Published in final edited form as:

Arch Toxicol. 2015 May ; 89(5): 743–755. doi:10.1007/s00204-015-1509-6.

MPTP's Pathway of Toxicity Indicates Central Role of Transcription Factor SP1

Alexandra Maertens¹, Thomas Luechtefeld¹, Andre Kleensang¹, and Thomas Hartung^{1,2,3}

Thomas Hartung: THartung@jhsph.edu

¹Center for Alternatives to Animal Testing, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

²Center for Alternatives to Animal Testing Europe, University of Konstanz, Konstanz, Germany

³Doerenkamp-Zbinden Professor and Chair for Evidence-Based Toxicology, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe St., W7035, Baltimore, MD 21205, USA

Abstract

Deriving a Pathway of Toxicity from transcriptomic data remains a challenging task. We explore the use of weighted gene correlation network analysis (WGCNA) to extract an initial network from a small microarray study of MPTP toxicity in mice. Five modules were statistically significant; each module was analyzed for gene signatures in the Chemical and Genetic Perturbation subset of the Molecular Signatures Database as well as for over-represented transcription factor binding sites and WGCNA clustered probes by function and captured pathways relevant to neurodegenerative disorders. The resulting network was analyzed for transcription factor candidates, which were narrowed down via text-mining for relevance to the disease model, and then combined with the large-scale interaction FANTOM4 database to generate a genetic regulatory network. Modules were enriched for transcription factors relevant to Parkinson's disease. Transcription factors significantly improved the number of genes that could be connected in a given component. For each module, the transcription factor that had, by far, the highest number of interactions was SP1, and it also had substantial experimental evidence of interactions. This analysis both captures much of the known biology of MPTP toxicity and suggests several candidates for further study. Furthermore, the analysis strongly suggests that SP1 plays a central role in coordinating the cellular response to MPTP toxicity.

Keywords

MPTP; Parkinson's disease; SP1; WGCNA; GRN; Pathway of toxicity

Introduction

In order to bring toxicology into the twenty-first century, the field is undergoing a profound paradigm change: away from animal-based, black-box models toward a systems toxicology approach based on higher-throughput testing (Hartung 2009). The necessary mapping of pathways of toxicity as a new emerging concept (Kleensang et al. 2014) often involves using high-dimensional datasets, which are traditionally analyzed by looking for a few differentially expressed genes. Cellular pathways leading to toxicity, however, may involve subtle perturbations in many genes rather than drastic alterations in a few. In addition, microarrays are intrinsically noisy and often show poor reproducibility, which only increases the difficulty of extracting meaningful, system-level insights into biology from the data.

The Pathway of Toxicity concept (Hartung and McBride 2011) is at the core of the NIH Human Toxome project (<http://humantoxome.com>, Bouhifd et al. 2014). In line, we used an approach that derives a *de novo* network from a small dataset, clusters genes into modules by network topology and uses the resulting modules for further analysis with text-mining and other sources of high-throughput data (ChIP experiments and siRNA perturbation studies), ultimately producing a more specific genetic regulatory network (GRN). Using a WGCNA approach offers, in essence, a dimensionality reduction technique that can be used to produce a more detailed genetic regulatory network based on known and predicted transcription factor interactions, bringing us a small step closer to a wiring diagram of the cell.

MPTP (methyl-4-phenyl-1,2,3,6-tetrahydropyridine) toxicity offers an excellent “proof of concept” for the ability to derive a Pathway of Toxicity from high-throughput data, since the broad outlines of the Pathway of Toxicity are understood. MPTP exposure is used widely as an animal model for the relatively data-rich Parkinson’s disease (Schober 2004) since MPTP poisoning, like Parkinson’s, is highly selective for dopaminergic neurons in the *substantia nigra* and the clinical symptoms are highly similar to Parkinson’s (Snyder and D’Amato 1986).

MPTP is not itself toxic, but owing to its high lipophilicity it is able to cross the blood–brain barrier, where it is metabolized in astrocytes by monoamine oxidase B (MOA-B) to MPP⁺. MPP⁺ is then transported selectively by the dopamine transporter into neurons. Once inside the neuron, it is thought to exert its primary action through targeting Complex I in the mitochondria, which results in disruption of the electron transport chain (ETC). While MPTP disruption of the ETC causes a loss of ATP, it is not a critical failure of Complex I and oxidative phosphorylation that causes pathology, as MPTP typically only causes a mild decrease in ATP levels and falls short of levels required to cause significant energy depletion (Perier and Vila 2012), and deficiency in a component of Complex I does not lead to selective dopaminergic neural death (Sterky et al. 2012). Therefore, MPTP neurodegeneration is not necessarily caused by energy depletion. More likely, a shift in energy balance is a contributing factor to Parkinson’s disease (Krug et al. 2014), as we recently showed identifying the pathways of defense of dopaminergic neurons in response to MPP⁺ before cytotoxicity manifests.

Another consequence of the ETC disruption is increased ROS generated by impaired mitochondria. This may in turn cause oxidative damage to Complex I, initiating a spiral of decreased mitochondrial efficiency and increased ROS. ROS can cause peroxidation of the lipids, disrupting the normal binding of cytochrome *c* to the mitochondrial membrane, and facilitates the pro-apoptotic release of cytochrome *c* to the cytosol (Perier and Vila 2012). Mitochondria-derived ROS has also been shown to damage lysosomal membranes in MPTP-intoxicated mice, leading to an impairment of lysosomal function and defective autophagic activity (Dehay et al. 2010), including mitochondrial autophagy (Ivatt and Whitworth 2014). In addition to proteins and lipids, MPTP-intoxicated mice also exhibit oxidative damage to nuclear and mitochondrial DNA (Hoang et al. 2009). Despite the centrality of the intracellular, mitochondrial-generated ROS from neurons, there may be other contributors to ROS in the context of PD/MPTP toxicity—for example, astrocytes or microglia.

Another key component of MPTP toxicity is microtubule disruption. MPP+ is believed to lead to hyperphosphorylation of microtubule-associated protein tau (MAPT), which leads to microtubule instability (Cappelletti et al. 2001). Depolymerization of MTs is one suggested reason for the selective vulnerability of dopaminergic (DA) neurons by toxins such as MPTP, paraquat and rotenone, as dopaminergic neurons require axonal transport of neurotransmitters to the striatum for dopamine release (Ren et al. 2005). The traffic along the axonal length of DA neurons requires intricate coordination between MTs and the motor proteins to ensure that dopamine is transported successfully through vesicle transport. Depolymerization—or, less acutely, an impairment of coordinated traffic—can lead to an impairment of neural function. Furthermore, in neurons, mitochondria are actively transported throughout the cell body; the combination of impaired mitochondrial activity and impaired transport is likely key to the toxic outcome (Sterky et al. 2012).

The final step of MPTP toxicity, apoptosis, is likely the result of several pathways that combine to produce cell death. Apoptosis is thought to be generated through a mitochondrial-initiated, BAX-dependent process. Complex I inhibition does not directly trigger mitochondrial cytochrome *c* release but instead increases the “releasable” pool of cytochrome *c* in the mitochondrial membrane—increasing the magnitude of the signal that can be released when activated by BAX (Perier and Vila 2012).

In summary, while MPTP toxicity has an agreed-upon origin (mitochondrial disruption), there is still much to be learned about the exact Pathway of Toxicity, and the toxicity mechanism likely involves alterations of several pathways along key points (Krug et al. 2014). Here, we demonstrate for the first time, how a Pathway of Toxicity can be deduced with bioinformatics approaches from a rather limited omics dataset.

Materials and methods

Data

Dataset GDS2053, which represented a small study of 12 samples based on the Affymetrix Murine Genome U74A Array from MPTP-treated mice (Miller et al. 2004), was downloaded from GEO with GEOQuery (Davis and Meltzer 2007) and checked for outliers

via the IAC function in WGCNA (Langfelder and Horvath 2008). The top 5000 genes were filtered using the rank means function in WGCNA.

WGCNA uses correlation to determine the strength of the network connection typically, β can be chosen to fit the network to a scale-free topology $A = [a_{ij}] = [|\text{cor}(x_i, x_j)|]^\beta$. Here, β was chosen as 7 based on the lowest value that produced a scale-free topology in the network. A Topological Overlap Metric (TOM) was calculated as described in (Yip and Horvath 2007), and probes were clustered and assigned to modules using the “blockwisemodule” function with a signed Spearman rank correlation with $\beta = 7$, and a deep split level of 2 (which represents a medium level of sensitivity in terms of how modules are detected), a minimum module size of 40, and clustering based on the Dynamic Tree Cut algorithm (Langfelder et al. 2008). Eigengenes were calculated from each module, and p values were calculated based on the functions in the WGCNA package (Langfelder and Horvath 2007). The network was based on the TOM calculated from an unsigned network with the same settings used for module detection, filtered down to only genes in the statistically significant modules and with an edge weight >0.20 .

Enrichment analysis

Probes for each module were entered into DAVID (Dennis et al. 2003) and analyzed for enrichment with a stringency set to high and all other settings at default.

Visualization

All networks were visualized in Cytoscape 3.0 (Shannon et al. 2003) with the yFiles circular layout or spring-embedded bio-layout.

Genetic regulatory network

All probes mapped to each module were entered into Molecular Signatures Database (MSigDB) (Subramanian et al. 2005), and all the top 100 motifs and microRNA binding sites with a false discovery corrected p value <0.01 were retrieved. Statistically significant curated gene sets were retrieved as well.

Text-mining was performed in PubMed with the name of the transcription factors and Parkinson’s disease as a MeSH term and “1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine” OR “1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine” OR “mptp.” In the case of fewer than 5 abstracts returned results were manually inspected for relevance.

Gene symbols were entered into FANTOM4 EdgeEx-pressDB (Severin et al. 2009); cases of ambiguity were resolved manually. ChIP interactions, siRNA experiments or published interactions were considered as experimental evidence. Predicted evidence was either the transcription factor binding predictions or mirRNA predictions.

Results

WGCNA clustered probes by function and captured the relevant pathways

While correlation networks—often referred to as “guilt-by-association” analysis (Quackenbush 2003)—are commonly used to derive networks *de novo* from microarray data, weighted gene correlation network analysis (WGCNA) offers several advantages. Unweighted correlation networks typically establish a hard cut-off for a link, but WGCNA links each gene by a weight, and this network is used to derive a Topological Overlap Metric, which is most simply thought of as a measurement of gene interconnectivity. This combines the advantages of a correlation network with the insights gleaned from a graph-theoretical approach; it is typically more sensitive to “weaker” connections among genes that may be significant, while at the same time it is somewhat more robust to noise (Langfelder and Hovarth 2008). We chose MPTP toxicity, a commonly used toxicity model for Parkinson’s disease, and located a publicly available GEO dataset produced from tissue isolated from the *substantia nigra* of male C57BL/6J mice dosed at 10 weeks of age with a total of three doses of 30 mg/kg MPTP dosed via i.p. or saline control and killed either 24 h or 7 days after the final dose of neurotoxin. Biological replicates were pooled, and twelve arrays in total were used with four arrays per group. The initial dataset was downloaded as normalized data from the GEO dataset file, filtered for the top 5000 probes by rank mean expression and used to produce the initial network, which was divided into modules based on the Topological Overlap Metric as clustered by the Dynamic Tree Cut algorithm (See Supplemental Figure 1). The modules were summarized as “eigengenes”—essentially, the first principal component of all genes’ expression for that module, which represents an “expression signature.” The eigengenes are then correlated with the phenotypic label, in this case time (Control, Day 1 and Day 7). Five modules were statistically significant, with the Midnight Blue module having the highest correlation (Table 1). Unassigned genes had no significant correlation, as would be expected. Therefore, WGCNA identified in an untargeted approach total of 1247 genes in five clusters that were significantly correlated with the phenotype label.

One of the underlying premises of WGCNA is that genes with a similar function will cluster together. In order to ensure both that the clusters produced were biologically meaningful and that they captured the known biological processes involved in MPTP toxicity, we analyzed the modules using DAVID for over-represented annotations. All significant modules except one, the Midnight Blue module, were significantly enriched for terms when investigated by DAVID, and the DAVID enrichment clusters captured the known biology of MPTP toxicity (e.g., apoptosis—Magenta module; oxidative phosphorylation/Parkinson’s disease—Brown module) (Supplemental Table 1). The resulting network was visualized in Cytoscape (Fig. 1). One advantage of WGCNA is that it is a dimensionality reduction technique allowing for insight into the interrelationship among the modules. As can be seen from the network (Fig. 1), three modules (Brown, Salmon and Magenta) were fairly tightly interconnected, while the Midnight Blue module appeared as a sparse module, which connected the Cyan module with the other three. This suggests that the Midnight Blue module may act to coordinate the distinct functions of the other three modules, which may be mediated by transcription factor TCF3 (see discussion below).

Correlation networks typically have a high rate of false-positives and provide no insight into regulatory mechanisms. Therefore, to bring this approach closer to a mechanistically specified network and to better characterize the underlying biology, each module was analyzed for gene signatures in the Chemical and Genetic Perturbation subset of MSigDB as well as for over-represented transcription factor binding sites. Each module, except for the Magenta module, was substantially enriched for genes involved in Alzheimer's disease (Table S2), and while Alzheimer's disease has a different mechanism of neuronal degeneration compared to Parkinson's, this does indicate that the approach selected genes that are involved in neurodegenerative disease. Furthermore, it indicates that while the Midnight Blue module had no annotations to establish the functional significance of the cluster, the genes identified are related to neurodegeneration.

Modules were enriched for transcription factors relevant to Parkinson's disease

One biological reason for correlation of gene expression is common transcription factors or microRNAs. Therefore, each module was also analyzed in MSigDB for enriched transcription factor binding sites with an FDR corrected p value of <0.01 . This generated a list of 114 candidate transcription factor binding motifs that were enriched in the modules (of which 25 had no known transcription factor) and 23 microRNA binding sites. All modules had more than 10 predicted enriched motifs, and there was substantial overlap between enriched motifs among the modules. Candidate transcription factors and microRNA were text-mined for association with either Parkinson's or MPTP toxicity, and any transcription factor with more than two articles for Parkinson's and/or MPTP toxicity was considered relevant for building a genetic regulatory network (Table 2). This methodology found transcription factors that were well known for Parkinson's—JUN and NRF2, as well as ELK1, which both had literature evidence for Parkinson's and was a transcription factor in the Parkinson's Pathway in the PANTHER Database (Mi et al. 2013). Additionally, one of the transcription factor binding sites—SP1—had relatively few articles for Parkinson's disease, but did have binding motifs enriched in each of the modules (Table 2). SP1 was the only transcription factor with annotations for Parkinson's that was identified by MSigDB as relevant to the module with the strongest correlation with time, Midnight Blue module. The Cyan module had many transcription factors that were not shared with other modules, while the Brown module, in keeping with its size, had the largest number of potential transcription factors.

Transcription factors significantly improved the number of genes that could be connected in a component

For each module, all of the genes that could be located in the FANTOM4 database were analyzed with and without the subset of transcription factors both significant for that module and identified as being relevant to Parkinson's disease to form the basis of a genetic regulatory network for that module. All modules except one—Midnight Blue—contained a subset of genes that were connected by experimentally verified regulatory interactions in FANTOM4 (ChIP data, siRNA or published interactions), indicating that the modules consisted of genes that could be connected to each other with experimental data (Table 3).

However, the “connected component”—that is to say, the largest subset of genes and proteins that were interconnected with each other—grew substantially with the addition of the predicted transcription factors as identified by MSigDB and text-mining, even when restricted to experimental evidence; the percent of the modules connected by experimentally verified interactions ranged from a low of 70–80 % for each module and was 100 % for the Midnight Blue module (Table 3). For each module, the transcription factor that had, by far, the highest number of interactions was SP1, and it also had substantial experimental evidence of interactions (see Figs. 2, 4, 5, S2, S3). Within the Brown module, a subnetwork centered on SP1 and JUN indicated that it not only activated JUN but was connected to several downstream components as well (see Fig. 3). The Cyan module showed SP1 as a main hub and SREBF1 as a smaller hub when restricted to experimental evidence (see Fig. 4). Within the Midnight Blue module, even when restricted to evidence of 4 ChIP experiments, SP1 remained a significant hub (see Fig. 5).

SP1 is a ubiquitously expressed transcription factor that regulates a sweeping number of genes during development and other cellular functions. SP1 is known to play a key role in tissue differentiation; knockout mice are embryo-lethal and have multiple abnormalities (OMIM). SP1 is also known to play a role in cell-cycle inhibition (Deniaud et al. 2009), and over-expression leads to apoptosis (Chuang et al. 2009). Furthermore, SP1 is known to regulate the dopamine transporter (Wang and Bannon 2005) and is involved in several neurodegenerative diseases (Qiu et al. 2006) (Santpere et al. 2006). SP1 appears to be acetylated in neurons in response to oxidative stress and works in tandem with histone deacetylases to prevent cell death (Ryu et al. 2003); acetylation is but one of many post-translational modifications that expand SP1’s response repertoire.

SP1 was not present in the modules, nor was it among the genes differentially expressed, even with the most generous of cut-off values for significance. However, SP1 protein and mRNA levels have been shown to increase following MPP+ dosing in PC12 cells by approximately 1.5-fold, which was blocked by antioxidant treatment (Ye et al. 2013). The lack of appearance of SP1 among the genes differentially expressed or in the modules may simply reflect that SP1 mRNA rises only modestly (or perhaps briefly); alternatively, it is regulated by means other than an increase in mRNA levels and the signal increase is therefore nonlinear compared to mRNA levels (Courey et al. 1989). As SP1 is constitutively expressed rather than inducible, it may also act as a preliminary sensor that initiates the cascade.

Genes from all modules were combined into a genetic regulatory network based on FANTOM4 interactions as follows: (1) evidence limited to published interactions and siRNA perturbation data, (2) published/perturbation data with the addition of ChIP data and (3) all evidence, including predicted transcription factor binding sites. Even when restricted to published evidence, the resulting genetic regulatory network consisted of a connected component of 256 genes with several hubs (Fig. 6). Including ChIP data extended it to 782 and predicted transcription factor bind sites to 830. In addition to SP1, the network hubs consist of some candidates well known for their role in Parkinson’s (STAT3, JUN) but also produced other candidates that have been implicated in Parkinson’s: SREBF1 has previously been identified as a risk locus for sporadic Parkinson’s disease (Do et al. 2011), and in a

recent RNAi screening study, it was implicated in the control of the PTEN-induced kinase 1 (PINK1)/Parkin pathways that control the autophagic destruction of mitochondria (Ivatt and Whitworth 2014).

One hub identified in the reconstructed GRN from FANTOM4, HDAC1, has been implicated in cell survival in neurotoxicity to dopaminergic neurons in vitro and ischemia in vivo (Kim et al. 2008); HDAC1 was also a hub in the WGCNA network, and all the genes connected to HDAC1 in the FANTOM4 GRN by two degrees were connected in WGCNA. The WGCNA network also suggested a protein, LANCL1, that was connected to both HDAC1 and STAT3. LANCL1 binds glutathione and is believed to play a role in neuronal survival following oxidative insult (Zhong et al. 2012), and its connection to HDAC1 and STAT3 seems plausible.

One of the smaller hubs, ZNF148 (a zinc-binding transcription factor), was a hub within both the FANTOM4 and the Midnight Blue module. Of the 18 genes directly connected to ZNF148 in the original WGCNA network (which included MAPT), four were also linked by predicted interactions in FANTOM4. ZNF148 (also referred to as ZPB89) is not present on any pathway in Panther or KEGG and has a relatively sparse literature base with no indication of any role in Parkinson's. Of the four genes (CBX3, DDX6, SYNGR1 and C20orf27) predicted to interact with ZNF148 in FANTOM4 and also in the WGCNA network, only SYNGR1 has evidence of involvement with neuroplasticity (Janz et al. 1999). ZNF148, however, is known to play a role in apoptosis (Zhang et al. 2010), is verified by proteomic studies as expressed in the *substantia nigra* (Chen et al. 2012) and would be an interesting candidate for further study. ATF4, which has recently been identified by us and others in independent high-throughput studies as a key transcriptional factor in MPTP toxicity (Ye et al. 2013, Krug et al. 2014), was also present as a small hub containing mostly protein-protein interaction connections in the network when restricted to experimentally verified interactions. Similarly, TCF3 had relatively few experimentally verified reactions and is thus relatively small in the graph; however, an expanded subnetwork that included predicted transcription factor binding sites, even when restricted to a high stringency level, would have been substantially larger. TCF3 was in the Midnight Blue module, and the Cyan, Salmon and Brown modules were all enriched for TCF3 binding motifs. This is likely a case where the relative importance of a gene is underestimated based on the lack of available experimental data in comparison with the better-studied SP1. One pitfall of this methodology is that it will be biased toward comparatively well-characterized, GC-rich transcription factor binding sites. In theory, examining all highly co-ordinated genes in the WGCNA network would not have such a bias, but given the high rate of false-positives there would have to be another method of confirming the causative agent of co-ordinated co-expression.

One of the transcription factor binding sites that were consistently ranked by MSigDB across all modules (PAX4) had no textual evidence for involvement with Parkinson's, although it does appear to be expressed in the brain. However, there was no experimental evidence in FANTOM4 that it bound to any of the targets in the modules, the predicted targets were quite sparse, and its inclusion would not have fundamentally changed the architecture of the network. Similarly, binding sites for LEF1 were also statistically over-

represented in all the modules, and LEF1 was present in one of the modules. While it was therefore a “hub” in the FANTOM4 GRN if both predicted and published interactions were included, it was not a hub in the WGCNA network, likely indicating that it is of minimal importance to the underlying physiological network. This indicates that the dimensionality reduction achieved by clustering the genes into modules helps eliminate spurious transcription factor binding sites and that using text-mining (or restricting to published interactions) helps to narrow down the candidate transcription factors and improves the results. Furthermore, the resultant genetic regulatory network does not merely reflect nonspecific predicted transcription factor binding motifs that are enriched in any subset of genes, but is likely enriched for biologically relevant targets.

Discussion

Current analysis of microarray data in toxicology typically does not take advantage of the data-mining and bioinformatics tools available to interpret the underlying mechanisms but remains at the level of “biomarker” or signature identification, either generating a relatively small list of genes differentially expressed using inferential statistics, or over-representation analysis, which is highly dependent on pathway annotations. We chose an existing dataset, originally used to identify a few genes as signatures of MPTP toxicity in vivo, in order to explore an alternative method that would offer more insight into dynamics of gene expression compared to inferential statistics and would not be dependent on pathway annotations. WGCNA offers many advantages for analyzing microarray data: It is unsupervised, and, unlike correlation networks that are based solely on a Pearson or Spearman correlation, it preserves weak links—capturing interactions that may be small, but nonetheless be biologically interesting. This may be especially relevant to toxicology, as the effects may be subtle and distributed among many pathways. The dimensionality reduction achieved has the advantage of preserving the connection between the clusters; the HDAC1 subnetwork in Fig. 7 includes almost equal parts of genes from the Magenta and Brown modules.

As this represented a fairly unsophisticated approach to text-mining transcription factor candidates, it is quite probable that the proposed regulatory network is only a “10,000-foot” view; many of the transcription factors may have had textual evidence of being involved in physiological processes that are relevant to MPTP toxicity—e.g., oxidative stress or apoptosis. Although extending the text-mining in such a way would likely have increased the false-positives, it could also have fine-tuned the map in some of the “neighborhoods.” Objective literature mining by search engines is a key for an unbiased use of the existing literature, clearly preferable over the cherry-picking of studies in interpreting individual results. MPTP’s Pathway of Toxicity is aided markedly by the fact that it serves as an animal model for a relatively well-researched disease such as Parkinson’s; depending on MPTP/MPP+ literature would have produced a much smaller subset of candidate transcription factors and no microRNAs, perhaps reflecting the relatively immature literature base from microRNAs compared to transcription factors.

Furthermore, just as the “connected component” likely contains some regulatory connections that are artifacts, the unconnected component contains both genes that are

spurious correlations and genes that are unconnected due to lack of data about the probable regulatory mechanism. Disappointingly, the microRNAs that were identified did not have any regulatory connections; this may reflect the fact that microRNAs simply have an inadequate dataset, and it is likely that multiple microRNAs are involved but are simply invisible in this analysis. Surprisingly, one of the “unconnected” genes in the Brown module was MAO-A (monoamine oxidase A). Although MPTP is metabolized much more efficiently by MAO-B, MAO-A is possibly involved in dopaminergic cell death in neurons (Naoi et al. 2012) and there is evidence that SP1 binds to the promoter of MAO-A (Zhu et al. 1994).

Interestingly, within the Midnight Blue module, two proteins that had relatively weak evidence of connection to SP1, AQP4 (Aquaporin-4) and TUB (Tubby protein) were not in the final FANTOM4 genetic regulatory network, but were examined for evidence of involvement in MPTP toxicity-related processes, as both have knockout mice models. Aquaporin-4 knockout mice are more prone to MPTP toxicity (Fan et al. 2008), and although Aquaporin-4 may or may not be regulated by SP1, it likely plays a role in the ultimate phenotypic consequences of MPTP toxicity (and perhaps Parkinson’s as well). Tubby protein knockout mice have a primary phenotype of obesity but also display neurodegeneration. There is some evidence that TUB is a regulator of microglial phagocytosis through the MerTK receptor (Caberoy et al. 2012). The exact nature and role of TUB in MPTP toxicity, however, remain speculative. Nonetheless, neither the genes unconnected to the larger network nor the weaker links in the network that lack substantial experimental evidence should be discarded wholesale.

Mitochondrial disruption is a commonality for a variety of neurotoxins and neurodegenerative diseases, but the exact route between mitochondrial disruption and the phenotype is unclear. MPTP, like other toxins, may work primarily to disrupt the mitochondria, but the disruption likely has pleiotropic effects that differ from other toxins and disease states. Depending on annotations to reveal physiological function (or, alternatively, discarding a cluster because of lack of annotations) may overlook useful information about toxic processes. In this case, the Midnight Blue module contained genes known or strongly suspected to be involved in Parkinson’s or MPTP toxicity (MAPT, SYNGR1), as well as genes known to be involved in neuropathology (THOP1, which cleaves amyloid precursor protein) (Pollio et al. 2008). It also suggested novel candidates that are plausibly involved in the degenerative process (AQP4 and TUB), neither of which were on existing Parkinson’s pathways (Panther, KEGG) and both of which had an inadequate literature depth on which to base enrichment analysis.

The pronounced promiscuousness of SP1 binding sites entails that many, if not most, of the predicted interactions are spurious and the experimentally verified interactions may be irrelevant within the particular context of MPTP toxicity in dopaminergic neurons. However, given the statistically significant over-representation of SP1 motifs in all the modules, the centrality of SP1 to the predicted network, the literature evidence of involvement in Parkinson’s, dopamine regulation and MPTP toxicity, and the experimental evidence of interactions with known signaling networks (such as JUN) involved in Parkinson’s and MPTP, SP1 is likely necessary (though not sufficient) for MPTP toxicity

and acts to integrate multiple signaling pathways in a combinatorial and complex manner. The proposed genetic regulatory network offers an advantage compared to a correlation network insofar as it offers a direction of action, an estimate of transcription factor binding site strength, multiple lines of evidence and an estimate of the dynamics of gene expression. Therefore, it can act as scaffolding which further experiments, both in silico and in vitro, can refine.

This study shows that a relatively small gene array study allows for pinpointing mechanistic information by a combination of correlation approaches with both text-mining and large-scale interaction datasets such as FANTOM. Each independently has strengths and weaknesses and is error-prone. Combining the approaches to narrow down potential candidates, however, is promising. Nonetheless, any data-mining approach—especially those that tend to generate false-positives—must go hand-in-hand with confirmation of the (patho-)physiological sense of the distilled information. These emerging approaches for Pathway of Toxicity identification can become even more powerful when several orthogonalomics technologies are employed and different experimental models are combined.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for “Mapping the Human Toxome by Systems Toxicology” was provided by an NIH Transformation Research Grant (RO1 ES 020750) and Food and Drug Administration Grant “DNTox-21c Identification of Pathways of Developmental Neurotoxicity for High Throughput Testing by Metabolomics” (U01 FD 004230) to Johns Hopkins Bloomberg School of Public Health (PI Thomas Hartung). Alexandra Maertens was supported by an NIEHS diversity grant.

References

- Bouhifd M, Hogberg HT, Kleensang A, Maertens A, Zhao L, Hartung T. Mapping the human toxome by systems toxicology. *Basic Clin Pharmacol Toxicol*. 2014; 115:1–8.
- Cabero NB, Alvarado G, Li W. Tubby regulates microglial phagocytosis through MerTK. *J Neuroimmunol*. 2012; 252:40–48. [PubMed: 22884297]
- Cappelletti G, Pedrotti B, Maggioni MG, Maci R. Microtubule assembly is directly affected by MPP(+) in vitro. *Cell Biol Int*. 2001; 25:981–984. [PubMed: 11589612]
- Chen S, Lu FF, Seeman P, Liu F. Quantitative proteomic analysis of human substantia nigra in Alzheimer’s Disease, Huntington’s disease and Multiple sclerosis. *Neurochem Res*. 2012; 37:2805–2813. [PubMed: 22926577]
- Chuang JY, Wu CH, Lai MD, Chang WC, Hung JJ. Overexpression of Sp1 leads to p53 dependent apoptosis in cancer cells. *Int J Cancer*. 2009; 125:2066–2076. [PubMed: 19588484]
- Courey AJ, Holtzman DA, Jackson SP, Tjian R. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell*. 1989; 59:827–836. [PubMed: 2512012]
- Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor. *Bioinformatics*. 2007; 23:1846–1847. [PubMed: 17496320]
- Dehay B, Bove J, Rodriguez-Muela N, Perier C, Recasens A, Boya P, Vila M. Pathogenic lysosomal depletion in Parkinson’s disease. *J Neurosci*. 2010; 30:12535–12544. [PubMed: 20844148]
- Deniaud E, Baguet J, Chalard R, Blanquier B, Brinza L, Meunier J, Michallet M-C, Laugraud A, Ah-Soon C, Wierinckx A, Cas-tellazzi M, Lachuer J, Gautier C, Marvel J, Leverrier Y. Overexpression

of transcription factor Sp1 leads to gene expression perturbations and cell cycle inhibition. *PLoS One*. 2009; 4:e7035. [PubMed: 19753117]

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003; 4:P3. <http://genomebiology.com/2003/4/9/R60>. [PubMed: 12734009]

Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW, Wojcicki A, Eriksson N. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet*. 2011; 7:e1002141. [PubMed: 21738487]

Fan Y, Kong H, Shi X, Sun X, Ding J, Wu J, Hu G. Hypersensitivity of aquaporin 4-deficient mice to 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine and astrocytic modulation. *Neurobiol Aging*. 2008; 29:1226–1236. [PubMed: 17353068]

Hartung T. Toxicology for the twenty-first century. *Nature*. 2009; 460:208–212. [PubMed: 19587762]

Hartung T, McBride M. Food for thought... on mapping the human toxome. *Altex*. 2011; 28:83–93. [PubMed: 21625825]

Hoang T, Choi D-K, Nagai M, Wu D-C, Nagata T, Prou D, Wilson GL, Vila M, Jackson-Lewis V, Dawson VL, Dawson TM, Chessel M-F, Przedborski S. Neuronal NOS and cyclooxygenase-2 contribute to DNA damage in a mouse model of Parkinson disease. *Free Radical Biol Med*. 2009; 47:1049–1056. [PubMed: 19616617]

Ivatt R, Whitworth AJ. SREBF1 links lipogenesis to mitophagy and sporadic Parkinson's disease. *Autophagy*. 2014; 10:33–34.

Janz R, Sudhof TC, Hammer RE, Unni V, Siegelbaum SA, Bolshakov VY. Essential roles in synaptic plasticity for synaptogyrin I and synaptophysin I. *Neuron*. 1999; 24:687–700. [PubMed: 10595519]

Kim D, Frank CL, Dobbin MM, Tsunemoto RK, Wu D, Peng PL, Guan J-S, Lee B-H, Moy LY, Giusti P, Broodie N, Mazitschek R, Delalle I, Haggarty SJ, Neve RL, Lu YM, Tsai L-H. Deregulation of HDAC1 by p25/Cdk5 in neurotoxicity. *Neuron*. 2008; 60:803–817. [PubMed: 19081376]

Kleinsang A, Maertens A, Rosenberg M, Fitzpatrick S, Lamb J, Auerbach S, Brennan R, Crofton KM, Gordon B, Fornace AJ Jr, Gaido K, Gerhold D, Haw R, Henney A, Ma'ayan A, McBride M, Monti S, Ochs MF, Pandey A, Sharan R, Stierum R, Tugendreich S, Willett C, Wittwehr C, Xia J, Patton GW, Arvidson K, Bouhifd M, Hogberg HT, Luechtefeld T, Smirnova L, Zhao L, Adeleye Y, Kanehisa M, Carmichael P, Andersen EM, Hartung T. Pathways of Toxicity. *Altex*. 2014; 31:53–61. [PubMed: 24127042]

Krug AK, Gutbier S, Zhao L, Pörtl D, Kullmann C, Ivanova V, Förster S, Jagtap S, Meiser J, Lepar G, Schildknecht S, Adam M, Hiller K, Farhan H, Brunner T, Hartung T, Sachinidis A, Leist M. Transcriptional and metabolic adaptation of human neurons to the mitochondrial toxicant MPP(+). *Cell Death Dis*. 2014; 5:e1222. [PubMed: 24810058]

Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol*. 2007; 1:54. [PubMed: 18031580]

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma*. 2008; 9:559.

Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinforma*. 2008; 24:719–720.

Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013; 8:1551–1566. [PubMed: 23868073]

Miller RM, et al. Dysregulation of gene expression in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine-lesioned mouse substantia nigra. *J neurosci*. 2004; 24:7445–7454. [PubMed: 15329391]

Naoi M, Maruyama W, Inaba-Hasegawa K. Type A and B monoamine oxidase in age-related neurodegenerative disorders: their distinct roles in neuronal death and survival. *Curr Top Med Chem*. 2012; 12:2177–2188. [PubMed: 23231395]

Perier C, Vila M. Mitochondrial biology and Parkinson's disease. *Cold Spring Harb Perspect Med*. 2012; 2:a009332. [PubMed: 22355801]

- Pollio G, Hoozemans JJ, Andersen CA, Roncarati R, Rosi MC, van Haastert ES, Seredenina T, Diamanti D, Gotta S, Fiorentini A, Magnoni L, Raggiaschi R, Rozemuller AJ, Casamenti F, Carica-sole A, Terstappen GC. Increased expression of the oligo-peptidase THOP1 is a neuroprotective response to Abeta toxicity. *Neurobiol Dis.* 2008; 31:145–158. [PubMed: 18571100]
- Qiu Z, Norflus F, Singh B, Swindell MK, Buzescu R, Bejarano M, Chopra R, Zucker B, Benn CL, DiRocco DP, Cha JH, Ferrante RJ, Hersch SM. Sp1 is up-regulated in cellular and transgenic models of Huntington disease, and its reduction is neuroprotective. *J Biol Chem.* 2006; 281:16672–16680. [PubMed: 16595660]
- Quackenbush J. Genomics. Microarrays-guilt by association. *Science.* 2003; 302:240–241. [PubMed: 14551426]
- Ren Y, Liu W, Jiang H, Jiang Q, Feng J. Selective vulnerability of dopaminergic neurons to microtubule depolymerization. *J Biol Chem.* 2005; 280:34105–34112. [PubMed: 16091364]
- Ryu H, Lee J, Olofsson BA, Mwidau A, Dedeoglu A, Escudero M, Flemington E, Azizkhan-Clifford J, Ferrante RJ, Ratan RR. Histone deacetylase inhibitors prevent oxidative neuronal death independent of expanded polyglutamine repeats via an Sp1-dependent pathway. *Proc Natl Acad Sci USA.* 2003; 100:4281–4286. [PubMed: 12640146]
- Santpere G, Nieto M, Puig B, Ferrer I. Abnormal Sp1 transcription factor expression in Alzheimer disease and tauopathies. *Neurosci Lett.* 2006; 397:30–34. [PubMed: 16378688]
- Schober A. Classic toxin-induced animal models of Parkinson's disease: 6-OHDA and MPTP. *Cell Tissue Res.* 2004; 318:215–224. [PubMed: 15503155]
- Severin J, Waterhouse AM, Kawaji H, Lassmann T, van Nimwegen E, Balwierz PJ, de Hoon MJ, Hume DA, Carninci P, Hayashizaki Y, Suzuki H, Daub CO, Forrest AR. FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microR-NAs, expression dynamics and regulatory interactions. *Genome Biol.* 2009; 10:R39. [PubMed: 19374773]
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–2504. [PubMed: 14597658]
- Snyder SH, D'Amato RJ. MPTP: a neurotoxin relevant to the pathophysiology of Parkinson's disease —The 1985 George C. Cotzias Lecture. *Neurology.* 1986; 36:250. [PubMed: 3080696]
- Sterky FH, Hoffman AF, Milenkovic D, Bao B, Paganelli A, Edgar D, Wibom R, Lupica CR, Olson L, Larsson NG. Altered dopamine metabolism and increased vulnerability to MPTP in mice with partial deficiency of mitochondrial complex I in dopamine neurons. *Hum Mol Gen.* 2012; 21:1078–1089. [PubMed: 22090423]
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression pro-files. *Proc Natl Acad Sci USA.* 2005; 102:15545–15550. [PubMed: 16199517]
- Wang J, Bannon MJ. Sp1 and Sp3 activate transcription of the human dopamine transporter gene. *J Neurochem.* 2005; 93:474–482. [PubMed: 15816870]
- Ye Q, Zhang X, Huang B, Zhu Y, Chen X. Astaxanthin suppresses MPP-induced oxidative damage in PC12 cells through a Sp1/NR1 signaling pathway. *Mar Drugs.* 2013; 11:1019–1034. [PubMed: 23538867]
- Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinforma.* 2007; 8:22.
- Zhang CZ, Chen GG, Lai P. Transcription factor ZBP-89 in cancer growth and apoptosis. *Biochim Biophys Acta Rev Cancer.* 2010; 1806:36–41.
- Zhong WX, Wang YB, Peng L, Ge XZ, Zhang J, Liu SS, Zhang XN, Xu ZH, Chen Z, Luo JH. Lanthionine synthetase C-like protein 1 interacts with and inhibits cystathionine beta-synthase: a target for neuronal antioxidant defense. *J Biol Chem.* 2012; 287:34189–34201. [PubMed: 22891245]
- Zhu QS, Chen K, Shih JC. Bidirectional promoter of human monoamine oxidase A (MAO A) controlled by transcription factor Sp1. *J Neurosci.* 1994; 14:7393–7403. [PubMed: 7996184]

Fig. 1.
Network generated by WGCNA, *colored* by module, using spring-embedded bio-layout based on edge strength

Fig. 2. Brown module, identified transcription factors in grey. *Legend* the Brown module formed a dense network of regulatory interactions centered on SP1. Self-loops indicate a gene interacts with itself

Fig. 3.

SP1, JUN and STAT1 subnetwork from the Brown module. *Legend green* indicates ChIP data; *red* indicates perturbation experiment; *yellow*, published protein–DNA interactions, and *purple* indicates protein–protein interaction. Node size is proportional to predicted dynamics of the gene, and *darker* nodes indicate higher scaled expression levels. Because the FANTOM4 database gives an estimate of the dynamics of gene expression, the resulting gene regulatory network can be used as the foundation for building a dynamic model

Fig. 4. Cyan module with TFs identified as indicated in *grey*; SP1 is in the middle. *Legend* edges represent experimentally verified interactions; straight lines are ChIP experiments or protein-protein interactions, wavy lines are siRNA preturbation experiments. All nodes were connected when predicted interactions were included (data not shown)

Fig. 5.
Midnight blue module; SP1 interactions verified with 4 ChIP experiments

Fig. 6. Genetic regulatory network based on published interactions. *Legend* node label is proportionate to hub status as determined by edge count. Self-interactions were deleted for visual clarity

Fig. 7.

HDAC1 subnetwork from FANTOM4; all genes within one or two degrees of HDAC1 in the FANTOM4 network (not including SP1); single leaves deleted for visual clarity. In the WGCNA network, all the above were linked to HDAC

Table 1

Modules correlated with time

Module	Correlation	<i>p</i> value
Magenta	0.7996246	1.80E-03
Salmon	0.76605824	3.67E-03
Brown	0.58916781	4.38E-02
Cyan	0.69331419	1.24E-02
Midnight Blue	0.94604195	3.29E-06
Unassigned	0.13829676	6.68E-01

Five of the modules produced were significantly correlated; significance is calculated via a permutation test

Table 2

Candidate transcription factors associated with Parkinson's and MPTP via text-mining; *p* values based on enrichment in MSigDB C3 gene sets

Transcription factor	Abstract for Parkinson's	MPTP/MPP±	Module	FDR corrected <i>p</i> value
JUN	4451	729	Brown	3.44E-08
NRFR2	59	25	Salmon	6.23E-03
FOXF2	21	1	Brown	1.69E-07
SP1	12	2	Brown	4.25E-26
			Cyan	2.51E-06
			Magenta	8.08E-05
			Midnight Blue	2.53E-03
			Salmon	7.85E-05
ATF4	12	2	Brown	9.14E-07
TCF3	11	6	Brown	9.20E-07
			Cyan	5.82E-04
			Salmon	2.52E-02
ELK1	3		Brown	3.14E-15
			Magenta	8.08E-05
AP1	3	1	Brown	3.44E-08
			Salmon	2.52E-02
STAT1	7	2	Brown	3.76E-06
NRF1	6	3	Brown	2.34E-06
			Cyan	7.59E-04
SRY	6		Cyan	4.84E-03
MIR-132	5		Brown	3.76E-06
SREBF1	5		Cyan	7.59E-04
ATF3	4	1	Cyan	4.84E-03
SRY	6		Cyan	4.84E-03
MIR221	2		Cyan	4.95E-03
MEF2A	2		Cyan	9.52E-03
			Magenta	0.000809
ELK1	2		Magenta	8.08E-05
			Brown	3.14E-15

Table 3

Addition of predicted transcription factors substantially increased connected component

Module	Genes in fantom	Connected component—without TFS		Connected component with TFS	
		Experimental	Predicted	Experimental	Predicted
Salmon	163	14	47	125	132
Midnight Blue	105	0	14	75	105
Cyan	150	16	41	121	130
Brown	463	31	163	381	409
Magenta	106	14	26	82	91

For each module, gene symbols were entered into FANTOM4 EdgeExpressDB and a predicted regulatory network was drawn based on experimental evidence (ChIP, published interactions and siRNA experiments), with and without the addition of predicted evidence (predicted transcription factor binding and microRNA). Transcription factors were added based on evidence of significantly over-represented motifs in MSigDB and textual evidence of involvement in Parkinson's.

* In the case of the Brown module, STAT1 and JUN were already in the module. "Connected component" consists of all genes that were not singletons in the predicted regulatory network