JOURNAL OF PROTEOMICS 90 (2013) 107–114

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Parkinson's disease plasma biomarkers: An automated literature analysis followed by experimental validation☆

Tiziana Alberio[a,b,1], Enrico M. Bucci[c,1], Massimo Natale[c,d], Dario Bonino[c], Marco Di Giovanni[e], Edo Bottacchi[e], Mauro Fasano[a,b,*]

[a]Division of Biomedical Research, Department of Theoretical and Applied Sciences, University of Insubria, Busto Arsizio, Italy
[b]Center of Neuroscience, University of Insubria, Busto Arsizio, Italy
[c]BioDigitalValley s.r.l., Pont Saint Martin (AO), Italy
[d]Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy
[e]Department of Neurology, Ospedale Regionale di Aosta, Aosta, Italy

ARTICLE INFO

Available online 4 February 2013

Keywords:
2-DE
Plasma
Parkinson's disease
Biomarkers
Meta-analysis

ABSTRACT

Diagnosis of Parkinson's disease (PD) is currently assessed by the clinical evaluation of extrapyramidal signs. The identification of specific biomarkers would be advisable, however most studies stop at the discovery phase, with no biomarkers reaching clinical exploitation. To this purpose, we developed an automated literature analysis procedure to retrieve all the background knowledge available in public databases. The bioinformatic platform allowed us to analyze more than 51,000 scientific papers dealing with PD, containing information on 4121 proteins. Out of these, we could track back 35 PD-related proteins as present in at least two published 2-DE maps of human plasma. Then, 9 different proteins (haptoglobin, transthyretin, apolipoprotein A-1, serum amyloid P component, apolipoprotein E, complement factor H, fibrinogen γ, thrombin, complement C3) split into 32 spots were identified as a potential diagnostic pattern. Eventually, we compared the collected literature data to experimental gels from 90 subjects (45 PD patients, 45 non-neurodegenerative control subjects) to experimentally verify their potential as plasma biomarkers of PD.
This article is part of a Special Issue entitled: From Genome to Proteome: Open Innovations.

1. Introduction

The identification of specific biomarkers for neurodegenerative disorders is one of the main goal of the current clinical research. Diagnosis of the second most common neurological disorder, Parkinson's disease (PD), is currently assessed by 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 peripheral PD biomarkers would be critical for the differential diagnosis between PD and other neurodegenerative diseases that share some clinical features with PD. Moreover, early diagnosis remains a primary aim, needed also for the assessment of disease-modifying drugs [3–5]. Eventually, biomarkers of disease progression would allow a better patients follow-up and an objective measurement in clinical trials [6]. Non-motor signs frequently precede the onset of typical PD motor dysfunctions but they lack the specificity to be informative for

☆ This article is part of a Special Issue entitled: From Genome to Proteome: Open Innovations.
* *Corresponding author at:* Division of Biomedical Research, Department of Theoretical and Applied Sciences, University of Insubria, via Alberto da Giussano 12, I-21052 Busto Arsizio, Italy. Tel.: +39 0331 339450; fax: +39 0331 339459.
E-mail address: mauro.fasano@uninsubria.it (M. Fasano).
[1] These authors contributed equally.

1874-3919/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jprot.2013.01.025

the clinician; on the other hand, instrumental investigations (polysomnograpy or functional imaging) are characterized by high costs and the use of radioactive tracers. For these reasons, a biochemical marker to be evaluated in the periphery would be highly preferred [3–5].

Several PD biomarker candidates were discovered using unbiased proteomic approaches [4,7]. 2-DE is able to resolve thousands of spots simultaneously and can visualize protein processing and post-translational modifications [8]. For these reasons, it has been widely used in biomarker discovery studies. In regard to the tissue where to look for biomarkers, human plasma was broadly exploited because potentially informative about almost any disease state and very easy to obtain with little discomfort for patients. Moreover, plasma seems to be suitable for proteomic analyses to be applied in the clinical practice [9]. Nevertheless, the proteomic investigation of human plasma, especially by 2-DE, has proved to be challenging, because of the very complex nature of the sample and the presence of few highly-abundant proteins [10].

Meta-analysis of proteomics data may represent a valuable tool to extract information from datasets present in the literature, thus increasing the sample size with respect to a single proteomics study [11]. In particular, meta-analysis of 2-DE data available in the literature has proved to have a great potential to retrieve valuable information related to Parkinson's disease, which otherwise would remain hidden [12]. Plenty of information is available online in public domain databases. However, these datasets need to be filtered in order to sensibly and specifically find proteins that are associated to PD as biomarkers [13]. Here again, the meta-analysis of 2-DE experiments present in the scientific literature or in pertinent datasets can help to identify biologically relevant biomarkers by filtering out proteins involved in generic processes (e.g., cell-cycle control, signal transduction, and stress responses) [14]. Eventually, the data obtained from the analysis of the literature should be further validated in an independent manner in a cohort large enough to ensure statistical significance.

In this study, we developed a bioinformatic platform for the automatic analysis of proteomics literature data (Parkinson Informative System, ParIS, http://www.biodigitalvalley.com/researchprojects.html). We used a meta-analysis approach to filter literature data and identify a list of potential PD biomarkers, proposed by different authors and visible in 2-DE maps of human plasma. We further verified these markers in a cohort of 45 PD patients compared to 45 control subjects with neurological, non neurodegenerative disorders. Some of the markers were confirmed, while others did not reach statistical significance. Eventually, we built a predictive model able to significantly stratify PD patients by the identification of all the spots whose level was significantly different in the two groups.

2. Methods

2.1. 2-DE gel database and proteomic meta-analysis

In order to extract human plasma 2-DE images from the literature, we parsed PDF files of selected articles using the BFO Java library (http://bfo.com) as previously described [12]. Briefly, the 2-DE images extracted from proteomic papers were annotated for all the protein spots, according to identifications provided by the article authors, matching the symbols and labels on the gels to the information given in the tables or in the text of the paper. Spots were further associated to their pixel Cartesian position and to pI and MW data. To get putative spot identification, we aligned the 22 literature-retrieved 2-DE gels to the human plasma reference 2-DE map (PLASMA_HUMAN) from Expasy SWISS-2DPAGE (http://world-2dpage.expasy.org/swiss-2dpage/). To store all the retrieved gels and their annotations, a human plasma 2-DE gel database was built using a MySQL environment on a Red-Hat server.

2.2. Subjects

Ninety subjects were enrolled by the Department of Neurology at the Aosta Regional Hospital, after approval by the local Institutional Review Board. All patients signed an informed consent before being recruited for the present study, according to the *Declaration of Helsinki*. Every subject was associated to an alphanumeric code by the medical personnel involved in this study, to ensure that his/her identity was not apparent to investigators. Among the enrolled individuals, 45 subjects were idiopathic PD patients, varied in terms of age, age at onset, pharmacological treatment, with no familiarity and no other co-occurring neurological disease. Beside personal details, for each subject affected by Parkinson's disease we registered the presumptive onset year, the Hoehn & Yahr score and the dopaminergic therapy (Table 1). The 45 control subjects were recruited by the same Hospital Department among neurological patients not affected by neurodegeneration or multiple sclerosis. Gender and age distributions were similar in different groups (Table 1). Plasma samples were collected according to the protocol of the Parkinson's progression markers initiative (http://www.ppmi-info.org/wp-content/uploads/2012/03/PPMI-Biologics-Manual.pdf). Immediately after the blood collection the tubes were gently mixed and within 30 min the samples were centrifuged at 1500 ×*g* at 4 °C for 15 min. Plasma was collected into 1.5 ml aliquots and immediately stored at −80 °C.

Table 1 – Summary of enrolled subjects.

	Controls (n=45)	PD (n=45)
Anamnestics		
Age±SD (years)	69.1±8.7	70.3±7.7
Gender (males)	23 (51%)	26 (58%)
PD duration±SD (years)		7.3±4.3
Medications		
Unmedicated		0
L-DOPA		5
DA agonists		12
L-DOPA+DA agonists		28
Hoehn and Yahr stage		
1		14 (31%)
2		14 (31%)
3		13 (29%)
4		3 (7%)
5		1 (2%)