

Merging Mouse Transcriptome Analyses with Parkinson's Disease Linkage Studies

Daniel GHERBASSI, Lavinia BHATT, Sandrine THURET[†] and Horst H. SIMON*

Department of Neuroanatomy, Interdisciplinary Center for Neuroscience, University of Heidelberg, Im Neuenheimer Feld 307, 69120 Heidelberg, Germany

(Received 5 September 2006; revised 26 March 2007; published online 23 May 2007)

Abstract

The hallmark of Parkinson's disease (PD OMIM #168600) is the degeneration of the nigral dopaminergic system affecting approximately 1% of the human population older than 65. In pursuit of genetic factors contributing to PD, linkage and association studies identified several susceptibility genes. The majority of these genes are expressed by the dopamine-producing neurons in the substantia nigra. We, therefore, propose expression by these neurons as a selection criterion, to narrow down, in a rational manner, the number of candidate genes in *orphan* PD loci, where no mutation has been associated thus far. We determined the corresponding human chromosome locations of 1435 murine cDNA fragments obtained from murine expression analyses of nigral dopaminergic neurons and combined these data with human linkage studies. These fragments represent 19 genes within *orphan* OMIM PD loci. We used the same approach for independent association studies and determined the genes in neighborhood to the peaks with the highest LOD score value. Our approach did not make any assumptions about disease mechanisms, but it, nevertheless, revealed *α-synuclein*, *NR4A2* (*Nurr1*), and the *tau* genes, which had previously been associated to PD. Furthermore, our transcriptome analysis identified several classes of candidate genes for PD mutations and may also provide insight into the molecular pathways active in nigral dopaminergic neurons.

Key words: dopaminergic neurons; substantia nigra; neurodegenerative disease; candidate genes

1. Introduction

The neuropathological hallmark of Parkinson's disease (PD) is the progressive degeneration of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNpc), affecting about 1–2% of the human population older than 65 years.¹ It is characterized by the clinical symptoms of resting tremor, muscular rigidity, postural instability, a positive response to the administration of L-DOPA, and the presence of cytoplasmic inclusions

in postmortem brains, Lewy Bodies.² Despite its mostly sporadic onset and a high discordance rate in monozygotic twins,³ several human linkage studies had been initiated to determine susceptibility genes for this disease.⁴ In the Online Mendelian Inheritance in Man (OMIM) database, 13 PD loci have been recorded: PARK1,⁵ PARK2,^{6–9} PARK3,¹⁰ PARK4,^{11,12} PARK5,¹³ PARK6,^{14,15} PARK7,^{16,17} PARK8,¹⁸ PARK9,^{19,20} PARK10,²¹ PARK11,^{22,23} PARK12,^{23,24} and PARK13.²⁵ Furthermore, genome-wide analyses of multiplex PD families provided evidence for linkage to regions on different chromosomes.^{21,22,24,26–29} The PARK loci are sometimes larger than 10 Mb and can contain hundreds of genes. In case of the genome-wide linkage studies for a complex, multifactorial disease such as PD, the regions with high LOD scores are rarely smaller than 20 cM.²⁹ The differences among independent studies and the size of the suggested

Communicated by Shoji Tsuji

* To whom correspondence should be addressed. Tel. +49-6221-548342. Fax. +49-6221-545605. E-mail: horst.simon@urz.uni-heidelberg.de

[†] Present address: King's College London, Centre for the Cellular Basis of Behaviour, MRC Centre for Neurodegeneration Research, Institute of Psychiatry, P039, 1-2 WW Ground, Denmark Hill, London SE5 8AF, UK.

© The Author 2007. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

susceptibility regions make the searches for the underlying mutations irremediably a time-consuming process.

For several PARK loci, the searches have been successful. Mutations in *α-synuclein* (PARK1 and PARK4), *DJ-1* (PARK7), *parkin* (PARK2), *PINK1* (*PTEN-induced putative kinase*) (PARK6), *LRRK2* (*leucine-rich repeat kinase 2*) (PARK8), *UCHL1* (*ubiquitin carboxy-terminal-hydrolase-L1*) (PARK5), and *ATP13A2* (*ATPase type 13A2*) (PARK9) have been identified.^{5,30–37}

Other studies have revealed the cytoskeletal protein *tau* (*MAPT*)^{36,38} and the ligand-independent nuclear receptor *NR4A2*^{30,39,40} (*Nurr1*) as susceptibility genes. Although the definite role in PD of many of these genes is still discussed and controversial (especially for *NR4A2* and *UCHL1*) and the known mutations account for less than 10% of all PD cases, the investigation into the functions of the underlying genes has generated an insight into the fundamental disease pathogenesis. For example, *α-synuclein* and *parkin* turned out to be major protein components of Lewy bodies in sporadic PD.⁴¹ Mutations in *parkin*, *UCHL1*, and *DJ-1* suggest that abnormal protein folding and protein degradation through the ubiquitin-proteasome system is an important factor in the etiology of the disease.^{42,43} *PINK1* may be involved in the phosphorylation of mitochondrial proteins in response to cellular stress, thus protecting against mitochondrial dysfunction.³⁵ Interestingly, mitochondria are also the site, where the known neurotoxins for DA neurons operate, suggesting that their malfunctioning could be a major contributor to PD pathogenesis.⁴⁴

Current or future searches for the underlying mutations in the remaining *orphan* Parkinson loci could be accelerated and widened to promoter regions and to haplotype variations, if the number of candidate genes is narrowed down by other criteria. At least seven out of the nine PD-associated genes are expressed by nigral DA neurons,^{45–50} with different expression levels and specificity. These are *α-synuclein*, *NR4A2*, *parkin*,⁴⁶ *PINK1*, *tau*, *UCHL1*, and *LRRK1* (<http://www.brain-map.org>). For this reason, we propose expression (specific or non-specific) by mesDA neurons as a selection criterion to identify candidate genes in those PD loci where the underlying gene is still unknown (*orphan*). Such an approach does not make any presumption with respect to disease mechanisms. Conceptually, the same method was applied on five large PD loci using serial analysis of gene expression for a comparative expression analysis of SNpc and adjacent mesencephalon in postmortem brains.⁵¹ As cell-specific expression in mouse and human is very similar, we took three murine expression studies which employed fluorescent-activated cell sorting (FACS) and two unrelated subtractive methods for the identification of genes expressed by mesDA neurons.^{52–54} We collected the cDNA sequences of these expression analyses from public databases, determined the underlying genes and the corresponding gene ontology annotations

[Gene Ontology (GO)] to obtain insight into their function. Then, we established their genetic locations and their syntenic positions on the human genome. Finally, we combined these data with existing human PD linkage studies.^{5–11,13–24,26–29,55,56}

2. Material and methods

2.1. Transcriptome analysis

All nucleotide sequences used in this study are publicly available at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide> and derived from three expression analyses in mouse: (i) Barrett *et al.*⁵² published 779 sequences (Accession Nos.: BE824469–BE824504, BE824506–BE824519, BE824521–BE824561, BE824563–BE824823, BE824825–BE825045, BE825047–BE825132, CK338036–CK338155). (ii) Stewart *et al.*^{53,57,58} published 496 cDNA sequences (Accession Nos.: AA008736, W33210–W33212, W33214–W33289, W35421–W35480, W36130–W36269, W39787–W40005, W40007–W40008, W40010–W40023, W45732). (iii) We published 160 sequences (Accession Nos.: CO436137–CO436293).⁵⁴

Each nucleotide sequence was employed for a nucleotide-nucleotide BLAST (blastn) (basic local alignment search tool) on the nr database (non-redundant) (<http://www.ncbi.nlm.nih.gov/BLAST/>) and on the mouse genome (<http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html>). We then recorded those alignments with the highest scores, the lowest e-values, and highest number of hits in a single locus. BLAST results were categorized into four groups: (1) no significant alignments on mouse genome (None), (2) significant alignments with mitochondrial DNA (Mitochondrial Genes), (3) multiple high-scoring alignments on mouse genome (Multiple Hits) for ambiguous results, and (4) significant alignments on mouse genome for single hits or otherwise unambiguous results (Table 1). The latter group was further subdivided into: ‘Genes’, ‘ESTs’, and ‘genomic Sequences’. The group ‘Genes’ comprises the results with high-scoring alignments in exons of single genes. In some cases, where the alignment lay in the region after the last exon or, according to the chromosome map view, in an intron of a given gene, we termed it also ‘Gene’, if the hit was in a UniGene cluster which was linked to the gene in the locus. With those alignments that we were unable to associate to a gene, we performed a blastn on the MmEST database. If we could associate the sequence to a previously described EST, we termed it ‘EST’; otherwise, it was termed ‘Genomic Sequence’.

For all the ‘Annotated Genes’, ‘Hypothetical Genes’, and mitochondrial genes, the following data were collected from the locus link feature (<http://www.ncbi.nlm.nih.gov/LocusLink> this was replaced by <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene> during the

Table 1. BLAST results on mouse genome

No significant alignments on mouse genome	262	None
Significant alignments with mitochondrial genes	104	Mitochondrial genes
Multiple high-scoring alignments on mouse genome	19	Multiple hits
Significant alignments on mouse genome	1050	Genes (940)
		Annotated genes (793)
		Hypothetical genes (147)
		ESTs (47)
		Genomic sequences (63)

cDNA sequences are separated into four different categories based on the types of alignments generated. Alignments on the mouse genome were subdivided into Genes, ESTs, and genomic sequences. For the category ‘Genes’, we differentiated further between ‘annotated’ and ‘hypothetical’ depending on the gene RefSeq status recorded at NCBI.⁷²

course of this study): the gene name, gene symbol, accession number, Gene ID, and the MGI link number, if available. The latter provides a relational link to the GO library and the information related to ‘biological processes’, ‘cellular components’, and ‘molecular functions’.

For all cDNA sequences categorized by ‘Significant Alignments on Mouse Genome’, we also registered the exact chromosomal position in kilobases (starting from the top of the short arm).

2.2. Mapping the murine cDNA sequences to the human genome

For most of the murine genes, a human homolog has already been determined, normally carrying the same name and symbol. This information is registered on the Entrez Gene page together with the cytogenetic locations. When this information did not exist, we used the mouse protein sequence of the identified gene for a translated BLAST (tblastn), or the nucleotide sequence of the cDNA fragment or the GenBank accession number of the corresponding gene for a blastn on the human genome. We registered the position in kilobases on the chromosome and verified each position on the human genome by comparing the neighboring genes to those in the mouse genome and recorded the human position only if the neighboring genes also matched.

When the cytogenetic position on the human genome was determined, we compared this information with the positions of the recorded PARK loci. We aligned the human chromosome map view with the map for ‘morbid/disease’, described in OMIM (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>). When the genes, or the estimated human locations, and the cytogenetic disease locations co-localized, we called the gene a PD

candidate gene. For the loci suggested by genome-wide studies, we selected those genes, which were situated ± 3 Mb from the chromosome marker (single nucleotide polymorphism (SNP)) with the highest LOD score (Table. 5). We are aware that this approach reduces the numbers of genes in an arbitrary manner. However, if preferred, the range can be widened with the provided data (see Supplementary Data) in order to more accurately consider asymmetry or size of each specific linkage peak.

The entire data set was collected and processed using the database program, Filemaker Pro 7.0. The latest update was in February 2007. This database is available upon request.

3. Results

We obtained 1435 sequences from three independent studies, which had the original aim to identify genes expressed by mesDA neurons. Barrett *et al.*⁵² had isolated DA neurons from E13 ventral midbrain by FACS. This library contains genes expressed by mesDA neurons with a preference for abundant genes. The other two studies used subtractive methods to enrich for rare RNA transcripts expressed by mesDA neurons. Stewart *et al.*^{53,57,58} had created a single-stranded directional cDNA library from substantia nigra of 8-week-old mice subtracted with a cDNA library from cerebellum. We had used a PCR-based differential display method⁵⁴ employing cDNA from engrailed-1/2 double-mutant and wild-type ventral midbrain during the embryonic stages when mesDA neurons disappear in the mutants.^{59,60} The amplified sequences were compared to the expression profile of adult olfactory bulb, a source of DA neurons unrelated to those in the ventral midbrain. Only differentially expressed cDNA fragments were isolated and sequenced. As the original sequence analyses of the former two studies had been performed when a smaller nucleotide data set was available and in order to update our own expression analysis, we subjected the sequence data from all three screens to new BLAST searches and determined their association to genes and published ESTs, and their location on the mouse genome. The 1435 cDNA fragments generated 1050 unambiguous murine genomic hits, 19 ambiguous multiple hits, and 104 alignments with mitochondrial DNA. Two hundred and sixty-two cDNA sequences produced no significant alignments (see Table 1 for definitions and the entire analysis, and Table 2 for the individual libraries).

Out of 1050 cDNA fragments, which generated unambiguous alignments on the mouse genome, 1020 were in gene loci. Most of them aligned to exons of those genes (72.6%; 741 of 1020). Out these 1020 cDNA fragments, 181 (17.8%) lay 3’ to the last annotated exon, suggesting that substantial amounts of mRNAs isolated from brain tissue are longer at their 3’ end than mRNAs from other

Table 2. Classification of BLAST results from each library

	Total analysis	Barrett ⁵²	Stewart ⁵³	Thuret ⁵⁴
A. Number of unique alignments per individual library				
Genes ^a	423	150	218	77
Hyp. genes ^b	80	23	39	19
ESTs ^c	32	16	12	3
Genomic	44	15	8	21
Mitochondria	11	8	2	2
Multiple hits ^d	14	6	4	4
None ^e	185	67	111	8
Total ^f	789	285	394	134
B. Total number of fragments ^g				
Genes ^a	793	403	293	97
Hyp. genes ^b	147	71	55	21
ESTs ^c	46	28	15	5
Genomic	62	30	11	21
Mitochondria	104	100	2	2
Multiple hits ^d	19	9	4	6
None ^e	262	138	116	8
Total	1435	779	496	160

^aAnnotated mouse genes.

^bHypothetical genes determined by EST clustering or predicted by automated computational genome analysis with a large open reading frame.

^cExpressed sequencing tags.

^dUnderlying gene not identifiable, due to multiple alignments with low e-values.

^eNo hit in mouse and human genome.

^fNumber of unique alignments. Five hundred and seventy-nine unique tags were on the mouse genome (excluding mitochondria).

^gNumber of fragments that represent genes, hypothetical genes, ESTs, genomic sequences, multiple alignments, and mitochondrial genes, listed per individual library.

tissues (Table 3). Finally, 9.6% (98 of 1020) of the alignments lay in regions designated as introns, suggesting that they are parts of unrecorded splice variants, possibly specific for mesDA neurons.

The 1050 cDNA fragments represented 503 genes (423 annotated and 80 hypothetical genes), 32 ESTs, and 44 unique genomic hits with no otherwise described ESTs. Additionally, the 104 sequences that aligned to the mitochondrial DNA represented 11 mitochondrial genes (Table 2). To these cDNA sequences, we associated the corresponding MGI numbers, if available. This provided us with insight into their molecular function, the cellular locations of the proteins, and the associated biological process (see Supplementary Data for the entire transcriptome analysis). Several protein classes were over-represented, like, for example, those, which take part in mitochondria-related processes, in fatty acid chain metabolism, in ubiquitination, in the MAPK signaling pathways,

Table 3. Alignments in relation to gene loci

		Total	Genomic sequences	ESTs	Genes
In gene loci	Only in last exon	471			471
	In last and other exon(s)	132			132 ^a
	Not in last exon	138			138
	After 3' end	181	6	16	159
	Intron	98	40	10	48
	Subtotal	1020	46	26	948
Outside gene loci		30	17	13	
Total		1050	63	39	948

Genomic alignments were divided into three groups: 'ESTs' (3.7%), 'genomic sequences' (6.0%), and 'genes' (90.3%). Majority of the cDNA fragments that aligned with genes are aligned with the last exon. A significant number of the cDNAs aligned with the region 3' to the last exon. See Material and Methods for details.

^aForty-four hits are in genes with only one exon.

or which are chaperones. Some of these molecular pathways were previously linked to the death of mesDA neurons, to PD, and other human neurodegenerative disorders.

The majority of the mutations, which are associated to PD, is in genes that are expressed in mesDA neurons. We, therefore, joined these expression analyses with human PD linkage and association studies,^{5-11,13-24,26-29,55,56} where no mutation has been associated thus far. For each unique mouse cDNA sequencing tag, we determined its human homolog and the corresponding cytogenetic and physical positions on the human chromosomes. We verified each locus on the human genome by identifying the neighboring genes on the mouse genome and recorded the human position only if the adjacent genes were the same. We then determined whether these positions were within OMIM (Table 4) and other suggestive (non-OMIM) PD loci (Table 5). In case of the OMIM *orphan PD* loci, we projected on the human chromosome view the map for 'morbid diseases'. In case of non-OMIM loci, we identified the genes ± 3 Mb to the SNP marker with the highest LOD score. Totally, we linked the mouse transcriptome analyses to 569 unique locations on the human genome. Nineteen of these are within *orphan* PARK loci (Table 6) and 51 in non-OMIM PD loci (Table 7).

The experimental design of the three different transcriptome analyses, we used for our study, were such that they included both highly and rarely expressed transcripts. Our analysis confirmed the complementary nature of the three screens. Only 7.2% (104 out of 1435) of the cDNA sequences of these libraries represent genes, hypothetical genes, or EST clusters, which are found in

Table 4. PARK loci

Locus	OMIM identifier	Gene	Cytogenetic location	From (kb)	To (kb)	Mb	Number of genes
PARK1	163890	SNCA	4q21.1-4q21.3				
PARK2	602544	Parkin	6q25.3-6q26				
PARK3	602404		2p13.3-2p13.1	68.075	75.307	7.2	106
PARK4	605543		4p15.33-4p15.1	13.424	37.324	23.9	60
PARK5	191342	UCHL1	4p14				
PARK6	605909	PINK1	1p36.33-1p35.1				
PARK7	602533	DJ1	1p36.23-1p36.22				
PARK8	607060		12q11.2-12q13.13	27.908	55.637	27.7	351
PARK9	606693	ATP13A2	1p36.33-1p36.11				
PARK10	606852		1p33-1p32.2	47.651	55.380	7.7	76
PARK11	607688		2q36.1-2q37.3	219.844	243.416	23.6	216
PARK12	300557		Xq21-q25	75.950	129.900	40.0	356
PARK13	610297		2p13.1-2p11.2	75.450	84.130	8.7	39
	601828	NR4A2	2q22.1-2q23.3				
	603779	SNCAIP	5q23.1-q23.3				
	260540	MAPT	17q21.1				

Genomic location of PARK loci as recorded in the OMIM databank. For seven of the PARK loci, the mutated genes were identified. The number of genes is the current GenBank estimation of all annotated and predicted genes in the corresponding PARK locus. For the PARK10 locus, we used the narrow definition 1p33-1p32.2 as determined by the two genetic markers D1S2134 and D1S200, and not the entire shorter arm of chromosome 1 (1p) which contains 1232 genes.²¹

Table 5. Association studies not recorded at OMIM

	Cytogenetic location	Genetic marker	Mb	CM Marshfield	LOD score
Bertoli-Avella (03) ²⁷	19p13.13	D19S221	12.6	36	2.26
	19p13.13	D19S840	13.7	38	
DeStefano (01) ²⁸	9q34.11	D9S1825	123.3	136	1.3
	10q22.1	GATA121A08	70.2	88	1.07
DeStefano (02) ⁵⁶	9q32	D9S930	110.6	120	1.86
	20q11.2	D20S478	37.9	54	1.82
	21q21	D21S2052	27.7	24	2.21
Hicks (02) ⁵⁵	5q23.3	D5S666	120–137	135	1.6
Li (02) ²¹	10q25.3	D10S1237	116.1	134	2.62
	6p21.1	D6S1017	41.7	63	1.88
	5q15	D5S1462	96.4	105	
	5q21.1	Peak	100	108	1.65
	5q21.3	D5S1453	105–109	115	
Martinez (04) ²⁹	17p13.1	D17S1303	10.8	24	1.93
	2p12–q22	D2S2216	88	111	1.24
	2p11–q12	Peak	102	117	2.04
	2q12	D2S160	107	123	1.77
	5q23	D5S471	117.5	130	1.05
	6p12	D6S257	56	80	1.37
	6q11–q13	Peak	69–73	85	1.41
	6q14	D6S460	~82	90	1.14
	7p22	D7S531	3	5	1.51

Continued

Table 5. continued

	Cytogenetic location	Genetic marker	Mb	CM Marshfield	LOD score
Pankratz (03) ²⁴	11q14	D11S4175	89.9	91	1.6
	19q13.3	D19S902	53.6	73	1.05
	Xq22.3	DXS8055	113.4	71	3.1
	10q11.2	D10S196	51.5	70.0	2.3
Scott (01) ²⁶	5q31.1	D5S816	135.4	139	2.39
	17p11.2	D17S921	14.5	36	1.92
Two-point and multipoint LOD	17q11.2	D17S1293	32.7	56	2.28
	17p11.2	D17S921	14.5	36	2.02
	17q11.2	D17S1293	32.7	56	2.62
Multipoint LOD	9q33.1	-10 cM	117.8	130	
	9q33.3	D9S301 66 cM	126.3	140	2.59
	9q34.2	+10 cM	132.3	150	
	3q13.32	D3S2460	118.7	135	1.62

For each individual study, the highest LOD scores with the associated genetic markers are listed. In these studies, the peak positions and the flanking genetic markers were given in centiMorgan on the Marshfield genetic map. We determined, when possible, the exact position in Mb on the corresponding chromosome. The average distance between the two adjacent genetic markers in each study varied between 5 and 11 cM.

Table 6. Candidate genes in Orphan PARK loci

No. of cDNA fragments aligning with the gene	Mouse ID	Human ID	Symbol	Human gene name	Position	Locus
1	NM_146169	XM_376062	KIAA1155	<i>KIAA1155 protein</i>	2p13.3	Park3
1	NM_008717	NM_014497	ZFML	<i>Zinc finger, matrin-like</i>	2p13.2-p13.1	Park3
1	NM_183138	XM_371501	MGC22014	<i>cDNA sequence BC037432</i>	2p13.1	Park3
3	NM_080555	NM_003713	PPAP2B	<i>Phosphatidic acid phosphatase type 2B</i>	1p32	Park10
1	AA819910	Estimated	FAF1	<i>In locus of Fas-associated factor 1</i>	1p33	Park10
6	NM_009129	NM_003469	SCG2	<i>Secretogranin II</i>	2q35-q36	PARK11
3	AK052241	NM_005544	IRS1	<i>Insulin receptor substrate 1</i>	2q36	PARK11
1	NM_152915	NM_139072	DNER	<i>Delta/notch-like EGF-related receptor</i>	2q37.1	PARK11
1	NM_008440	NM_004321	KIF1A	<i>Kinesin family member 1A</i>	2q37.3	PARK11
2	NM_024197	NM_004544	NDUFA10	<i>NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10</i>	2q37.3	PARK11
1	NM_025437	NM_001412	EIF1AX	<i>Eukaryotic translation initiation factor 1A, X-linked</i>	Xp22.13	PARK12
2	NM_019768	NM_012286	MORF4L2	<i>Mortality factor 4 like 2</i>	Xq22	PARK12
3	NM_011123	NM_000533	PLP1	<i>Proteolipid protein 1</i>	Xq22	PARK12
3	NM_013898	NM_004085	TIMM8A	<i>Translocase of inner mitochondrial membrane 8 homolog a</i>	Xq22.1	PARK12
3	NM_016783	NM_006667	PGRMC1	<i>Progesterone receptor membrane component 1</i>	Xq22-q24	PARK12
7	NM_030688	estimated	IL1RAPL2	<i>After 3' of interleukin 1 receptor accessory protein-like 2</i>	Xq22.2-q22.3	PARK12
1	NM_133196	NM_001325	CSTF2	<i>Cleavage stimulation factor, 3' pre-RNA, subunit 2</i>	Xq22.1	PARK12
1	NM_025893	NM_173798	ZCCHC12	<i>Zinc finger, CCHC domain containing 12</i>	Xq24	PARK12
2	NM_172782	NM_018698	NXT2	<i>Nuclear transport factor 2-like export factor 2</i>	Xq23	PARK12

Table 7. Candidate genes for non-OMIM PARK loci

GenBank ID	Human ID	Symbol	Human location	In kb ^b	Gene name	
Scott (01) D3S2460 ²⁶						
5	NM_008083	NM_002045	GAP43	3q13.1–13.2	116700	<i>Growth-associated protein 43</i>
3	BB626331	EST	Lsmp	3q13.2–q21	117200	<i>Limbic system-associated membrane protein</i>
2	NM_177093	XM_057296	LRRC58	3q13.33	121300	<i>Leucine-rich repeat containing 58</i>
2	NM_008047	NM_007085	FST1	3q13.32–q13.3	121460	<i>follistatin-like 1</i>
Martinez (04) D5S471 ²⁹						
1	XM_283496	NM_005509	DMXL1	5q22	118600	<i>Dmx-like 1</i>
1	Genomic	Estimated	FEM1C	5q22	114939	<i>fem-1 homolog c</i>
3	NM_152809	NM_004384	CSNK1G3	5q23	123000	<i>Casein kinase 1, gamma 3</i>
Li (02) D5S1462 D5S1453 ²¹						
1	NM_172827	EST	LNPEP	5q15	96440	<i>Leucyl/cystinyl aminopeptidase</i>
Hicks (02). Scott (01) D5S666. D5S816 ^{26,55}						
1	NM_173753	NM_001008738	FNIP1	5q31.1	131060	<i>Folliculin interacting protein 1</i>
1	NM_144823	NM_015256	ACSL6	5q31	131400	<i>Acyl-CoA synthetase long-chain family member</i>
1	NM_033144	XM_034872	SEPT8	5q31	132180	<i>Septin 8</i>
1	AK011363	NM_003337	UBE2B	5q23–q31	133800	<i>Ubiquitin-conjugating enzyme E2B, RAD6 homology</i>
Scott (01) D5S816 ²⁶						
1	NM_029518	NM_016604	JMJD1B	5q31	137810	<i>Jumonji domain containing 1B</i>
3	NM_010771	NM_018834	MATR3	5q31.3	138730	<i>Matrin 3</i>
Li (02) D6S1017 ²¹						
1	NM_025365	NM_013397	C6ORF49	6p21.31	41800	<i>Chromosome 6 open reading frame 49</i>
1	NM_020493	NM_003131	SRF	6p21.1	43200	<i>Serum response factor (c-fos serum response element-binding transcription factor)</i>
5	NM_008302	NM_007355	HSP90AB1	6p12	44300	<i>Heat shock protein 90 kDa alpha (cytosolic), class B member 1</i>
Martinez (04) D6S257 D6S460 ²⁹						
1	Genomic	Estimated		6q12–q13	72500	
8	NM_010106	NM_001402	EEF1A1	6q14.1	74224	<i>Eukaryotic translation elongation factor 1 alpha 1</i>
Martinez (04) D7S531 ²⁹						
1	NM_028469	NM_032350	MGC11257	7p22.3	850	<i>Hypothetical protein MGC11257</i>
1	NM_010302	NM_007353	GNA12	7p22–p21	2510	<i>Guanine nucleotide binding protein (G protein) alpha 12</i>
6	NM_007393	NM_001101	ACTB	7p15–p12	5300	<i>Actin beta</i>
1	NM_026050	NM_032706	MGC12966	7p22.2	6110	<i>Hypothetical protein MGC12966</i>
1	NM_009007	NM_006908	RAC1	7p22	6170	<i>ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac)</i>
DeStefano (01) D9S1825, ⁵⁶ Scott (01) D9S301 ²⁶						
7	NM_026434	NM_033117	RBM18	9q34.11	120400	<i>RNA binding motif protein 18</i>
2	NM_022310	NM_005347	HSPA5	9q33–q34.1	123370	<i>Heat shock 70 kD protein 5</i>
1	NM_025709	NM_015635	GAPVD1	9q34.11	123450	<i>GTPase-activating protein and VPS9 domains 1</i>
1	NM_172661	XM_497080	KIAA0515	9q34.1	129650	<i>KIAA0515 gene</i>
DeStefano (01) GATA121A08 ⁵⁶						
1	NM_183295	NM_015634	KIAA1279	10q22.1	70100	<i>KIAA1279 gene</i>
Martinez (04) D11S4175 ²⁹						
1	NM_025844	NM_012124	CHORDC1	11q14.3	89650	<i>Cysteine and histidine-rich domain (CHORD)-containing, zinc-binding protein 1</i>

Continued

Table 7. continued

GenBank ID	Human ID	Symbol	Human location	In kb ^b	Gene name
Li (02) D10S1239 ²¹					
1	NM_172523	NM_003054	VMAT2	10q25	118680 Solute carrier family 18
Li (02) D17S1303 ²¹					
1	NM_018768	NM_004853	STX8	17p12	9350 Syntaxin 8
Scott (01) D17S921, D17S1293 ²⁶					
1	NM_011664	NM_018955	UBB	17p12-p11.2	16470 Ubiquitin B
1	NM_011480	NM_004176	SREBF1	17p11.2	17950 Sterol regulatory element binding factor 1
1	XM_110937	NM_145809	USP32	17p11.2	18621 Ubiquitin-specific protease 32
1	NM_026389	NM_015584	POLDIP2	17q11.2	26800 Polymerase delta interacting protein 2
1	NM_174852	NM_020889	PHF12	17q11.1	27400 PHD finger protein 12
1	NM_010897	NM_000267	NF1	17q11.2	29700 Neurofibromatosis 1
1	NM_010161	NM_014210	EVI2A	17q11.2	29800 Ecotropic viral integration site 2A
1	NM_010716	NM_002311	LIG3	17q11.2-q12	33450 Ligase III, DNA, ATP-dependent
Bertoli-Avella (03) D19S221 ²⁷					
2	NM_008319	NM_003259	ICAM5	19p13.2	10260 Intercellular adhesion molecule 5, telencephalin
16	NM_016742	NM_007065	CDC37	19p13.2	10370 Cell division cycle 37 homolog (<i>S. cerevisiae</i>)-like
1	NM_145624	NM_016264	ZNF44	19p13.2	12200 Zinc finger protein 44
1	NM_010906	NM_002501	NFIX	19p13.3	13030 Nuclear factor I/X
1	NM_183097	Estimated		19p13.13	14060 Progesterin and adipoQ receptor family member
DeStefano (02) D20S478 ⁵⁶					
1	BQ927659	Estimated		20q11.2-q12	35330
1	NM_013865	NM_022477	NDRG3	20q11.21-q11.23	36000 <i>n-myc</i> downstream regulated 3
1	NM_010658	NM_005461	MAFB	20q11.2-q13.1	40000 <i>v-maf</i> musculoaponeurotic fibrosarcoma oncogene family. protein B
2	NM_021464	NM_007050	PTPRT	20q12-q13	40500 Protein tyrosine phosphatase. receptor type T
DeStefano (02) D21S2052 ⁵⁶					
2	NM_11782	Estimated	ADAMTS5	21q21.2	27170 A disintegrin-like and metalloprotease (repolysin-type) with thrombospondin type 1 motif 5 (aggrecanase-2') 3'
Pankratz (03) DXS8055 ²⁴					
1	NM_016783	NM_006667	PGRMC1	Xq24	116713 Progesterone receptor membrane component 1

Listed genes are situated ± 3 Mb to peak with the highest LOD score, except for D10S196 where we used ± 8 Mb.

^aNumber of cDNA fragments aligning with the gene.

^bkb from the top of the short arm of the chromosome.

^cHuman chromosome location was estimated by comparing the flanking regions of mouse and man.

more than one of them (Table 8). Moreover, the libraries also contained two cDNA fragments for α -synuclein, three for *NR4A2*, and one for the *tau* genes. Mutations in all three genes have been previously associated to PD.^{5,30,36} Assuming that all 30 000 genes in the human genome⁶¹ were equally likely detected, the probability to identify three of nine PD susceptibility genes by chance out of a pool of 569 was less than 3.4×10^{-3} . If we exclude the controversial *NR4A2* and *UCHL1*, the probability was less than 1.5×10^{-2} .

4. Discussion

The entire human and mouse genome sequences have been available for more than 3 years.^{61,62} Therefore, the chromosomal locations of most genes have been determined and as a consequence also those genes within a given disease locus. In order to identify potential PD susceptibility genes, we projected the sequence data of three murine transcriptome studies for mesDA neurons onto the human genome and compared them with previously

Table 8. cDNA library comparison

	Barrett ⁵²	Stewart ⁵³	Thuret ⁵⁴
Barrett		45 (22)	11 (2)
Stewart	35 (22)		5 (4)
Thuret	3 (2)	5 (4)	

Of 1435, 104 (7.2%) cDNA fragments overlap with sequences also present in one other library. This number includes not only fragments that align with each other, but also those which align with the same annotated gene, hypothetical gene, mitochondrial gene, EST, or genomic position. These overlapping 104 cDNA fragments represent 28 of 781 (3.6%) unique tags (Table 2).

identified PD loci. We determine the human homologs of 1435 murine cDNA fragments which corresponded to 579 unique *mouse* chromosomal locations; 423 annotated genes, 80 hypothetical genes, 32 ESTs, and 44 genomic locations, which are not linked to any genes or otherwise reported cDNA sequences. Of the 569 unique locations on the human genome, 19 were positioned in OMIM PARK loci and 51 within genomic regions that have a weaker linkage to PD, which are not recorded in the OMIM database and need further confirmation.

Multiple studies are on the way to determine the underlying mutations of *orphan* PARK loci⁶³; however, the length of putative regulatory regions of most gene, their unpredictable position, and the common presence of SNPs have thus far restricted such studies to nucleotide variation in the coding region and in 5' and 3' UTR. Disparities in the promoter–enhancer–silencer regions were only the aim if the targeted gene had been previously linked to PD.^{64,65} A nucleotide variation in the *α-synuclein* promoter, for example, was associated to the disease.^{12,66} Variability on the level of gene expression is far more common than nucleotide variations which alter protein sequences⁶⁷ and it is believed that these haplotype variations determine individual traits and predispositions for common diseases such as PD. Narrowing down the number of candidate genes in identified loci in a rational manner may encourage the inclusion of the promoter regions in future studies aiming to identify mutations associated to PD.

Among the candidate genes that we found, the most interesting is *VMAT2* (*vesicular monoamine transporter 2*) (10q25). Reduced expression of *VMAT2* could be correlated with a higher sensitivity to environmental factors. For example, *VMAT2* heterozygote mice (+/–) are remarkably more sensitive than wild-type to the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine.^{68,69} Furthermore, we identified two genes in the ubiquitination pathway, *Ube2b* [*ubiquitin-conjugating enzyme E2B, RAD6 homology (S. cerevisiae)*] and *Ubb* (*Ubiquitin B, member of the HSP90 family*) and *Hspa5* (heat shock 70 kDa protein 5, member of the *HSP70* family).

Finally, 26 mitochondrial genes encoded by nuclear DNA are present in our transcriptome analysis. Of these, an unexpected high proportion of genes, namely four, are located within *orphan* OMIM PARK loci. There is increasing evidence that impairment of mitochondrial functions and oxidative stress are contributing factors to PD⁷⁰ supported by the recent finding of a mutation in *PINK1*.³⁵ Furthermore, the functional deficiencies induced by several of the other PD mutations seem to converge onto the mitochondria.⁷¹ Our finding confirms a central role of the mitochondria in PD and suggests the possibility that a misregulation of some of these four mitochondrial genes may be a contributing factor for the disease.

We conclude that our transcriptome analysis, along with being applicable for the identification of PD candidate genes, may also be a useful tool for future genome-wide association studies with newer resources, such as HapMap (<http://www.hapmap.org/>), where tagSNPs can be chosen close to loci of genes expressed by mesDA neurons. Furthermore, new GO annotations are constantly added and with time it may turn out that many of the identified genes are part of shared metabolic pathways. Our data set may give new insight into ligand/receptor interactions and/or intracellular signaling pathways acting in mesDA neurons, allowing novel studies into the molecular etiology of PD.

Acknowledgements: This work was supported by a grant from the German Federal Secretary for Education and Research (BMBF) Biofuture 98.

Supplementary Data: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

References

- Olanow, C. W. and Tatton, W. G. 1999, Etiology and pathogenesis of Parkinson's disease, *Annu. Rev. Neurosci.*, **22**, 123–144.
- Spillantini, M. G., Schmidt, M. L., Lee, V. M., et al. 1997, Alpha-synuclein in Lewy bodies, *Nature*, **388**, 839–840.
- Nussbaum, R. L. and Polymeropoulos, M. H. 1997, Genetics of Parkinson's disease, *Hum. Mol. Genet.*, **6**, 1687–1691.
- Gasser, T. 2001, Genetics of Parkinson's disease, *J. Neurol.*, **248**, 833–840.
- Polymeropoulos, M. H., Lavedan, C., Leroy, E., et al. 1997, Mutation in the *alpha-synuclein* gene identified in families with Parkinson's disease, *Science*, **276**, 2045–2047.
- Matsumine, H., Saito, M., Shimoda-Matsubayashi, S., et al. 1997, Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27, *Am. J. Hum. Genet.*, **60**, 588–596.
- Tassin, J., Durr, A., de Broucker, T., et al. 1998, Chromosome 6-linked autosomal recessive early-onset Parkinsonism: linkage in European and Algerian families, extension of the clinical spectrum, and evidence of a small homozygous deletion in one family. The French Parkinson's Disease Genetics Study Group, and the

- European Consortium on Genetic Susceptibility in Parkinson's Disease, *Am. J. Hum. Genet.*, **63**, 88–94.
8. Jones, A. C., Yamamura, Y., Almasy, L., et al. 1998, Autosomal recessive juvenile parkinsonism maps to 6q25.2-q27 in four ethnic groups: detailed genetic mapping of the linked region, *Am. J. Hum. Genet.*, **63**, 80–87.
 9. Saito, M., Matsumine, H., Tanaka, H., et al. 1998, Refinement of the gene locus for autosomal recessive juvenile parkinsonism (AR-JP) on chromosome 6q25.2-27 and identification of markers exhibiting linkage disequilibrium, *J. Hum. Genet.*, **43**, 22–31.
 10. Gasser, T., Muller-Myhsok, B., Wszolek, Z. K., et al. 1998, A susceptibility locus for Parkinson's disease maps to chromosome 2p13, *Nat. Genet.*, **18**, 262–265.
 11. Waters, C. H. and Miller, C. A. 1994, Autosomal dominant Lewy body parkinsonism in a four-generation family, *Ann. Neurol.*, **35**, 59–64.
 12. Singleton, A. B., Farrer, M., Johnson, J., et al. 2003, alpha-Synuclein locus triplication causes Parkinson's disease, *Science*, **302**, 841.
 13. Leroy, E., Boyer, R., Auburger, G., et al. 1998, The ubiquitin pathway in Parkinson's disease, *Nature*, **395**, 451–452.
 14. Valente, E. M., Bentivoglio, A. R., Dixon, P. H., et al. 2001, Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36, *Am. J. Hum. Genet.*, **68**, 895–900.
 15. Valente, E. M., Brancati, F., Ferraris, A., et al. 2002, PARK6-linked parkinsonism occurs in several European families, *Ann. Neurol.*, **51**, 14–18.
 16. van Duijn, C. M., Dekker, M. C., Bonifati, V., et al. 2001, Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36, *Am. J. Hum. Genet.*, **69**, 629–634.
 17. Bonifati, V., Breedveld, G. J., Squitieri, F., et al. 2002, Localization of autosomal recessive early-onset parkinsonism to chromosome 1p36 (PARK7) in an independent data set, *Ann. Neurol.*, **51**, 253–256.
 18. Funayama, M., Hasegawa, K., Kowa, H., et al. 2002, A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1, *Ann. Neurol.*, **51**, 296–301.
 19. Najim al-Din, A. S., Wriekat, A., Mubaidin, A., Dasouki, M. and Hiari, M. 1994, Pallido-pyramidal degeneration, supranuclear upgaze paresis and dementia: Kufor-Rakeb syndrome, *Acta Neurol. Scand.*, **89**, 347–352.
 20. Hampshire, D. J., Roberts, E., Crow, Y., et al. 2001, Kufor-Rakeb syndrome, pallido-pyramidal degeneration with supranuclear upgaze paresis and dementia, maps to 1p36, *J. Med. Genet.*, **38**, 680–682.
 21. Li, Y. J., Scott, W. K., Hedges, D. J., et al. 2002, Age at onset in two common neurodegenerative diseases is genetically controlled, *Am. J. Hum. Genet.*, **70**, 985–993.
 22. Pankratz, N., Nichols, W. C., Uniacke, S. K., et al. 2003, Significant linkage of Parkinson disease to chromosome 2q36–37, *Am. J. Hum. Genet.*, **72**, 1053–1057.
 23. Pankratz, N., Nichols, W. C., Uniacke, S. K., et al. 2002, Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations, *Am. J. Hum. Genet.*, **71**, 124–135.
 24. Pankratz, N., Nichols, W. C., Uniacke, S. K., et al. 2003, Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families, *Hum. Mol. Genet.*, **12**, 2599–2608.
 25. Strauss, K. M., Martins, L. M., Plun-Favreau, H., et al. 2005, Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease, *Hum. Mol. Genet.*, **14**, 2099–2111.
 26. Scott, W. K., Nance, M. A., Watts, R. L., et al. 2001, Complete genomic screen in Parkinson disease/evidence for multiple genes, *JAMA*, **286**, 2239–2244.
 27. Bertoli-Avella, A. M., Giroud-Benitez, J. L., Bonifati, V., et al. 2003, Suggestive linkage to chromosome 19 in a large Cuban family with late-onset Parkinson's disease, *Mov. Disord.*, **18**, 1240–1249.
 28. DeStefano, A. L., Golbe, L. I., Mark, M. H., et al. 2001, Genome-wide scan for Parkinson's disease/the Gene PD Study, *Neurology*, **57**, 1124–1126.
 29. Martinez, M., Brice, A., Vaughan, J. R., et al. 2004, Genome-wide scan linkage analysis for Parkinson's disease: the European genetic study of Parkinson's disease, *J. Med. Genet.*, **41**, 900–907.
 30. Le, W. D., Xu, P., Jankovic, J., et al. 2003, Mutations in NR4A2 associated with familial Parkinson disease, *Nat. Genet.*, **33**, 85–89.
 31. Kitada, T., Asakawa, S., Hattori, N., et al. 1998, Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism, *Nature*, **392**, 605–608.
 32. Wintermeyer, P., Kruger, R., Kuhn, W., et al. 2000, Mutation analysis and association studies of the UCHL1 gene in German Parkinson's disease patients, *Neuroreport*, **11**, 2079–2082.
 33. Harhangi, B. S., Farrer, M. J., Lincoln, S., et al. 1999, The Ile93Met mutation in the ubiquitin carboxy-terminal-hydrolase-L1 gene is not observed in European cases with familial Parkinson's disease, *Neurosci. Lett.*, **270**, 1–4.
 34. Marx, F. P., Holzmann, C., Strauss, K. M., et al. 2003, Identification and functional characterization of a novel R621C mutation in the synphilin-1 gene in Parkinson's disease, *Hum. Mol. Genet.*, **12**, 1223–1231.
 35. Valente, E. M., Abou-Sleiman, P. M., Caputo, V., et al. 2004, Hereditary early-onset Parkinson's disease caused by mutations in PINK1, *Science*, **304**, 1158–1160.
 36. Martin, E. R., Scott, W. K., Nance, M. A., et al. 2001, Association of single-nucleotide polymorphisms of the tau gene with late-onset Parkinson disease, *JAMA*, **286**, 2245–2250.
 37. Di Fonzo, A., Rohe, C. F., Ferreira, J., et al. 2005, A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease, *Lancet*, **365**, 412–415.
 38. Zhang, J., Song, Y., Chen, H. and Fan, D. 2005, The tau gene haplotype h1 confers a susceptibility to Parkinson's disease, *Eur. Neurol.*, **53**, 15–21.
 39. Zheng, K., Heydari, B. and Simon, D. K. 2003, A common NURR1 polymorphism associated with Parkinson disease and diffuse Lewy body disease, *Arch. Neurol.*, **60**, 722–725.
 40. Xu, P. Y., Liang, R., Jankovic, J., et al. 2002, Association of homozygous 7048G7049 variant in the intron six of Nurr1 gene with Parkinson's disease, *Neurology*, **58**, 881–884.
 41. Kahle, P. J., Haass, C., Kretschmar, H. A. and Neumann, M. 2002, Structure/function of alpha-synuclein in health and disease/rational development of animal models for Parkinson's and related diseases, *J. Neurochem.*, **82**, 449–457.
 42. Giasson, B. I. and Lee, V. M. 2001, Parkin and the molecular pathways of Parkinson's disease, *Neuron*, **31**, 885–888.

43. Shendelman, S., Jonason, A., Martinat, C., Leete, T. and Abeliovich, A. 2004, DJ-1 is a redox-dependent molecular chaperone that inhibits alpha-synuclein aggregate formation, *PLoS Biol.*, **2**, e362.
44. von Bohlen Und Halbach, O. 2004, Synucleins and their relationship to Parkinson's disease, *Cell Tissue Res.*, **318**, 163–174.
45. Zetterstrom, R. H., Williams, R., Perlmann, T. and Olson, L. 1996, Cellular expression of the immediate early transcription factors Nurr1 and NGFI-B suggests a gene regulatory role in several brain regions including the nigrostriatal dopamine system, *Brain Res. Mol. Brain Res.*, **41**, 111–120.
46. Solano, S. M., Miller, D. W., Augood, S. J., Young, A. B. and Penney, J. B. Jr. 2000, Expression of alpha-synuclein, parkin, and ubiquitin carboxy-terminal hydrolase L1 mRNA in human brain/genes associated with familial Parkinson's disease, *Ann. Neurol.*, **47**, 201–210.
47. Abeliovich, A., Schmitz, Y., Farinas, I., et al. 2000, Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system, *Neuron*, **25**, 239–252.
48. Murray, I. J., Medford, M. A., Guan, H. P., et al. 2003, Synphilin in normal human brains and in synucleinopathies/studies with new antibodies, *Acta Neuropathol. (Berl)*, **105**, 177–184.
49. Xie, T., Tong, L., Barrett, T., et al. 2002, Changes in gene expression linked to methamphetamine-induced dopaminergic neurotoxicity, *J. Neurosci.*, **22**, 274–283.
50. Unoki, M. and Nakamura, Y. 2001, Growth-suppressive effects of BPOZ and EGR2, two genes involved in the PTEN signaling pathway, *Oncogene*, **20**, 4457–4465.
51. Hauser, M. A., Li, Y. J., Takeuchi, S., et al. 2003, Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage, *Hum. Mol. Genet.*, **12**, 671–677.
52. Barrett, T., Xie, T., Piao, Y., et al. 2001, A murine dopamine neuron-specific cDNA library and microarray/increased COX1 expression during methamphetamine neurotoxicity, *Neurobiol. Dis.*, **8**, 822–833.
53. Stewart, G. J., Savioz, A. and Davies, R. W. 1997, Sequence analysis of 497 mouse brain ESTs expressed in the substantia nigra, *Genomics*, **39**, 147–153.
54. Thuret, S., Bhatt, L., O'Leary, D. D. and Simon, H. H. 2004, Identification and developmental analysis of genes expressed by dopaminergic neurons of the substantia nigra pars compacta, *Mol. Cell. Neurosci.*, **25**, 394–405.
55. Hicks, A. A., Petursson, H., Jonsson, T., et al. 2002, A susceptibility gene for late-onset idiopathic Parkinson's disease, *Ann. Neurol.*, **52**, 549–555.
56. DeStefano, A. L., Lew, M. F., Golbe, L. I., et al. 2002, PARK3 influences age at onset in Parkinson disease/a genome scan in the GenePD study, *Am. J. Hum. Genet.*, **70**, 1089–1095.
57. Davies, R. W., Roberts, A. B., Morris, A. J., et al. 1994, Enhanced access to rare brain cDNAs by prescreening libraries/207 new mouse brain ESTs, *Genomics*, **24**, 456–463.
58. Savioz, A. and Davies, R. W. 1995, Discovering genes with localised expression in the mouse brain/cDNAs specific to the substantia nigra, *Gene*, **154**, 225–230.
59. Alberi, L., Sgado, P. and Simon, H. H. 2004, Engrailed genes are cell-autonomously required to prevent apoptosis in mesencephalic dopaminergic neurons, *Development*, **131**, 3229–3236.
60. Simon, H. H., Saueressig, H., Wurst, W., Goulding, M. D. and O'Leary, D. D. 2001, Fate of midbrain dopaminergic neurons controlled by the engrailed genes, *J. Neurosci.*, **21**, 3126–3134.
61. Venter, J. C., Adams, M. D., Myers, E. W., et al. 2001, The sequence of the human genome, *Science*, **291**, 1304–51.
62. Gregory, S. G., Sekhon, M., Schein, J., et al. 2002, A physical map of the mouse genome, *Nature*, **418**, 743–750.
63. West, A. B., Zimprich, A., Lockhart, P. J., et al. 2001, Refinement of the PARK3 locus on chromosome 2p13 and the analysis of 14 candidate genes, *Eur. J. Hum. Genet.*, **9**, 659–666.
64. West, A., Farrer, M., Petrucelli, L., et al. 2001, Identification and characterization of the human parkin gene promoter, *J. Neurochem.*, **78**, 1146–1152.
65. Holzmann, C., Kruger, R., Saecker, A. M., et al. 2003, Polymorphisms of the alpha-synuclein promoter/expression analyses and association studies in Parkinson's disease, *J. Neural. Transm.*, **110**, 67–76.
66. Farrer, M., Maraganore, D. M., Lockhart, P., et al. 2001, alpha-Synuclein gene haplotypes are associated with Parkinson's disease, *Hum. Mol. Genet.*, **10**, 1847–1851.
67. Stamatoyannopoulos, J. A. 2004, The genomics of gene expression, *Genomics*, **84**, 449–457.
68. Reveron, M. E., Savelieva, K. V., Tillerson, J. L., et al. 2002, L-DOPA does not cause neurotoxicity in VMAT2 heterozygote knockout mice, *Neurotoxicology*, **23**, 611–619.
69. Takahashi, N., Miner, L. L., Sora, I., et al. 1997, VMAT2 knockout mice: heterozygotes display reduced amphetamine-conditioned reward, enhanced amphetamine locomotion, and enhanced MPTP toxicity, *Proc. Natl Acad. Sci. USA*, **94**, 9938–9943.
70. Beal, M. F. 2003, Mitochondria, oxidative damage, and inflammation in Parkinson's disease, *Ann. NY Acad. Sci.*, **991**, 120–131.
71. Greenamyre, J. T. and Hastings, T. G. 2004, Biomedicine: Parkinson's—divergent causes, convergent mechanisms, *Science*, **304**, 1120–1122.
72. Pruitt, K. D., Tatusova, T. and Maglott, D. R. 2003, NCBI Reference Sequence project: update and current status, *Nucleic Acids Res.*, **31**, 34–37.