SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MOLECULAR NEUROSCIENCE
DIAGNOSTIC MARKERS
PARKINSON'S DISEASE
PROTEOMICS

Received
22 October 2012

Accepted
19 November 2012

Published
11 December 2012

Correspondence and requests for materials should be addressed to M.F. (mauro.fasano@uninsubria.it)

Discovery and verification of panels of T-lymphocyte proteins as biomarkers of Parkinson's disease

Tiziana Alberio[1,2], Agnese C. Pippione[1,2], Maurizio Zibetti[3], Simone Olgiati[1], Daniela Cecconi[4], Cristoforo Comi[5], Leonardo Lopiano[3] & Mauro Fasano[1,2]

[1]Division of Biomedical Sciences, Department of Theoretical and Applied Sciences, University of Insubria, Busto Arsizio, Italy, [2]Centre of Neuroscience, University of Insubria, Busto Arsizio, Italy, [3]Department of Neuroscience, University of Torino, Torino, Italy, [4]Proteomics and Mass Spectrometry Laboratory, Department of Biotechnology, University of Verona, Verona, Italy, [5]Division of Neurology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy.

The diagnosis of Parkinson's disease (PD) is currently based on the clinical evaluation of extrapyramidal signs with a considerable error rate. The identification of specific markers might allow PD diagnosis before the onset of classical motor symptoms. By two-dimensional electrophoresis we identified proteome alterations in T-lymphocytes of 17 control subjects and 15 PD patients. The observed changes were used to build predictive models that were verified by the leave-one-out cross-validation. We further built two functions able to stage the subjects. We chose to verify by Western blotting the identity of spots corresponding to β-fibrinogen and transaldolase, two recurrent proteins in six out of 20 spots. β-Fibrinogen levels are lowered in PD patients, whereas a heavy transaldolase set of isoforms was more abundant. Eventually, we identified a list of seven proteins showing different levels in early-onset with respect to late-onset PD patients.

Diagnosis of Parkinson's disease (PD) is currently based on the clinical evaluation of extrapyramidal signs, such as tremor, rigidity and bradykinesia, when the degeneration of dopaminergic nigral neurons has raised over 70%[1,2]. The identification of specific biomarkers is critical for the early diagnosis and also to monitor PD progression. Additionally, the assessment of disease-modifying drugs requires the identification of early-stage patients to be included in clinical studies[3–5]. Non-motor signs frequently precede the onset of PD but they may be unspecific, whereas instrumental investigations (polysomnography or functional imaging) are characterized by high cost and use of radioactive tracers that hamper their application population-wide[3–5].

Molecular biomarkers in body fluids are likely to meet the expectation of unprecedented specificity, in particular when a panel of biomarkers is concerned, together with costs that are lower than those of imaging/functional biomarkers. An alternative to body fluids is represented by blood cells, in particular T-lymphocytes[4,6,7]. T-lymphocytes express all dopamine receptors (D1–D5), each of which exerts different actions on the regulation of T-cell functions[8]. In addition, immune mechanisms may contribute to neuronal damage in PD and peripheral blood cells have been shown to share some of the changes exhibited by nigral neurons[9,10].

Several PD biomarker candidates were discovered using unbiased proteomic approaches[4]. Two-dimensional electrophoresis (2-DE) is able to resolve thousands of spots simultaneously and to discriminate protein processing isoforms and post-translational modifications[11]. As a proof of principle, a preliminary investigation on peripheral blood lymphocytes (PBL) of advanced PD patients has demonstrated that a proteomic profiling based on 2-DE highlights differences at the peripheral level between patients and control subjects, thus supporting the rationale for protein level alterations in T-lymphocytes of PD subjects[12].

In this study, we used 2-DE to identify alterations in the protein expression profile of T-lymphocytes of PD patients. These alterations were used to build predictive models that are able to significantly stratify PD patients. We further analyzed the selected proteins to search for correlations with Hoehn and Yahr score and years from disease onset. By linear regression, we built two discriminant functions able to stage the subjects in terms of disease duration and severity. We measured levels of β-fibrinogen and transaldolase by Western blotting to confirm the quantification performed on 2-DE gels. Eventually, we obtained another predictive model that was able to discriminate early-onset (EO) from late-onset (LO).

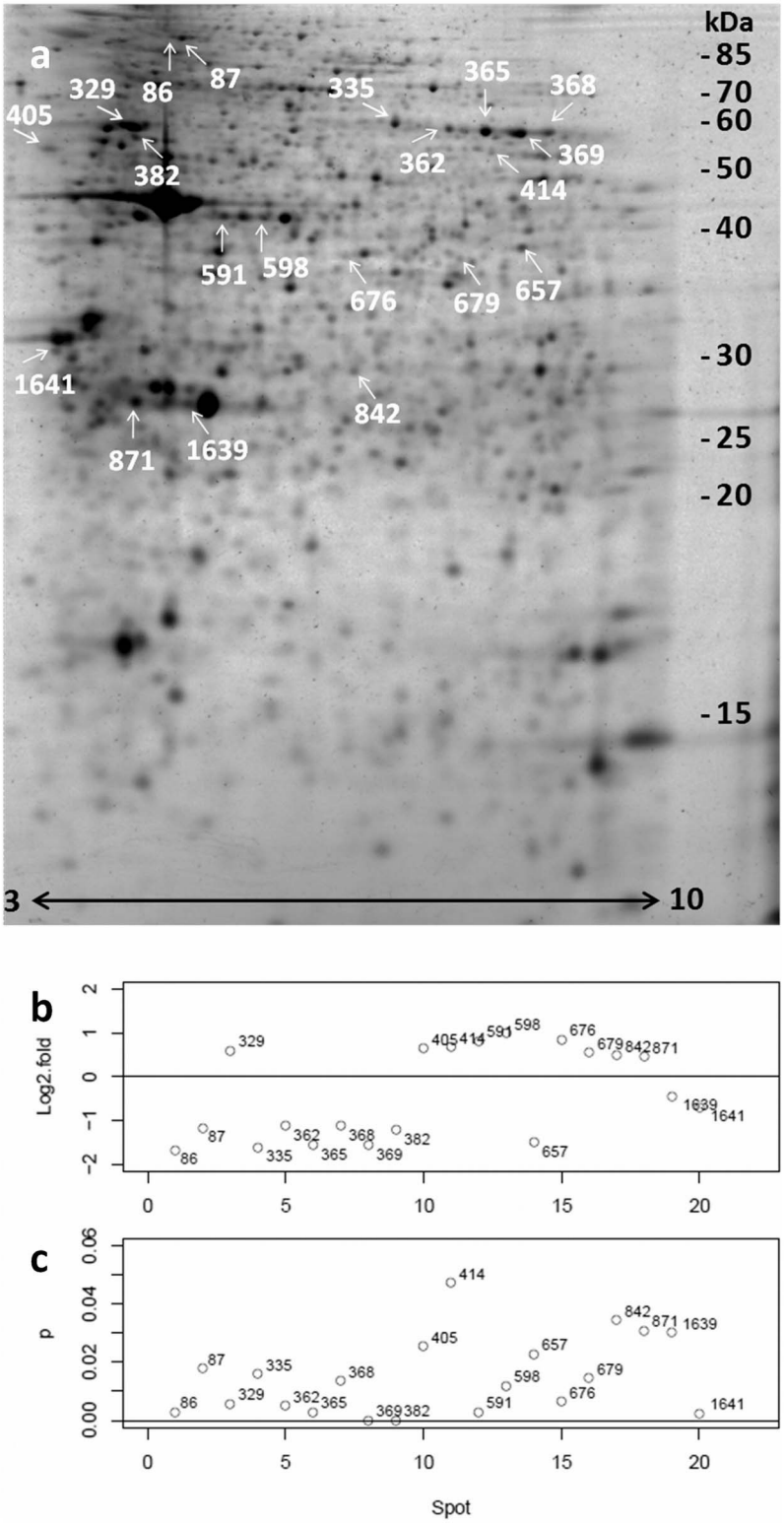

Figure 1 | Identification of 20 spots that discriminate PD patients from control subjects. Their position in the 2-DE map of T-cells from a control subject (a), the fold of change in Log_2 scale (b) and the Wilcoxon test p value (c) are reported.

Results

Two-dimensional electrophoresis profiling of T-cell proteins. We obtained T-cell protein expression profiles from the enrolled groups by 2-DE. First, we screened the profiles to identify proteins or protein modifications whose changes were linked to confounding factors such as therapy and age[13]. Spots showing linear Pearson correlation with age (evaluated in control subjects only) or daily L-DOPA dose, or showing significant differences between patients treated or not with dopamine agonists (Wilcoxon test, $p<0.05$) were excluded from the further analysis. By comparing 2-DE maps from 15 PD patients to 17 control subjects, we selected 20 protein spots showing different levels in the two groups (Wilcoxon test, $p<0.05$) (**Fig. 1** and **Supplemental Fig. 1**). Proteins corresponding to selected spots were identified by LC-MS/MS as reported in **Table 1**

Table 1 | Identity of protein spots

Spot No.	Protein	Uniprot Id	Correlation	p
A. PD patients vs. control subjects (Wilcoxon)				
86	Vinculin	P18206	↓ PD	0.003
87	Vinculin	P18206	↓ PD	0.018
329	Vimentin	P08670	↑ PD	0.006
335	Talin-1	Q9Y490	↓ PD	0.016
362	Beta-fibrinogen	P02675	↓ PD	0.005
365	Beta-fibrinogen	P02675	↓ PD	0.003
368	Beta-fibrinogen	P02675	↓ PD	0.014
369	Beta-fibrinogen	P02675	↓ PD	<0.001
382	Filamin-A	P21333	↓ PD	<0.001
405	Lymphocyte-specific Protein 1	P33241	↑ PD	0.025
414	Septin-6	Q14141	↑ PD	0.047
591	Vimentin	P08670	↑ PD	0.003
598	Moesin	P26038	↑ PD	0.012
657	Gelsolin	P06396	↓ PD	0.023
676	Transaldolase	P37837	↑ PD	0.007
679	Transaldolase	P37837	↑ PD	0.014
842	Twinfilin-2	Q6IBS0	↑ PD	0.034
871	Rho GDP dissociation inhibitor isoform 2	P52566	↑ PD	0.031
1639	beta actin fragment	P60709	↓ PD	0.030
1641	14-3-3 epsilon	P62258	↓ PD	0.002
B. LOPD patients vs. EOPD patients (Wilcoxon)				
351	Beta-tubulin	P07437	↑ LOPD	0.047
363	Protein disulfide isomerase A3	P30101	↑ LOPD	0.013
392	Vimentin	P08670	↓ LOPD	0.046
505	Plastin-2	P13796	↓ LOPD	0.013
792	Purine nucleoside phosphorylase	P00491	↑ LOPD	0.013
940	Glutathione S-transferase P	P09211	↑ LOPD	0.016
1787	PDCD6 interacting protein	Q8WUM4	↑ LOPD	0.047

(see **Supplemental Table 1** for details on protein identification by mass spectrometry). Multiple identifications were refined by matching with reference 2-DE maps available in the SWISS-2DPAGE database (http://world-2dpage.expasy.org/swiss-2dpage/). In particular, the LYMPHOCYTE_HUMAN map was used for disambiguation. In the case ambiguous assignment was not resolved by matching (see **Supplemental Results**), the correct assignment was verified by Western blotting (**Supplemental Fig. 2** and **Supplemental Fig. 3**). All peptides attributable to keratins or trypsin (common contaminants) were eliminated from the mascot search.

Correlation of selected spot volumes with Hoehn and Yahr score and disease duration. We evaluated the occurrence of linear correlation of selected spot volumes with the Hoehn and Yahr score and the duration of the disease, to test if some of the proposed biomarkers changed with time or disease severity. Correlation with the Hoehn and Yahr score was observed for spots 86, 362, 365, 368, 369, 405, 598 and 1641 (**Supplemental Fig. 4**). These spot volumes were linearly combined to afford a function that significantly tracks the Hoehn and Yahr score. Multiple regression on PD patients allowed us to determine a set of coefficients for the eight spots listed above (**Supplemental Table 2**). In this way, the spot volume combination linearly correlates with the Hoehn and Yahr score ($r = 0.674$, $p = 0.006$) (**Fig. 2a**). The significance of the correlation is improved by adding control subjects at $x = 0$ ($r = 0.603$, $p = 3\times10^{-4}$) (**Fig. 2b**).

Correlation with the duration of the disease was observed for spots 369, 591, 598 and 1641 (**Supplemental Fig. 5**). These spot volumes were linearly combined to afford a function that significantly tracks disease duration. Multiple regression on PD patients allowed us to determine a set of coefficients for the four spots listed above (**Supplemental Table 2**). In this way, the spot volume combination linearly correlates with the disease duration ($r = 0.516$, $p = 0.049$) (**Fig. 2c**). Also in this case, the correlation is improved by adding control subjects at $x = 0$ ($r = 0.548$, $p = 0.001$) (**Fig. 2d**). Worth of note, spots 369, 598 and 1641 correlate with both parameters.

Linear discriminant analysis of selected spot levels. We analyzed all spots ($n = 20$) showing different levels in PD patients by LDA so to obtain a likelihood score (PD Score) expressed as a linear combination of relative spot volumes (**Supplemental Table 3**). The spots were selected on the basis of uncorrected $p < 0.05$ so to avoid the possible exclusion of factors that could contribute to the model even if their change should not be considered as significant in multiple testing. Spot combinations were significantly different in PD patients with respect to control subjects (Wilcoxon test, $p < 10^{-8}$), with a cutoff value of 0.67. To effectively test the performance of the model, each subject was iteratively excluded from the training set and predicted on the basis of the other subjects. According to the "leave-one-out" procedure we obtained 87% sensitivity and 81% specificity. Predictions so obtained were used to build a ROC curve with an area under curve of 0.906 (**Fig. 3**).

A simplification of the model was achieved by ranking the 20 spots in terms of their ability to discriminate PD patients from control subjects. Thus, the six spots showing the worst contribution (weight < 0.3) were discarded and the LDA was performed on the remaining 14 spots (**Supplemental Table 3**). Likelihood scores (PD Score) were significantly different in PD patients with respect to control subjects (Wilcoxon test, $p < 10^{-8}$), with a cutoff value of -0.31. In this case, the "leave-one-out" cross-validation procedure of the model allowed us to obtain 100% sensitivity and 94% specificity. Predictions obtained so far were used to build a ROC curve with an area under curve of 0.996 (**Fig. 3**).

We further simplified the model by removing spots with weight $<$ 1, thus obtaining a 9-spot model (**Supplemental Table 3**). Again, LDA yielded likelihood scores (PD Score) significantly different in

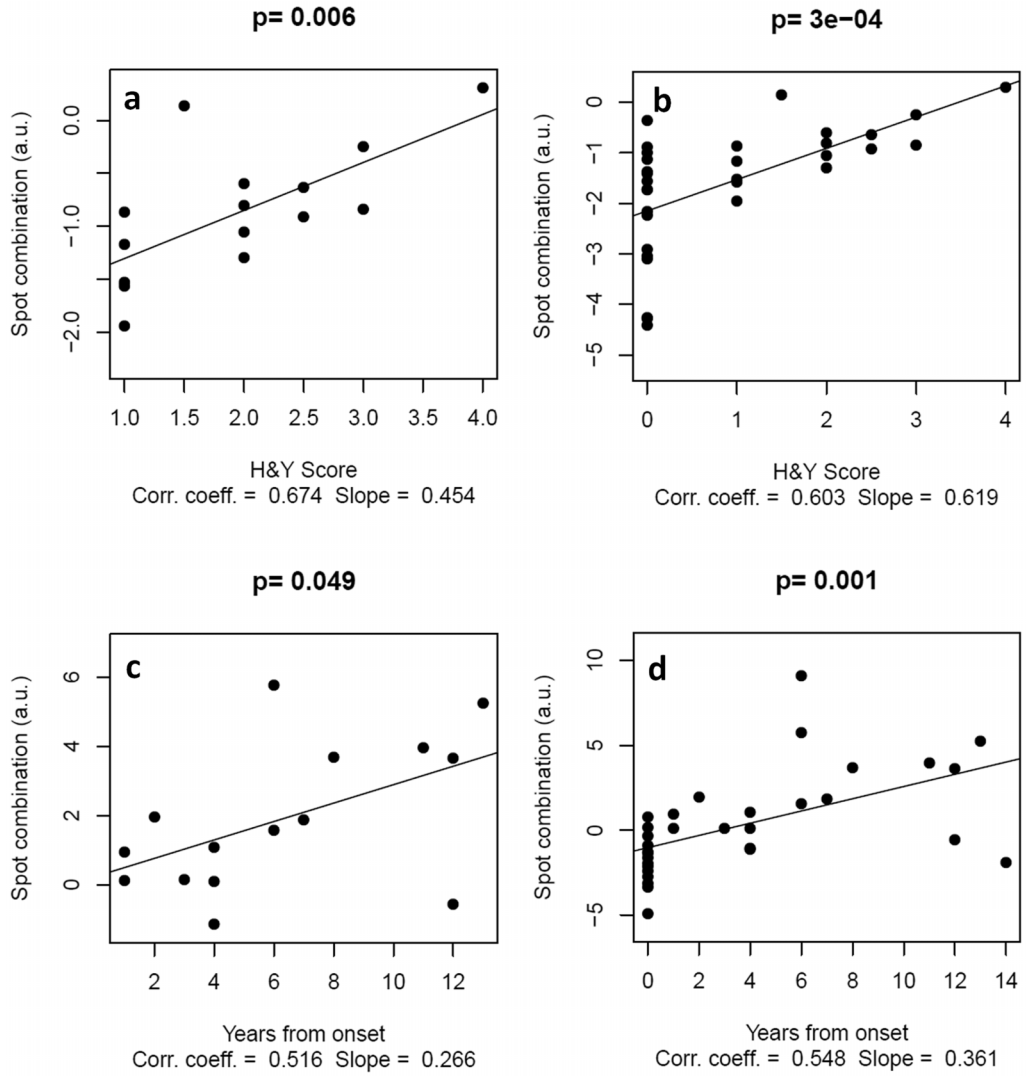

Figure 2 | Predictors of PD severity and duration. Panels a and b: a linear combination of spots 86 (vinculin), 362 (vinculin), 365 (β-fibrinogen), 368 (β-fibrinogen), 369 (β-fibrinogen), 405 (lymphocyte-specific protein 1), 598 (moesin) and 1641 (14-3-3 epsilon) correlates with the Hoehn and Yahr score. Panels c and d: a linear combination of spots 369 (β-fibrinogen), 591 (vimentin), 598 (moesin) and 1641 (14-3-3 epsilon) correlates with the disease duration. Linear correlations are significant both excluding (a, c) and including (b, d) control subjects. P values refer to slope different from zero.

PD patients with respect to control subjects (Wilcoxon test, $p < 10^{-8}$), with a cutoff value of −2.90. The "leave-one-out" cross-validation procedure of the model allowed us to obtain 100% sensitivity and 88% specificity, and the ROC curve built with these predictions had an area under curve of 0.992 (**Fig. 3**). For comparison, results of the three proposed predictive models are summarized in **Table 2**.

To test the power of our analysis we evaluated the intra-group variance and the difference of mean values for each spot included in the three models (equation (4)). The minimum number of subjects required for a significant verification is in the range from 10 to 30 for most spots (**Supplemental Table 4**).

Total beta-fibrinogen levels are reduced in PD patients. We clustered spot groups by aggregative nesting according to Pearson correlation (**Supplemental Fig. 6**). Spots 362, 365, 368 and 369, all corresponding to β-fibrinogen, were clustered in a single outgroup, showing the most stringent similarity among spots considered in each model. To verify their proper identification and confirm the differences observed by 2-DE with another technique, we measured total β-fibrinogen levels by Western blotting in three control subjects and three PD patients, taken from the training set, and correlated them to the sum of relative volumes for spots 362, 365, 368 and 369 ($r = 0.889$; $p = 1.7\times10^{-4}$). Downregulation of β-fibrinogen involves all 2-DE isoforms, without an appreciable qualitative change of the pattern (**Fig. 4**).

Total transaldolase levels are increased in PD patients. Spots 676 and 679, both corresponding to transaldolase, showed increased levels in PD patients by 2-DE quantification and close correlation by aggregative nesting (**Supplemental Fig. 6**). To confirm this finding we evaluated transaldolase expression by two-dimensional Western blotting. **Fig. 5** shows the characteristic 2-DE pattern of transaldolase, with two trains of spots at different apparent molecular weight. In particular, its pattern was altered in three PD patients (two LO PD and one EO PD) with respect to three control subjects (two healthy subjects and one with atypical parkinsonism), patients showing a higher abundance of the higher molecular weight isoforms.

Linear discriminant analysis of spots showing differences between late-onset and early-onset patients. To test the possibility to discriminate PD subtypes, we compared LO PD patients to EO PD patients (**Fig. 6**). We selected 7 protein spots (Wilcoxon test, $p < 0.05$) and identified them by LC-MS/MS (**Table 1**; see **Supplemental**

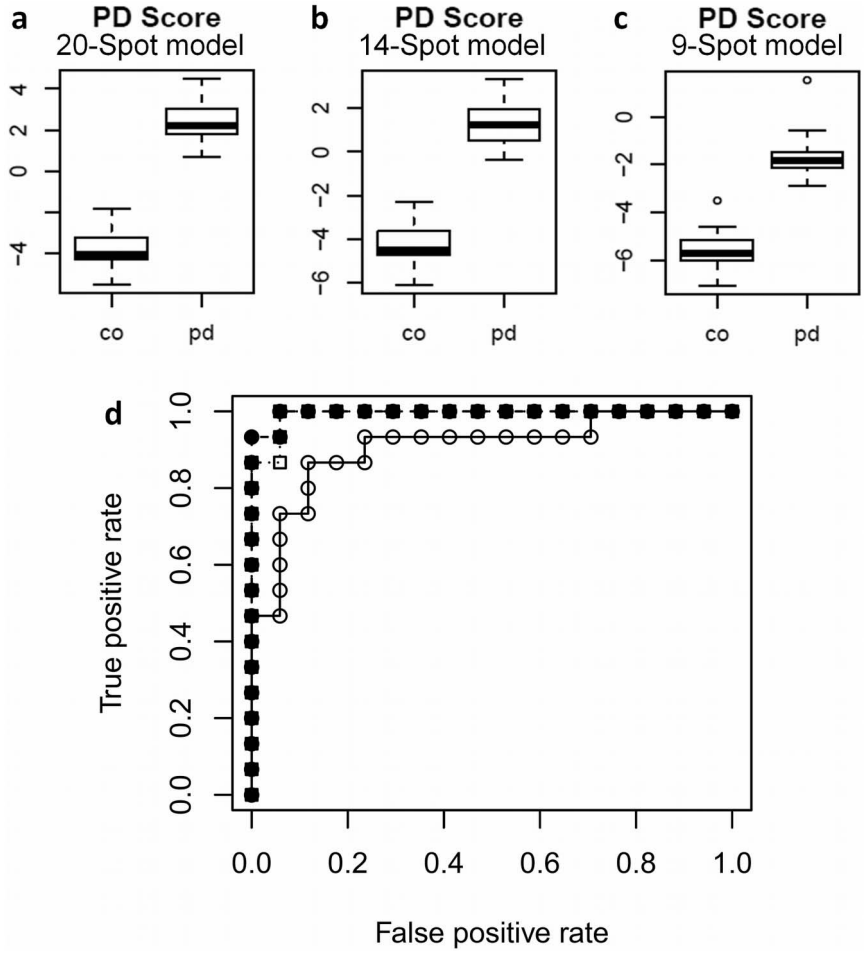

Figure 3 | **Definition of three predictive models by linear discriminant analysis.** Panels a, b and c: Scoring functions built on distinct sets of 20, 14 and 9 spots, respectively. Panel d: ROC curves for the leave-one-out cross-validation (open circles: 20-spot model; filled squares: 14-spot model; open squares: 9-spot models).

Table 1 for details on protein identification). We performed a linear discriminant analysis on selected spots in LO patients with respect to EO patients so to obtain a likelihood score (LO vs. EO Score) expressed as a linear combination of relative spot volumes (**Supplemental Table 5**). Spot combinations were significantly different in LO PD patients with respect to EO PD patients (Wilcoxon test, $p = 0.006$), with a cutoff value of 11.0. To effectively test the performance of the model, we performed the "leave-one-out" procedure described above and obtained 71% sensitivity and 100% specificity. Predictions obtained so far were used to build a ROC curve with an area under curve of 0.911 (**Fig. 7**).

Table 2 | Summary of performance parameters for the three predictive models [a]

	20-spot model	14-spot model	9-spot model
Cutoff	0.67	−0.31	−2.90
Sensitivity (%)	87	100	100
Specificity (%)	81	94	88
ROC AUC	0.906	0.996	0.992
p value	$< 10^{-8}$	$< 10^{-8}$	$< 10^{-8}$

[a]Cutoff values are expressed in terms of the likelihood PD score obtained as a linear combination of selected spot relative volumes; ROC = receiver operating characteristic; AUC = area under curve; p values refer to Wilcoxon test on PD scores. Sensitivity, specificity and ROC AUC were obtained by leave-one-out cross-validation.

Discussion

In this study we showed that T-lymphocyte proteome changes may be a valid tool to classify PD patients. Furthermore, we demonstrated that linear combinations of selected spots of the 2-DE gel display linear correlation with parameters such as disease duration (expressed as years from onset) and Hoehn and Yahr score. Eventually, we found a set of seven spots that can differentiate LO from EO PD patients.

The main advantage of this investigation is to merge the powerful ability of 2-DE to identify quantitative changes in single protein isoforms or post-translational modifications with a multiparametric approach where distinct 2-DE spots are combined in a predictive function. A similar result might not be reached by antibody-based techniques such as ELISA, where total protein levels are measured[4]. The combination of several spot volumes increased the specificity of the proposed biomarkers, while maintaining high sensitivity. Indeed, we were able to obtain ROC AUC values higher than 0.99 both in the 14-spot and in the 9-spot model. The 20-spot dataset was selected on the basis of the non-parametric Wilcoxon test, since the distribution of spot volumes deviates from normality and intra-group variance values are dissimilar. Therefore, LDA does not grant the best classification results when all 20 spots are included in the discriminant model. In an original way, we calculated the weight of each spot in the model (equation (3)), so to exclude those with a limited contribution. In this way, performance parameters of the models were markedly improved (**Table 2**).

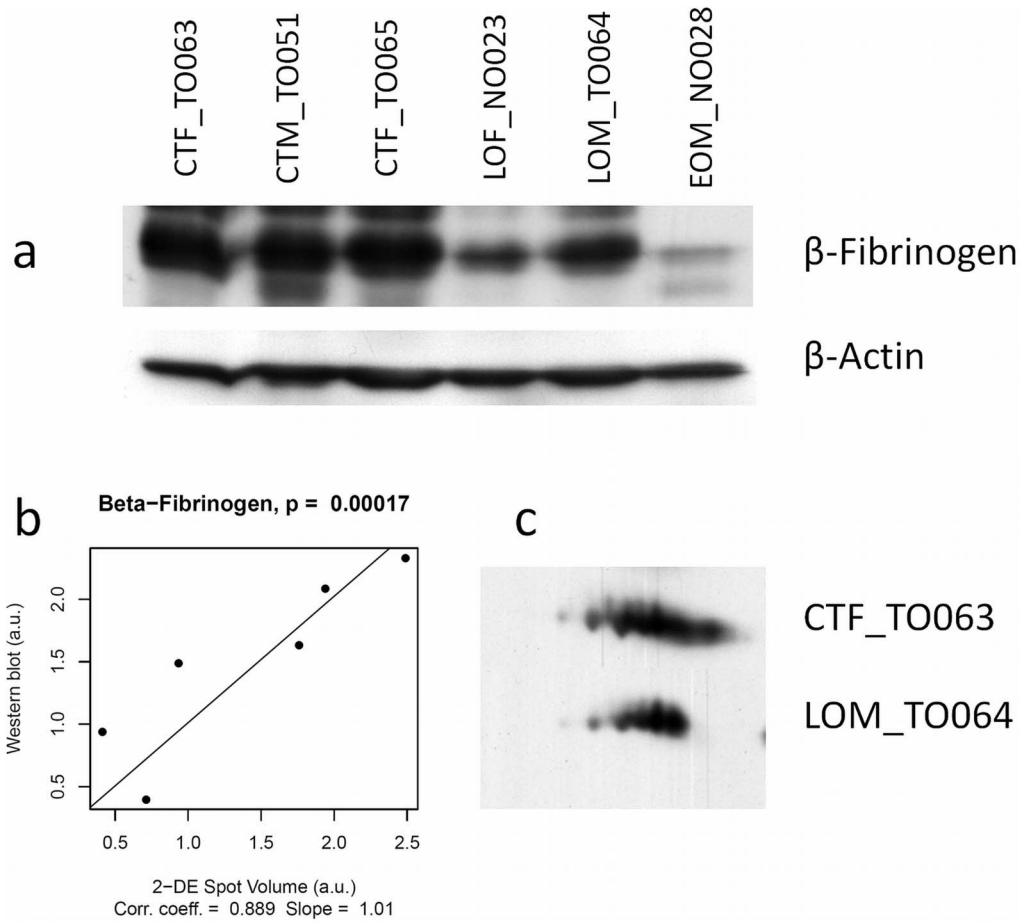

Figure 4 | **Western blotting validation of β-fibrinogen.** Panel a: Western blotting of total β-fibrinogen from six subjects selected among those enrolled in the biomarker discovery study. Panel b: Linear correlation of total β-fibrinogen levels as measured by 2-DE spot volume and by Western blotting. The correlation coefficient is calculated according to equation (1). Panel c: Two representative 2-DE Western blotting experiments.

Performance parameters obtained in this investigation need to be compared to similar biomarker discovery and validation studies appeared in the literature so far[3,4,5,14]. Targeted studies usually evaluate single biomarker alterations at the peripheral level that were first observed in the affected tissue[15]. A marked improvement is achieved when two or more biomarkers are measured in order to predict a disease probability based on multiple observations. Recently, a panel of four gene transcripts, selected on the basis of their altered levels in the substantia nigra, was measured in total blood cells of a large group of sporadic PD patients with respect to a smaller group of control subjects, but verified on the training set itself[16]. Measurement of α-synuclein, tau and total protein concentrations in CSF of patients with PD, Alzheimer disease, Dementia with Lewy Body, multiple system atrophy and non-neurodegenerative neurological disorders allowed Mollenhauer and coworkers to define a very accurate predictive model (ROC AUC = 0.909 for the training set); in a subsequent validation phase this model displayed a AUC = 0.706 for the classification of PD patients and control subjects[17].

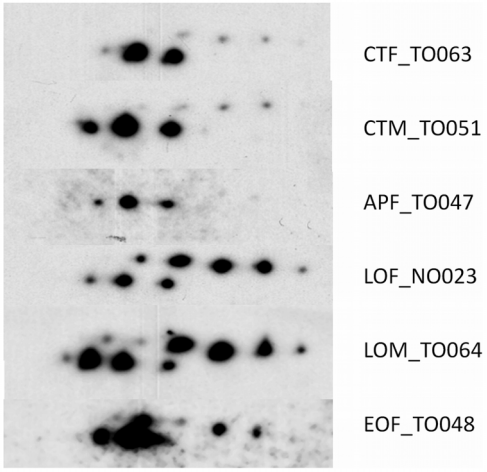

Figure 5 | **Two-dimensional Western blotting of transaldolase, showing two distinct pI distributions at different MW.** Samples have been selected in the training set among those showing larger discrepancies in the 2-DE quantification to validate the correct assignment of protein spots.

Unbiased discovery studies that rely on proteomics[4] or metabolomics[18] have the great advantage to identify a larger panel of targets to be included in the predictive model and to explore a greater universe of candidates without limiting to those that have been related to pathogenesis. We intentionally avoided to focus on the identity of selected proteins. Indeed, every single protein is far from being a significant element for PD early diagnosis[4,15]. What we wanted to point out was a protein fingerprinting of the disease state, more than a functional correlation between proteins altered at the central level and those that mirror neurodegeneration at the peripheral level. Nevertheless, we chose to verify the identity of spots corresponding to β-fibrinogen and transaldolase, not only in order to validate the results with an independent technique, but also for the predominant role of these spots in the models. Although little is known on the role of β-fibrinogen in T-lymphocytes, its downregulation complements

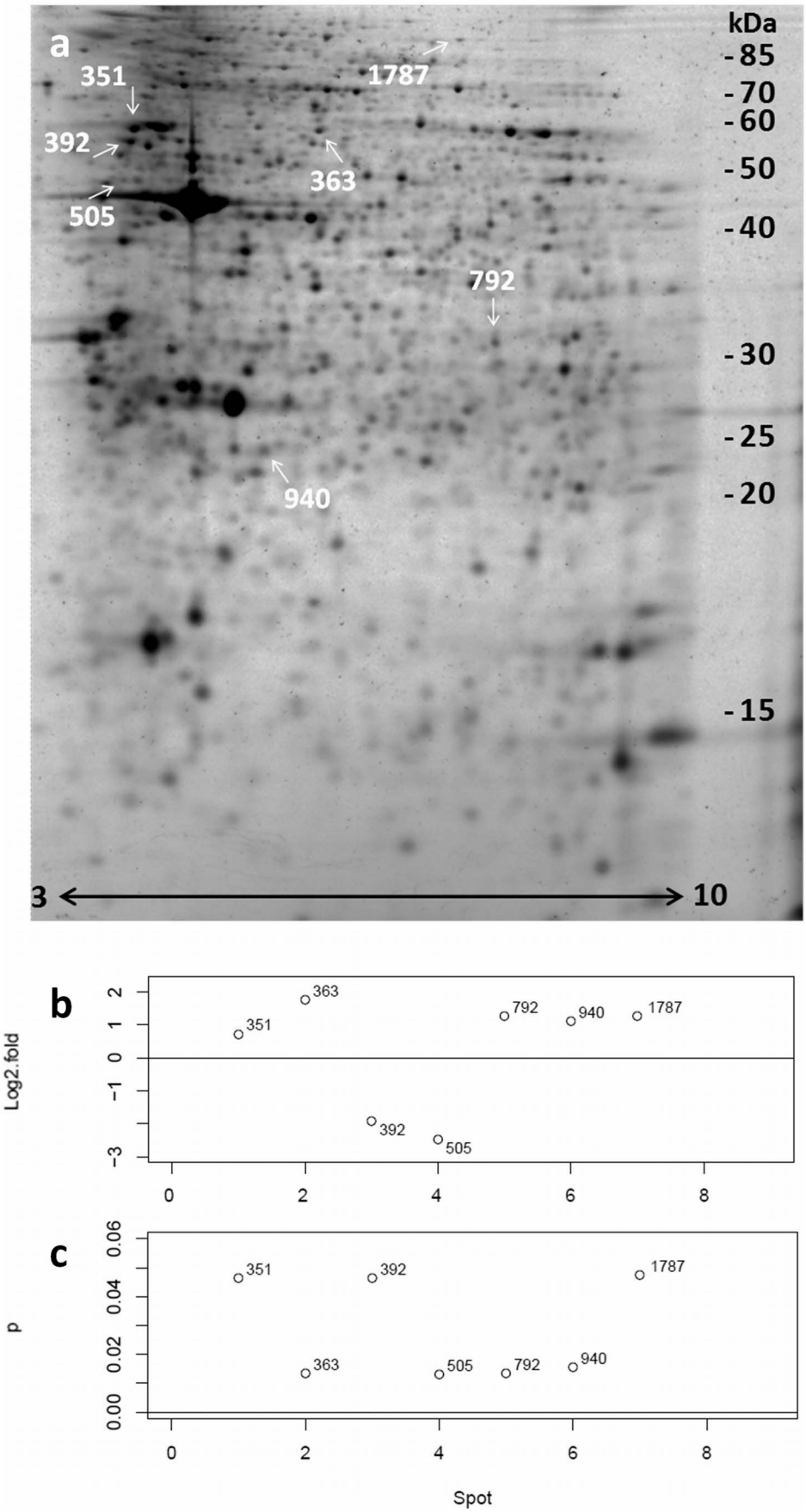

Figure 6 | Identification of 7 spots that discriminate late-onset (LO) from early-onset (EO) patients. For each of the 7 spots, the position in the 2-DE map of T-cells from a control subject (a), the fold of change in Log_2 scale (b) and the Wilcoxon test p value (c) are reported.

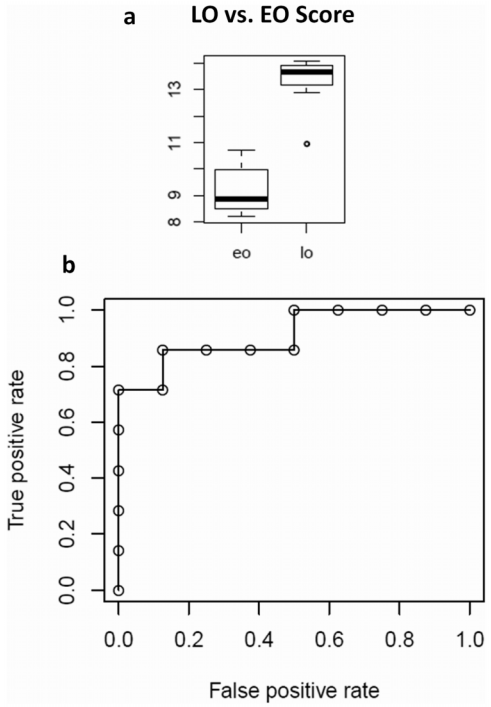

Figure 7 | Definition of a clinical subtype predictive model by linear discriminant analysis. Panel a: LO vs. EO classification scores. Panel b: ROC curve for the leave-one-out cross-validation.

the preliminary evidence that two distinct isoforms of γ-fibrinogen either correlate with the disease state or with disease duration in PBL of PD patients[12]. Noticeably, the 2-DE β-fibrinogen pattern is not significantly modified in PD patients with respect to control subjects and a lower total β-fibrinogen level was observed in lymphocytes of PD patients. On the other hand, transaldolase, a key enzyme of the pentose phosphate pathway (PPP), has a characteristic pattern in 2-DE[19] and has been involved in apoptotic cell death both in cancer and in neurodegeneration[20]. Pattern modifications have been linked to PPP regulation[19]. When comparing PD patients with control subjects, it appeared that PD patients have a remarkably higher abundance of the heavy isoforms. Although the experiment was limited to a small number of subjects to confirm proteomic results, the ratio between the two forms could contribute to the diagnostic panel.

A limitation that might be envisaged in the present investigation is the limited number of cases and control subjects included in the study. It should be noticed, however, that we have enrolled a sufficient number of subjects (about 15 per each group) based on established guidelines for biomarker discovery studies[15]. Being aware that, in the ideal biomarker pipeline, discovery, verification and validation should be decoupled steps of a phased approach[15], we have attempted to evaluate intra-group variance and the difference of mean values for each spot included in the models and to perform a power analysis of our study. On this basis, we assured that the minimum number of subjects required to verify the observed variation is in the range from 10 to 30 for almost every single spot, thus permitting an appropriate verification on the same data set by the leave-one-out validation procedure. Worthy of note, we measured protein levels in the verification step in the same cellular population where the discovery procedure was conducted, thus avoiding issues frequently emerging when using plasma and CSF samples (*i.e.*, high dilution of the candidates in the body fluid and mixture with components coming from different body compartments). In this context, enrolment of larger cohorts in the verification phase of the biomarker pipeline would not add much value to the significance of our results[15].

Table 3 | Summary of enrolled subjects

	PD (n = 17)	Controls (n = 15)
Age ± SD (years)	54.5 ± 9.2	59.9 ± 10.2
Male %	67	41
PD duration ± SD (years)	6 ± 4	
Medications		
Unmedicated	2	
L-DOPA	3	2
DA agonists	1	
L-DOPA + DA agonists	8	

The ability to diagnose PD long before the onset of motor signs and the availability of a neuroprotective treatment will eventually allow to slow down disease progression. Given the common features between T-cells and dopaminergic neurons, T-lymphocyte proteome biomarkers may reasonably provide information on the effectiveness of future (or under development) drugs designed to interfere with specific cell death mechanisms targeting nigral neurons. Our findings suggest the possibility to discriminate PD patients in terms of disease subtype (EO vs. LO) and to identify patients with different clinical features, in terms of disease evolution and treatment response[21,22]. Furthermore, the fact that several proposed biomarkers in this study correlate with the Hoehn and Yahr score opens the possibility to track "proteomic" PD progression.

Noticeably, studies aimed at the identification of candidate disease biomarkers have to correlate findings not only with disease state but also with patients' treatment, as T-lymphocytes are modified by dopaminergic therapy[13].

Further work is needed to lead the current finding to the clinical practice. Unfortunately, most biomarker studies stop at the discovery phase, so that a huge number of candidates have been proposed with few or no biomarkers reaching practical exploitation[23]. Main clinical validation issues are the selection of large cohorts of subjects with high heterogeneity in terms of gender, age, ethnicity, clinical phenotype[4]. This aim is likely to be reached only through large collaborative networks. Eventually, a general issue is the economic effort of the high-risk phase of validation of candidate biomarkers. Indeed, the absence of an objective measure that accurately tracks disease progression or allows early diagnosis strongly impacts on the costs that biotech- and pharma-companies will sustain in order to develop disease-modifying therapies that effectively extend patients quality of life.

Methods

Study subjects. Thirty-two subjects were enrolled by the Parkinson's Disease Center at the Department of Neuroscience, University of Torino, and by the Neurology Division at the Department of Translational Medicine, University of Eastern Piedmont. Every subject was associated to an alphanumeric code to ensure that his/her identity was not disclosed to investigators. Among them, 15 subjects were PD patients, varied in terms of age, age at onset, pharmacological treatment and familiarity. This population was intentionally enriched in early-onset patients (eight of 15) to highlight possible effects due to the genetic background. Similarly, we recruited 17 control subjects. This cohort included also three patients affected by atypical degenerative parkinsonisms. **Supplemental Table 6** reports demographic and clinical data for all enrolled subjects (for a summary see **Table 3**). Gender and age distributions were similar in different groups. Absolute inclusion criteria for PD patients were: idiopathic PD, absence of atypical signs and a good response to L-DOPA. Supportive criteria were: asymmetry of symptoms or signs at onset, clinical course of more than five years without atypical signs, L-DOPA induced motor fluctuations or dyskinesias. Exclusion criteria for patients and control subjects were: use of neuroleptic drugs, focal cerebral lesions and a history of encephalitis[1,24–26]. Subjects suffering from inflammatory or infectious diseases and subjects that took drugs capable of interfering with T-lymphocytes at the time of enrolment were excluded from the study. All patients signed an informed consent before being recruited for the present study, following approval by the Institutional Review Boards of the University Hospitals where subjects were enrolled and according to the *Declaration of Helsinki*.

T-lymphocyte isolation. All subjects underwent a venous blood sampling (20 ml) from the antecubital vein, between 9 and 10 a.m., after an overnight fast. Whole blood was collected into vacuum tubes containing EDTA, diluted with 50 ml of

phosphate-buffered saline (PBS) and stratified in two 50 ml tubes on top of 15 ml of Lympholyte®-H (Cedarlane) each. After centrifugation (800×g, 20 min, 20°C) peripheral blood mononuclear cells (PBMC) were removed and pelleted through centrifugation (400×g, 15 min, 20°C). Pellets were washed twice with 10 ml of magnetic-activated cell sorting (MACS) buffer (Miltenyi Biotec). The isolation of T lymphocytes was achieved by MACS with the Pan T-cell isolation kit (Miltenyi Biotec) using the manufacturer's protocol.

Two-dimensional electrophoresis and image analysis. T-lymphocytes were resuspended in 120 µl UTC (7 M urea, 2 M thiourea and 4% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS)) with a protease inhibitors cocktail (Sigma-Aldrich). Cells were left in this solution for 30 min to allow a complete cell lysis, sonicated (3×5 sec) and centrifuged (12000×g, 20 min, 10°C) to precipitate the cellular debris. Protein concentration in the extracts was determined according to Bradford. Total proteins (200 µg) were separated through two-dimensional electrophoresis (2-DE) using 18 cm IPG DryStrips with a nonlinear 3–10 pH gradient (GE Healthcare) followed by 14%T SDS-PAGE. The resulting maps were stained with Ru(II) tris(bathophenanthroline disulfonate) (Serva). Images were acquired (12 bit grayscale) with the GelDoc-It Imaging System (UVP) and analyzed with ImageMaster 2D Platinum (GE Healthcare). A common set of about 300 spots was selected for statistical analysis. Spots that were missing in more than 25% of gels were not taken into account.

Statistical analysis. Spot volumes were normalized by the total volume of a subset of spots common to all gels. Missing spot values were replaced by the mean value of the spot volume of the group or, if the mean was lower than the 98th percentile, by the minimum value observed in the group[27]. Relative volumes were analyzed by the non-parametric Wilcoxon test to find significant differences ($p < 0.05$) in patients with PD with respect to control subjects, in early-onset patients with respect to late-onset patients and in PD patients treated or not with dopamine agonists[28].

The Pearson linear correlation coefficient r was evaluated according to equation (1), where x is the independent variable (*i.e.*, Hoehn and Yahr score, age, years from onset, daily L-DOPA dose), y is the relative spot volume, $\overline{x}$ is the mean x value, $\overline{y}$ is the mean y value, xy is the x × y product, $\overline{xy}$ is the product of $\overline{x}$ and $\overline{y}$ mean values[28].

$$r = \frac{\frac{\sum xy}{n} - \overline{xy}}{\sqrt{\frac{\sum x^2}{n} - \overline{x}^2}\sqrt{\frac{\sum y^2}{n} - \overline{y}^2}} \quad (1)$$

Predictive models for the classification of PD patients with respect to control subjects and of early-onset patients with respect to late-onset were built by linear discriminant analysis (LDA) of the spots identified as described above[28]. In this case, missing values were set to half of the minimum value observed in the gel. A likelihood score was assigned to each subject by linear combination of relative spot volumes according to equation (2).

$$\text{PD } Score = \sum\nolimits_i c_i Vol_i \quad (2)$$

Simplified models were obtained by progressively removing groups of spots in terms of their discriminating weight (W), calculated according to equation (3),

$$W = \left| c_i \left(\overline{Vol_{i,CO}} - \overline{Vol_{i,PD}} \right) \right| \quad (3)$$

where c_i is the LDA coefficient for spot i and $\left| \left(\overline{Vol_{i,CO}} - \overline{Vol_{i,PD}} \right) \right|$ is the absolute separation of the mean values of spot i in control subjects (CO) and PD. Each predictive model has been tested with the leave-one-out method[28]. The performance of predictive models has been quantified by measuring the area under the ROC curve.

Aggregative nesting of spots included in the predictive model was based on pair-wise Pearson correlation (equation (1)), where x and y are relative spot volumes.

The minimum number of subjects to be included in validation studies was calculated according to equation (4),

$$n = \frac{(N_{CO} - 1)\cdot\sigma_{CO}^2 + (N_{PD} - 1)\cdot\sigma_{PD}^2}{(N_{CO} + N_{PD} - 2)\cdot\left(\frac{\mu_{CO} - \mu_{PD}}{2.2}\right)^2} \quad (4)$$

where σ_{CO} and σ_{PD} are standard deviations of relative spot volumes in CO and PD groups, N_{CO} and N_{PD} are numbers of subjects in the groups, and μ_{CO} and μ_{PD} are mean values of relative spot volumes in the groups[28].

All procedures for data analysis and graphics were written in ***R***, an open-source environment for statistical computing[29].

In-gel digestion, mass spectrometry and protein identification. Spots showing $p < 0.05$ (Wilcoxon test) and not correlating with therapy[13] were manually excised and destained (50% ethanol), dehydrated with acetonitrile (2 × 20 min 100 µl) and then dried at 37°C by vacuum centrifugation. The gel pieces were then swollen in 10 µl digestion buffer containing 50 mM NH_4HCO_3 and 12.5 ng/µl modified porcine trypsin (sequencing grade, Promega). After 10 min, 30 µl of 50 mM NH_4HCO_3 were added to the gel pieces and digestion allowed to proceed at 37°C overnight. The supernatants were collected and peptides were extracted in an ultrasonic bath for 10 min (twice 100 µl 50% acetonitrile, 50% H_2O, 1% formic acid; twice 50 µl acetonitrile). All the supernatants were collected in the same tube, dried by vacuum centrifugation and dissolved in 5 µl 2% acetonitrile, 0.1% formic acid in water.

Peptides from each sample were then separated by reversed phase nano-HPLC-Chip technology (Agilent Technologies) online-coupled with a 3D ion trap mass spectrometer (model Esquire 6000, Bruker Daltonics). The chip was composed of a Zorbax 300SB-C18 (43 mm×75 µm, with a 5 µm particle size) analytical column and a Zorbax 300SB-C18 (40 nL, 5 µm) enrichment column. The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) softwares. The scan range used was from 300 to 1800 m/z. For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides of each m/z were selected to be further isolated and fragmented. The MS/MS scanning was performed in the normal resolution mode at a scan rate of 13000 m/z per second. A total of five scans were averaged to obtain an MS/MS spectrum. Protein identification was manually performed by searching the National Center for Biotechnology Information non-redundant database using the Mascot 2.3 MS/MS Ion Search program (http://www.matrixscience.com). The following parameters were set: specific trypsin digestion, up to one missed cleavage; complete carbamidomethylation of cysteines, partial oxidation of methionines, partial protein N-terminal acetylation, peptide mass tolerance ± 0.9 Da, fragment mass tolerance ± 0.9 Da, 1+ to 3+ peptide charge, species restriction to human. All identified proteins had a Mascot score corresponding to a statistically significant ($p < 0.05$) confident identification according to Fisher's test. At least 2 different peptides had to be assigned. Peptide and protein identifications corresponding to keratins or trypsin were not taken into account.

Quantitative Western blotting analysis. For one-dimensional Western blotting, cell lysates (20 µg) were denatured in Laemmli sample buffer for 5 min at 98°C and electrophoresed on 10% SDS-PAGE gel. 2-DE was performed as described above. Gels were transferred to polyvinylidene difluoride (PVDF) membranes at 1 mA/cm², 1.5 h (TE77pwr, Hoefer). Membranes were saturated in 5% non-fat milk in TBS-T (0.1 M Tris-HCl pH 7.4, 1.5 M NaCl and 0.5% Tween-20) and incubated in the same buffer at 4°C overnight with goat anti-fibrinogen polyclonal antibody (Abnova PAB11318), 1 : 10000 dilution, with mouse anti-transaldolase polyclonal antibody (Abcam ab67467), 1 : 500 dilution, or with mouse anti-β-actin monoclonal antibody (GeneTex GTX23280), 1 : 3000 dilution. Membranes were then washed with TBS-T and incubated with peroxidase-conjugated anti-goat-IgG antibody (Millipore AP106P), 1 : 8000 in 5% milk-TBS-T, or anti-mouse-IgG antibody (Millipore 12–349), 1 : 3000 in 5% milk-TBS-T, respectively, for chemiluminescence detection (Millipore). Photographic films were scanned with a Epson Perfection V750 Pro transmission scanner (Epson) and images (16 bit grayscale) were analyzed using the ImageJ software (http://rsb.info.nih.gov/ij/). Images showing dynamic range saturation were discarded. Signal intensities were corrected for protein loading by normalization to actin intensity.

1. Jankovic, J. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry.* **79**, 368–376 (2008).
2. Shulman, J. M., De Jager, P. L. & Feany, M. B. Parkinson's disease: genetics and pathogenesis. *Annu Rev Pathol.* **6**, 193–222 (2011).
3. Morgan, J. C., Mehta, S. H. & Sethi, K. D. Biomarkers in Parkinson's disease. *Curr Neurol Neurosci Rep.* **10**, 423–430 (2010).
4. Alberio, T. & Fasano, M. Proteomics in Parkinson's disease: An unbiased approach towards peripheral biomarkers and new therapies. *J Biotechnol.* **156**, 325–337 (2011).
5. Gerlach, M. *et al.* Biomarker candidates of neurodegeneration in Parkinson's disease for the evaluation of disease-modifying therapeutics. *J Neural Transm.* **119**, 39–52 (2012).
6. Fasano, M., Alberio, T. & Lopiano, L. Peripheral biomarkers of Parkinson's disease as early reporters of central neurodegeneration. *Biomark Med.* **2**, 465–478 (2008).
7. Alberio, T. *et al.* Proteomic characterization of Jurkat T leukemic cells after dopamine stimulation: A model of circulating dopamine-sensitive cells. *Biochimie.* **93**, 892–898 (2011).
8. Pacheco, R., Prado, C. E., Barrientos, M. J. & Bernales, S. Role of dopamine in the physiology of T-cells and dendritic cells. *J Neuroimmunol.* **216**, 8–19 (2009).
9. Besser, M. J., Ganor, Y. & Levite, M. Dopamine by itself activates either D2, D3 or D1/D5 dopaminergic receptors in normal human T-cells and triggers the selective secretion of either IL-10, TNF alpha or both. *J Neuroimmunol.* **169**, 161–171 (2005).
10. Blandini, F. *et al.* Calcium homeostasis is dysregulated in parkinsonian patients with L-DOPA-induced dyskinesias. *Clin Neuropharmacol.* **32**, 133–139 (2009).
11. Jacob, A. M. & Turck, C. W. Detection of post-translational modifications by fluorescent staining of two-dimensional gels. *Methods Mol Biol.* **446**, 21–32 (2008).
12. Mila, S. *et al.* Lymphocyte proteomics of Parkinson's disease patients reveals cytoskeletal protein dysregulation and oxidative stress. *Biomark Med.* **3**, 117–128 (2009).
13. Alberio, T. *et al.* Dopaminergic therapies modulate the T-cell proteome of patients with Parkinson's disease. *IUBMB Life* **64**, 846–852 (2012).
14. Nyhlén, J., Constantinescu, R. & Zetterberg, H. Problems associated with fluid biomarkers for Parkinson's disease. *Biomark Med.* **4**, 671–681 (2010).

15. Surinova, S. *et al.* On the development of plasma protein biomarkers. *J Proteome Res.* **10**, 5–16 (2011).
16. Grünblatt, E. *et al.* Pilot study: peripheral biomarkers for diagnosing sporadic Parkinson's disease. *J Neural Transm.* **117**, 1387–1393 (2010).
17. Mollenhauer, B. *et al.* alpha-Synuclein and tau concentrations in cerebrospinal fluid of patients presenting with parkinsonism: a cohort study. *Lancet Neurol.* **10**, 230–340 (2011).
18. Bogdanov, M. *et al.* Metabolomic profiling to develop blood biomarkers for Parkinson's disease. *Brain.* **131**, 389–396 (2008).
19. Lachaise, F. *et al.* Relationship between posttranslational modification of transaldolase and catalase deficiency in UV-sensitive repair-deficient xeroderma pigmentosum fibroblasts and SV40-transformed human cells. *Free Radic Biol Med.* **30**, 1365–1373 (2001).
20. Samland, A. K. & Sprenger, G. A. Transaldolase: from biochemistry to human disease. *Int J Biochem Cell Biol.* **41**, 1482–1494 (2009).
21. Calne, S. M. & Kumar, A. Young onset Parkinson's disease. Practical management of medical issues. *Parkinsonism Relat Disord* **14**, 133–142 (2008).
22. Merola, A. *et al.* Subthalamic nucleus deep brain stimulation outcome in young onset Parkinson's disease: a role for age at disease onset? *J Neurol Neurosurg Psychiatry.* **83**, 251–257 (2012).
23. Frasier, M., Chowdhury, S., Eberling, J. & Sherer, T. Biomarkers in Parkinson's disease: a funder's perspective. *Biomark Med.* **4**, 723–729 (2010).
24. Litvan, I. *et al.* Task force appraisal of clinical diagnostic criteria for parkinsonian disorders. *Mov. Disord.* **18**, 467–486 (2003).
25. Litvan, I. *et al.* Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop. *Neurology* **47**, 1–9 (1996).
26. Gilman, S. *et al.* Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* **71**, 670–676 (2008).
27. Albrecht, D., Kniemeyer, O., Brakhage, A. A. & Guthke, R. Missing values in gel-based proteomics. *Proteomics.* **10**, 1202–1211 (2010).
28. McDonald, J. H. Handbook of Biological Statistics. 2nd ed. Baltimore: Sparky House Publishing; 2009.
29. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2009.

Acknowledgments

This work was supported by a grant from "Associazione Amici Parkinsoniani Piemonte ONLUS" to LL and MF. TA is a recipient of a postdoctoral research contract (UNIRE) from Regione Lombardia. Authors are indebted to Dr. Piergiorgio Gili for the continuous support, to Dr. Aristide Merola for assistance in patients enrolment, and to Prof. Riccardo Fesce for helpful discussion.. The authors deeply appreciate those who have donated their blood for this study.

Author contributions

TA, ACP and SO performed the 2-DE and Western blot experiments. CC and MZ enrolled the patients and collected clinical/demographic data. DC identified the proteins by mass spectrometry. MF wrote the software routines in ***R*** and analyzed the data. TA and MF designed the experiments. TA, MF and LL wrote the manuscript. All authors critically revised the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: TA, MF and LL are listed as inventors in the patent application IT/TO2011A001241 entitled "Method for the in vitro diagnosis of Parkinson's disease".

How to cite this article: Alberio, T. *et al.* Discovery and verification of panels of T-lymphocyte proteins as biomarkers of Parkinson's disease. *Sci. Rep.* **2**, 953; DOI:10.1038/srep00953 (2012).