The commonality of protein interaction networks determined in neurodegenerative disorders (NDDs)

Vachiranee Limviphuvadh\textsuperscript{1}, Seigo Tanaka\textsuperscript{2}, Susumu Goto\textsuperscript{1}, Kunihiro Ueda\textsuperscript{3} and Minoru Kanehisa\textsuperscript{1,}\* \\
\textsuperscript{1}Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, \textsuperscript{2}Laboratory of Molecular Clinical Chemistry, Faculty of Pharmacy, Osaka Ohtani University, Tondabayashi, Osaka, 584-8540 and \textsuperscript{3}Kobe Tokiwa Junior College, Nagata-ku, Kobe, Hyogo, 653-0838, Japan

Received on February 15, 2007; revised on May 21, 2007; accepted on June 1, 2007

Advance Access publication June 6, 2007

Associate Editor: Limsoon Wong

ABSTRACT

\textbf{Motivation:} Neurodegenerative disorders (NDDs) are progressive and fatal disorders, which are commonly characterized by the intracellular or extracellular presence of abnormal protein aggregates. The identification and verification of proteins interacting with causative gene products are effective ways to understand their physiological and pathological functions. The objective of this research is to better understand common molecular pathogenic mechanisms in NDDs by employing protein–protein interaction networks, the domain characteristics commonly identified in NDDs and correlation among NDDs based on domain information.

\textbf{Results:} By reviewing published literatures in PubMed, we created pathway maps in Kyoto Encyclopedia of Genes and Genomes (KEGG) for the protein–protein interactions in six NDDs: Alzheimer’s disease (AD), Parkinson’s disease (PD), amyotrophic lateral sclerosis (ALS), Huntington’s disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA) and prion disease (PRION). We also collected data on 201 interacting proteins and 13 compounds with 282 interactions from the literature. We found 19 proteins common to these six NDDs. These common proteins were mainly involved in the apoptosis and MAPK signaling pathways. We expanded the interaction network by adding protein interaction data from the Human Protein Reference Database and gene expression data from the Human Gene Expression Index Database. We then carried out domain analysis on the extended network and found the characteristic domains, such as 14-3-3 protein, phosphotyrosine interaction domain and caspase domain, for the common proteins. Moreover, we found a relatively high correlation between AD, PD, HD and PRION, but not ALS or DRPLA, in terms of the protein domain distributions.

\textbf{Availability:} http://www.genome.jp/kegg/pathway/hsa/hsa01510.html (KEGG pathway maps for NDDs)

\textbf{Contact:} kanehisa@kuicr.kyoto-u.ac.jp

*To whom correspondence should be addressed.

1 INTRODUCTION

Neurodegenerative disorders (NDDs) are progressive and fatal disorders, and their causes and pathogenic mechanisms remain to be clarified. No efficient medical treatments are currently available for NDDs. NDDs consist of several diseases that present a distinct neuropathology in particular brain regions. A common molecular feature of NDDs is intracellular or extracellular occurrence of protein aggregates in fibrillar structures, known as ‘amyloid’ (Bossy-Wetzel et al., 2004; Ross and Poirier, 2004), suggesting some common molecular mechanisms in NDDs (Bacciantini et al., 2002; Kayed et al., 2003). Also, identifying similarities in the pathogenesis of NDDs enables certain genes to be targeted for intervention of multiple diseases (Mathisen, 2003). Although the research area of each NDD is very active, only a few studies have focused on their protein–protein interaction networks (Giorgini and Muchowski, 2005; Goehler et al., 2004) and no comparative analysis combining two or more diseases has been performed. Analyzing the global picture of NDDs by combining the protein–protein interaction networks of each NDD should add new insights into the common pathogenesis in NDDs.

In this article, we focused on protein–protein interaction networks associated with causative proteins of six well-known NDDs: Alzheimer’s disease (AD), Parkinson’s disease (PD), amyotrophic lateral sclerosis (ALS), Huntington’s disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA) and prion disease (PRION). The objective is to better understand the molecular pathogenesis of these NDDs by revealing any common molecular mechanisms found in the interaction network. First, we collected protein–protein interactions associated with protein aggregation in these six NDDs. These interactions are represented as KEGG pathway maps as a consensus based on data from review articles (see reference list in KEGG pathway pages) (Kanehisa et al., 2006). Therefore, information within this data set directly represents relation to the known mechanisms of NDDs. Second, we built a data set of protein–protein interactions related to 13 causative gene products of the six NDDs, which were manually extracted from the published literature after comprehensive keyword searching. This second data set contains protein–protein interactions that remain to be
fully understood or confirmed to be related to pathogenesis of NDDs. Third, we combined the protein–protein interaction data from the Human Protein Reference Database (HPRD) (Peri et al., 2003) as well as gene expression data from the Human Gene Expression Index Database (Hugnet) (Haverty et al., 2002; Hsiao et al., 2001) to find further possible protein interactions linked to causative genes of the six NDDs. We also used protein domains to investigate the common mechanisms from the molecular point of view. These analyses revealed the common proteins and domain characteristics of the protein–protein interaction networks of NDDs.

2 METHODS

2.1 Data set for KEGG PATHWAY

First, we chose six well-known NDDs: AD, PD, ALS, HD, DRPLA and PRION. To construct the data set directly associated with the six NDDs, we examined 80 review articles to extract protein interactions concerning pathological mechanisms, especially protein aggregation and/or neuronal cell loss. From the extracted interactions, we determined known causative genes of the six NDDs. Next, we created a wiring diagram (graphical diagram) consisting of nodes to indicate proteins and chemical compounds, and edges to represent many types of interactions. Moreover, we defined a KEGG Orthology (KO) identifier for each gene in our data set. KO is a common identifier for linking the genomic information in the KEGG GENES database and the network information in the KEGG PATHWAY database. Organism-specific pathways can be computationally generated by matching KO's in the genome with those in the pathways. Here, we defined ortholog genes for Homo sapiens (human), Mus musculus (mouse) and Rattus norvegicus (rat).

2.2 Network analysis

We expanded the protein networks of KEGG reference maps for NDDs by using two sets of protein–protein interactions. One was those manually extracted from literature in PubMed (see the next section for details). The other was protein–protein interaction data from HPRD. We combined protein–protein interaction data from literature in PubMed and HPRD to expand the protein–protein interaction network. Then, we used Hugnet to limit protein interactions in HPRD to only those containing proteins expressed in the brain. Finally, the protein–protein interactions linked from the causative genes of six NDDs up to two steps away were extracted.

2.3 Literature search for protein–protein interaction

To construct an additional data set of protein–protein interactions, we searched all articles in PubMed using the causative gene names and keywords ‘bind* OR interact*’ as queries. Among the approximately 4000 articles returned, we manually selected 233 articles describing protein–protein interactions and/or other types of interactions, such as cleavage, activation and phosphorylation, in NDDs. We extracted the following information: gene names or gene product names involved in an interaction, the type of interaction (direct binding, indirect interaction, activation, attenuation, cleavage, co-regulation, inactivation, induction, inhibition, phosphorylation, production, promote expression, repress expression and ubiquitination), the experimental method used to identify the interaction (yeast two hybrid assay, co-immunoprecipitation, pull-down assay, gel-shift assay and affinity column chromatography), any sequence domain or motif involved in the interaction and the reference information (PubMed ID). We included only those interactions that were confirmed in neuronal cells or with proteins derived from mammalian brain. Moreover, we collected protein–compound interactions and protein–protein interactions involving indirect interactions with a causative protein via an intermediate. We also collected information associated with each interaction.

2.4 Pfam domain analysis

We extracted Pfam domains from the protein–protein interaction data and analyzed the function of the domains related to NDDs by the following methods.

2.4.1 Extraction of the domains significantly appearing in the common protein interaction network

(1) Domains from Pfam release 19.0 (Finn et al., 2006) were assigned to each of the 146 proteins contained in the common protein interaction network (see Results section below) with HMMER version 2.3.2 (Eddy, 1998), using E-value cutoff 0.1, 0.01 and 0.001. We also counted the frequency of each domain contained in the 146 proteins.

(2) From each cutoff, we extracted domains associated with at least three proteins and calculated the ratio (%) (which we call the ‘rate of content A’) of the number of proteins with the domain to the total number of proteins in the common protein interaction network (146 proteins).

(3) In order to find domains that appear significantly in the common protein interaction network, we used proteins from HPRD (4734 proteins) as control. After assigning domains to the proteins and calculating the percentage (that we call here ‘rate of content B’) as in step (1) and (2), respectively, the value of R = (‘rate of content A’/‘rate of content B’) was calculated. We use the value of R as an indicator of the relative frequency of the characteristic domains in the 146 proteins relative to the entire HPRD.

2.4.2 Method for investigating the correlation of the six NDDs based on the type of domains included in the network of each NDD

(4) The protein–protein interactions linked from the causative gene products of each disease up to two steps away were extracted. The number of proteins in AD, PD, ALS, HD, DRPLA and PRION are as follows: 237, 137, 46, 145, 55 and 104. Then, Pfam domains were assigned to the proteins with HMMER, using E-value cutoff 0.1, 0.01 and 0.001 like (1). The number of different domains contained in each disease following E-value cutoff 0.1, 0.01 and 0.001 were; AD (312, 277 and 266), PD (186, 164 and 157), ALS (61, 55 and 54), HD (222, 199 and 194), DRPLA (104, 93 and 89) and PRION (164, 142 and 137).

(5) For each disease, we constructed a profile vector representing domain distribution. In this study, each element in the profile vector was defined as the ratio (%) of the number of proteins containing the domain against the number of all proteins related to the disease. For each Pfam cutoff, we computed a correlation coefficient between diseases and y based on the following formula using the R package:

\[
\text{cor}(x, y) = \frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}}
\]

Note that the number of unique domains depends on the cutoff in the Pfam database. In this case, the numbers of unique domains are 420, 386 and 375, for the E-value cutoff 0.1, 0.01 and 0.001, respectively.

(6) We used a t-test to find the P-value for the correlation coefficient in (5).
3 RESULTS

3.1 The pathway maps of NDDs

We created KEGG pathway maps for the six NDDs and stored them in the ‘Neurodegenerative Disorders’ subcategory of the ‘Human Diseases’ category in KEGG PATHWAY. In the pathway map, protein names in red designate causative genes or risk factors (see Fig. 1 for the Parkinson’s disease pathway map).

We used 13 causative genes identified in the six NDDs: amyloid precursor protein (APP), presenilin 1 (PSEN1), presenilin 2 (PSEN2), apolipoprotein E4 (APOE), parkin (PARK2), α-synuclein (SNCA), ubiquitin carboxyl-terminal esterase L1 (UCHL1), DJ-1 protein (PARK7), superoxide dismutase 1 (SOD1), alsin (ALS2), huntingtin (HD), atrophin 1 (DRPLA) and prion protein (PRNP). The list of causative genes or risk factors is updated in the neurodegenerative disorders pathway when we have new information from reviews in PubMed. We used specific symbols to represent various kinds of interactions, such as ‘+u’ for ubiquitination, ‘+g’ for glycosylation and ‘+p’ for phosphorylation. Each protein on the map is linked to its KEGG GENES entry. Pathway maps for mouse and rat can also be seen by changing the organism selection in the popup menu above the map.

3.2 Common proteins linking disease and normal pathways

Table 1 lists the number of protein–protein interactions identified from the literature. We also collected 13 other compounds and 16 further interactions involving them. By superimposing these interaction data onto the six pathway maps, we identified 19 common proteins, listed in Table 2. In this article, we define the terms ‘common protein’ as a protein which is contained in the intersection of at least two different NDD networks. These proteins link the six NDDs as illustrated on the super-map in Figure 2, in which each NDD is designated in a different color. We also defined ‘common interaction’ in the same way. For example, the interaction between PNUTL1 and STX1A is contained in the intersection of both AD and PD protein interaction networks, therefore, it is the ‘common interaction’ of AD and PD, although the common interactions were found only in the extended network (see Section 3.3).

Table 1. NDD related proteins and their interactions

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of proteins</th>
<th>Number of interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>96</td>
<td>144</td>
</tr>
<tr>
<td>PD</td>
<td>34</td>
<td>36</td>
</tr>
<tr>
<td>ALS</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>HD</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>DRPLA</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>PRION</td>
<td>21</td>
<td>29</td>
</tr>
<tr>
<td>Total*</td>
<td>201</td>
<td>282</td>
</tr>
</tbody>
</table>

*Total number of unique proteins and unique interactions.

3.3 Commonality of protein networks in NDDs

In this article, we define the terms ‘common protein’ as a protein which is contained in the intersection of at least two different NDD networks. These proteins link the six NDDs as illustrated on the super-map in Figure 2, in which each NDD is designated in a different color. We also defined ‘common interaction’ in the same way. For example, the interaction between PNUTL1 and STX1A is contained in the intersection of both AD and PD protein interaction networks, therefore, it is the ‘common interaction’ of AD and PD, although the common interactions were found only in the extended network (see Section 3.3).
According to the KEGG PATHWAY database, these common proteins are most often involved in apoptosis (8 proteins) and the MAPK signaling pathway (4 proteins), as well as focal adhesion (3 proteins) and the Jak-STAT signaling pathway (3 proteins). These and other pathways containing the common proteins were linked from this super-map. CASP8, an initiator caspase, was the most common protein in the protein–protein interaction data set, and related to

<table>
<thead>
<tr>
<th>Common proteins</th>
<th>Gene names</th>
<th>Related KEGG pathways</th>
<th>Common diseases</th>
<th>Interacting proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>APBA1</td>
<td>Amyloid beta (A4) precursor protein-binding, family A member 1</td>
<td>Apoptosis, Focal adhesion, Insulin signaling pathway, VEGF signaling pathway</td>
<td>AD, ALS</td>
<td>APP, CCS, PSEN1</td>
</tr>
<tr>
<td>APLP1</td>
<td>Amyloid beta (A4) precursor-like protein 1</td>
<td>–</td>
<td>AD, PRION</td>
<td>APBB1, APBB3, DAB1, HMOX1, HMOX2, PRNP</td>
</tr>
<tr>
<td>BAD</td>
<td>BCL2-antagonist of cell death</td>
<td>Apoptosis, Focal adhesion, Insulin signaling pathway, VEGF signaling pathway</td>
<td>PD, ALS</td>
<td>SNCA, SOD1</td>
</tr>
<tr>
<td>BAX</td>
<td>BCL2-associated X protein</td>
<td>Apoptosis</td>
<td>ALS, PRION</td>
<td>SOD1, PRNP</td>
</tr>
<tr>
<td>BCL2</td>
<td>B-cell CLL/lymphoma 2</td>
<td>Apoptosis, Focal adhesion</td>
<td>AD, ALS, PRION</td>
<td>PRNP, PSEN1, SOD1</td>
</tr>
<tr>
<td>BCL2L1</td>
<td>BCL2-like 1</td>
<td>Apoptosis, Jak-STAT signaling pathway</td>
<td>AD, ALS</td>
<td>PSEN1, PSEN2, SOD1</td>
</tr>
<tr>
<td>CASP1</td>
<td>Caspase 1</td>
<td>Apoptosis, MAPK signaling pathway, Natural killer cell mediated cytotoxicity</td>
<td>AD, DRPLA</td>
<td>DRPLA, PSEN1, PSEN2</td>
</tr>
<tr>
<td>CASP3</td>
<td>Caspase 3</td>
<td>Apoptosis, MAPK signaling pathway, Natural killer cell mediated cytotoxicity</td>
<td>AD, HD, DRPLA</td>
<td>CSEN, DRPLA, HD, PSEN1, PSEN2</td>
</tr>
<tr>
<td>CASP6</td>
<td>Caspase 6</td>
<td>Apoptosis</td>
<td>AD, HD</td>
<td>HD, PSEN1, PSEN2</td>
</tr>
<tr>
<td>CASP7</td>
<td>Caspase 7</td>
<td>Apoptosis</td>
<td>AD, DRPLA</td>
<td>DRPLA, PSEN1, PSEN2</td>
</tr>
<tr>
<td>CASP8</td>
<td>Caspase 8</td>
<td>Apoptosis, Toll-like receptor signaling pathway</td>
<td>AD, HD, DRPLA, PRION</td>
<td>DRPLA, ESRRBL1, HIP1, NGFR, PSEN1, PSEN2</td>
</tr>
<tr>
<td>CREBBP</td>
<td>CREB binding protein</td>
<td>Adherens junction, Cell cycle, Jak-STAT signaling pathway, Long-term potentiation, Notch signaling pathway, TGF-beta signaling pathway, Wnt signaling pathway</td>
<td>HD, DRPLA</td>
<td>DRPLA, HD</td>
</tr>
<tr>
<td>FBXW7</td>
<td>F-box and WD-40 domain protein 7</td>
<td>Ubiquitin mediated proteolysis</td>
<td>AD, PD</td>
<td>PARK2, PSEN1</td>
</tr>
<tr>
<td>GAPD</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Glycolysis/Gluconeogenesis</td>
<td>AD, HD, DRPLA</td>
<td>APP, DRPLA, HD</td>
</tr>
<tr>
<td>GFAP</td>
<td>Glial fibrillary acidic protein</td>
<td>Fc epsilon RI signaling pathway, Focal adhesion, Gap junction, GnRH signaling pathway, Insulin signaling pathway, Jak-STAT signaling pathway, MAPK signaling pathway, Natural killer cell mediated cytotoxicity, T cell receptor signaling</td>
<td>AD, PRION</td>
<td>PRNP, PSEN1</td>
</tr>
<tr>
<td>GRB2</td>
<td>Growth factor receptor-bound protein 2</td>
<td>–</td>
<td>AD, HD, PRION</td>
<td>APP, HD, PRNP</td>
</tr>
<tr>
<td>HSPA5</td>
<td>Heat shock 70kDa protein 5</td>
<td>Antigen processing and presentation, MAPK signaling pathway</td>
<td>AD, PRION</td>
<td>APP, PRNP</td>
</tr>
<tr>
<td>MAPT</td>
<td>Microtubule-associated protein tau</td>
<td>MAPK signaling pathway</td>
<td>AD, PD</td>
<td>APOE, APP, PRKACA, PSEN1, SNCA</td>
</tr>
<tr>
<td>NGFR</td>
<td>Nerve growth factor receptor</td>
<td>Cytokine–cytokine receptor interaction</td>
<td>AD, PRION</td>
<td>APP, CASP8, PRNP</td>
</tr>
</tbody>
</table>
Four diseases: AD, HD, DRPLA and PRION. Four proteins, BCL2, CASP3, GAPD and GRB2, were found to be common in three diseases. We could not find any common protein that was related to all the diseases.

We found that 8 out of the 19 common proteins were related to apoptosis, including members of the CASP family (CASP3, CASP6, CASP7 and CASP8) and the Bcl-2 family (BAD, BAX, BCL2 and BCL2L1). There is a consensus that the pathophysiology of a variety of NDDs involves excessive apoptosis in distinct brain areas. Bcl-2 family proteins and CASP7 are involved in ER-stress-induced cell death, which is known to protect cells against the toxic buildup of misfolded proteins (Rao and Bredesen, 2004). Misfolded proteins, and the associated ER stress, are emerging as a clue to understanding common mechanisms of NDDs. We also found four proteins (GRB2, HSPA5, MAPT and CASP3) related to MAPK signaling. GRB2 is known to be involved in cell cycle, oncogenic proliferation, neuronal development, cell differentiation and apoptosis, while we found that it is related to AD, HD and PRION. Recent studies of NDDs have discovered that bona fide cell cycle regulators, i.e. GRB2 and MAPT, are expressed in postmitotic neurons of affected or susceptible brain regions and suggested that aberrant activation of the cell cycle in certain NDDs leads to neuronal cell death (Vincent et al., 2003). Our approach thus clarifies relationships between disease and normal pathways. Understanding these normal pathways may be a key issue to elucidate the biological functions of common mechanisms of NDDs.

### 3.3 Common proteins and interactions in the extended network

The analysis thus far was limited to protein–protein interactions already reported to be linked to the causative gene products, we further attempted to search for unidentified proteins by using comprehensive experimental data sets. We used human protein–protein interaction data from HPRD version 1.0. By parsing 10 285 XML files of HPRD, we obtained 4734 proteins (14 023 interactions) which have protein–protein interaction information and corresponding Entrez Gene IDs. In addition, we downloaded brain-selective genes, i.e. those that are highly expressed only in the brain (567 genes), and housekeeping genes that are expressed in all tissue types (391 genes) from HugeIndex (http://zlab.bu.edu/HugeIndex/HugeIndexCompendiumSupplement.shtml), which is a repository for gene expression data on normal human tissues using high-density oligonucleotide arrays. Finally, we obtained unique 894 genes that were highly expressed in brain or in all tissue types from HugeIndex.

We combined the 266 interactions in our protein–protein interaction data set with 14 023 interactions in HPRD. Then, we extracted interactions that may be involved with brain function by using both the 894 gene products of HugeIndex and 201 proteins in our protein–protein interaction data set. As a result, we obtained the extended network containing 528 proteins and 1041 interactions.
Table 3. Summary of data set within two links from causative genes of extended protein–protein interaction network and the number of common proteins and common interactions when superimposed on the six NDDs’ networks

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of proteins</th>
<th>Number of interactions</th>
<th>Number of common proteins and common interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>385</td>
<td>703</td>
<td>174 common proteins, 202 common interactions</td>
</tr>
</tbody>
</table>

When our network is examined with the 13 causative gene products, a total of 196 proteins and 247 interactions were directly linked, and 385 proteins and 703 interactions were found within two links (Table 3). Twenty-eight proteins, including nine causative genes, are hubs in a global network consisting of these 703 interactions (Fig. 3). It is not surprising that the top five hubs are causative gene products, i.e. APP, PSEN1, HD, SNCA and PSEN2, which have 47, 46, 32, 28 and 27 interaction partners, respectively. Other hubs, such as CTNNB1, DLG4 and BCL2, are not causative gene products but are related to signal transduction and apoptosis.

We then examined this global network distinguishing the six NDDs, starting with their causative gene products. After superimposing the six networks, we found 174 proteins and 202 common interactions (146 proteins) that were found in at least two NDDs (Table 3 and orange lines in Fig. 3). Of the 174 common proteins, 19 were listed in Table 2. The other 155 proteins were identified from the extended protein–protein interaction network, 74 of which were already included in our protein–protein interaction data set. However, from the extended protein–protein interaction network, we could find more interactions, and hence more common proteins. We could not find any interactions common to all six NDDs, but we found that the interaction between BCL2 and CASP3, which is well known in apoptosis, was common to five NDDs (except PD). We also found that five interactions (BAD-BCL2, BCL2-CASP8, BCL2L1-CASP8, CASP3-HSPD1 and CASP3-NFE2L2) were common to four NDDs.

PSEN1, one of the causative genes in AD, is a common protein that was linked to all six NDDs. Another 81 proteins were found as new proteins in the common network. Four (BCAP31, FYN, JUND and TCF4) of these 81 proteins were directly linked to causative genes. We could only find five proteins (BCAP31, DNCL1, HSPCA, NR3C1 and PIN1) among these 81 which link to five NDDs and none that link to all six. According to the KEGG pathways, focal adhesion, regulation of actin cytoskeleton, adherens junction, axon guidance, tight junction and leukocyte transendothelial migration are the functions in which most of these 81 proteins were involved.

The KEGG pathways found to include more than five common interactions are apoptosis, Wnt signaling pathway and focal adhesion. These pathways are known to exist in the brain and thought to have functions related to neuronal survival, cell death and oxidative stress. There is emerging evidence that focal adhesion and Wnt signaling are associated with some NDDs (Caltagarone et al., 2007; Caricasole et al., 2005). Alteration or impairment of such pathways and their downstream pathways may contribute to the common pathogenic mechanisms in NDDs.

3.4 Domain analysis of the extended network

Further, we investigated the domains of the 146 proteins making up the common interaction network, extracting the domains in each protein from Pfam Release 19.0. We calculated R (the relative frequency of the characteristic domains in the 146 proteins relative to the entire HPRD data set) (Table 4). We also extracted the domains contained in the common protein interaction network from Pfam. As shown in Table 4, domains characteristic to the common protein network compared to the HPRD data set are 14-3-3 protein (pfam:14-3-3), phosphotyrosine interaction domain (PTB/PID) (pfam:PID), caspase domain (pfam:Peptidase C14), apoptosis regulator proteins Bcl-2 family (pfam:Bcl-2) and WW domain (pfam:WW).

The Bcl-2 family, caspase domain and PTB/PID domains were also found in the 19 common proteins from our protein–protein interaction data set (Table 2). This might suggest that other proteins containing these domains are related to the common mechanisms of NDDs. The 14-3-3 protein is a key regulator of cell division, signal transduction and stress response. Brain tissue contains the highest concentration of 14-3-3 proteins, representing ~1% of total soluble brain protein. Evidence is accumulating that 14-3-3 might contribute to AD, PD and CJD (Dougherty and Morrison, 2004). The WW domain is involved in the cytoskeleton, embryonic development and differentiation of the central nervous system. The protein kinase C terminal domain functions in the phosphorylation of proteins in order to regulate their activity. The 14-3-3 protein domain, PTB/PID domain and WW domain were found only in eukaryotes, which might explain why domains evolved in eukaryotes are more involved in NDDs.

In order to evaluate the correlation between each pair of these six NDDs, we also calculated the correlation coefficient of the domain distributions between the two diseases using three
E-value cutoffs 0.1, 0.01 and 0.001. The total number of domains at E-value cutoff 0.1, 0.01 and 0.001 were 429, 386 and 375, respectively. We found that PD and HD showed the highest correlation at all cutoffs (E-value ≤ 0.1, \( r = 0.85, P = 6.21 \times 10^{-121} \); E-value ≤ 0.01, \( r = 0.87, P = 5.5 \times 10^{-120} \); E-value ≤ 0.001, \( r = 0.87, P = 1.33 \times 10^{-116} \)) and AD and PD showed the next highest correlation for E-value ≤ 0.1 and 0.01. For E-value ≤ 0.001, the next highest correlation were observed between AD and PD, and also between AD and HD. The results for E-value cutoff 0.001 are shown in Table 5. All correlation coefficients were statistically significant (\( P < 0.01 \)).

AD, PD, HD and PRION showed relatively high correlation each other, but ALS and DRPLA seem to differ from the other NDDs. Currently, the protein–protein interactions of ALS and DRPLA are not well studied compared to other four NDDs, and it is natural that the number of identified domains of ALS and DRPLA is relatively lower than the others. This may explain why the correlation coefficients between these two NDDs and other four NDDs are lower.

4 DISCUSSION

4.1 Common proteins in NDDs

We found 19 common proteins in the protein–protein interaction data from literature and 81 new common proteins from the common protein interaction network in our extended network. It is reasonable that these 19 common proteins obtained from literature are well-known proteins related to NDDs. For example, HSPA5 is one of the chaperones associated with the folding and assembly of proteins in the ER. In the study of Kudo et al. (2002), HSPA5 was shown to be down regulated in

![Fig. 3. A protein interaction network consisting of two steps from 13 causative genes in six NDDs. A total of 385 proteins and 703 interactions were found in this network. The thickness of an edge represents the frequency of the interactions that were confirmed in the literature, and a larger node indicates that it is involved in 10 or more interactions (i.e. nodes that are directly connected to a large number of other nodes, called ‘hubs’). Twenty-eight proteins including 9 causative genes are hubs. A red node refers to a protein with the characteristic domain in the common protein interaction network compared to the HPRD data set. An orange edge refers to a common interaction. The network was created by Pajek software version 1.00.](https://academic.oup.com/bioinformatics/article-abstract/23/16/2129/198594)
brains of AD patients and likely to be associated with the pathology of AD. Although glycolytic function is the only well known functional category for GAPD designated in both KEGG and HPRD, accumulated data suggest that it is a multifunctional protein, including a role in apoptosis. Other new activities of GAPD include regulation of the cytoskeleton, membrane fusion and transport, glutamate accumulation into presynaptic vesicles and binding to low-molecular-weight G proteins (Chuang et al., 2005; Kim and Dang, 2005; Mazzola and Sirover, 2002). In addition, several parallel investigations have indicated a relationship between GAPD and NDDs. As NDDs are characterized by diverse perturbations of cell function, common proteins and their functions may be key factors in any common mechanisms.

From the 81 new common proteins, we found BCAP31 (B-cell receptor-associated protein 31), DNCL1 (dynein, cytoplasmic, light polypeptide 1), HSPCA (heat shock 90 kDa protein 1, alpha), NR3C1 [nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)] and PIN1 [protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting 1] are the most common in the six NDDs. BCAP31 is one of several candidate proteins that might contribute to the regulation of neuronal apoptosis. It is also a member of a novel class of sorting proteins regulating cellular anterograde transport (Zen et al., 2004). PIN1 is already known to be associated with AD and to function in neuronal dedifferentiation and apoptosis (Butterfield et al., 2006). DNCL1, a cytoplasmic dynein, physically interacts with neuronal nitric oxide synthase and inhibits its activity. In axons, retrograde transport is mediated mainly by cytoplasmic dynein, and dysfunction of motors results in NDDs (Hirokawa and Takemura, 2004). HSPCA is a molecular chaperone that plays an important role in conformational protein regulation and cell signaling (Barral et al., 2004). NR3C1 is an intracellular receptor for glucocorticoids. Marchetti et al. suggested that crosstalk between glucocorticoid and nitric oxide is a pivotal factor to organize neuroprotection in PD (Marchetti et al., 2005). It is already known that inflammation and oxidative stress have been closely associated with the pathogenesis of NDDs.

### 4.2 Common domains in NDDs

Among domains that appeared more than six times in the common network, the 14-3-3 family is involved in protein chaperones that bind to specific phosphorylated sites on diverse target proteins, thereby triggering events that promote cell survival (Mackintosh, 2004). Although most of their binding partners remain to be identified in the brain, they are suggested to be involved in brain development, memory and learning. It is likely that 14-3-3s are implicated in NDDs. Interestingly, α-synuclein, a causative gene product of PD, shares physiological and functional homology with 14-3-3 proteins, and binds to 14-3-3 proteins as well as some proteins associated with 14-3-3, including protein kinase C, BAD and ERK, to activate or inhibit their function (Ostrerova et al., 1999).

WW domains are small protein modules composed of approximately 40 amino acids (Macias et al., 2002). They are found in many different signaling and structural proteins, and may have pathogenic roles of NDDs. Both huntingtin and atrophin-1 bind specific subsets of WW domain-containing proteins (Wood et al., 1998). Some members of WW domain-containing proteins are closely related to a family of Nedd-4 like ubiquitin ligases, suggesting that huntingtin and atrophin-1 are degraded in a controlled fashion via the ubiquitin system, or that both could be acting as scaffold proteins for the assembly of a functional ubiquitylation complex.

Activated growth factor receptors bind several signaling proteins in a phosphotyrosine-dependent fashion. The phosphotyrosine interaction (PI) domain, also known as the phosphotyrosine binding (PTB) domain, binds peptides with phosphorylated tyrosine residues. In this context, the PI domains of XI1 and FE65, two neuronal proteins, bind to the cytoplasmic domain of the APP. These interactions are suggested to be involved in the processing of APP to produce amyloid protein in AD brain (Borg et al., 1996).

### 4.3 Common features of NDDs

Increasing evidence has been reported that reveals the considerable overlap of the clinicopathological features amongst NDDs. For example, Armstrong et al. (2005) discussed about the factors that contribute to disease overlap, including historical factors, disease heterogeneity, Apo ε genotype and the coexistence of more than a single disorder.
in the same patient. We investigated the overlap among six NDDs using domain distribution (Table 5). We found that PD and HD showed the highest correlation at all E-value. The common domains between PD and HD were the Cation transport ATPase C-terminus domain, Cation transporter/ATPase N-terminus domain, E1-E2 ATPase domain, Helix-loop-helix DNA-binding domain and haloacid dehalogenase-like hydrolase. KEGG pathways related to proteins with these domains are tight junction, long-term potentiation, calcium signaling pathway, phosphatidyl inositol signaling system, olfactory transduction and insulin signaling pathway. We collected clinical symptoms for each NDD and found that motor symptoms, i.e. tremor, rigidity and bradykinesia are common between PD and HD. Also, subcortical dementia and depression were common symptoms. After we examined literature associated with PD and HD, we found that the tight junction pathway has not previously been associated with the mechanism of either disease.

AD and PD also showed a high correlation coefficient. Seventeen common domains (pfam:AAA, pfam:Band_41, pfam:C2-set_2, pfam:CTNNB1_binding, pfam:ERM, pfam:F-box, pfam:GTP_CDC, pfam:Integrin_B_tail, pfam:Integrin_beta, pfam:LRRNT, pfam:RGS, pfam:SNARE, pfam:Sec1, pfam:Syndecan, pfam:Syntaxin and pfam:Tubulin-binding) were identified, suggesting the presence of similar etiologies between these diseases as reported previously (Armstrong et al., 2005). Signaling pathways such as SNARE interactions in vesicular transport, TGF-beta signaling pathway and ubiquitin mediated proteolysis were suggested to be the common mechanisms between AD and PD.

The high correlation between AD and PRION corresponds to several similarities in the pathology of AD and PRION such as an extracellular accumulation of Aβ in AD and PrPSc in PRION (Barnham et al., 2006). Also, common symptoms, including amnesia, aphasia, agnosia, apraxia, disorientation and acalculia, are found in both diseases. BAR domain, casein kinase II regulatory subunit domain, Tis11B like protein N terminus domain and basic region leucine zipper domain were found in common between AD and PRION. We found six proteins associated with the long-term potentiation, Wnt signaling pathway, adherens junction and GnRH signaling pathway.

Based on these results, we expect that the proteins and interactions found in the present study in silico might be factors, that function in the common pathogenic mechanisms among NDDs.

5 CONCLUSION

We first investigated the commonality among the six NDDs from the molecular point of view, i.e. the protein–protein interaction networks, the domain characteristics and domain distributions. PD and HD showed the highest correlation in terms of domain distributions and we found the commonality in the tight junction pathway, which has not previously been associated with the mechanism of either disease. For further studies to understand the molecular mechanisms and possibly lead to diagnosis and treatment of the NDDs, we plan to incorporate chemical information such as signaling molecules, known environmental factors and pharmacological targets into the analysis of the protein–protein interaction network.

ACKNOWLEDGEMENTS

We especially thank Dr Masanori Takehashi, Dr Masahiro Hattori and Dr Kiyoko F. Aoki-Kinoshi for helpful discussions. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology and the Japan Science and Technology Agency. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

Conflict of Interest: none declared.

REFERENCES


