Neurobiology of Disease 74 (2015) 1–13

Contents lists available at ScienceDirect

Neurobiology of Disease

journal homepage: www.elsevier.com/locate/ynbdi

Comparative pathway and network analysis of brain transcriptome changes during adult aging and in Parkinson's disease

Enrico Glaab *, Reinhard Schneider

Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg

ARTICLE INFO

Article history:
Received 31 May 2014
Revised 23 September 2014
Accepted 3 November 2014
Available online 12 November 2014

Keywords:
Parkinson's disease
Aging
Neurodegeneration
Biomarker
Omics
Systems biology
Pathway analysis
Network analysis
Meta-analysis

ABSTRACT

Aging is considered as one of the main factors promoting the risk for Parkinson's disease (PD), and common mechanisms of dopamine neuron degeneration in aging and PD have been proposed in recent years. Here, we use a statistical meta-analysis of human brain transcriptomics data to investigate potential mechanistic relationships between adult brain aging and PD pathogenesis at the pathway and network level. The analyses identify statistically significant shared pathway and network alterations in aging and PD and an enrichment in PD-associated sequence variants from genome-wide association studies among the jointly deregulated genes. We find robust discriminative patterns for groups of functionally related genes with potential applications as combined risk biomarkers to detect aging- and PD-linked oxidative stress, e.g., a consistent over-expression of metallothioneins matching with findings in previous independent studies. Interestingly, analyzing the regulatory network and mouse knockout expression data for NR4A2, a transcription factor previously associated with rare mutations in PD and here found as the most significantly under-expressed gene in PD among the jointly altered genes, suggests that aging-related NR4A2 expression changes may increase PD risk via downstream effects similar to disease-linked mutations and to expression changes in sporadic PD. Overall, the analyses suggest mechanistic explanations for the age-dependence of PD risk and reveal significant and robust shared process alterations with potential applications in biomarker development for pre-symptomatic risk assessment or early stage diagnosis.

Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative disorders and a disease-modifying therapy is still not available. With an average age of onset of 60 years and a risk of developing sporadic PD known to increase significantly with age, the disease has been linked with aging by several studies (Bender et al., 2006; Collier et al., 2011; Frey et al., 2004; Hindle, 2010; Levy, 2007; Kaasinen and Rinne, 2002; Naoi and Maruyama, 1999). Previous hypotheses have suggested a combination of age-related neuronal attrition and environmental factors as a major cause for sporadic PD (Calne et al., 1986) or that aging may influence the clinical progression of the disease (Levy, 2007). More recently, PD has also been proposed to represent a form of premature or accelerated aging (Collier et al., 2011). Independent of the type and extent of association between aging and PD pathogenesis, various shared molecular hallmarks have been observed, including a gradual decline in dopamine synthesis (Scatton et al., 1983; Ota et al., 2006), reduced striatal density of the type 2 vesicular monoamine transporter (Frey et al., 2004) and increased levels of deleted mitochondrial DNA (Bender et al., 2006). These common features suggest that a more comprehensive investigation of shared/interlinked cellular process changes in aging and PD could provide new insights on the disease etiology and progression and facilitate the discovery of pre-symptomatic risk biomarkers for PD or general neurodegeneration.

In recent years, large-scale transcriptomic measurements from research studies on brain aging and complex neurodegenerative disorders have been made available in public data repositories (Barrett et al., 2009; Kang et al., 2011; Jones et al., 2009; Lein et al., 2006). Although these data sources have been analyzed individually (Kang et al., 2011; Johnson et al., 2009; Zhang et al., 2005; Lesnick et al., 2007; Kumar et al., 2013), the potential for a joint pathway- and network-analysis of high-throughput gene expression data for aging and PD has not yet been exploited, despite PD being regarded as one of the prime examples of an age-related disease (Hindle, 2010).

Here, we investigate relations between brain transcriptome changes in PD patients (as compared to age-matched, non-demented control subjects) and transcriptome changes associated with adult brain aging in a separate group of unaffected individuals, exploiting new cross-study data integration, pathway and network analysis methods. Specifically, we first apply a recent statistical meta-analysis approach (Marot et al., 2009) to 8 public microarray gene expression data sets (Zhang et al., 2005; Lesnick et al., 2007; Moran et al., 2006; Simunovic et al., 2009; Zheng et al., 2010), using *post mortem* samples from the midbrain

* Corresponding author at: Luxembourg Centre for Systems Biomedicine, 7, avenue des Hauts-Fourneaux, L-4362, Esch-sur-Alzette, Luxembourg. Fax: +352 466644 6949.
E-mail address: enrico.glaab@uni.lu (E. Glaab).
Available online on ScienceDirect (www.sciencedirect.com).

http://dx.doi.org/10.1016/j.nbd.2014.11.002

substantia nigra region in patients and age- and gender-matched controls, and then compare the differentially expressed genes in PD from the meta-analysis to genes associated with adult brain aging. These aging-associated genes are derived from a statistical analysis of *post mortem* microarray samples from the Human Brain Transcriptome (HBT) project (Kang et al., 2011), by determining significant brain gene expression changes across different age groups during adulthood. By integrating these data to identify shared and associated gene expression alterations during natural brain aging and in PD, we aim at two main goals: (1) obtaining a more detailed molecular-level understanding of how aging contributes to the risk for PD and (2) finding robust shared alterations in PD and aging for further evaluation as candidate early risk biomarkers for PD or general neurodegeneration. The potential of the jointly altered genes for biomarker applications is investigated using public data to determine their expression in peripheral tissues, to identify their previously reported PD/aging-linked peripheral changes and to evaluate the specificity of their alterations in PD as compared to Alzheimer's disease.

These analyses at the single-gene level are complemented by a statistical assessment of cellular processes changes, using our previously developed pathway and network analysis method EnrichNet (Glaab et al., 2012) to identify shared significant pathway and sub-network deregulations in PD and aging. The transcription factors most relevant for the regulation of these affected sub-networks are predicted using an over-representation analysis for transcription factor binding sites among the altered genes in PD/aging. Finally, to investigate possible relations between genetic variations and transcriptome alterations linked to PD/aging, we test the enrichment of PD-associated single-nucleotide sequence variants (SNPs) from public genome-wide association studies (GWAS) among the altered genes in PD/aging and report the genes found significant in both transcriptomics and GWAS analyses.

Materials and methods

Microarray data collection, pre-processing and differential expression analysis

In order to exploit the synergies of the available transcriptomics data for a joint analysis of PD pathogenesis and brain aging, we collected *post mortem* samples in the *substantia nigra* midbrain region from public PD case–control microarray data sets, as well as *post mortem* microarray samples from the Human Brain Transcriptome (HBT) project for 3 adult age periods (20 to 40 years, 40 to 60 years, and 60 years onwards; used to identify aging-associated changes in brain gene expression during adulthood, see below).

For the meta-analysis of *substantia nigra* brain samples from PD case–control studies, raw microarray data were obtained from 8 published data sets (Zhang et al., 2005; Lesnick et al., 2007; Moran et al., 2006; Simunovic et al., 2009; Zheng et al., 2010). Importantly, the samples were already age- and gender-matched to prevent biases in downstream analyses. All microarray data sets were pre-processed using the GC-RMA procedure for background correction, normalization and probe replicate summarization (Wu et al., 2004), and only samples from the *substantia nigra* brain region were retained for further analysis. Since platform-specific biases can lead to artifacts when directly integrating microarray data from different studies via cross-study normalization methods, we instead used a meta-analysis to integrate statistical results obtained on the individual data sets. First, differential expression statistics were computed on each data set separately using the empirical Bayes moderated *t*-statistic (Smyth, 2004), and then the *p*-value significance scores were combined via the weighted meta-analysis approach by Marot et al. (2009). In contrast to the commonly used unweighted Fisher method for *p*-value combination, this approach involves data set-specific weights reflecting the relative number of samples collected in each study. The obtained meta-analysis *p*-values were adjusted for multiple hypothesis testing using the approach by Benjamini and Hochberg (1995) and a false discovery rate (FDR) threshold of 0.05 to determine the final gene selection. Since two of the microarray data sets were derived using laser-capture microdissection (LCM), we confirmed the consistency between LCM and non-LCM data by determining the Spearman correlation between the median fold changes across the LCM- and the non-LCM data sets for the genes considered in this study (Spearman's rho = 0.634) and the significance of the linear regression fit between these two data series ($p = 2.19E{-}09$).

For the aging-related microarray data from the Human Brain Transcriptome project (Kang et al., 2011) (HBT), all samples covering the 3 main adult age periods 20 to 40 years, 40 to 60 years and 60 years onwards were collected in order to identify differentially expressed genes across these age groups. Importantly, as mentioned in the original publication for the HBT project, none of the individuals included in this study suffered from any known neurological or psychiatric disorder, severe head injuries or signs of neurodegeneration (Kang et al., 2011). The significance of differential expression across the age groups was computed using a dedicated multiclass-analysis method designed for microarray data (Tusher et al., 2001). We chose this specific approach in order to identify increases or decreases in gene expression variance related to aging in addition to positive or negative correlations with aging (the correlation with aging is additionally reported in Tables 1 and 2, and for the full-length gene ranking table in the Supplementary Material). Finally, the heat map visualizations in Figs. 1, 2 and 3 were generated using a Pearson correlation hierarchical clustering (i.e., the distance metric is 1-correlation; larger versions of these heat maps including the gene names, as well as heat maps for a Euclidean distance metric and additional sample clustering are provided in the Supplementary Material, see Fig. S1–S8).

Network-based enrichment analysis of PD and aging transcriptomics data

To analyze associations between the deregulated genes in PD/aging and cellular pathways and exploit additional information from public molecular interaction data, we used our algorithm *EnrichNet* for network-based gene/protein set enrichment analysis (see Glaab et al., 2012 for a detailed description and the publicly available web-application www.enrichnet.org). Briefly, EnrichNet consists of a 3-step procedure: First, a gene or protein set of interest (the target gene set) as well as gene/protein sets representing cellular pathways from public databases (the reference gene sets) are mapped onto a genome-scale protein–protein interaction network. Next, a deterministic procedure for simulating random walks in a network (the Random Walk with Restart algorithm Tong et al., 2008) is applied to score the network distances and multiplicity of interactions between the target and reference gene/protein sets. In order to obtain final association scores for the pathway reference sets, the combined interconnectivity/distance scores are compared to a background score distribution using the XD-statistic (Olmea et al., 1999; Glaab et al., 2012) (larger XD-scores reflect stronger associations, and the algorithm determines an XD-score significance threshold corresponding to a false-discovery rate of 0.05).

Here, in order to identify and score network associations of the deregulated genes in aging and PD with known cellular pathways, we applied EnrichNet on a target gene set given by the intersection of the significant genes from the differential expression analyses of the aging and PD transcriptome data (FDR < 0.05, see above). The pathway-representing reference gene sets were obtained from the public databases Gene Ontology (Ashburner et al., 2000), KEGG (Kanehisa and Goto, 2000), WikiPathways (Pico et al., 2008) and Reactome (Joshi-Tope et al., 2005). To assemble the genome-scale protein–protein interaction network only experimentally verified, direct physical interactions from public data repositories including tissue-specificity annotations (Bossi and Lehner, nd.) were used. In addition to the network association scores obtained from the graph-based statistic, we also performed a conventional over-representation analysis, scoring the significance of the overlap between target and