

①9 RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
PARIS

①1 N° de publication : 2 935 712

(à n'utiliser que pour les
commandes de reproduction)

②1 N° d'enregistrement national : 08 55997

⑤1 Int Cl⁸ : C 12 Q 1/68 (2006.01), C 07 H 21/00

⑫

DEMANDE DE BREVET D'INVENTION

A1

②2 Date de dépôt : 05.09.08.

③0 Priorité :

④3 Date de mise à la disposition du public de la
demande : 12.03.10 Bulletin 10/10.

⑤6 Liste des documents cités dans le rapport de
recherche préliminaire : *Se reporter à la fin du
présent fascicule*

⑥0 Références à d'autres documents nationaux
apparentés :

⑦1 Demandeur(s) : CENTRE HOSPITALIER REGIONAL
UNIVERSITAIRE DE LILLE — FR et UNIVERSITE DU
DROIT ET DE LA SANTE DE LILLE II — FR.

⑦2 Inventeur(s) : CHARTIER HARLIN MARIE
CHRISTINE, DESTEE ALAIN, LARVOR LYDIE, LE
RHUN EMILIE, MOUROUX VINCENT et MUTEZ
EUGENIE.

⑦3 Titulaire(s) : CENTRE HOSPITALIER REGIONAL UNI-
VERSITAIRE DE LILLE, UNIVERSITE DU DROIT ET
DE LA SANTE DE LILLE II.

⑦4 Mandataire(s) : CABINET BEAU DE LOMENIE.

⑤4 METHODE DE DETECTION DE L'EXPRESSION DIFFERENTIELLE D'UN ENSEMBLE DE MARQUEURS
MOLECULAIRES ASSOCIEE A LA MALADIE DE PARKINSON.

⑤7 La présente invention a pour objet une méthode pour
détecter l'expression différentielle d'un ensemble de mar-
queurs moléculaires qui est associée à la présence ou au
risque de développer la maladie de Parkinson, et en parti-
culier d'une forme sporadique de cette maladie. L'invention
a également pour objet un kit ou nécessaire pour la mise en
oeuvre de cette méthode.

FR 2 935 712 - A1



La présente invention est relative au domaine du diagnostic de la maladie de Parkinson chez un sujet.

Plus précisément, la présente invention a essentiellement pour objet une méthode pour détecter l'expression différentielle d'un ensemble de marqueurs moléculaires qui est associée à la présence ou au risque de développer la maladie de Parkinson, et en particulier d'une forme sporadique de cette maladie.

La maladie de Parkinson est une maladie neurodégénérative commune caractérisée par une dégénérescence des neurones de l'axe nigrostrié et, qui suit une progression lente et irréversible. L'âge moyen de diagnostic de la maladie se situe vers 57 ans, à une période où les individus sont encore en activité professionnelle. La maladie de Parkinson est la deuxième maladie neurodégénérative après la maladie d'Alzheimer et affecte environ 2% de la population des pays industrialisés au delà de 65 ans. La prévalence de cette maladie augmente avec l'âge de la population et l'évolution démographique actuelle.

La majorité des patients atteints par la maladie de Parkinson ne présente pas d'antécédents familiaux et développe une forme de la maladie dite sporadique. L'étiologie de leur maladie est mal comprise. On suppose que son développement serait la résultante d'une multitude de facteurs environnementaux et génétiques dont la plupart ne sont pas identifiés à ce jour.

En revanche, certaines formes héréditaires (appelées aussi formes familiales) ont été identifiées au cours des dix dernières années et sont en cours d'études. Ces formes familiales représentent entre 5 et 15% des cas. Elles ont un mode de transmission autosomique dominant ou récessif. Plus de 13 loci ont été liés à la maladie et les inventeurs ont identifié un nouveau locus sur le chromosome 3 dans une forme familiale de transmission autosomique dominante. Le gène de ce locus responsable de la maladie intervient dans le mécanisme de traduction des ARNm et s'appelle EIF4G1. Il vient s'ajouter aux 10 autres gènes déjà décrits dans la maladie de Parkinson porteurs de mutations délétères dont voici les gènes les plus fréquemment mutés. Le gène de la Parkine (PARK2 ou PRKN) est responsable de 50% des formes précoces (développement des symptômes avant 40 ans) de la maladie. Toutefois, les mutations de ce gène ne représentent que 0,4% de l'ensemble des cas. Les mutations les plus fréquentes interviennent dans les formes autosomiques dominantes et concernent les gènes SNCA (alpha-synucléine), SCA2 et surtout LRRK2. Ce dernier serait responsable de 3 % des formes familiales caucasiennes et d'environ 1% des cas sporadiques de la maladie. Dans de rares exceptions, cette proportion peut atteindre 18% des patients de certaines

populations Arabes-Nord Africaines et 39% des populations Juives Ashkenazes. Ces pourcentages élevés sont reliés aux habitudes culturelles et à un pourcentage important de consanguinité. Le gène LRRK2 code une kinase qui interviendrait à un carrefour métabolique dans différents syndromes parkinsoniens et maladies apparentées (synucléopathies et tauopathies, qui sont caractérisées respectivement par des agrégats de protéines alpha-synucléine et Tau). Cette connexion entre syndromes parkinsoniens et autres maladies neurodégénératives est également illustrée par les mutations touchant le gène SCA2 ; certaines mutations étant responsables d'une Ataxie Cérébelleuse typique (SCA2-C) alors que d'autres donnent naissance à une maladie de Parkinson (SCA2-P). De la même manière le gène SNCA code l'alpha-synucléine qui peut s'agréger et former des inclusions comme les corps de Lewy ; son agrégation anormale caractérise la famille de maladies appelées les synucléopathies dont font parties, notamment, la maladie de Parkinson, la Démence à Corps de Lewy (DLB), et l'Atrophie Multi-Systématisée (MSA).

15 Bien que 10 gènes responsables de formes rares de la maladie de Parkinson aient déjà été identifiés et que leur identification soit importante pour mieux comprendre les mécanismes menant à la maladie, ces gènes n'expliquent pas la majorité des cas sporadiques de la maladie. En fait l'impact des facteurs environnementaux est prépondérant dans la maladie de Parkinson comme le montre l'absence de différence de concordance entre jumeaux monozygotes et dizygotes dans les études épidémiologiques réalisées. Si le taux de concordance était plus important chez les jumeaux monozygotes que dizygotes, cela montrerait un impact des facteurs génétiques dans le développement de la maladie de Parkinson. Ces études soulignent le rôle important joué par les facteurs environnementaux pour la majorité des patients. Plusieurs modèles animaux de la maladie ont d'ailleurs été réalisés en utilisant les pesticides comme agent causal environnemental de la maladie. Toutefois, il est possible que bien que moindre, l'impact des facteurs génétiques puissent agir grâce à des interactions gène-environnement. En effet, les inventeurs ont démontré par le passé que de telles interactions étaient possibles. Par exemple, certains sujets exposés aux pesticides avaient un risque augmenté de développer la maladie lorsqu'ils étaient porteurs du génotype CYP2D6, alors que ce risque était nul pour les sujets porteurs d'un génotype différent exposés eux aussi aux mêmes facteurs. Au cours de sa vie, l'individu va être exposé à de nombreux facteurs environnementaux, qui en fonction de son patrimoine génétique, pourront donc lui être favorables ou délétères. La maladie de Parkinson étant une maladie du

sujet âgé, de nombreux facteurs environnementaux interviendront sur certaines voies métaboliques avant l'apparition des premiers symptômes.

De plus, lorsque ces premiers symptômes émergent, plus de 50% des neurones dopaminergiques de la substance noire sont morts et le déficit en
5 dopamine dans le striatum excède 70%. Cette affection chronique est donc caractérisée par une longue période silencieuse au cours de laquelle l'individu ne présente aucun symptôme. Cependant sur le plan moléculaire et cellulaire, d'infimes modifications se produisent et s'additionnent notamment au sein de la voie nigro-striée.

10 Le diagnostic de la maladie de Parkinson ne s'effectue qu'à l'apparition insidieuse de symptômes asymétriques ayant progressés lentement pendant plusieurs années. Il n'existe pas de diagnostic au stade précoce de la maladie. On parle, à ce stade-ci, souvent de syndromes parkinsoniens. En effet, le développement de la DLB ou de la MSA par exemple peut mimer la maladie de
15 Parkinson ou le début de cette maladie.

Trois signes cardinaux définissent un syndrome parkinsonien :

- le tremblement de repos
- l'akinésie/bradykinésie
- l'hypertonie/rigidité musculaire

20 Le syndrome parkinsonien est présent dans la maladie de Parkinson mais peut également se retrouver dans d'autres affections comme la DLB, certaines formes de MSA et de Paralyse Supranucléaire Progressive (PSP)...

Toutefois la maladie de Parkinson se distingue notamment des autres syndromes parkinsoniens par un quatrième signe cardinal : l'instabilité posturale
25 d'apparition tardive et par une réponse positive au traitement par la levodopa (L-dopa). En effet, la plupart des sujets atteints d'un syndrome parkinsonien présentent une résistance à ce traitement. De plus plusieurs critères d'exclusion ont été définis afin de poser le diagnostic de la maladie de Parkinson.

Ces similitudes de symptômes contribuent à la difficulté d'établir un
30 diagnostic initial de maladie de Parkinson.

Une évolution des manifestations cliniques sur une période d'au moins cinq ans est nécessaire pour affirmer ou infirmer le diagnostic initial de la maladie de Parkinson. Un diagnostic défini de cette maladie ne peut être posé qu'après un examen neuropathologique post-mortem.

35 Plusieurs ensembles de critères cliniques ont été proposés pour tenter de poser un diagnostic de la maladie de Parkinson du vivant du patient. Il faut

cependant noter qu'il n'existe pas, à l'heure actuelle, de critères internationaux communs pour ce diagnostic. Les critères diagnostiques les plus utilisés en pratique clinique sont ceux de l'*UK Parkinson's Disease Society Brain Bank* (UK PDSBB), Hughes A.J., *Journal of Neurology, Neurosurgery, and Psychiatry*, 1992, **55**, 181-184 ; Gelb D.J *Arch Neurol.* 1999 Jan; **56**(1), 33-9. Ils classent souvent la certitude du diagnostic selon trois niveaux dits : défini, probable et possible.

Un diagnostic **défini** retient les quatre critères cardinaux (bradykinésie/akinésie, tremblement de repos, rigidité, instabilité posturale) et une confirmation par autopsie.

10 Un diagnostic **probable** retient trois critères sur les quatre précédents avec bien souvent la présence de critères prospectifs supplémentaires. En effet, ce diagnostic reposant sur des arguments cliniques est difficile à établir avec certitude, notamment au stade précoce de la maladie en raison de la ressemblance des symptômes entre la maladie de Parkinson et les syndromes parkinsoniens. Par
15 conséquent, la plupart des études scientifiques réalisées utilise des critères stricts de diagnostic de la maladie de Parkinson retenant trois des quatre critères dont la bradykinésie ainsi que la réponse positive à la levodopa et l'absence des 16 critères d'exclusion définis par Hughes (*supra*).

Un diagnostic **possible** débute généralement avec deux des quatre signes
20 cardinaux de la maladie, le critère de bradykinésie étant requis pour retenir le diagnostic comme possible.

Afin de quantifier l'amélioration thérapeutique lors du suivi des patients, l'échelle d'évaluation « *Unified Parkinson's Disease Rating Scale* » (UPDRS) est la plus utilisée par les cliniciens. Cette échelle vise une évaluation objective de la
25 sévérité, de la fluctuation et de la progression des symptômes cliniques ainsi que leur retentissement sur la vie quotidienne du patient. Cette échelle n'est pas indispensable au diagnostic et au suivi mais peut se révéler nécessaire lors de certaines décisions thérapeutiques notamment chirurgicales. Cette échelle multidimensionnelle comporte quatre parties auxquelles leur sont attribués des
30 scores :

- l'état mental du patient et son comportement, (partie I : score maximal 16)
 - ses activités de la vie quotidienne en période de traitement pharmacologique (période ON) et en l'absence de traitement pharmacologique (période OFF) (partie II, score maximal : 52)
- 35

- un examen moteur pratiqué avec et plus rarement en l'absence du traitement (ON et OFF) (partie III, score maximal : 108)
- les complications du traitement. (partie IV : score maximal 23)

L'échelle UPDRS comporte, en plus de ces quatre parties, un classement des patients selon l'échelle Hoehn et Yahr (HY). L'échelle HY comporte 8 stades (Tableau 1) qui estiment la gravité de la maladie de Parkinson chez un malade.

Stade	Manifestations cliniques
Stade 0	Aucun signe de la maladie
Stade 1	Signes unilatéraux n'entraînant pas de handicap dans la vie quotidienne
Stade 1,5	Atteinte unilatérale et atteinte axiale
Stade 2	Signes à prédominance unilatérale entraînant un certain handicap
Stade 2,5	Atteinte bilatérale légère à modérée ; une certaine instabilité posturale
Stade 3	Atteinte bilatérale avec une instabilité posturale ; malade autonome
Stade 4	Handicap sévère mais possibilité de marche ; perte partielle de l'autonomie
Stade 5	Malade en chaise roulante ou alité ; n'est plus autonome.

Tableau 1 : Echelle de Hoehn et Yahr

10 L'âge de début et l'importance de la gêne fonctionnelle sont les deux principaux facteurs qui guident les choix thérapeutiques. Bien qu'il n'existe pas à ce jour de traitement permettant de stopper ou ralentir la progression de cette maladie, de nouveaux traitements à but protecteur sont en cours d'évaluation en essais cliniques (antioxydants, antagonistes glutamatergiques, facteurs neurotrophiques, 15 ligands de neuro-immunophilines ou encore agents anti-apoptiques, etc). Idéalement, ces nouvelles thérapeutiques devraient être adaptées à chaque stade de la maladie pour viser une efficacité optimale. Il est donc très important de cibler de manière efficace les patients susceptibles de bénéficier de traitements neuroprotecteurs. De ce fait, le diagnostic de la maladie de Parkinson lors de sa 20 phase précoce est essentiel et permettrait de proposer un traitement adapté au patient.

Le développement des nouvelles technologies de génomique fonctionnelle a permis la réalisation de premières études transcriptomiques à partir de tissus cérébraux de patients atteints de maladies neurodégénératives et de témoins. Ces

données montrent notamment qu'il existe un certain nombre de gènes dont l'expression est spécifiquement dérégulée chez les patients atteints de la maladie de Parkinson (WO2005/067391). Toutefois ces études ne sont le reflet que des modifications finales de la maladie puisque ces données sont obtenues à partir d'échantillons de tissu cérébral de patients décédés. Ces études incluent également dans leurs résultats les nombreuses perturbations moléculaires liées à la mort cellulaire et aux délais post-mortem. Dans ces conditions, il est difficile de dire que les variations observées d'expression des gènes sont le miroir d'un stade précoce de la maladie de Parkinson.

10 Le diagnostic de la maladie de Parkinson est donc complexe et peut être remis en cause au cours de l'évolution de la maladie. Il n'existe actuellement aucune signature biologique spécifique de la maladie de Parkinson susceptible d'établir un diagnostic, ou de prédire un développement, de cette pathologie à un stade précoce. La mise à disposition d'un test de diagnostic efficace et simple, pour un stade précoce, permettrait aux patients d'être pris en charge dès le début de leur maladie et de bénéficier ainsi d'un traitement plus efficace, plus adapté à leurs besoins et de leur proposer des stratégies de neuroprotection.

15 Il existe donc un besoin important de disposer d'un test de diagnostic/pronostic permettant de déterminer, de leur vivant, si des sujets sont atteints de la maladie de Parkinson et/ou de prédire si des sujets ont un risque de développer cette maladie. Il existe également un besoin important de pouvoir distinguer les patients susceptibles de développer la maladie de Parkinson de ceux qui sont susceptibles de développer un autre syndrome parkinsonien en particulier de type DLB, MSA qui peuvent développer en phase précoce les mêmes symptômes cliniques.

20 La présente invention apporte des solutions à ces besoins. Grâce à une analyse de l'expression différentielle d'un ensemble de marqueurs moléculaires qui sont différentiellement exprimés spécifiquement dans des formes familiales de la maladie, on dispose d'un outil pour détecter la présence ou le risque de développer cette maladie, même chez des sujets asymptomatiques ou pauci-symptomatiques, ou de confirmer un diagnostic chez des patients ayant des symptômes. L'invention permet également d'identifier les patients atteints d'un syndrome parkinsonien de ceux atteints de la maladie de Parkinson.

30 La demande internationale publiée sous le numéro WO2006/050475 décrit une méthode de diagnostic de maladies neurodégénératives à partir d'un échantillon sanguin dans laquelle l'expression d'un ou plusieurs marqueurs génétiques est

mesurée puis comparée à celle de ce(s) marqueur(s) chez un sujet sain pour déterminer si le sujet est atteint d'une maladie neurodégénérative. Une liste de 8 gènes permettrait de différencier des sujets ayant un risque de développer une forme sporadique de la maladie de Parkinson. Comme dans toutes les études
5 actuelles pour déterminer quels gènes sont dérégulés dans les formes sporadiques de maladies neurodégénératives, le critère familial a fait l'objet d'une exclusion de diagnostic de maladie de Parkinson. En effet, les auteurs cherchent à identifier des marqueurs moléculaires qui seraient impliqués dans le développement de cette maladie dont l'étiologie est inconnue. Les formes héréditaires familiales ne rentrent
10 pas dans cette définition puisque par essence, leur cause est connue ; le développement de la maladie étant lié à la présence d'une mutation.

Il est en effet communément admis par la communauté scientifique que chaque porteur de mutation pathogène développe un cas particulier (une variante) de maladie de Parkinson. En effet, une mutation pathogène d'un gène va générer un
15 gain ou une perte de fonction de la protéine codée. Ces modifications vont cibler préférentiellement un mécanisme biologique et entraîner des modifications spécifiques dans les réseaux biologiques. Pour simplifier, on pourrait dire que chaque gène muté va entraîner une forme de la maladie distincte par rapport à un autre gène muté dans le sens où il ne met en exergue qu'une petite partie des
20 mécanismes potentiellement importants pour comprendre l'ensemble des cas de la maladie.

Par ailleurs, il est admis que le rôle des facteurs environnementaux est prépondérant pour l'apparition des cas sporadiques et que les formes familiales ne représentent que 5 à 15% des formes totales de la maladie de Parkinson. De plus,
25 les mutations des gènes SNCA, LRRK2, SCA2 et PRKN représentent moins de 2% de l'ensemble des cas. Les formes familiales sont donc des formes rares apparaissant à un âge souvent précoce (souvent avant 40 ans) alors qu'on observe chez les formes sporadiques un développement plus tardif (en moyenne après 57 ans). Par conséquent, statistiquement les modifications observées pour chacune de ces formes
30 distinctes de la maladie de Parkinson ne peuvent pas totalement expliquer l'ensemble des cas, en particulier les formes sporadiques.

Les inventeurs réfutent cette idée reçue et ils montrent au contraire, et ceci constitue le fondement de la présente invention, que la prise en compte des données combinées des gènes dont l'expression est dérégulée chez des sujets atteints de
35 formes familiales, donc des formes rares de la maladie de Parkinson, permet d'identifier une série de gènes dont l'expression est dérégulée chez les sujets

malades ou à risque de développer la maladie de Parkinson, quelque soit le stade de gravité de la maladie, et même des cas asymptomatiques, et en particulier des formes sporadiques.

5 La méthode mise au point par les inventeurs qui se base sur l'ensemble des gènes dont l'expression est dérégulée dans des formes familiales dues à des mutations rares dans la population générale caucasienne permet d'identifier des sujets atteints de la maladie de Parkinson présentant ou non des symptômes avec un taux d'erreur faible. De préférence, la méthode selon l'invention permet de déterminer la présence ou le risque de développer la maladie de Parkinson avec un
10 taux d'erreur inférieur ou égal à 30 %, et en particulier d'identifier des cas sporadiques de cette maladie.

Les critères diagnostiques de la maladie de Parkinson selon l'invention sont basés sur ceux acceptés en la matière par la Fédération Française de Neurologie, Agence Nationale d'Accréditation et de Santé, 2000. Plus précisément, les critères
15 cumulatifs 1) 2) et 3) suivants ont été utilisés :

- 1) au moins 3 des 4 signes cardinaux de la maladie de Parkinson, à savoir tremblement de repos, bradykinésie, rigidité musculaire ou instabilité posturale; la bradykinésie faisant partie obligatoirement d'un des 3 signes cardinaux retenus pour le diagnostic,
- 20 2) la réponse positive au traitement « anti-parkinsonien » levodopa, et
- 3) l'absence des 15 critères d'exclusion définis par UK PDSBB, Hughes A.J., *Journal of Neurology, Neurosurgery, and Psychiatry*, 1992, **55**, 181-184. Le 16^{ème} critère d'antécédent familiale n'a pas été retenu.

Ainsi, au sens de la présente invention, on entend désigner par :

- 25 - « sujet atteint d'une forme familiale ou cas familial », tout sujet dont au moins un autre membre de sa famille est atteint par la maladie de Parkinson. La notion d'antécédents familiaux est celle qui est définie dans Hughes *et al.*, 1992, (*supra*).
- « sujet atteint d'une forme sporadique ou cas sporadique », tout sujet
30 n'ayant aucune forme héréditaire de la maladie de Parkinson.
- « sujet asymptomatique », tout sujet, atteint de forme héréditaire ou sporadique, n'ayant aucun symptôme de la maladie de Parkinson.
- « sujet pauci-symptomatique », personne présentant des troubles légers
35 mais insuffisants pour poser un diagnostic de maladie de Parkinson selon les critères diagnostiques rappelés ci-dessus.

- « sujet symptomatique », personne ayant développée la maladie de Parkinson dont les symptômes sont visibles et sans équivoque pour poser un diagnostic selon ces mêmes critères.

5 - stade précoce de la maladie de Parkinson, la phase silencieuse (sujet asymptomatique) et la phase pauci symptomatique de la maladie.

Dans le cadre de l'invention, les inventeurs n'ont donc pas retenu la notion de forme familiale de la maladie de Parkinson comme étant un critère d'exclusion.

10 Par personnes à risque, on entend personnes ayant des antécédents familiaux, ou personnes ayant des facteurs de risque reconnus de développer la maladie de Parkinson (ex : exposition aux pesticides, dérèglement de marqueurs biologiques etc).

15 Par expression dérégulée ou gènes dérégulés, on entend tout gène dont le niveau d'expression chez un sujet est augmenté ou diminué par rapport à celui d'un sujet témoin. Le seuil de variation d'expression considéré comme significatif dépend de chaque technique de comparaison. Par exemple, pour une analyse par micro-
20 puces à ADNc, un seuil de variation d'expression d'au moins 1,2 avec un test T de Welch lors d'un filtrage en Volcano plot est classiquement considéré comme étant significatif. Ainsi, l'homme du métier est à même de déterminer le seuil de variation d'expression à prendre en compte en fonction de la technique de détection et de
20 comparaison d'expression utilisée.

25 Dans le cadre de la présente invention, on entend par mutation, toute affection génique consistant en le remplacement d'une base nucléotidique, la délétion d'une ou plusieurs base(s) nucléotidique(s), la délétion ou la multiplication d'exons (1, 2, 3 ou d'avantages de copies d'exons), etc, mais également des variations du nombre de copies des gènes telles que la présence d'une copie (délétion), d'un nombre de copies supérieur à 3 (une duplication) ou 4 copies (triplication) ou plus d'une région chromosomique.

30 Selon un premier aspect, la présente invention a donc pour objet une méthode pour détecter dans un échantillon biologique à tester l'expression différentielle d'un ensemble de marqueurs moléculaires associée à la présence ou au risque de développer la maladie de Parkinson, caractérisée en ce qu'elle comprend les étapes suivantes :

35 a1) extraction *in vitro* des marqueurs moléculaires à partir de l'échantillon biologique à tester,

a2) extraction *in vitro* des marqueurs moléculaires à partir des échantillons biologiques de sujets atteints de différentes formes familiales de la maladie de

Parkinson, de sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens et de sujets sains,

5 b) déterminer l'ensemble des marqueurs moléculaires dont l'expression est significativement dérégulée en commun chez les sujets atteints de différentes formes familiales quel que soit le stade de la maladie mais dont l'expression n'est pas significativement dérégulée chez les sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens, par rapport à celle mesurée chez les sujets sains,

10 c) vérifier que l'expression de l'ensemble de ces marqueurs moléculaires est bien dérégulée dans l'échantillon à tester.

L'originalité de la présente invention réside donc dans l'utilisation de données (gènes dont l'expression est dérégulée par rapport à des individus sains) qui sont issues de patients atteints de formes familiales de la maladie de Parkinson pour déterminer un ensemble de gènes dont l'expression dérégulée est spécifiquement associée à cette maladie, par comparaison aux méthodes décrites dans l'art antérieur et en particulier dans WO2006/050475 et Scherzer *et al.*, PNAS 2007, vol. 104, **3**, 15 955-960 qui ne tiennent précisément pas compte des données issues de ces types de patients pour prédire l'existence ou le risque de développer une forme sporadique de la maladie de Parkinson. En effet, les patients qui font partie de ces études 20 répondent aux critères diagnostiques cliniques de la *United Kingdom Parkinson's Disease Society Brain Bank* qui exclut les patients atteints d'une forme familiale de la maladie.

Ainsi, les inventeurs pensent que cette notion est à remettre en cause, et c'est pourquoi ils n'ont pas retenu ce dernier critère d'exclusion (exclusion des 25 formes familiales) dans la présente invention. Par contre, ils ont utilisé des critères stricts de maladie de Parkinson incluant la présence de trois des quatre critères cardinaux de la maladie dont la bradykinésie, et la réponse positive au traitement à la levodopa et l'absence de 15 critères d'exclusion autres que celui concernant l'exclusion des formes familiales.

30 Outre le diagnostic général de la maladie de Parkinson, l'intérêt majeur de la méthode selon l'invention est de pouvoir détecter la présence ou le risque de développer une forme sporadique de la maladie de Parkinson. Dans le cas d'un diagnostic positif, la distinction entre une forme familiale et une forme sporadique de la maladie de Parkinson sera déterminée sur la base des informations déjà 35 disponibles concernant les autres membres de la famille (c'est-à-dire savoir si au moins un autre membre de la famille a déjà développé ou pas la maladie).

La mise en œuvre de la présente invention permet également de déterminer la présence de la maladie de Parkinson même chez des personnes atteintes de formes asymptomatiques de la maladie.

De préférence, la méthode selon l'invention permet de déterminer la présence ou le risque de développer la maladie de Parkinson avec un taux d'erreur inférieur ou égal à 30 %.

Par sujets atteints de différentes formes familiales de la maladie de Parkinson, on entend désigner des patients dont au moins un membre de leur famille est atteint de la maladie de Parkinson. Les données d'expression des gènes significativement dérégulés chez des sujets atteints de formes familiales de la maladie de Parkinson susceptibles d'être utilisées dans le cadre de la présente invention peuvent en particulier être issues de patients atteints de formes familiales différentes, et de préférence porteurs de mutation dans le gène SNCA et/ou de patients porteurs de mutation dans le gène PRKN et/ou de patients porteurs de mutation dans le gène SCA2 et/ou de patients porteurs de mutation dans le gène LRRK2.

Selon un mode de réalisation particulièrement avantageux, les données d'expression des gènes proviennent de sujets atteints des quatre formes familiales dont chacun est causée par mutation dans les gènes SNCA, PRKN, SCA2 et LRRK2 respectivement.

De préférence, lorsque les données d'expression des gènes dérégulés sont notamment issues de patients porteurs d'une mutation dans le gène LRRK2, les patients possèdent de préférence la mutation G2019S.

De même, lorsque les données d'expression des gènes dérégulés sont notamment issues de patients porteurs de mutations responsables de la maladie de Parkinson dans le gène SNCA, ces patients possèdent de préférence une multiplication du gène SNCA, et de préférence une duplication du gène SNCA.

De préférence encore, lorsque les données d'expression des gènes dérégulés sont notamment issues de patients porteurs d'une mutation dans le gène SCA2, les patients possèdent de préférence une mutation de type 39 triplets répétés CAG/CAA interrompus.

Enfin, lorsque les données d'expression des gènes dérégulés sont notamment issues de patients porteurs d'une mutation dans le gène PRKN, les patients possèdent de préférence une mutation de type homozygote ou une mutation de type hétérozygote composite Trp445stop ou ex3,4dupl ou ex6del/c.1385insA ou ex6dupl/c.225delA.

Par ailleurs, les données d'expression des gènes dérégulés chez des sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens peuvent notamment être obtenues à partir de patients atteints de différentes maladies neurodégénératives ou différents syndromes parkinsoniens tels que l'Ataxie Spino-Cérébelleuse (SCA), la MSA, la DLB, la PSP, la sclérose latérale amyotrophique (SLA), et en particulier ces données d'expression proviennent à la fois de patients atteints de SCA, de patients atteints de MSA et de patients atteints de DLB. De manière particulière encore, ces données d'expression proviennent de patients atteints de SCA, de patients atteints de MSA, de patients atteints de DLB, de patients atteints de PSP et de patients atteints de SLA.

Selon un mode de réalisation particulièrement préféré, l'ensemble des marqueurs biologiques dont l'expression est dérégulée est constitué des marqueurs qui sont communs à au moins 1 patient porteur d'une multiplication du gène SNCA, au moins 4 patients porteurs d'une mutation dans le gène PRKN, au moins 6 patients parkinsoniens porteurs d'une mutation dans le gène SCA2 et au moins 9 patients porteurs d'une mutation dans le gène LRRK2 (en ce qui concerne les patients atteints de formes familiales) et non communs à au moins 6 patients atteints de SCA, au moins 1 patient atteint de MSA et au moins 4 patients atteints de DLB (en ce qui concerne les patients atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens).

L'échantillon biologique qui est utilisé dans le cadre de la présente invention consiste plus particulièrement en tout matériel issu du patient, susceptible de contenir un marqueur biologique qui permette de détecter l'expression d'un gène ou de son produit. En particulier, l'échantillon biologique peut être un fluide biologique (du sang total, du sérum, de la salive, des frottis des muqueuses, des cellules circulantes du patient (sanguines ou immunitaires comme par exemple les plaquettes ou des cellules mononucléées du sang) ou encore un tissu comme de la peau ou des cellules issues de la peau comme les fibroblastes, des cellules issues de muscle, de cheveux, ou poils), du moment que celui-ci contient des marqueurs moléculaires susceptibles de refléter l'expression de gènes, par exemple tels que des ARNs ou des protéines. De préférence, l'échantillon biologique issu de la personne à tester est du même type que celui qui est issu des sujets de références (sujets atteints de différentes formes familiales de la maladie de Parkinson, sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens et sujets sains).

Différents travaux, et notamment ceux des inventeurs, ont montré que les cellules mononucléées peuvent être utilisées pour étudier les changements

intervenants au niveau du tissu cérébral des sujets atteints de maladies dégénératives, et en particulier de la maladie de Parkinson. Ainsi, l'échantillon biologique utilisé dans le cadre de la méthode selon l'invention est de préférence constitué de cellules mononucléées issues du sang total.

5 Les marqueurs biologiques selon la présente invention sont représentés par toute molécule dont la détection correspond ou permet de mettre en évidence l'expression de gènes, comme notamment un acide ribonucléique ou une protéine. Selon un mode de réalisation préféré de l'invention, les marqueurs moléculaires utilisés sont des ARNs (en particulier des ARNm) ou des protéines.

10 Dans le cadre de la présente invention, on préférera tout particulièrement utiliser comme échantillon biologique des cellules mononucléées isolées à partir du sang, desquelles on pourra extraire en particulier les ARNs qui seront utilisés en tant que marqueur biologique pour la mise en œuvre de la méthode. Dans ce cas, l'échantillon biologique est obtenu de préférence d'un patient à jeun.

15 Les marqueurs biologiques peuvent être extraits des échantillons biologiques par tous les moyens classiques bien connus de l'homme du métier et spécifiquement adaptés à la nature du marqueur biologique choisi.

Par ailleurs, toute technique de mesure quantitative de l'expression des gènes et de comparaison des niveaux d'expression bien connue de l'homme du
20 métier peut être utilisée pour la mise en œuvre de la présente invention. Les techniques de mesure et de comparaison seront choisies pour être spécifiquement adaptées au type de marqueurs biologiques choisi.

A titre d'exemples de méthode de quantification, on citera notamment les techniques basées sur des micro-puces, des PCR en temps réel ou toute autre
25 technique permettant une quantification (séquençage, Western blot etc...). Cependant, selon une variante préférée de la présente invention, l'expression dérégulée des marqueurs moléculaires est déterminée par une mesure quantitative de l'expression des gènes ou de la quantité des protéines choisie parmi les micro-puces et la RT-PCR quantitative.

30 Ainsi, l'étude de l'expression différentielle des marqueurs moléculaires réalisée à l'étape c) de la méthode fait appel aux connaissances de l'homme du métier qui saura préparer les réactifs spécifiques (par exemple conception d'amorces, de puces, d'anticorps etc) ou choisir ces réactifs parmi ceux qui sont déjà disponibles dans le commerce et analyser l'expression différentielle de ces marqueurs
35 moléculaires.

En particulier, lors de la mise en œuvre de l'étape c) de la méthode selon l'invention, les réactifs spécifiques des marqueurs moléculaires identifiés lors de l'étape b) peuvent détecter des homologues de ces marqueurs. Par homologues on entend tout marqueur moléculaire dont l'homologie de séquence est d'au moins 70 %
5 %, et de préférence d'au moins 80 %, d'au moins 85 % ou même d'au moins 90 % ou encore d'au moins 95 % avec la séquence du marqueur moléculaire concerné qui aura été identifié à l'étape b).

La méthode selon la présente invention peut également comprendre en outre une étape de validation des résultats obtenus à l'aide d'une analyse prédictive afin de
10 déterminer le taux d'erreur de prédiction d'être atteint ou de développer la maladie de Parkinson.

Cette étape qui peut être réalisée à l'aide de toute méthode connue de l'homme du métier est capable de prédire au mieux le phénotype de nouveaux individus à partir de leurs données d'expressions. Parmi les méthodes connues on
15 peut citer, de manière non limitative, la méthode KNN (*K Nearest Neighbors*, ou méthode des plus proches voisins), la méthode des centroïdes les plus proches (*nearest Shrunk centroid* méthode aussi appelée *Prediction Analysis for Microarray PAM*), la méthode SVM (*Support Vector Machine*), la méthode d'analyse discriminante linéaire (LDA), ...

20 Selon un mode de réalisation particulièrement avantageux, l'ensemble des gènes dont l'expression dérégulée est associée à la maladie de Parkinson est mentionné dans le tableau 2 suivant :

N°	Code puce	Nom du gène	Code Genbank
1	A_23_P216712	Récepteur potentiel transitoire du canal à cation, sous famille M, membre 6 (TRPM6)	NM_017662
2	A_24_P245838	Mannosyl(beta-1,4-)-glycoprotéine beta-1,4-N-acetylglucosaminyltransferase (MGAT3)	AK125361
3	A_23_P369316	Vang-like 1 (VANGL1)	NM_138959
4	A_23_P1904	Domaine transmembranaire 4, sous famille A, membre 2 (MS4A2)	NM_000139
5	A_23_P349321	Xylosyltransferase de type I (XYLT1)	NM_022166

6	A_23_P117662	Histidine decarboxylase (HDC)	NM_002112
7	A_23_P133916	Composant C2 du complément (C2)	NM_000063
8	A_32_P162187	Composant C2 du complément (C2)	NM_000063
9	A_23_P8961	Interleukine 7 (IL7)	NM_000880
10	A_32_P175313	EST PM1-NN1207-071200-003-a12 NM1207	BF960555
11	A_32_P6433	EST yr57d04.r1 ADNcomplémentaire du clone IMAGE:209383 soares foie foetal rate, similaire à la protéine de la famille de choc thermique humaine de 71KD	H64096
12	A_32_P162183	Composant C2 du complément (C2)	NM_000063
13	A_32_P104572	EST UI-E-DW1-ahe-d-15-0-UI.s1 clone de l'ADNcomplémentaire UI-E-DW1-ahe-d-15-0-UI 3'	BM670971
14	A_32_P234485	EST UI-1-BB1p-auq-g-04-0-UI.s1 NCI_CGAP_PI6 ADNcomplémentaire du clone UI-1-BB1p-auq-g-04-0-UI 3'	BQ024427
15	A_23_P60047	protéine 1 contenant des structures acides bi-spiralées (TACC1)	NM_006283
16	A_32_P45875	EST ws88f04.x1 Homo sapiens, ADNcomplémentaire du clone IMAGE:2505055 3', similaire à la séquence voisine de la phospholipase C TR:Q62084 Q62084	AW009873
17	A_23_P352435	Régulateur des protéines G de signalisation 12 (RGS12)	NM_002926
18	A_23_P387649	DNAJA4 homologue à DnaJ (Hsp40), sous-famille A, membre 4 (DNAJA4)ou PRO1472	AF116663
19	A_23_P200252	Cadre ouvert de lecture 57, chromosome 1 (C1orf57)	NM_032324
20	A_23_P66174	Domaine de liaison SH3 de la protéine kinase 1(SBK1)	NM_00102440
21	A_23_P33927	FKSG43	AF334945
22	A_23_P333640	Papiline (PAPLN) glycoprotéine sulfatées similaire à un protéoglycane	NM_173462

23	A_24_P323682	région chromosomique de la sclérose latérale amyotrophique 2 (juvénile)	NM_020919
----	--------------	---	-----------

Tableau 2 : Exemple d'un ensemble de gènes dont l'expression dérégulée est associée à la maladie de Parkinson.

5 Selon un second aspect, la présente invention concerne l'utilisation de réactifs spécifiques d'un ensemble de marqueurs moléculaires identifié selon l'étape b) de la méthode décrite ci-dessus pour déterminer la présence ou le risque de développer la maladie de Parkinson. Dans le cadre de cette utilisation, l'identification de l'ensemble de marqueurs moléculaires, ainsi que l'utilisation des réactifs spécifiques de ces marqueurs afin d'étudier leur expression différentielle sont mises
10 en œuvre *ex-vivo* donc non appliquées au corps humain ou animal.

De préférence, on utilisera des réactifs spécifiques de l'ensemble des 23 gènes mentionnés dans le tableau 2 ci-dessus. Dans cette variante de l'invention, les sondes indiquées dans le tableau 3 ci-dessous peuvent en particulier être utilisées.

Selon un troisième aspect, la présente invention a pour objet un kit ou nécessaire pour la mise en œuvre de la méthode décrite ci-dessus comprenant au moins les éléments suivants :

5 - des réactifs spécifiques (tels que des amorces ou des sondes) de l'ensemble des marqueurs moléculaires tel qu'identifié selon l'étape b) de la méthode de l'invention décrite ci-dessus,

10 - éventuellement les réactifs et/ou dispositifs nécessaires et adaptés en fonction du type de marqueurs moléculaires choisi pour effectuer l'extraction des marqueurs moléculaires à analyser et/ou la détection et/ou la quantification relative de ces marqueurs par rapport à la quantité des marqueurs moléculaires du même type dans un échantillon issu d'un sujet sain (témoin négatif), et

- éventuellement un échantillon contenant des marqueurs moléculaires du type choisi issu d'un patient atteint de la maladie de Parkinson de façon avérée (témoin positif).

15 Selon un mode de réalisation préférée, l'invention concerne un kit ou nécessaire pour détecter dans un échantillon biologique à tester l'expression différentielle d'un ensemble de marqueurs moléculaires associée à la présence ou le risque de développer la maladie de Parkinson, comprenant au moins :

20 - des réactifs spécifiques des 23 gènes indiqués dans le tableau 2 ci-dessus,
- éventuellement les réactifs et/ou dispositifs nécessaires et adaptés en fonction du type de marqueurs moléculaires choisi pour effectuer l'extraction des marqueurs moléculaires à analyser et/ou la détection et/ou la quantification relative de ces marqueurs par rapport à la quantité des marqueurs moléculaires du même type dans un échantillon issu d'un sujet sain (témoin négatif), et

25 - éventuellement un échantillon contenant des marqueurs moléculaires du type choisi issu d'un patient atteint de la maladie de Parkinson de façon avérée (témoin positif).

30 En particulier, l'invention concerne un kit ou nécessaire pour détecter dans un échantillon biologique à tester l'expression différentielle d'un ensemble de marqueurs moléculaires associée à la présence ou le risque de développer la maladie de Parkinson, comprenant au moins :

35 - des réactifs spécifiques des 23 gènes indiqués dans le tableau 2 ci-dessus,
- éventuellement les réactifs et/ou dispositifs nécessaires pour effectuer l'extraction des ARNs et/ou la détection et/ou la quantification relative d'ARN par rapport à la quantité d'ARN dans un échantillon issu d'un sujet sain (témoin négatif), et

- éventuellement un échantillon d'ARN issu d'un patient atteint de la maladie de Parkinson de façon avérée (témoin positif).

De préférence, les réactifs spécifiques des marqueurs moléculaires identifiés selon l'étape b) de la méthode ci-dessus sont choisis parmi des amorces, des sondes,
5 des anticorps, etc.

De préférence, le témoin positif fourni dans les kits selon l'invention est un échantillon issu d'un sujet présentant une mutation pathogène responsable de la maladie de Parkinson à un stade HY de la maladie supérieur ou égal à 3.

De préférence également, le témoin négatif fourni dans les kits selon
10 l'invention est un échantillon issu d'un sujet indemne de tout antécédent personnel ou familial de la maladie de Parkinson, sans autre maladie apparente (et en particulier sans autre maladie neurodégénérative ou syndrome parkinsonien), ni prise de traitement médical.

Selon un mode de réalisation particulièrement avantageux, le kit selon
15 l'invention contient des réactifs spécifiques des 23 gènes du tableau 2, et en particulier les sondes qui sont indiquées dans le tableau 3 ci-dessus.

La présente invention sera décrite plus en détails dans les exemples non limitatifs qui suivent et qui se réfèrent aux figures annexées dans lesquelles :

La **figure 1** représente la corrélation entre les mesures d'expression
20 obtenues par des micro-puces et celles obtenues par RT-PCR quantitative. L'axe des ordonnées représente la différence d'expression ($FC = \text{Fold Change}$) des données des micro-puces exprimées en \log_2 , et l'axe des abscisses celle obtenue en RT-PCR quantitatives des données exprimées en \log_2 . Les noms des gènes dont l'expression a été mesurée par ces deux techniques sont indiqués sur le graphique.

25 La **figure 2A** représente le classement des échantillons du groupe « d'entraînement » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 233 gènes significativement dérégulés chez les patients porteurs de mutations du gène PRKN.

30 Le groupe d'entraînement comprend des sujets sains (T1 à T20) et des sujets porteurs de mutations dans le gène PRKN et développant la maladie de Parkinson (P13 à P16).

L'axe des abscisses représente le nombre de sujets dans le groupe d'entraînement.

35 L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

Les **figures 2B, 2C, 2D** représentent le classement des échantillons du groupe « test » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 233 gènes significativement dérégulés chez les patients porteurs de mutations du gène PRKN. La **figure 2B** représente le classement des patients atteints de maladies neurodégénératives autres que la maladie de Parkinson (A1 à A11). La **figure 2C** représente le classement des sujets atteints de formes familiales de maladie de Parkinson autres que des cas familiaux porteurs de mutation du gène PRKN (P1 à P12 et P17 à P29). La **figure 2D** représente le classement des sujets atteints de formes sporadiques de la maladie de Parkinson (S1 à S17). L'axe des abscisses représente le nombre de sujets (A, P ou S) dans le groupe « test ».

L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

La **figure 3A** représente le classement des échantillons du groupe « d'entraînement » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade m (carré noir) à partir de l'ensemble des 176 gènes significativement dérégulés chez les patients porteurs de mutations du gène SCA2.

Le groupe d'entraînement comprend des sujets sains (T1 à T20) et des sujets porteurs de mutations dans le gène SCA2 et développant la maladie de Parkinson (P17 à P22).

L'axe des abscisses représente le nombre de sujets dans le groupe d'entraînement.

L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

Les **figures 3B, 3C, 3D** représentent le classement de l'échantillon « test » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 176 gènes significativement dérégulés chez les patients porteurs de mutations du gène SCA2. La **figure 3B** représente le classement des patients atteints de maladies neurodégénératives autres que la maladie de Parkinson (A1 à A11). La **figure 3C** représente le classement des sujets atteints de formes familiales de maladie de Parkinson autres que des cas familiaux porteurs de mutation du gène SCA2 (P1 à P16 et P23 à P29). La **figure 3D** représente le classement des sujets atteints de formes sporadiques de la maladie de Parkinson (S1 à S17).

L'axe des abscisses représente le nombre de sujets (A, P ou S) dans le groupe « test ».

L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

5 La **Figure 4A** représente le classement des échantillons du groupe « d'entraînement » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 438 gènes significativement dérégulés chez des patients porteurs de la mutation LRRK2, excepté les sujets P6 et P12.

10 Le groupe d'entraînement comprend des sujets sains (T1 à T20) et des sujets porteurs de la mutation LRRK2 (P2 à P5 et P7 à P11).

L'axe des abscisses représente le nombre de sujets dans le groupe d'entraînement.

15 L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

20 Les **figures 4B, 4C, 4D** représentent le classement des échantillons du groupe « test » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 438 gènes significativement dérégulés chez des patients porteurs de la mutation LRRK2. La **figure 4B** représente le classement des patients atteints de maladies neurodégénératives autres que la maladie de Parkinson (A1 à A11). La **figure 4C** représente le classement des sujets atteints de formes familiales de maladie de Parkinson autres que des cas familiaux porteurs de mutation du gène LRRK2 (excepté pour deux d'entre eux P6 et P12) (P1, P6, P12 à P29). La **figure 4D**

25 représente le classement des sujets atteints de formes sporadiques de la maladie de Parkinson (S1 à S17).

L'axe des abscisses représente le nombre de sujets (A, P ou S) dans le groupe « test ».

30 L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

35 La **Figure 5A** représente le classement des échantillons du groupe « d'entraînement » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 23 gènes significativement dérégulés chez des patients atteints de différentes formes familiales de la maladie de Parkinson.

Le groupe « d'entraînement » comprend des sujets sains (T1 à T20) et des sujets porteurs de mutations des gènes PRKN, SCA2, LRRK2 ou SNCA développant des formes familiales et héréditaires de la maladie de Parkinson (P1 à P22, excepté les sujets P6 et P12).

5 L'axe des abscisses représente le nombre de sujets dans le groupe d'entraînement.

L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

10 Les **figures 5B, 5C, 5D** représentent le classement des échantillons du groupe « test » en fonction de la probabilité que des sujets puissent appartenir au groupe témoin (t, rond blanc) ou au groupe malade (m, carré noir) à partir de l'ensemble des 23 gènes significativement dérégulés chez des patients atteints de différentes formes familiales de la maladie de Parkinson. La **figure 5B** représente le classement des patients atteints de maladies neurodégénératives autres que la
15 maladie de Parkinson (A1 à A11). La **figure 5C** représente le classement des sujets atteints de formes familiales de maladie de Parkinson autres que des cas familiaux porteurs de mutation des gènes PRKN, SCA2, LRRK2 (excepté pour deux d'entre eux P6 et P12) ou SNCA (P6, P12, P23 à P29). La **figure 5D** représente le classement des sujets atteints de formes sporadiques de la maladie de Parkinson (S1 à S17).

20 L'axe des abscisses représente le nombre de sujets (A, P ou S) dans le groupe « test ».

L'axe des ordonnées représente le taux de probabilité qu'un sujet appartienne au groupe témoin ou au groupe malade.

EXEMPLES

25 **A - POPULATIONS DE PATIENTS**

Les patients ont été recrutés au sein du Centre Hospitalier Universitaire de Lille dans le cadre de l'étude PARKFANORD (*PARKinson FAmilial NORD*).

30 Les sujets dits « malades traités ou symptomatiques » atteints de maladie de Parkinson ont été diagnostiqués selon les critères cumulatifs 1) 2) et 3) suivants acceptés en la matière, à savoir

1) au moins 3 des 4 signes cardinaux de la maladie de Parkinson, à savoir tremblement de repos, bradykinésie, rigidité musculaire ou instabilité posturale; la bradykinésie faisant partie obligatoirement d'un des 3 signes cardinaux retenus pour le diagnostic,

35 2) la réponse positive au traitement « anti-parkinsonien » levodopa, et

3) l'absence des 15 critères d'exclusion définis par UK PDSBB, Hughes A.J., *Journal of Neurology, Neurosurgery, and Psychiatry*, 1992, **55**, 181-184. Le 16^{ème} critère d'antécédent familiale n'a pas été retenu.

Trois types de populations de patients ont été recrutés :

- 5 - les sujets (P) ayant un lien de parenté avec au moins une personne atteinte de la maladie de Parkinson (Tableau 4), le caractère familial ou non (lien de parenté) ayant été déterminé en fonction de l'arbre généalogique (formes héréditaires ou familiales). Dans cette population, il existe à la fois des sujets qui
- 10 présentent des symptômes (« malades traités », « symptomatiques » et « pauci-symptomatiques ») ou qui ne présentent pas de symptômes de la maladie de Parkinson (« asymptomatiques »). Les sujets dits « pauci-symptomatiques » sont des formes familiales de la maladie et qui présentent, en plus de la réponse positive au traitement « anti-parkinsonien » levodopa et de l'absence des 15
- 15 critères d'exclusion définis par UK PDSBB, au moins deux des quatre signes cardinaux dont la bradykinésie.
- les sujets (S) atteints de la maladie de Parkinson selon les critères 1) 2) et 3) tels que définis ci-dessus pour les sujets dits « malades traités ou symptomatiques », mais n'ayant pas de lien de parenté avec une autre personne atteinte de la maladie de Parkinson (Tableau 5),
- 20 - les sujets (A) développant un syndrome parkinsonien ou une autre maladie neurodégénérative apparentée (Tableau 6). Ces sujets ne répondent pas aux critères de diagnostic de la maladie de Parkinson tels que définis ci-dessus (sujets dits « malades traités ou symptomatiques ») et notamment en ce qui concerne la réponse au traitement levodopa et les critères d'exclusion.
- 25 En parallèle de ces trois types de populations, quarante individus sains indemnes d'antécédent personnel ou familial de maladie de Parkinson ou toute autre maladie neurodégénérative ont été également recrutés en tant que témoins (T). Dix-neuf échantillons d'ARN provenant de dix-neuf individus témoins ont été inclus dans
- 30 l'étude d'expression par micro-puces, ainsi qu'un pool comportant des quantités égales d'ARN provenant de vingt hommes et de vingt femmes, d'âge compris entre 22 et 82 ans (moyenne = 52,9 ans) afin d'avoir un échantillon représentatif de ces quarante témoins.

Les prélèvements sanguins ont été effectués sur chaque type de population (individus malades ou témoins) après recueil du consentement éclairé des sujets.

Cas familial	Nom du gène muté	Statut clinique	Stade HY	Tremblement de repos	Bradykinesie	Rigidité	Instabilité posturale	Début Asymétrique	Score UPDRS III
P1	SNCA	paucisymptomatique	1	0	1	1	0	1	5
P2	LRRK2	malade traitée	4	1	1	1	1	1	20
P3	LRRK2	asymptomatique	0	0	0	0	0	0	0
P4	LRRK2	malade traitée	2	1	1	1	0	1	18
P5	LRRK2	symptomatique	1,5	1	1	1	0	1	15
P6	LRRK2	malade traitée	4	1	1	1	1	1	34
P7	LRRK2	malade traité	2	1	1	1	0	1	34
P8	LRRK2	malade traité	2	1	1	1	1	1	20
P9	LRRK2	malade traitée	3	0	1	1	1	1	16
P10	LRRK2	malade traité	3	1	1	1	1	1	56
P11	LRRK2	malade traitée	2,5	1	1	1	1	1	36
P12	LRRK2	malade traité	2	1	1	1	0	1	33
P13	PRKN	malade traité	3	1	1	1	1	1	30
P14	PRKN	malade traitée	3	1	1	1	1	1	65
P15	PRKN	malade traitée	2	1	1	1	0	1	10
P16	PRKN	malade traité	2	1	1	1	0	1	15

cas sporadique	Statut clinique	Stade HY	Tremblement de repos	Brady-kinesie	Rigidité	Instabilité posturale	Début Asymétrique	Score UPDRS III
S1	malade traité	3	1	1	1	1	1	47
S2	malade traité	4	0	1	1	1	1	32
S3	malade traité	2	1	1	1	1	1	21
S4	malade traité	3	1	1	1	1	1	26
S5	malade traité	3	0	1	1	1	0	34
S6	malade traité	2	1	1	1	0	1	20
S7	malade traité	2	0	1	1	1	0	18
S8	malade traité	2	1	1	1	0	1	11
S9	malade traité	3	0	1	1	1	1	21
S10	malade traité	2	1	1	1	0	1	13
S11	malade traité	2	0	1	1	1	1	19
S12	malade traité	2	1	1	1	0	1	12
S13	malade traité	3	1	1	1	1	1	26
S14	malade traité	2	1	1	1	1	1	31
S15	malade traité	2	1	1	1	0	1	16
S16	malade traité	2	1	1	1	1	1	16
S17	malade traité	2	1	1	1	0	1	24

Tableau 5 : Description des symptômes et du stade de gravité des cas sporadiques de la maladie de Parkinson

L'absence ou la présence de symptômes est indiquée par 0 et 1 respectivement

Le score de la partie III de l'échelle UPDRS est indiqué. Plus le score est élevé, plus l'examen moteur est défavorable et plus le patient présente des troubles au niveau de la parole, des gestes de la vie courante et une démarche perturbée, gênants pour sa vie courante.

cas	Nom du gène muté	Autre maladie neurodégénérative
A1	SCA2	SCA
A2	SCA2	SCA
A3	SCA2	SCA
A4	SCA2	SCA
A5	SCA2	SCA
A6	SCA2	SCA
A7	inconnu	MSA
A8	inconnu	DLB
A9	inconnu	DLB
A10	inconnu	DLB
A11	inconnu	DLB

Tableau 6 : Description des patients ayant une autre maladie neurodégénérative ou syndrome parkinsonien répondant aux critères en vigueur

- 5 Les patients A1 à A6 sont porteurs d'une mutation SCA2-C.

B - ECHANTILLONS BIOLOGIQUES ET MARQUEURS MOLECULAIRES

1. Récupération des cellules mononucléées sanguines

10 Les échantillons sanguins, prélevés dans des tubes CPT® (BD Vacutainer) contenant un gradient de FICOLL®, sont centrifugés pendant 20 minutes à 1750g à 23°C. Les cellules mononucléées sont recueillies. Après lavage, des aliquots contenant $0,5 \times 10^7$ cellules sont réalisés en ajoutant dans chaque tube 700µL de tampon de lyse (kit RNeasy mini de QIAGEN®) contenant 1% de β-mercaptoéthanol.

2. Extraction des ARN totaux

15 L'extraction de l'ARN est réalisée grâce au kit RNeasy mini et au kit RNase-free DNase de QIAGEN® selon les recommandations du fabricant. Brièvement, les cellules sont lysées mécaniquement. Un volume d'éthanol à 70% est ensuite incorporé à chaque échantillon et le mélange est homogénéisé à la pipette. Le mélange ainsi obtenu est déposé sur une colonne contenant une membrane en silice
20 qui va retenir les acides nucléiques. Après un lavage, un traitement à la DNase I est appliqué pour éliminer toute trace d'ADN résiduel. Trois autres lavages permettent

d'éliminer la DNase, l'ADN et les traces d'éthanol. L'ARN est récupéré dans un volume de 30µl d'eau RNase-free par une centrifugation d'une minute à 10000g. L'échantillon d'ARN total purifié est conservé à -80°C.

3. Quantification et vérification de la qualité de l'ARN

5 Le dosage des ARN est effectué grâce au spectrophotomètre Nano-Drop® ND-1000 par mesure de l'absorbance à 260 nm à partir d'1µl d'échantillon. Le calcul des rapports d'absorbance 260 nm/280 nm et 260 nm/230 nm permet d'évaluer la pureté de l'ARN.

10 L'analyse de la qualité de l'ARN est réalisée grâce au Bioanalyzer RNA 6000 nano d'Agilent®. Les échantillons à analyser (1µl) et un marqueur de poids moléculaire sont déposés dans chaque puits de la puce RNA 6000 nano composée d'un gel mélangé à un fluorophore (intercalant des acides nucléiques simple-brin). Les puits sont analysés par le Bioanalyzer 2100 d'Agilent®. Les données utilisées pour vérifier la qualité des ARN sont le rapport des fractions d'ARN ribosomal 15 28S/18S et le RIN pour *RNA Integrity Number* qui est un algorithme développé par Agilent® permettant de coter la dégradation de l'ARN sur une échelle allant de 1 à 10.

C - ETUDE DE L'EXPRESSION DES GENES

1. Pucés à ADN complémentaire

20 La technologie des pucés à ADNc permet d'analyser de façon simultanée l'ensemble des transcrits du génome. Les pucés d'expression utilisées sont développées par la firme Agilent® et couvre entièrement le génome humain grâce à près de 44 000 sondes (44K *whole human genome expression arrays* (Agilent®, 4 X 44K slide formats). Ces sondes reposent sur la technologie de synthèse *in situ* 25 d'oligonucléotides de 60-mers sur une lame de verre qui vont servir à capter des cibles marquées d'ADN complémentaire (ADNc). A partir de l'ARN total (300 ng), sont préparées les cibles d'ADNc par rétro-transcription. A chaque échantillon, ont été rajoutés des « *spike-in* » qui sont des ARN servant de contrôles positifs lors des étapes d'amplification et de marquage. Puis la T7 RNA polymérase amplifie 30 l'échantillon et incorpore des CTP marqués à la Cyanine 3. Les ADNc marqués sont ensuite fragmentés en des oligomères de 60 à 100 nucléotides permettant une hybridation optimale avec les sondes. L'hybridation des échantillons sur la puce s'effectue dans une chambre fermée hermétiquement et placée dans un four pendant 17h à 65°C. Les lames sont ensuite lavées puis séchées et lues au scanner. 35 Le scanner possède un laser qui excite les cyanines 3 afin de mesurer l'intensité de fluorescence ainsi émise. Le logiciel d'extraction des données (*Feature Extraction*

d'Agilent[®]) permet d'éliminer le bruit de fond et de vérifier la qualité des spots sur les lames. A chaque sonde sur la lame correspond une intensité de fluorescence. Une normalisation des données est nécessaire car des biais expérimentaux existent. En effet, il peut apparaître des différences dans le taux d'incorporation des cyanines dans l'ADNc, une mauvaise qualité de détection de la fluorescence par le scanner ; d'autres biais peuvent être occasionnés par le dépôt, l'hybridation.

2. Contrôles qualité

Chacun des échantillons d'ARN utilisés pour les puces a montré une qualité satisfaisante des rapports d'absorbance 260 nm/280 nm (compris entre 1,8 et 2), et de 28S/18S (>1,5) et de RIN (>8.9). L'analyse des rapports de qualité des puces par « *Feature Extraction* » a montré des intensités de fluorescence des spots qui respectaient une loi normale.

3. Traitement des données issues des puces

L'analyse des données s'effectue avec le logiciel GeneSpring GX 7.3 (Agilent[®]), regroupant un ensemble de fonctions d'analyse statistique.

Dans un premier temps, l'analyse des résultats des puces comprend, classiquement, une étape de normalisation des données par puce et par gène sur la médiane des intensités de fluorescence des sujets témoins. Les valeurs normalisées sont exprimées en log ratio.

Les données d'expression sont ensuite classées en fonction du stade HY de la maladie de Parkinson du sujet.

Série A : sujets atteints de la maladie de Parkinson dont le stade HY est de 0 à 1,5 inclus.

Série B : sujets atteints de la maladie de Parkinson dont le stade HY est de 2 à 2,5 inclus.

Série C : sujets atteints de la maladie de Parkinson dont le stade HY est de 3 à 5 inclus.

Chaque série (A, B, C) est comparée de manière indépendante à l'ensemble des données d'expression des sujets témoins. Cette comparaison s'effectue par un filtrage en Volcano plot en appliquant l'analyse statistique par le test T de Welch qui est définie par la formule :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

où \bar{X}_i , s_i^2 et N_i sont respectivement le n^{eme} de la moyenne de l'échantillon, variance de l'échantillon et taille de l'échantillon, avec un seuil de significativité $p=0,05$ et un seuil de variation d'expression (*Fold Change*) de 1,2. Ainsi les gènes dérégulés dont le seuil de variation d'expression est inférieur à 1,2 ne seront pas
 5 retenus pour le reste de l'analyse. A l'issue de cette étape, on obtient pour chaque série une liste de gènes spécifiquement dérégulés pour un stade HY donné (liste_A, liste_B, liste_C).

L'ensemble des données d'expression des sujets atteints de maladies neurodégénératives ou atteints d'un syndrome parkinsonien autres que la maladie de
 10 Parkinson est comparé avec l'ensemble des données d'expression des témoins en appliquant les mêmes critères que précédemment pour l'analyse statistique par le test T de Welch. On obtient une liste de gènes spécifiquement dérégulés chez des sujets atteints de maladies neurodégénératives ou atteints d'un syndrome parkinsonien autres que la maladie de Parkinson (liste_D).

15 Les gènes dont l'expression est significativement dérégulée en commun aux différents stades HY de la maladie de Parkinson (liste_E) sont sélectionnés grâce à une représentation graphique en diagrammes de Venn en croisant les listes : liste_A, liste_B et liste_C. La dernière étape consiste à exclure de cette liste (liste_E), les gènes qui sont communs avec ceux qui sont aussi significativement dérégulés chez
 20 des patients atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens (liste_D). On obtient ainsi un ensemble de gènes dérégulés spécifiquement chez les patients atteints de la maladie de Parkinson (liste_F).

Le mot « gène » est employé pour désigner les sondes Agilent[®] de la puce. Toutefois plusieurs sondes peuvent être présentes pour un même gène et des
 25 sondes peuvent correspondre encore à des gènes potentiels voire même des séquences dont le rôle n'est pas encore connu.

4. Validation éventuelle des résultats obtenus par une analyse prédictive avec le logiciel PAM

Les données d'expression correspondant à l'ensemble des gènes obtenus
 30 comme défini au paragraphe C 3. sont introduites dans le logiciel PAM (*Prediction Analysis for Microarray*) (version 2.1 ; <http://www-stat.stanford.edu/~tibs/PAM/>) qui applique l'analyse des centroïdes réduits les plus proches (« *nearest shrunken centroid analysis* »). Le nombre optimal de gènes est déterminé à partir de la méthode des validations croisées (10 fois) à partir du groupe « d'entraînement ».

35 Le groupe « d'entraînement » est constitué par les sujets pour lesquels l'appartenance à une classe est connue (la classe étant définie par groupe malade ou

groupe témoin). Les données d'expression des gènes de ces sujets sont fournies pour l'analyse. Le groupe « d'entraînement » permet de déterminer l'ensemble des gènes dont l'expression est la plus corrélée au diagnostic (malade/témoin).

5 Puis cet ensemble de gènes est utilisé pour prédire le diagnostic des sujets du groupe test. Le groupe « test » est un groupe constitué par des sujets pour lesquels l'appartenance à la classe sera prédite en fonction des règles apprises du groupe « d'entraînement ». La véritable appartenance de ces échantillons à la classe peut être connue ou non.

10 Ainsi concernant le groupe « d'entraînement », un sujet P sera correctement classé en tant que malade lorsque sa probabilité sera proche de 1 et qu'il sera représenté par le symbole qui correspond au groupe malade. Un sujet T sera correctement classé en tant que sujet témoin lorsque sa probabilité sera proche de 1 et qu'il sera représenté par le symbole qui correspond au groupe témoins.

15 Pour le groupe « test », un sujet ayant un risque de développer la maladie de Parkinson sera correctement classé en tant que malade lorsque sa probabilité sera proche de 1 et qu'il sera représenté par le symbole qui correspond au groupe malade.

20 Le taux d'erreur permet d'évaluer l'efficacité de l'ensemble des gènes utilisés dans la méthode de détection de marqueurs moléculaires pour déterminer la présence ou le risque de développer la maladie de Parkinson. Le taux d'erreur correspond au nombre de sujets mal classés sur le nombre total de sujet de cette classe; deux classes sont possibles : malade parkinsonien ou témoin (sain ou atteint d'un syndrome parkinsonien ou d'une autre pathologie neurodégénérative que la maladie de Parkinson).

25 **EXEMPLE 1 : Concordance des résultats obtenus par l'utilisation des micro-puces avec ceux obtenus par d'autres techniques de mesure d'expression telles que la RT-PCR quantitative (exemple de vérification technique)**

30 Une analyse par RT-PCR quantitative de l'expression de 12 gènes, dont certains sont choisis parmi ceux qui sont connus pour être dérégulés lorsque le gène LRRK2 est muté et présentant une expression différentielle d'au moins un facteur 2 (*Fold Change* >2 ou <0,5) significative en test T de Welch ($p < 0,05$), a été réalisée. Le niveau d'expression par PCR quantitative de certains gènes classiquement dérégulés dans les formes génétiques de maladie de Parkinson a été également étudié.

Cette analyse a été réalisée en parallèle de l'analyse par micro-puces afin de vérifier la concordance des résultats (c'est-à-dire du niveau d'expression des gènes) obtenus par ces deux techniques.

a. Etape de rétro-transcription (RT)

5 La rétro-transcription de l'ARN en ADNc est réalisée grâce au kit QuantiTect® Reverse Transcription (Qiagen) à partir de 250 ng d'ARN conformément au protocole du fabricant. L'ADN éventuel résiduel est d'abord éliminé par un traitement à la DNase avant d'effectuer la réaction de rétro-transcription. Les amorces utilisées sont un mélange d'hexamères et d'oligo-dT permettant la synthèse d'ADNc à partir de
10 l'ensemble de l'ARN.

b. Etape de PCR en temps réel

La PCR quantitative a été réalisée selon les recommandations du fabricant à l'aide du kit SYBR® Green PCR Master (Applied Biosystems). Cette étape permet de suivre en temps réel l'amplification d'un fragment d'ADNc cible grâce au fluorophore
15 SYBR® Green. Cet agent intercalant de l'ADN devient fluorescent lorsqu'il est lié à l'ADN double brin. La quantité de fluorescence est mesurée à la fin de l'étape d'élongation de chaque cycle de PCR. Cette mesure permet de détecter à quel cycle (Ct pour *cycle threshold*) la fluorescence dépasse le seuil déterminé par l'utilisateur, sur la courbe d'amplification. Cette technique permet de calculer l'expression
20 normalisée relative du gène cible. La normalisation se fait par rapport à un gène de référence (ARN ribosomal 18S, amorce sens CCTGGATACCGCAGCTAGGA (SEQ ID NO : 24); amorce anti-sens GCGGCGCAATACGAATGCCCC (SEQ ID NO : 25); température d'hybridation 60°C). La méthode de calcul est dérivée de la méthode $2^{-\Delta\Delta Ct}$ pour tenir compte du fait que l'efficacité n'est pas de 100%. L'expression normalisée est donc le
25 rapport : $(1/\text{efficacité du gène cible}^{Ct \text{ du gène cible}})/(1/\text{efficacité du gène référent}^{Ct \text{ du gène référent}})$.

Le mélange comporte 10µl d'eau, 0,75µl de chacune des amorces (10 pmol/µl) et 12,5µl de master (2X) pour un puits. Un µl d'ADNc est ajouté au mélange réactionnel. Les mesures des échantillons sont réalisées en double exemplaire. La
30 PCR en temps réel est réalisée sur un appareil ABI Prism® 7700 (Applied Biosystems). L'analyse statistique des rapports d'expression entre les individus malades et les témoins est réalisée pour chaque gène par un test de Mann et Whitney (seuil de significativité $p < 0,05$).

Les séquences des amorces utilisées pour réaliser la PCR sont indiquées dans le
35 tableau 7 suivant :

Gènes (numéro d'accion GenBank)	Description	Séquences des Amorces (5' -> 3')	Identificateurs de séquence	Taille du produit de PCR	Température d'hybridation (°C)
ARG1 (NM_000045)	Arginase1	sens ACCCATCTTTACACCAGC anti-sens TCCTGAGAGTAGCCCTGTTT	SEQ ID NO : 26 SEQ ID NO : 27	107 pb	58
ERAF (NM_016633)	Facteur associé à l'Erythroïde	sens AATCAGCAGGTCTTCAATG anti-sens GCTGCCTGTAATAGTTGATG	SEQ ID NO : 28 SEQ ID NO : 29	97	53
SNCA (NM_000345)	α Synucléine	sens CCAAAACCAAGGAGGGAGTG anti-sens ACACCCGTCACCACTGCTCC	SEQ ID NO : 30 SEQ ID NO : 31	97 pb	60
LRRK2 (NM_198578)	Kinase 2 riche en motif leucine répété	sens AGGTACACAAAAGCAGAAAAG anti-sens ACTTCAATGTGTTTTCTAA	SEQ ID NO : 32 SEQ ID NO : 33	96 pb	57
PINK1 (NM_032409)	Kinase 1 induite par PTEN	sens AGAGGCCAGCAAGAGACC anti-sens CTAACCTTCAGATTCTTCAGGGC	SEQ ID NO : 34 SEQ ID NO : 35	98 pb	58
DJ1 (NM_007262)	PARK 7 Parkinson disease (autosomal recessive, early-onset)7	sens CCATCTGTGCAGGTCCTA anti-sens CTCATTTCATCTTTGTCT	SEQ ID NO : 36 SEQ ID NO : 37	99 pb	55
IFNG (NM_000619)	Interféron gamma	sens GGAAGAGGAGAGTGACAGA anti-sens ATGTCTTCTTGATGGTCTC	SEQ ID NO : 38 SEQ ID NO : 39	124 pb	56
DUSP10 (NM_007207)	Phosphatase 10 à double spécificité	sens GGCTTTTGAGTTCATTGAGG anti-sens GTCATCCGAGTGTGCTTCAT	SEQ ID NO : 40 SEQ ID NO : 41	123 pb	64
LTF (NM_002343)	Lactotransferrine	sens GCTACTCTGGTGCCCTTCAA anti-sens AGGTCTCAAACTGTGC	SEQ ID NO : 42 SEQ ID NO : 43	82 pb	58
SELENBP1 (NM_003944)	Protéine 1 de liaison au sélénium	sens AAGAACGAGGGAGGTACATG anti-sens AGAGCAGGATGTCGGTGAT	SEQ ID NO : 44 SEQ ID NO : 45	112 pb	55
ALAS2 (NM_000032)	Aminolevulinatase delta synthétase 2	sens CTGCCAGGGTGCAGATTTA anti-sens GTGGTCAGGGTTCATTGTGCC	SEQ ID NO : 46 SEQ ID NO : 47	111 pb	55
UGT2B17 (NM_001077)	UDP glucuronosyl transférase 2 famille, polypeptide B17	sens CACAAAAGGTTCTATGGAG anti-sens AAGGTCATTCTGGGGTA	SEQ ID NO : 48 SEQ ID NO : 49	92 pb	58

Tableau 7 : séquences des amorces, tailles des amplimères et températures d'hybridation pour chaque gène étudié en PCR quantitative. Les séquences des amorces ont été conçues grâce au logiciel Oligo 5©.

Les résultats sont regroupés dans la **Figure 1** et montrent une forte corrélation ($r^2=0.94$) entre ces deux techniques. Ceci signifie que les augmentations ou les diminutions d'expression des gènes observées sur les puces peuvent être retrouvées en utilisant une autre technique de quantification avec le même niveau de différence d'expression. Par conséquent, ces variations d'expression ne sont pas liées à des artefacts de la technique de micro-puces mais sont bien le reflet des variations d'expression de gènes dans les cellules mononucléées des patients.

EXEMPLE 2 : Détermination d'une liste de gènes dont l'expression est dérégulée chez des patients atteints d'une forme familiale particulière de la maladie de Parkinson et étude de la capacité d'identification de sujets atteints de formes sporadiques à partir de cette liste

Afin de déterminer si à partir d'un seul type de forme familiale il est possible de prédire le risque de développer la maladie de Parkinson, la méthode décrite aux paragraphes A à C ci-dessus a été appliquée séparément pour chaque forme familiale telle qu'identifiées dans le tableau 4, à savoir pour les mutations des gènes PRKN, LRRK2, SCA2 et SNCA.

L'expression des gènes chez les sujets porteurs de ces différentes mutations est étudiée à partir de l'ARN extrait des cellules mononucléées sanguines de ces patients et analysée par des micro-puces Agilent® comme décrit aux paragraphes B et C ci-dessus.

La valeur de prédiction (taux d'erreur) de l'ensemble de gènes ainsi calculée pour déterminer la présence ou le risque de développer la maladie de Parkinson, et en particulier une forme sporadique de cette maladie est analysée grâce au logiciel PAM comme décrit ci-dessus au paragraphe C 4.

a. Etude des gènes dérégulés uniquement chez des patients porteurs de mutation du gène de la Parkine PRKN

L'analyse d'expression des gènes dérégulés des sujets porteurs de mutations du gène PRKN par rapport aux témoins sains a été réalisée par un filtrage en Volcano Plot en appliquant un test T de Welch avec un seuil de significativité de $p=0,05$ et un seuil de variation d'expression de 1,2 après classement des sujets selon leur stade Hoehn et Yahr. Cette analyse s'effectue en utilisant le logiciel Genespring GX 7.3.

2 775 gènes composent la liste liste_B des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation PRKN et ayant développé une maladie de Parkinson dont le stade HY est de 2 ou 2,5 (sujets P15, P16).

2 211 gènes composent la liste liste_C des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation PRKN et ayant développés une maladie de Parkinson dont le stade HY est de 3, 4 ou 5 (sujets P13, P14).

- 5 Les gènes spécifiquement dérégulés et communs aux différents stades HY sont sélectionnés dans la liste liste_E (270 gènes).

Cette liste de gènes dérégulés (Liste_E) a ensuite été croisée avec la liste des gènes dérégulés chez les sujets atteints d'autres maladies neurodégénératives ou syndrome parkinsonien (Liste_D qui comporte 5 645 gènes). Après exclusion des
10 gènes qui sont communs avec ceux qui sont aussi significativement dérégulés chez des patients atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens, on obtient ainsi un ensemble de 233 gènes dérégulés spécifiquement chez les sujets porteurs de la mutation PRKN (liste_F).

Cet ensemble de 233 gènes a été analysé par le logiciel PAM :

- 15 - le groupe « d'entraînement » est constitué des 4 sujets porteurs de mutation du gène PRKN (sujets P13, P14, P15, P16) et des 20 témoins (sujets T1 à T20),
- le groupe « test » est constitué des 25 autres sujets familiaux (sujets P1 à P12 et P17 à P29), des 17 sporadiques (sujets S1 à S17) et des 11 sujets atteints
20 d'un autre syndrome parkinsonien (sujets A1 à A11).

Les résultats obtenus à partir du groupe « d'entraînement » sont représentés dans le tableau ci-dessous et dans la **figure 2A**.

Résultats du groupe d'entraînement (seuil = 0,01)				
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %
Sujets P	Malades	4	0	0
Sujets T	Témoins	0	20	0

- 25 Grâce à la série de 233 gènes, tous les patients porteurs de mutation dans le gène PRKN sont représentés par un carré noir dans le groupe des malades avec une probabilité de 1. Par conséquent, ils sont bien identifiés comme des patients malades. De même, les échantillons issus des individus sains sont représentés par un
30 rond blanc avec une probabilité supérieure à 0,5 (ici égale à 1) et sont donc bien identifiés dans le groupe des témoins (**figure 2A**).

- En revanche dans le groupe « test », la **figure 2D** et le tableau ci-dessous montrent que tous les malades qui développent une forme sporadique (sujets S1 à S17) sont représentés par le symbole « témoin » (rond blanc) et avec une probabilité supérieure à 0,5 (ici égale à 1) dans le groupe malade (taux d'erreur de 100 %).
- 5 Pour être correctement classés, il aurait fallu qu'ils soient représentés par le symbole « malade » (carré noir) et classés dans le groupe malade avec une probabilité supérieure à 0,5.

Résultats du groupe test (seuil = 0,01)					
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %	Figures
Sujets A	Témoins	0	11	0	2B
Sujets P	Malades	3	22	88,00	2C
Sujets S	Malades	0	17	100,00	2D

- 10 Seuls les profils des patients P4, P5 et P26 sont reconnus comme appartenant à des patients parkinsoniens (tableau ci-dessus). Ils sont représentés par un carré noir avec une probabilité supérieure à 0,5 dans le groupe des malades (**figure 2C**). Le taux d'erreur de classement de ce groupe est de 88%.

- 15 Des analyses complémentaires ont été effectuées en diminuant dans PAM le nombre de gènes inclus dans la liste liste _F, mais les différents ensembles de gènes obtenus au cours de ces analyses ne permettent pas non plus de distinguer les cas sporadiques des sujets témoins.

- 20 Ces résultats montrent donc que l'ensemble des 233 gènes identifié uniquement à partir des informations des patients porteurs de mutations du gène PRKN ne permet pas de classer correctement les sujets sporadiques ni les autres cas familiaux.

b. Etude des gènes dérégulés uniquement chez des porteurs d'une mutation du gène SCA2 donnant naissance à une maladie de Parkinson

- 25 L'analyse d'expression des gènes dérégulés des sujets porteurs de mutation du gène SCA2 par rapport aux témoins sains a été réalisée par un filtrage en Volcano Plot en appliquant un test T de Welch avec un seuil de significativité de $p=0,05$ et un seuil de variation d'expression de 1,2 après classement des sujets selon leur stade Hoehn et Yahr.

4 757 gènes composent la liste liste_A des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation SCA2 et ayant développés une maladie de Parkinson dont le stade HY est de 0, 1 ou 1,5 (sujets P18, P22).

5 4 288 gènes composent la liste liste_B des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation SCA2 et ayant développés une maladie de Parkinson dont le stade HY est de 2 ou 2,5 (sujets P17, P19, P20).

10 6 446 gènes composent la liste liste_C des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation SCA2 et ayant développés une maladie de Parkinson dont le stade HY est de 3, 4 ou 5 (sujets P21).

Les gènes spécifiquement dérégulés et communs aux différents stades HY sont sélectionnés dans la liste liste_E (217 gènes).

15 Cette liste de gènes dérégulés (Liste_E) a ensuite été croisée avec la liste des gènes dérégulés chez les sujets atteints d'autres maladies neurodégénératives ou syndrome parkinsonien (Liste_D qui comporte 5 645 gènes). Après exclusion des gènes qui sont communs avec ceux qui sont aussi significativement dérégulés chez des patients atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens, on obtient ainsi un ensemble de 176 gènes dérégulés spécifiquement
20 chez les sujets porteurs de la mutation SCA2 (Liste_F).

Cet ensemble de gènes a été analysé à l'aide du logiciel PAM.

Le groupe « d'entraînement » est constitué des 6 sujets porteurs de mutation du gène SCA2 (sujets P17 à P22) et des 20 témoins (sujets T1 à T20).

25 Le groupe « test » est constitué des 23 autres sujets familiaux (sujets P1 à P16 et P23 à P29), des 17 sujets sporadiques (sujets S1 à S17) et des 11 sujets atteints d'un autre syndrome parkinsonien (sujets A1 à A11) :

Les résultats obtenus à partir du groupe « d'entraînement » sont représentés dans le tableau ci-dessous et dans la **figure 3A** :

Résultats du groupe d'entraînement (seuil=0,01)				
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %
Sujets P	Malades	3	3	50
Sujets T	Témoins	0	20	0

Grâce à cet ensemble de gènes, 3 sujets porteurs de mutation SCA2 sur les 6 sont bien identifiés comme des patients malades et tous les témoins sont identifiés comme des témoins.

- 5 En revanche dans le groupe « test », le tableau ci-dessous et la **figure 3C** montrent que 2 malades seulement sont reconnus dans la classe des malades (P16 et P26). Trois sujets (A1, A5, A10) présentant une autre maladie neurodégénérative sont mal classés (**figure 3B**) et aucun des sujets sporadiques n'a été identifié comme malade (**figure 3D**). Pour être correctement classés, il aurait fallu qu'ils
10 soient représentés par le symbole « malade » (carré noir) et classés dans le groupe malade avec une probabilité supérieure à 0,5.

Résultats du groupe test (seuil = 0,01)					
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %	Figures
Sujets A	Témoins	3	8	27,27	3B
Sujets P	Malades	2	21	91,30	3C
Sujets S	Malades	0	17	100	3D

- 15 Des analyses complémentaires ont été effectuées en diminuant dans PAM le nombre de gènes inclus dans la liste liste _F, mais les différents ensembles de gènes obtenus au cours de ces analyses ne permettent pas non plus de distinguer les cas sporadiques des témoins.

- 20 Ces résultats montrent donc que la série des 176 gènes identifiée uniquement à partir des informations des patients porteurs de mutations du gène SCA2 ne permet pas de classer correctement les sujets sporadiques ni les autres cas familiaux.

c. Analyse des gènes dérégulés uniquement chez les sujets porteurs d'une mutation du gène LRRK2

- 25 L'analyse d'expression des gènes dérégulés des sujets porteurs de mutation du gène LRRK2 par rapport aux témoins sains a été réalisée par un filtrage en Volcano Plot en appliquant un test T de Welch avec un seuil de significativité de $p=0,05$ et un seuil de variation d'expression de 1,2 après classement des sujets selon leur stade Hoehn et Yahr.

9 005 gènes composent la liste liste_A des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation LRRK2 et ayant développés une maladie de Parkinson dont le stade HY est de 0, 1 ou 1,5 (sujets P3, P5).

5 5 783 gènes composent la liste liste_B des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation LRRK2 et ayant développés une maladie de Parkinson dont le stade HY est de 2 ou 2,5 (sujets P4, P7, P8, P11). Le sujet P12 n'a pas été pris en considération dans cette analyse pour la sélection des gènes dérégulés car P12 présentait à la fois une maladie de
10 Parkinson et un syndrome maniaco-dépressif, ce qui risquait de biaiser la spécificité vis-à-vis de la maladie de Parkinson des gènes sélectionnés.

6 665 gènes composent la liste liste_C des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation LRRK2 et ayant développés une maladie de Parkinson dont le stade HY est de 3, 4 ou 5 (sujets P2,
15 P9, P10). Le prélèvement de P6 présentait une contamination importante en hémoglobine, ce qui risquait de biaiser la sélection des gènes ; il n'a donc pas été pris en compte dans cette analyse.

Les gènes spécifiquement dérégulés et communs aux différents stades HY sont sélectionnés dans la liste liste_E (759 gènes).

20 Cette liste de gènes dérégulés (Liste_E) a ensuite été croisée avec la liste des gènes dérégulés chez les sujets atteints d'autres maladies neurodégénératives ou syndrome parkinsonien (Liste_D). Après exclusion des gènes qui sont communs avec ceux qui sont aussi significativement dérégulés chez des patients atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens, on obtient ainsi un
25 ensemble de 438 gènes dérégulés spécifiquement chez les sujets porteurs de la mutation LRRK2 (Liste_F).

Cet ensemble de 438 gènes est analysé avec le logiciel PAM.

Le groupe « d'entraînement » est constitué des 9 sujets porteurs de mutation du gène LRRK2 (sujets P2 à P5 et P7 à P11) et des 20 témoins (sujets T1 à T20).

30 Le groupe test est constitué des 18 autres sujets familiaux (sujets P1 et P13 à P29) ainsi que des 2 autres sujets porteurs d'une mutation LRRK2 (sujets P6 et P12), des 17 sujets sporadiques (sujets S1 à S17) et des 11 sujets atteints d'un autre syndrome parkinsonien (sujets A1 à A11).

Les résultats obtenus à partir du groupe « d'entraînement » sont représentés
35 dans le tableau suivant et dans la **figure 4A** :

Résultats du groupe d'entraînement (seuil=0,01)				
	Vrai\Prédit	Porteur mutation	Sujet Sain (témoins)	Taux d'erreur de classement en %
Sujets P	Malades	9	0	0
Sujets T	Témoins	8	12	40

Grâce à cet ensemble de gènes, tous les sujets porteurs de mutation LRRK2 sont bien identifiés comme des patients malades et 12 témoins sur 20 sont identifiés comme des témoins dans le groupe d'entraînement (**figure 4A**), soit un taux d'erreur de classement de 40%.

En revanche dans le groupe test, le tableau ci-dessous montre que 17 malades seulement sont reconnus dans la classe des malades P1, P6, P12, P14, P15, P26, P27, P28, P29 (**figure 4C**) et S3, S5, S7, S8, S9, S10, S14, et S17 (**figure 4D**). Tous les sujets présentant une autre maladie neurodégénérative sont bien classés (**figure 4B**).

Résultats du groupe test (seuil = 0,01)					
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %	Figures
Sujets A	Témoins	0	11	0	4B
Sujets P	Malades	9	11	55,00	4C
Sujets S	Malades	8	9	52,94	4D

Des analyses complémentaires ont été effectuées en diminuant dans PAM le nombre de gènes inclus dans la liste liste_F, mais les différents ensembles de gènes obtenus au cours de ces analyses ne permettent pas non plus de distinguer les cas sporadiques des sujets témoins.

Ces résultats montrent donc que l'ensemble de 438 gènes obtenu uniquement à partir des résultats d'expression dérégulée des gènes chez des patients porteurs d'une mutation LRRK2 ne permet pas de classer plus de 8 cas sporadiques sur 17, soit un taux d'erreur de classement de 52,94%.

Des résultats similaires ont été obtenus lorsque les sujets P6 et P12 ont été exclus du groupe « test » lors de l'analyse par PAM (en plus d'être exclus lors de la détermination des listes liste_B et liste_C).

5 En conclusion, les trois analyses décrites ci-dessus aux points a., b. et c. de l'exemple 2 montrent qu'il n'est pas possible de prédire les sujets qui sont atteints de la maladie de Parkinson ou qui pourraient développer une forme sporadique de cette maladie en utilisant les informations provenant seulement d'un seul type de mutation délétère.

10 **EXEMPLE 3 : Détermination d'un ensemble de gènes dont l'expression est dérégulée chez des patients atteints de différentes formes familiales de la maladie de Parkinson et étude de la capacité d'identification de sujets atteints, notamment de formes sporadiques, à partir de cet ensemble**

15 Dans cet exemple, les données d'expression des gènes obtenues à partir d'un ensemble de 4 types de formes familiales de la maladie de Parkinson ont été prises en compte pour déterminer si l'ensemble des gènes ainsi obtenu permet de prédire la maladie de Parkinson, et en particulier les formes sporadiques de la maladie.

20 Ainsi, les données d'expression différentielle des gènes obtenues à partir de l'ensemble des patients présentant des mutations des gènes SNCA, LRRK2, SCA2 et PRKN responsables de ces 4 formes familiales de la maladie de Parkinson ont été prises en compte dans l'analyse par le logiciel GeneSpring, excepté pour les sujets P6 et P12 (voir *supra* exemple 2 c.).

25 Les autres patients parkinsoniens (patients P23, P24, P25, P26, P27, P28, P29) et sporadiques n'ont pas été retenus dans cette première analyse avec GeneSpring, mais ont été inclus dans l'analyse avec PAM. Ils serviront à vérifier la valeur prédictive des gènes sélectionnés dans cette analyse.

30 L'analyse d'expression des gènes dérégulés des sujets porteurs de mutations du gène SCA2, PRKN, SNCA et LRRK2 par rapport aux témoins sains a été réalisée par un filtrage en Volcano Plot en appliquant un test T de Welch avec un seuil de significativité de $p=0,05$ et un seuil de variation d'expression de 1,2 après classement des sujets selon leur stade Hoehn et Yahr.

35 4 415 gènes composent la liste liste_A des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation SCA2, PRKN, SNCA et LRRK2 et ayant développés une maladie de Parkinson dont le stade HY est de 0, 1 ou 1,5 (sujets P1, P3, P5, P18, P22).

2 297 gènes composent la liste liste_B des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation SCA2, PRKN, SNCA et

LRRK2 et ayant développés une maladie de Parkinson dont le stade HY est de 2 ou 2,5 (sujets P4, P7, P8, P11, P15, P16, P17, P19, P20). De même que pour l'exemple 2 c., le sujet P12 n'a pas été pris en considération pour cette analyse.

5 3 292 gènes composent la liste liste_C des gènes spécifiquement dérégulés par rapport aux témoins des sujets porteurs d'une mutation SCA2, PRKN, SNCA et LRRK2 et ayant développés une maladie de Parkinson dont le stade HY est de 3, 4 ou 5 (sujets P2, P9, P10, P13, P14, P21). De même que pour l'exemple 2 c., le sujet P6 n'a pas été pris en considération pour cette analyse.

10 Les gènes spécifiquement dérégulés et communs aux différents stades HY sont sélectionnés dans la liste liste_E (38 gènes).

15 Cette liste de gènes dérégulés (Liste_E) a ensuite été croisée avec la liste des gènes dérégulés chez les sujets atteints d'autres maladies neurodégénératives ou syndrome parkinsonien (Liste_D composée de 5 645 gènes). Après exclusion des gènes qui sont communs avec ceux qui sont aussi significativement dérégulés chez des patients atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens, on obtient ainsi un ensemble de 23 gènes (Liste_F).

Cet ensemble de 23 gènes regroupé dans le tableau 8 ci-dessous correspond aux gènes spécifiquement dérégulés chez les sujets porteurs de la mutation SCA2, PRKN, SNCA et LRRK2.

20

N°	Code puce	Nom du gène	Code Genbank
1	A_23_P216712	Récepteur potentiel transitoire du canal à cation, sous famille M, membre 6 (TRPM6)	NM_017662
2	A_24_P245838	Mannosyl(beta-1,4-)-glycoprotéine beta-1,4-N-acetylglucosaminyltransferase (MGAT3)	AK125361
3	A_23_P369316	Vang-like 1 (VANGL1)	NM_138959
4	A_23_P1904	Domaine transmembranaire 4, sous famille A, membre 2 (MS4A2)	NM_000139
5	A_23_P349321	Xylosyltransferase de type I (XYLT1)	NM_022166
6	A_23_P117662	Histidine decarboxylase (HDC)	NM_002112
7	A_23_P133916	Composant C2 du complément (C2)	NM_000063

8	A_32_P162187	Composant C2 du complément (C2)	NM_000063
9	A_23_P8961	Interleukine 7 (IL7)	NM_000880
10	A_32_P175313	EST PM1-NN1207-071200-003-a12 NM1207	BF960555
11	A_32_P6433	EST yr57d04.r1 ADNcomplémentaire du clone IMAGE:209383 soares foie foetal rate, similaire à la protéine de la famille de choc thermique humaine de 71KD	H64096
12	A_32_P162183	Composant C2 du complément (C2)	NM_000063
13	A_32_P104572	EST UI-E-DW1-ahe-d-15-0-UI.s1 clone de l'ADNcomplémentaire UI-E-DW1-ahe-d-15-0-UI 3'	BM670971
14	A_32_P234485	EST UI-1-BB1p-auq-g-04-0-UI.s1 NCI_CGAP_P16 ADNcomplémentaire du clone UI-1-BB1p-auq-g-04-0-UI 3'	BQ024427
15	A_23_P60047	protéine 1 contenant des structures acides bi-spiralées (TACC1)	NM_006283
16	A_32_P45875	EST ws88f04.x1 Homo sapiens, ADNcomplémentaire du clone IMAGE:2505055 3', similaire à la séquence voisine de la phospholipase C TR:Q62084 Q62084	AW009873
17	A_23_P352435	Régulateur des protéines G de signalisation 12 (RGS12)	NM_002926
18	A_23_P387649	DNAJA4 homologue à DnaJ (Hsp40), sous-famille A, membre 4 (DNAJA4)ou PRO1472	AF116663
19	A_23_P200252	Cadre ouvert de lecture 57, chromosome 1 (C1orf57)	NM_032324
20	A_23_P66174	Domaine de liaison SH3 de la protéine kinase 1(SBK1)	NM_00102440
21	A_23_P33927	FKSG43	AF334945
22	A_23_P333640	Papiline (PAPLN) glycoprotéine sulfatées similaire à un protéoglycane	NM_173462
23	A_24_P323682	région chromosomique de la sclérose latérale amyotrophique 2 (juvénile)	NM_020919

Pour l'analyse PAM, les échantillons ont été divisés de la manière suivante :

- le groupe « d'entraînement » correspond aux 20 témoins sains (sujets T1 à T20) et les 20 sujets porteurs d'une mutation LRRK2, PRKN, SNCA et SCA2 (sujets P1, P2, P3, P4, P5, P7, P8, P9, P10, P11, P13, P14, P15, P16, P17, P18, P19, P20, P21 et P22).
- le groupe « test » comprend 9 cas familiaux autres (c'est-à-dire des cas familiaux autres que ceux porteurs d'une mutation PRKN, LRRK2, SCA2 ou SNCA (sujets P), à savoir les sujets, P23, P24, P25, P26, P27, P28, P29 ainsi que de 2 sujets porteurs d'une mutation LRRK2 (P6 et P12), les 17 cas sporadiques (sujets S1 à S17), et les 11 sujets atteints d'un autre syndrome parkinsonien (sujets A1 à A11).

Les résultats de l'analyse du groupe « d'entraînement » sont résumés dans le tableau suivant et dans la **figure 5A** :

Résultats du groupe d'entraînement (seuil=0,01)				
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %
Sujets P	Malades	20	0	0
Sujets T	Témoins	2	18	10

- 15 Cet ensemble de 23 gènes permet de classer les malades et les témoins avec un taux d'erreur de 0% et 10% respectivement sur le groupe d'entraînement.

Les résultats du groupe « test » sont représentés dans le tableau suivant et dans les **figures 5B, 5C et 5D** :

Résultats du groupe test (seuil = 0,01)					
	Vrai\Prédit	Malades	Témoins	Taux d'erreur de classement en %	Figures
Sujets A	Témoins	4	7	36,36	5B
Sujets P	Malades	6	3	33,33	5C
Sujets S	Malades	15	2	11,76	5D

- 20 Ces résultats montrent que 5 des 26 malades parkinsoniens du groupe « test » sont mal classés. Il s'agit de 2 cas sporadiques (S1 et S13; **figure 5D**) et de 3 autres cas familiaux (P24, P25 ; P12 **figure 5C**) qui ont été classés comme

témoins représentés par un rond blanc au lieu de malades représentés par un carré noir dans le groupe des malades.

Les sujets familiaux pauci-symptomatiques (P28, P29) ont été bien classés comme patients à risque de développer la maladie de Parkinson (ainsi que le sujet porteur d'une mutation LRRK2 (P6) ayant une contamination en hémoglobine).

Les sujets ayant une autre pathologie neurodégénérative A5, A6, A8 et A11 sont mal classés car représentés par un carré noir au lieu d'un rond blanc dans le groupe des témoins (**figure 5B**).

Des résultats similaires ont été obtenus lorsque les sujets P6 et P12 ont été exclus du groupe « test » lors de l'analyse par PAM (en plus d'être exclus lors de la détermination des listes liste_B et liste_C).

En conclusion, ces résultats montrent que l'ensemble des gènes dérégulés dans les formes familiales de la maladie de Parkinson déterminé selon la méthode de l'invention permet non seulement de déterminer si des sujets sont atteints de la maladie de Parkinson avec un taux d'erreur faible, même pour des cas pauci-symptomatiques, mais également d'identifier des cas sporadiques de la maladie avec un taux d'erreur très faible (dans cet exemple de l'ordre de 11,76%). Par ailleurs, ces résultats démontrent qu'il est indispensable de prendre en considération les gènes dérégulés communs à plusieurs formes familiales incriminant des gènes différents pour obtenir un effet synergique des informations obtenues et permettre l'identification de ces cas sporadiques.

REVENDEICATIONS

1. Méthode pour détecter dans un échantillon biologique à tester l'expression différentielle d'un ensemble de marqueurs moléculaires associée à la présence ou au risque de développer la maladie de Parkinson, caractérisée en ce qu'elle comprend les étapes suivantes :

a1) extraction *in vitro* des marqueurs moléculaires à partir de l'échantillon biologique à tester

a2) extraction *in vitro* des marqueurs moléculaires à partir d'échantillons biologiques de sujets atteints de différentes formes familiales de la maladie de Parkinson, de sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens et de sujets sains,

b) déterminer l'ensemble des marqueurs moléculaires dont l'expression est significativement dérégulée en commun chez les sujets atteints de différentes formes familiales quel que soit le stade de la maladie mais dont l'expression n'est pas significativement dérégulée chez les sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens, par rapport à celle mesurée chez les sujets sains,

c) vérifier que l'expression de l'ensemble de ces marqueurs moléculaires est bien dérégulée dans l'échantillon à tester.

2. Méthode selon la revendication 1, pour détecter la présence ou le risque de développer une forme sporadique de la maladie de Parkinson.

3. Méthode selon la revendication 1 ou la revendication 2, pour détecter la présence d'une forme asymptomatique de la maladie de Parkinson.

4. Méthode selon l'une quelconque des revendications précédentes, caractérisée en ce que les sujets atteints de différentes formes familiales sont des patients porteurs de mutations dans les gènes SNCA, PRKN, SCA2 ou LRRK2.

5. Méthode selon l'une quelconque des revendications précédentes, caractérisée en ce que les sujets atteints d'autres maladies neurodégénératives ou syndromes parkinsoniens sont des patients atteints de SCA, MSA ou DLB.

6. Méthode selon l'une quelconque des revendications précédentes, caractérisée en ce que l'échantillon biologique à analyser est constitué d'ARN extraits des cellules mononucléées sanguines.

7. Méthode selon l'une quelconque des revendications précédentes, caractérisée en ce que l'expression dérégulée des marqueurs moléculaires est déterminée par une mesure quantitative de l'expression des gènes ou des protéines choisie parmi micro-puces et RT-PCR.

8. Méthode selon l'une quelconque des revendications précédentes, caractérisée en ce que les marqueurs moléculaires sont des ARN ou des protéines.

9. Méthode selon l'une quelconque des revendications précédentes, caractérisée en ce que l'ensemble des marqueurs moléculaires qui est déterminée selon l'étape b) est la liste des gènes suivants :

	Nom du gène	Code Genbank
1	Récepteur potentiel transitoire du canal à cation, sous famille M, membre 6 (TRPM6)	NM_017662
2	Mannosyl(beta-1,4-)-glycoprotéine beta-1,4-N-acetylglucosaminyltransferase (MGAT3)	AK125361
3	Vang-like 1 (VANGL1)	NM_138959
4	Domaine transmembranaire 4, sous famille A, membre 2 (MS4A2)	NM_000139
5	Xylosyltransferase de type I (XYLT1)	NM_022166
6	Histidine decarboxylase (HDC)	NM_002112
7	Composant C2 du complément (C2)	NM_000063
8	Interleukine 7 (IL7)	NM_000880
9	EST PM1-NN1207-071200-003-a12 NM1207	BF960555

10	EST yr57d04.r1 ADNcomplémentaire du clone IMAGE:209383 soares foie foetal rate, similaire à la protéine de la famille de choc thermique humaine de 71KD	H64096
11	EST UI-E-DW1-ahe-d-15-0-UI.s1 clone de l'ADNcomplémentaire UI-E-DW1-ahe-d-15-0-UI 3'	BM670971
12	EST UI-1-BB1p-auq-g-04-0-UI.s1 NCI_CGAP_P16 ADNcomplémentaire du clone UI-1-BB1p-auq-g-04-0-UI 3'	BQ024427
13	protéine 1 contenant des structures acides bi-spiralées (TACC1)	NM_006283
14	EST ws88f04.x1 Homo sapiens, ADNcomplémentaire du clone IMAGE:2505055 3', similaire à la séquence voisine de la phospholipase C TR:Q62084 Q62084	AW009873
15	Régulateur des protéines G de signalisation 12 (RGS12)	NM_002926
16	DNAJA4 homologue à DnaJ (Hsp40), sous-famille A, membre 4 (DNAJA4)ou PRO1472	AF116663
17	Cadre ouvert de lecture 57, chromosome 1 (C1orf57)	NM_032324
18	Domaine de liaison SH3 de la protéine kinase 1(SBK1)	NM_00102440
19	FKSG43	AF334945
20	Papiline (PAPLN) glycoprotéine sulfatées similaire à un protéoglycane	NM_173462
21	région chromosomique de la sclérose latérale amyotrophique 2 (juvénile)	NM_020919

10. Utilisation de réactifs spécifiques d'un ensemble de marqueurs moléculaires identifié selon l'étape b) de la méthode selon l'une quelconque des revendications précédentes pour déterminer la présence ou le risque de développer la maladie de Parkinson.

11. Utilisation selon la revendication 10, caractérisée en ce que lesdits réactifs spécifiques sont spécifiques de l'ensemble des gènes tels que définis dans la revendication 9.

12. Utilisation selon la revendication 11, caractérisée en ce que lesdits réactifs spécifiques sont les séquences SEQ ID NO : 1 à SEQ ID NO : 23.

13. Kit pour la mise en œuvre de la méthode selon l'une quelconque des revendications 1 à 9 comprenant au moins les éléments suivants :

- des réactifs spécifiques de l'ensemble des marqueurs moléculaires tels qu'identifiés selon l'étape b) de ladite méthode,

- éventuellement les réactifs et/ou dispositifs nécessaires et adaptés pour effectuer l'extraction des marqueurs moléculaires à analyser et/ou la détection et/ou la quantification relative de ces marqueurs par rapport à la quantité des marqueurs moléculaires du même type,

- éventuellement un échantillon contenant des marqueurs moléculaires du type choisi issu d'un patient atteint de la maladie de Parkinson de façon avérée (témoin positif).

14. Kit selon la revendication 13, caractérisé en ce que lesdits réactifs spécifiques sont spécifiques des gènes tels que définis à la revendication 9.

15. Kit selon la revendication 14, caractérisé en ce que les marqueurs moléculaires sont des ARNs.

16. Kit selon la revendication 13 ou la revendication 14, caractérisé en ce que lesdits réactifs spécifiques des gènes sont des amorces ou des sondes spécifiques.

17. Kit selon la revendication 16, caractérisé en ce que lesdites sondes spécifiques des gènes sont les séquences SEQ ID NO : 1 à SEQ ID NO : 23.

18. Kit selon l'une quelconque des revendications 13 à 17, caractérisé en ce que le témoin positif est un échantillon issu d'un sujet présentant une mutation pathogène responsable de la maladie de Parkinson à un stade HY de la maladie supérieur ou égal à 3.

19. Kit selon l'une quelconque des revendications 13 à 18, caractérisé en ce que le témoin négatif est un échantillon issu d'un sujet indemne de tout antécédent personnel ou familial de la maladie de Parkinson, sans autre maladie apparente, ni prise de traitement médical.

1/9

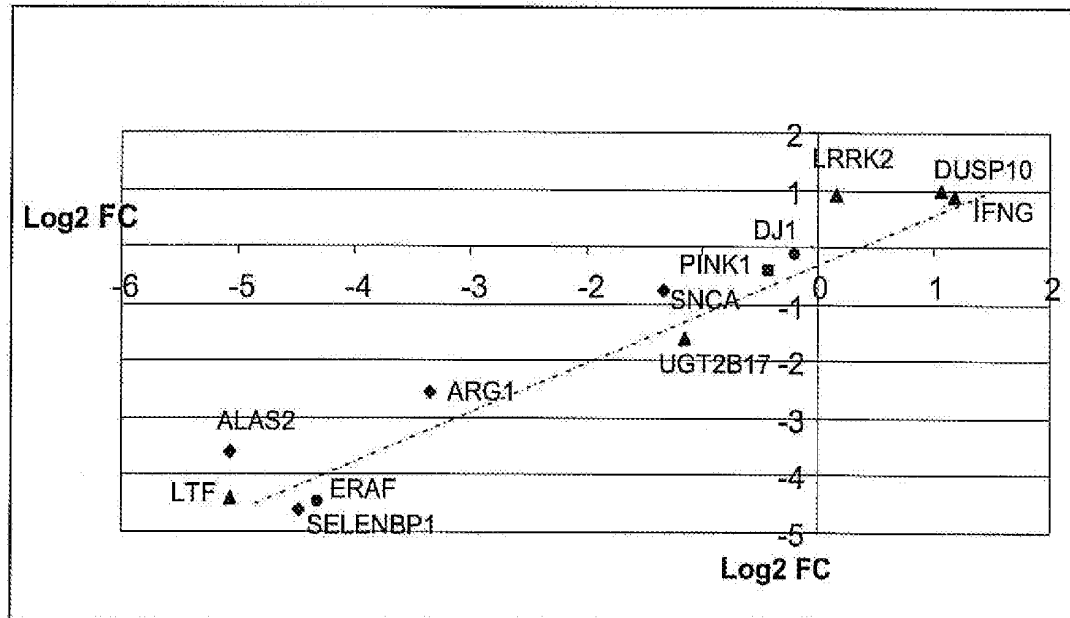


FIGURE 1

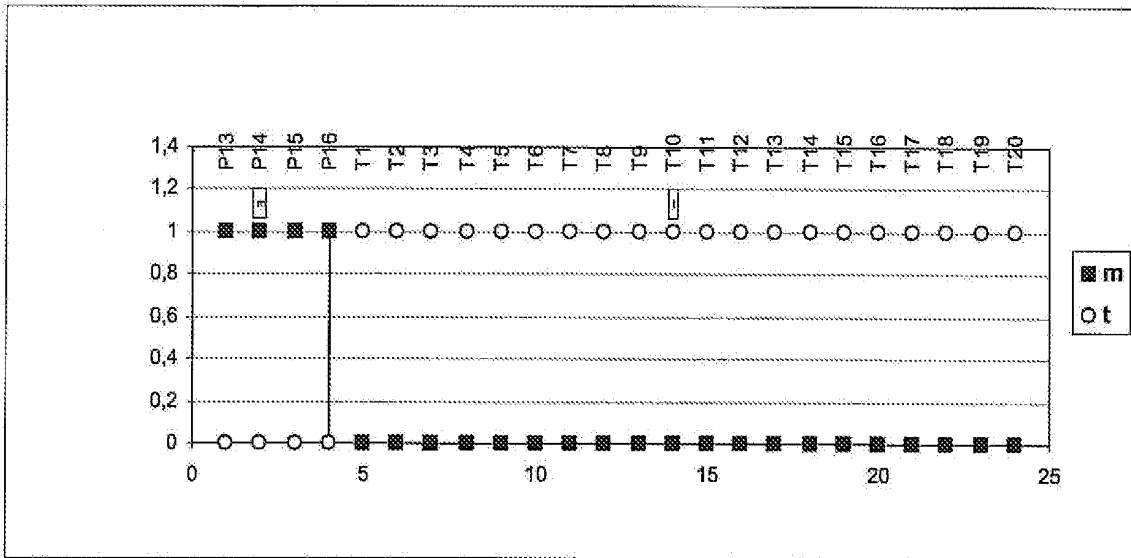


FIGURE 2A

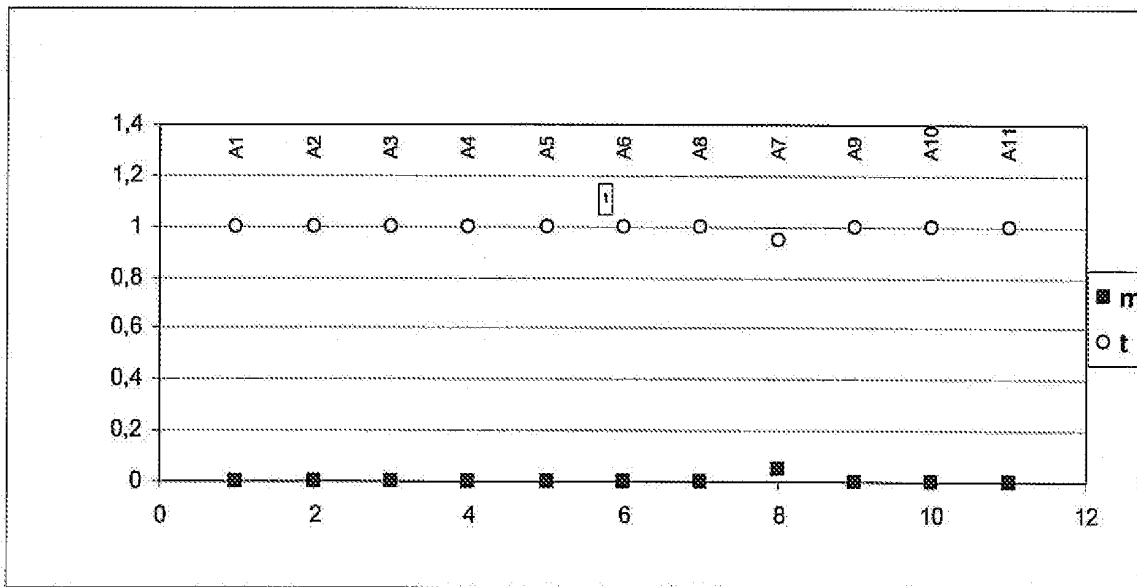


FIGURE 2B

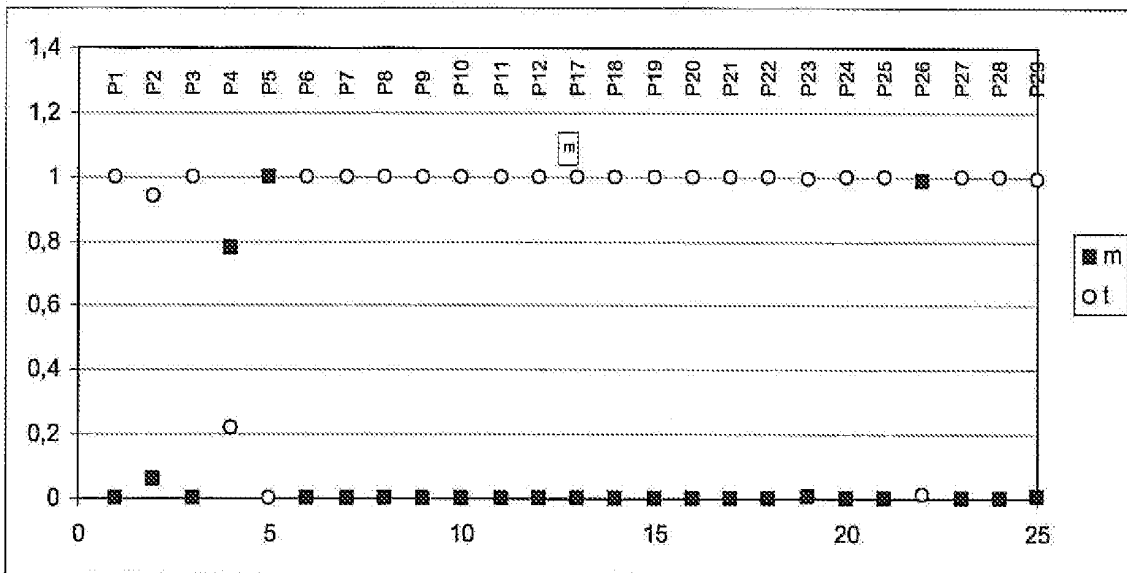


FIGURE 2C

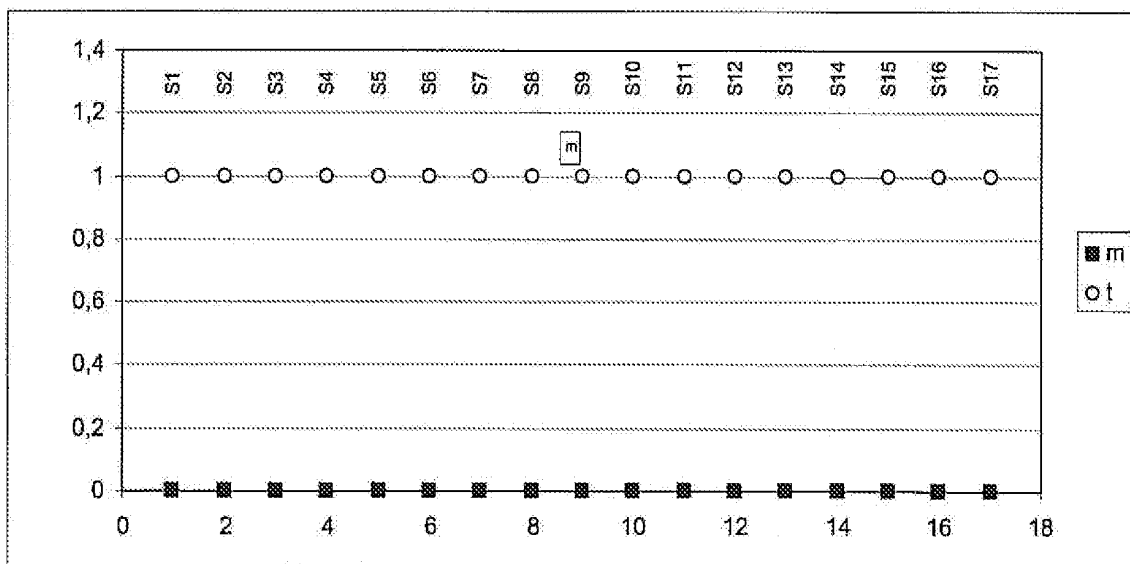


FIGURE 2D

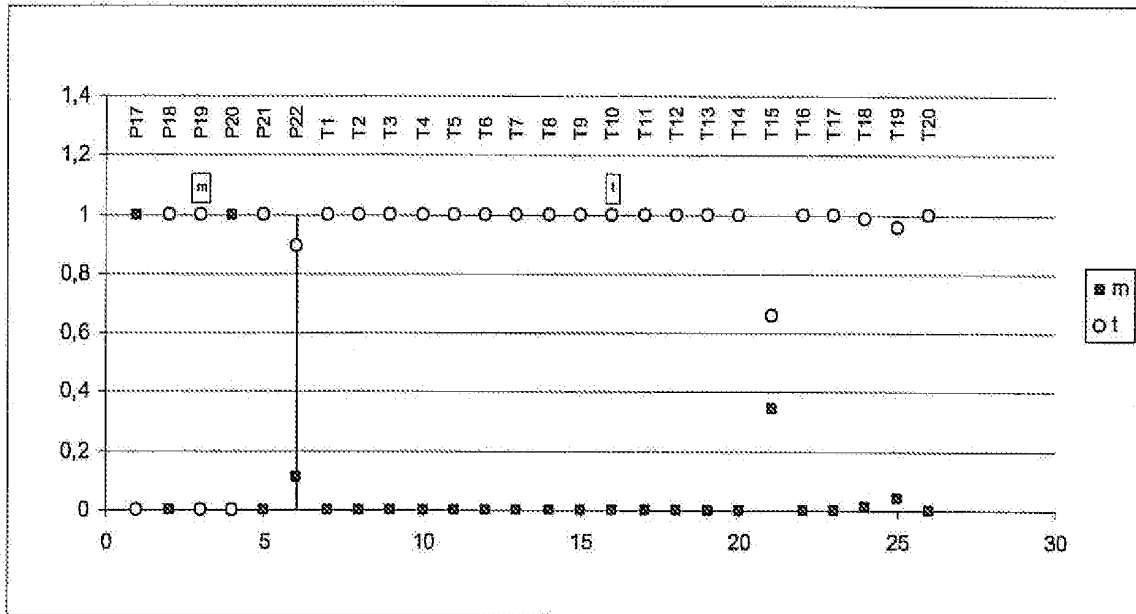


FIGURE 3A

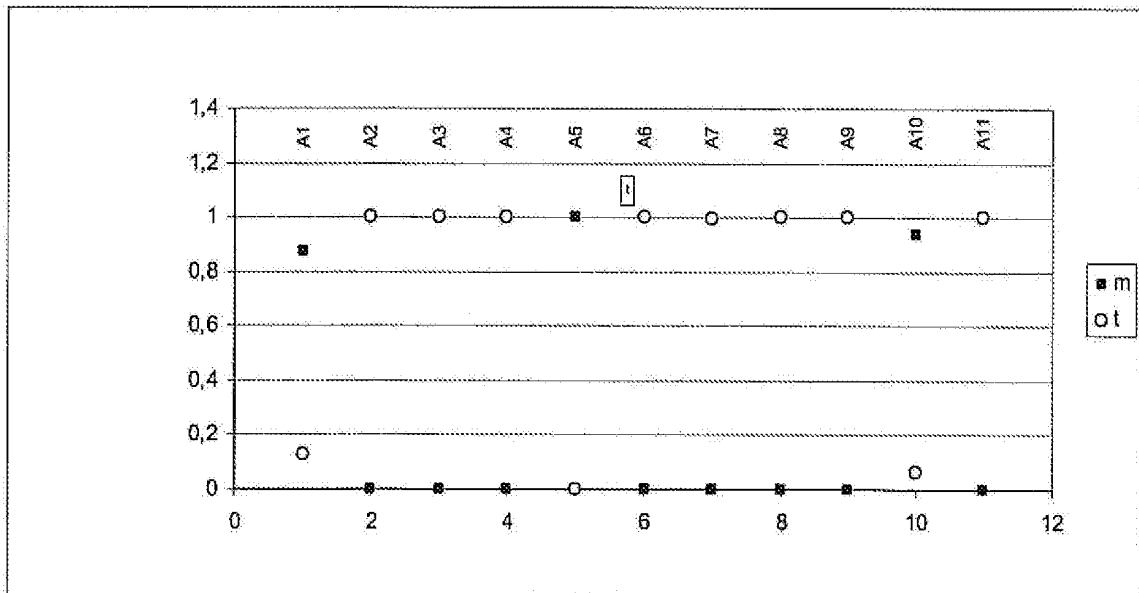


FIGURE 3B

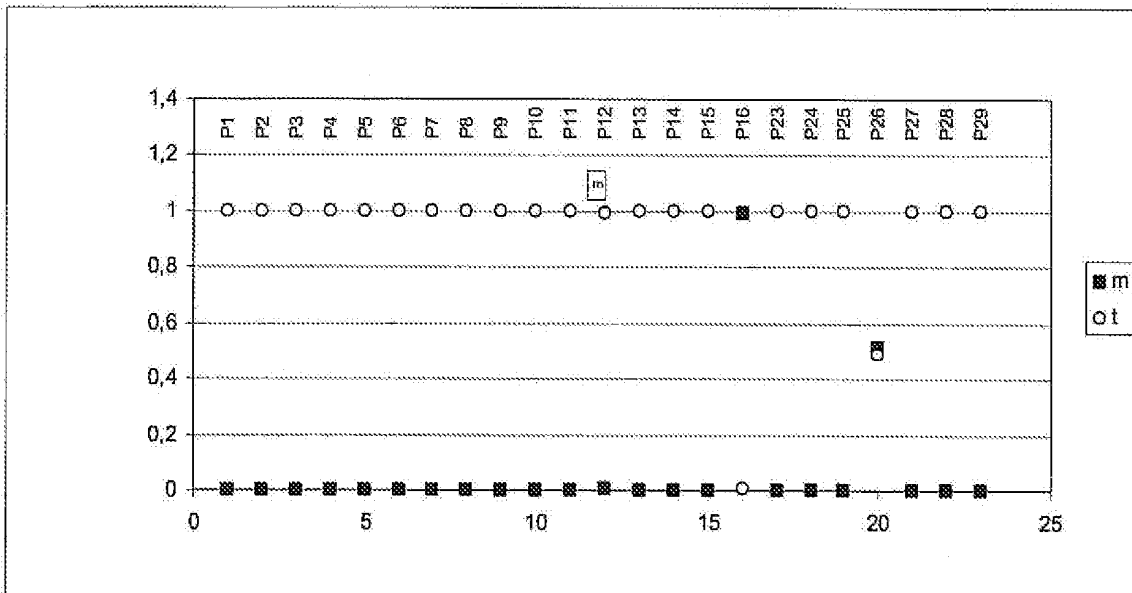


FIGURE 3C

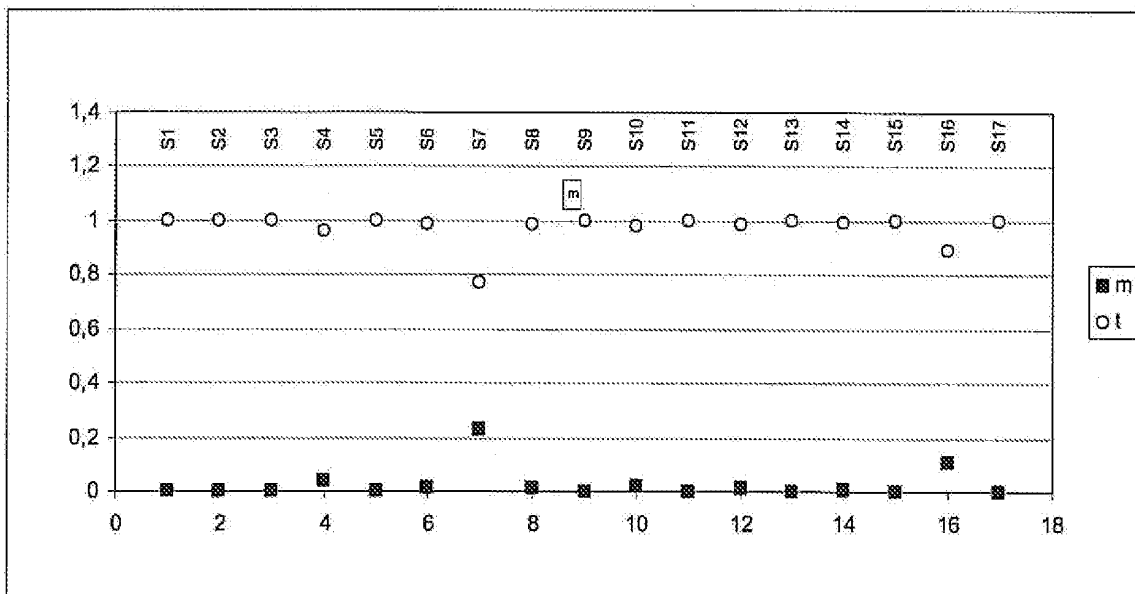


FIGURE 3D

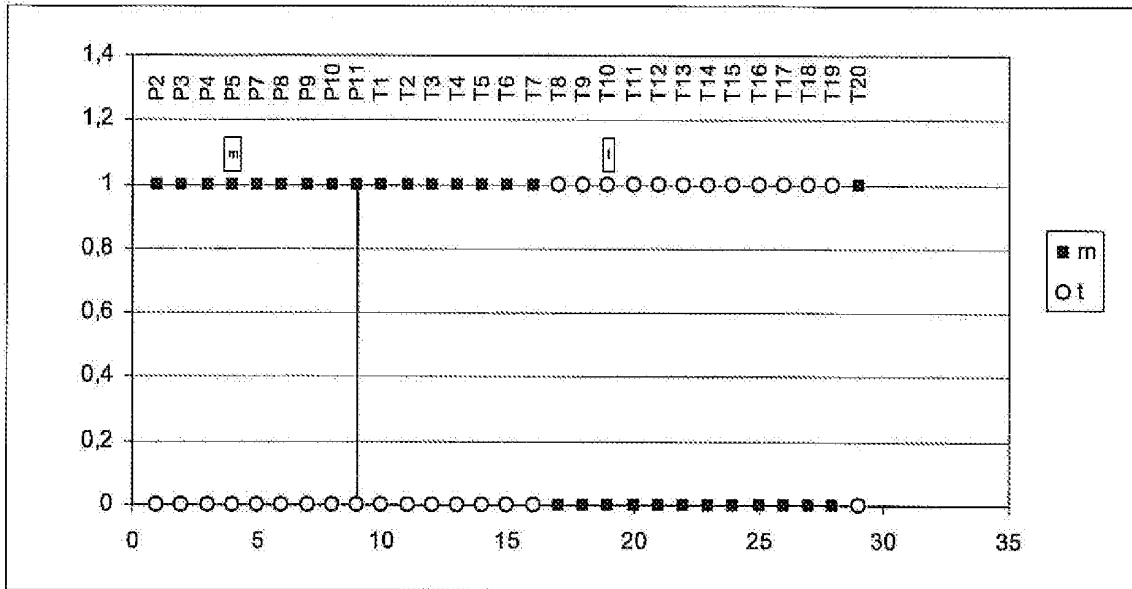


FIGURE 4A

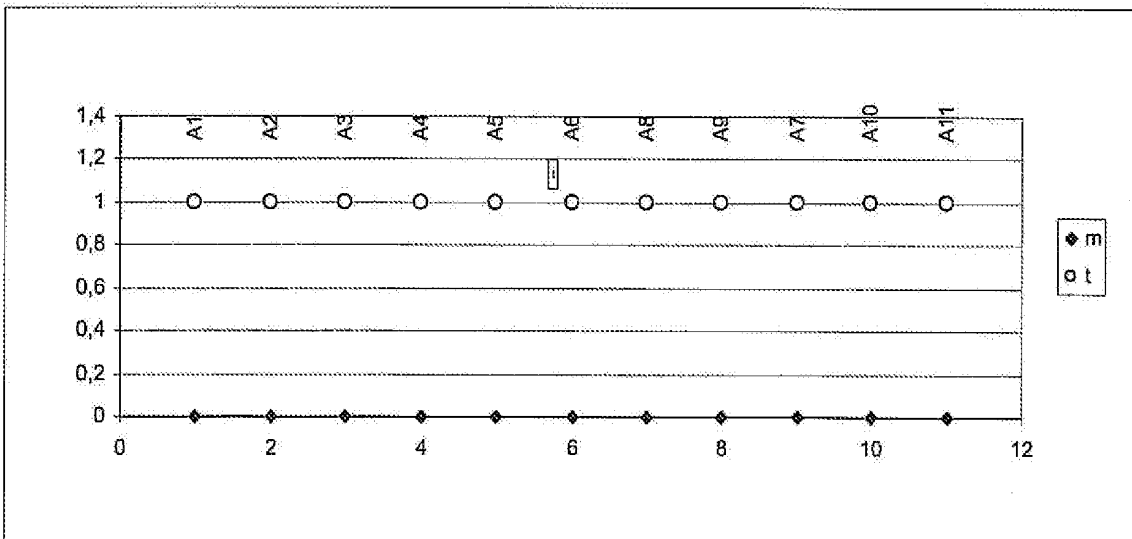


FIGURE 4B

7/9

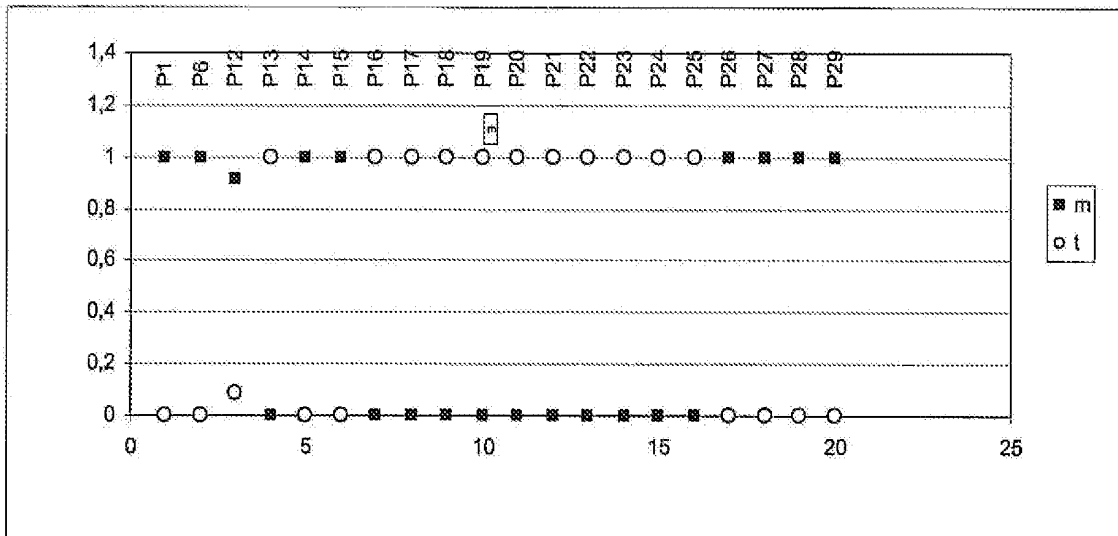


FIGURE 4C

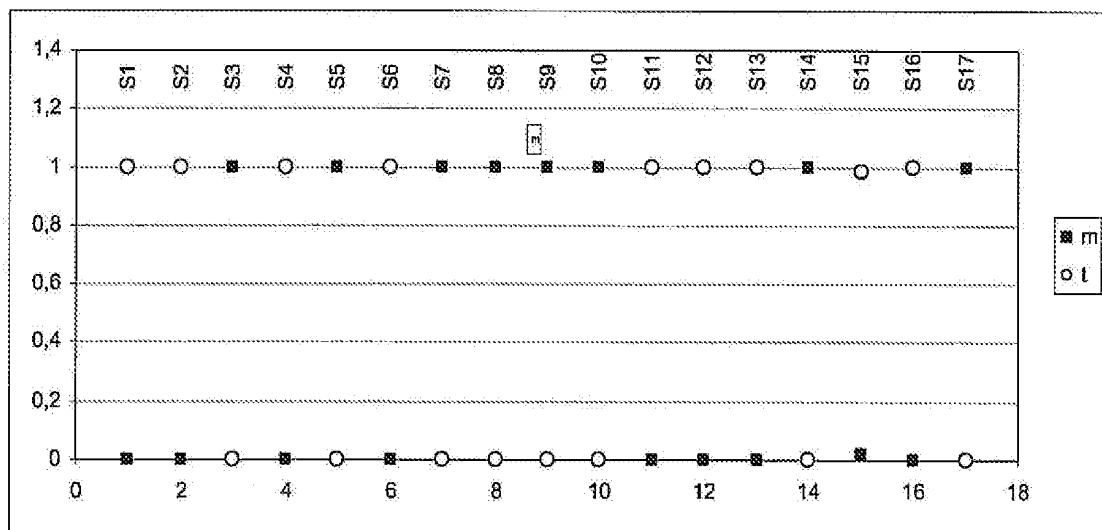


FIGURE 4D

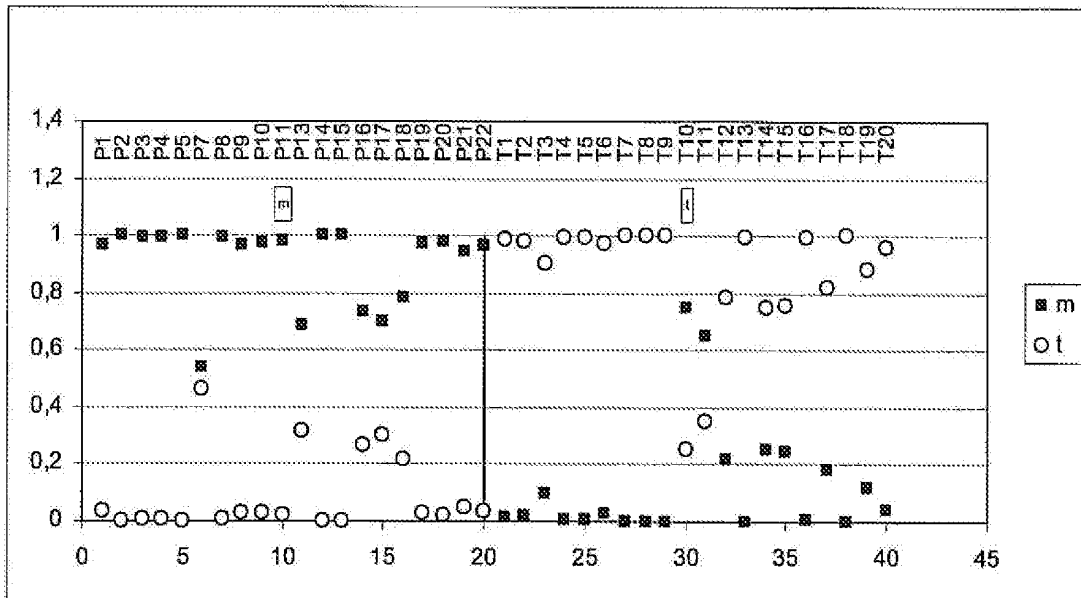


FIGURE 5A

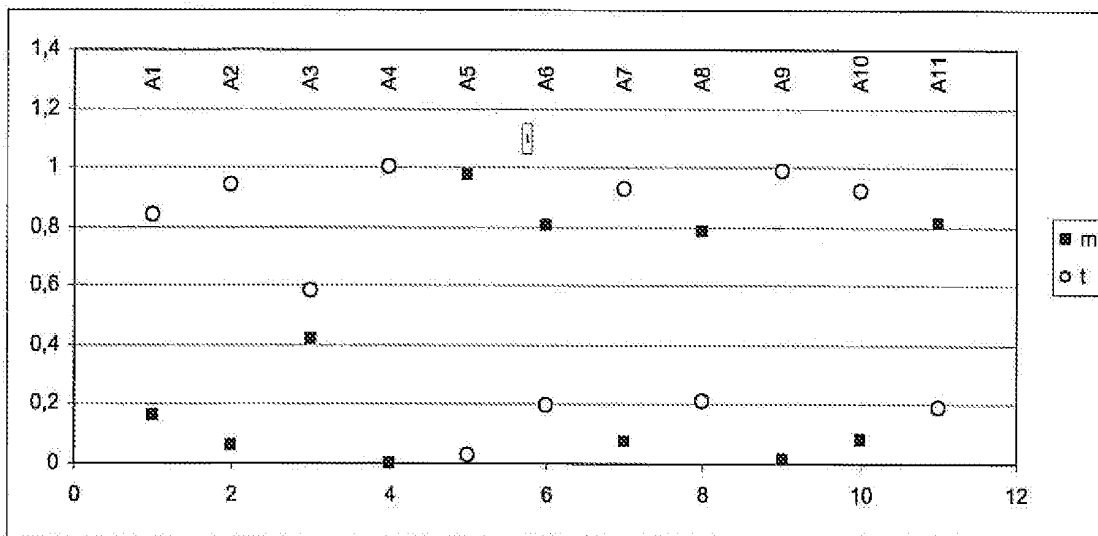


FIGURE 5B

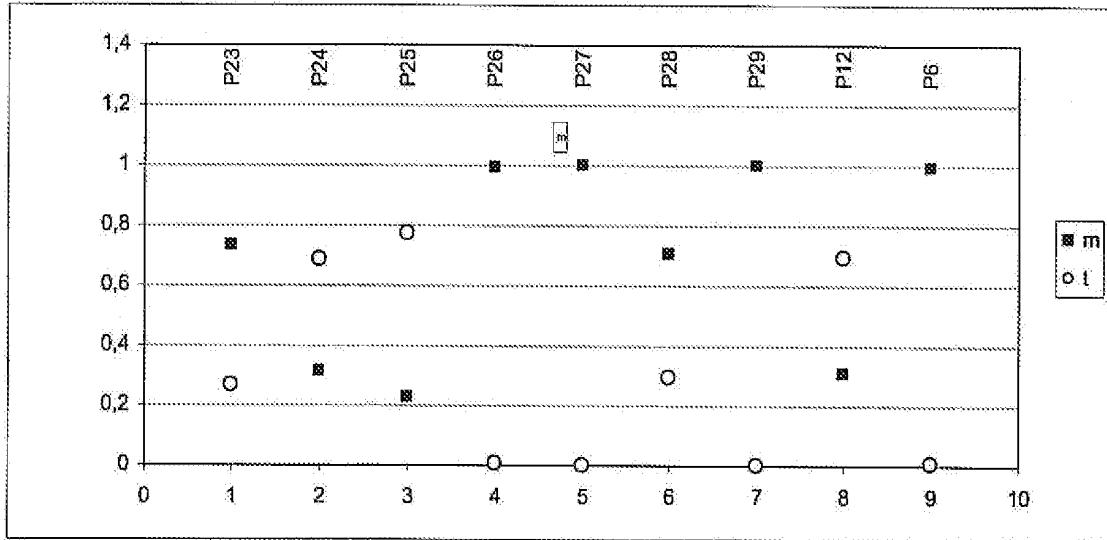


FIGURE 5C

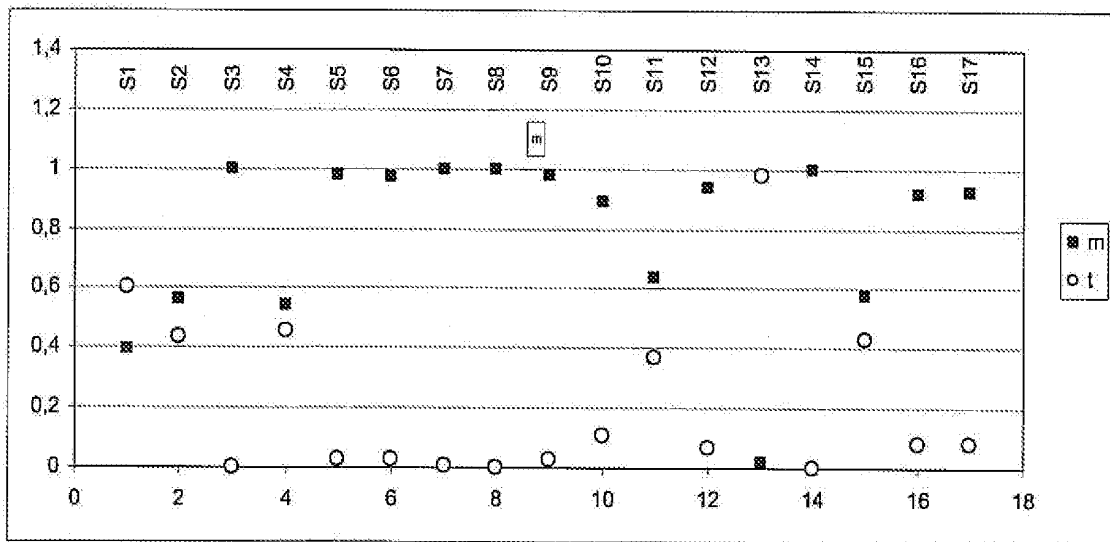


FIGURE 5D



**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

N° d'enregistrement national

établi sur la base des dernières revendications déposées avant le commencement de la recherche

FA 714869
FR 0855997

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	WO 2008/063521 A (GEN HOSPITAL CORP [US]; ELSON CONSTANCE M [US]; HAYDEN DOUGLAS [US]; S) 29 mai 2008 (2008-05-29) pages 18-22 -----	13,16, 18,19	C12Q1/68 C07H21/00
X	WO 2008/070311 A (GENE LOGIC INC [US]; UNIV MIAMI [US]; PAPAPETROPOULOS SPIRIDON [US]; F) 12 juin 2008 (2008-06-12) revendications; pages 30-32; tableaux -----	10,13, 16,18,19	
X	US 2004/248092 A1 (VANCE JEFFREY M [US] ET AL) 9 décembre 2004 (2004-12-09) page 4, [0049]; exemple 4 et revendications -----	10,13, 16,18,19	
X	US 2005/191652 A1 (VANCE JEFFERY M [US] ET AL) 1 septembre 2005 (2005-09-01) page 10, [0055]-[0060]; revendications -----	10,13, 16,18,19	
X	NOUREDDINE MAHER A ET AL: "Genomic convergence to identify candidate genes for Parkinson disease: SAGE analysis of the substantia nigra." MOVEMENT DISORDERS : OFFICIAL JOURNAL OF THE MOVEMENT DISORDER SOCIETY OCT 2005, vol. 20, no. 10, octobre 2005 (2005-10), pages 1299-1309, XP002517159 ISSN: 0885-3185 page 1301, dernier paragraphe; tableaux -----	10,13, 16,18,19	DOMAINES TECHNIQUES RECHERCHÉS (IPC) C12Q
Date d'achèvement de la recherche		Examineur	
27 février 2009		Hennard, Christophe	
CATÉGORIE DES DOCUMENTS CITÉS		T : théorie ou principe à la base de l'invention	
X : particulièrement pertinent à lui seul		E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.	
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie		D : cité dans la demande	
A : arrière-plan technologique		L : cité pour d'autres raisons	
O : divulgation non-écrite		
P : document intercalaire		& : membre de la même famille, document correspondant	

**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 0855997 FA 714869**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.

Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **27-02-2009**

Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 2008063521 A	29-05-2008	AUCUN	

WO 2008070311 A	12-06-2008	AUCUN	

US 2004248092 A1	09-12-2004	AUCUN	

US 2005191652 A1	01-09-2005	US 2006068428 A1	30-03-2006
