yee, G. Lacan, W. P. Melega, S. Cherry, R. Leahy, and D. Smith, *Genome Research* **12**, 868–884 (2002).
2. C. Cotta, and P. Moscato, *Journal of Computer and Systems Science* **67**, 686–690 (2003).
3. R. Berretta, A. Mendes, and P. Moscato, *Journal of Research and Practice in Information Technology,* **to appear** (2007).
4. M. Polymeropoulos, and et al., *Science* **274**, 1197–1199 (1996).
5. C. Cotta, C. Sloper, and P. Moscato, "Evolutionary search of thresholds for robust feature selection: application to microarray data," in *Proc. of the 2nd European Workshop in Evolutionary Computation and Bioinformatics (EvoBIO-2004), Universidade de Coimbra, Portugal*, edited by G. Raidl, and *et al.*, Springer, 2004, vol. 3005 of *Lecture Notes in Computer Science*, pp. 31–40, ISBN 3-540-21378-3.
6. P. Moscato, R. Berretta, M. Hourani, A. Mendes, and C. Cotta, "Genes related with Alzheimer's disease: A Comparison of Evolutionary Search, Statistical and Integer Programming Approaches," in *Applications of Evolutionary Computing*, edited by F. Rothlauf, and *et al.*, Springer-Verlag, Berlin, 2005, vol. 3449 of *Lecture Notes in Computer Science*, pp. 84–94.
7. R. Berretta, A. Mendes, and P. Moscato, "Integer Programming Models and Algorithms for Molecular Classification of Cancer from Microarray Data," in *Proceedings of the 28th Australasian Computer Science Conference (ACSC 2005), Newcastle, Australia*, 2005.
8. S. Davies, and S. Russell, "*NP*-Completeness of Searches for Smallest Possible Feature Sets," in *AAAI Symposium on Intelligent Relevance*, edited by R. Greiner, and D. Subramanian, AAAI Press, New Orleans, 1994, pp. 41–43.
9. J. Chang, and J. Nevins, *Bioinformatics* **22**, 2926–2933 (2006).
10. B. Desai, A. Monahan, P. Carvey, and B. Hendey, *Cell Transplantation* **16**, 285–299 (2007).
11. B. Steiner, S. Wolf, and G. Kempermann, *Regenerative Medicine* **1**, 15–28 (2006).
12. D. Pellegrino, F. Cicchetti, X. Wang, A. Zhu, M. Yu, M. Saint-Pierre, and A. Brownell, *Journal of Nuclear Medicine,* **to appear** (2007).
13. A. Plaitakis, and P. Shashidharan, *Journal of Neurology* **247**, II25–35 (2000).
14. M. Bak, C. Hansen, N. Tommerup, and L. Larsen, *Pharmacogenomics* **4**, 411–429 (2003).
15. J. Burbach, and M. Smidt, *Trends in Neurosciences* **29**, 601–603 (2006).
16. A. Quesada, H. Romeo, and P. Micevych, *The Journal of Comparative Neurology* **503**, 198–208 (2007).
17. M. Okouchi, N. Okayama, J. Alexander, and T. Aw, *Current Neurovascular Research* **3**, 249–261 (2006).
18. L. White, M. Toft, S. Kvam, M. Farrer, and J. Aasly, *Journal of Neuroscience Research* **85**, 1288–1294 (2007).
19. T. Nagatsu, and M. Sawada, *Cellular and Molecular Neurobiology* **26**, 781–802 (2006).
20. B. Blanchard-Fillion, D. Prou, M. Polydoro, D. Spielberg, E. Tsika, Z. Wang, S. Hazen, M. Koval, S. Przedborski, and H. Ischiropoulos, *The Journal of Neuroscience* **26**, 6124–6130 (2006).
21. M. Dodson, and M. Guo, *Current Opinion in Neurobiology* **17**, 331–337 (2007).
22. M. Schwarzschild, L. Agnati, K. Fuxe, J. Chen, and M. Morelli, *Trends in Neurosciences* **29**, 647–654 (2006).
23. X. Wang, N. Krinsky, and R. Russell, *Journal of Nutrition* **123**, 1277–1285 (1993).
24. J. Mey, and P. McCaffery, *Neuroscientist* **10**, 409–421 (2004).
25. M. Heafield, S. Fearn, G. Steventon, R. Waring, A. Williams, and S. Sturman, *Neuroscience Letters* **110**, 216–220 (1990).
26. A. Williams, L. Cartwright, and D. Ramsden, *QJM: monthly journal of the Association of Physicians* **98**, 215–226 (2005).
27. S. Kalinin, D. Feinstein, H. Xu, G. Huesa, D. Pelligrino, and E. Galea, *The European Journal of Neuroscience* **24**, 3393–3400 (2006).
28. J. Kwok, M. Hallupp, C. Loy, D. Chan, J. Woo, G. Mellick, D. Buchanan, P. Silburn, G. Halliday, and P. Schofield, *Annals of Neurology* **58**, 829–839 (2005).
29. R. Djaldetti, I. Ziv, and E. Melamed, *The Lancet Neurology* **5**, 796–802 (2006).