



Article

Separating the Wheat from the Chaff: The Use of Upstream Regulator Analysis to Identify True Differential Expression of Single Genes within Transcriptomic Datasets

Jeremiah Hadwen ^{1,2,*}, Sarah Schock ^{1,2}, Faraz Farooq ^{1,2}, Alex MacKenzie ^{1,2,*} and Julio Plaza-Diaz ^{2,3,4}

¹ Department of Cellular and Molecular Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON K1H8M5, Canada; sarah.schock@canada.ca (S.S.); faraztfarooq@gmail.com (F.F.)

² Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON K1H8L1, Canada; jrplaza@ugr.es

³ Department of Biochemistry and Molecular Biology II, School of Pharmacy, University of Granada, 18071 Granada, Spain

⁴ Instituto de Investigación Biosanitaria IBS.GRANADA, Complejo Hospitalario Universitario de Granada, 18014 Granada, Spain

* Correspondence: hadwen.jeremiah@gmail.com (J.H.); mackenzie@cheo.on.ca (A.M.)

Abstract: The development of DNA microarray and RNA-sequencing technology has led to an explosion in the generation of transcriptomic differential expression data under a wide range of biologic systems including those recapitulating the monogenic muscular dystrophies. Data generation has increased exponentially due in large part to new platforms, improved cost-effectiveness, and processing speed. However, reproducibility and thus reliability of data remain a central issue, particularly when resource constraints limit experiments to single replicates. This was observed firsthand in a recent rare disease drug repurposing project involving RNA-seq-based transcriptomic profiling of primary cerebrocortical cultures incubated with clinic-ready blood-brain penetrant drugs. Given the low validation rates obtained for single differential expression genes, alternative approaches to identify with greater confidence genes that were truly differentially expressed in our dataset were explored. Here we outline a method for differential expression data analysis in the context of drug repurposing for rare diseases that incorporates the statistical rigour of the multigene analysis to bring greater predictive power in assessing individual gene modulation. Ingenuity Pathway Analysis upstream regulator analysis was applied to the differentially expressed genes from the Care4Rare Neuron Drug Screen transcriptomic database to identify three distinct signaling networks each perturbed by a different drug and involving a central upstream modulating protein: levothyroxine (*DIO3*), hydroxyurea (*FOXM1*), dexamethasone (*PPARD*). Differential expression of upstream regulator network related genes was next assessed in in vitro and in vivo systems by qPCR, revealing 5× and 10× increases in validation rates, respectively, when compared with our previous experience with individual genes in the dataset not associated with a network. The Ingenuity Pathway Analysis based gene prioritization may increase the predictive value of drug-gene interactions, especially in the context of assessing single-gene modulation in single-replicate experiments.

Keywords: transcriptomics; differential expression analysis; rare disease; drug repurposing



Citation: Hadwen, J.; Schock, S.; Farooq, F.; MacKenzie, A.; Plaza-Diaz, J. Separating the Wheat from the Chaff: The Use of Upstream Regulator Analysis to Identify True Differential Expression of Single Genes within Transcriptomic Datasets. *Int. J. Mol. Sci.* **2021**, *22*, 6295. <https://doi.org/10.3390/ijms22126295>

Academic Editor: Catherine Coirault

Received: 3 May 2021

Accepted: 7 June 2021

Published: 11 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rare diseases including the muscular dystrophies system are significant contributors to human disability and illness. Although defined as less than 1 in 2000 people in Europe and less than 1 in 200,000 people in the United States, they are collectively common, carrying significant medical, societal, and economic costs [1–4].

There are comparatively few effective rare disease therapies; one approach to identifying new treatments involves genome-wide differential expression (DE) analysis (the

process through which transcriptomic data are converted to statistical measures of gene expression) by microarray and RNA-sequencing [5–12], as a means of potentially correcting the pathogenic protein expression which underlies so many diseases [13].

Although the DE pipeline has been extensively customized for individual research purposes [14], the diverse approaches generally share the same workflow [15]: raw data processing, DE analysis, and multiple testing correction [16,17], followed by choice of a statistical cut-off to identify differentially expressed genes (DEGs). The DEGs must then be classified and validated to extract biologically useful information. Research involving transcriptome-wide DE data analysis may either focus on gene enrichment patterns or DE of single genes, this is particularly so when studying rare diseases for it is the modulation of a specific gene that can hold the greatest therapeutic promise (e.g., utrophin for Duchenne muscular dystrophy). In the former analysis, which relies primarily on the gene annotation and enrichment analysis [18], a certain degree of statistical rigour can be achieved by virtue of the number of differentially expressed genes being studied. One such example is the analysis of drug-mediated transcriptome modulation to identify novel indications for established drugs, pioneered by Lamb et al. who first used connectivity mapping of transcriptomic signatures to pair clinically approved drugs with disease signatures [19].

In contrast, single-gene modulation observed in system-wide datasets is not as reproducible, and this is especially the case in single-replicate experiments. This can be seen in efforts to repurpose drugs to treat rare monogenic diseases by mining transcriptome-wide drug screen data for drug hits on rare disease genes of interest [12,20–23]. This work, unlike the broader assessment of groups of genes outlined above, focuses on single-gene DE resulting in suboptimal reproducibility, particularly in single-replicate experimental systems. As a result, although several genes from these studies have been validated at the in vitro level and in whole-animal experiments, the validation rates were only 20% in vitro and 5% in vivo [12,20].

DEGs are often analyzed in aggregate using gene enrichment bioinformatic tools [24] or, more recently, commercial pathway analysis tools such as Ingenuity Pathway Analysis (IPA) [25]. IPA and similar tools (e.g., iPathwayGuide [26], PathVisio [27], GeneGO MetaCore [28]) are similar to gene-set enrichment analysis platforms [29], with the added benefit of knowing the direction (up- or downregulation) of each pathway. Based on annotated knowledge of functionally related DEGs, IPA upstream regulator analysis (URA) identifies a common upstream regulator (e.g., transcription factor, kinase, drug) and analyzes the DE of downstream genes. The tool then generates an associated z-score and *p*-value of overlap that impart statistical significance and directionality (up- or downregulation) to each upstream regulator [25].

Here, by marrying the statistical robustness of gene network analysis to single-gene DE measurement, we propose a method of gauging the validity of single DEGs identified from in vitro transcriptomic data, helping identify those that warrant further investigation in cellular and animal systems. Data from the Neuron Screen Database (available at <http://bigbear.med.uottawa.ca:1000>, accessed on 7 June 2021) [12] were analyzed by IPA URA to identify three DEG groups that share statistically activated or inhibited upstream regulators: (i) thyroid hormone upregulation of deiodinases (DIO) network as positive control for the URA methodology, (ii) novel hydroxyurea (HU)-mediated inhibition of *FOXM1* to investigate in vitro validation rates, and (iii) dexamethasone (DEX)-mediated activation of *PPARD* to investigate in vivo validation rates. The use of the URA-based method to select DEGs resulted in a validation rate far superior to that achieved previously for single genes not related to a network. This suggests the utility of the IPA-directed approach in distinguishing true DEGs from stochastic artifacts.

2. Results

Pathway Analysis

URA of RNA-seq DE data from the recently published Neuron Screen of clinic-ready drugs [12] was conducted to assess whether an a priori “reliability” of an observed drug-

conferred single-replicate single-gene DE might be achieved. Samples of “activated” and “inhibited” networks elicited by the 50 drugs that modulated the greatest number of transcripts were identified by IPA using upstream regulator analysis of DE data. Six statistically significant activated and inhibited URA networks ($|z| > 2$, and a minimum of five downstream DEGs with $p\text{-adj} < 0.05$) were selected for further study: Levothyroxine (DIO3), hydroxyurea (FOXM1), DEX (PPARD), DEX (STAT4), vigabatrin (MKNK1), and pregabalin (PGR) (Table 1).

Table 1. Summary of six upstream regulators identified in the Neuron Screen data.

Drug Name	Upstream Regulator	Expression Fold Change	Predicted Activation State	Activation Z-Score	<i>p</i> -Value of Overlap	Number of Genes
Levothyroxine	DIO3	4.385	Inhibited	−2.975	1.79×10^{-7}	8
Hydroxyurea	FOX M1	−3.579	Inhibited	−3.245	1.63×10^{-10}	8
Dexamethasone	PPARD	2.522	Activated	3.126	4.58×10^{-2}	10
Dexamethasone	STAT4	NA	Activated	2.933	7.36×10^{-4}	12
Vigabatrin	MKNK1	1.187	Inhibited	−3	1.40×10^{-4}	5
Pregabalin	PGR	1.013	Activated	3.376	7.77×10^{-9}	13

The expression fold change indicated is derived from the original study reporting the RNA-seq results from the neuronal screen. The level of mRNA (i.e., RNA-seq read numbers) for a given gene in the presence of a drug is compared against the average mRNA level of that gene for all conditions while the upstream regulator is the central molecule in the network. The activation z-score serves as a statistical measure as well as a directional tool (negative/downregulated, positive/upregulated) while the *p*-value of overlap serves as an additional measure of statistical certainty pertaining to the interrelated nature of the molecules forming each network. NA, data not available as gene expression was not included in the Neuron Screen data [25].

The deiodinase 3 (*DIO3*) network, inhibited by the thyroid hormone analog levothyroxine (T4) in the Neuron Screen data (z-score = −2.975), was initially analyzed. The *DIO3* gene, known to be regulated by thyroid hormone, encodes type 3 iodothyronine deiodinase that inactivates thyroid hormone and downregulates thyroid hormone-responsive genes [30]. The postulated downstream targets of *DIO3* (Figure 1) are known transcriptional targets of thyroid hormone (e.g., *Shh*, *Sema7a*, *Hr*) and have been validated in mouse cerebrotectum and cortical cultures [12,31]. Thus, the *DIO3* network was labelled with a z-score of −2.975, since the IPA software assumed inhibition of *DIO3* (although *DIO3* was in fact upregulated) because of the upregulation of thyroid hormone-responsive genes (Figure 1). This analysis serves to highlight the possible discrepancy that exists between an upstream regulator and the downstream genes, while confirming the reliability of URA network-directed prioritization of DEGs since all the genes in the *DIO3* network are upregulated in response to thyroid hormone experimentally (from the Neuron Screen data) and in the published literature. Another example is the *STAT4* network, which has a robust activation z-score with 12 network-related genes. This occurs despite the fact that *STAT4* did not appear as a differentially expressed gene in the Neuron Screen data, and thus there is no fold expression present. This then is an example of using this approach to identify a potential upstream regulator in the absence of it showing up as a DEG itself.

The *FOX M1* network, inhibited by treatment with hydroxyurea (HU) (activation z-score of −3.245; *p*-value of overlap at 1.63×10^{-10}) was used to determine the rate of in vitro validation of network-associated DEGs. The *FOX M1* transcription factor has documented roles both in cell cycle gene regulation and DNA damage repair cascade activation [32]. However, the role of HU in the transcriptional modulation of the *FOX M1* network genes has not been well described. The IPA analysis showed that eight *FOX M1*-target genes involved in G2 (growth phase 2) and M phase (mitosis phase) of cell cycle progression were downregulated in response to hydroxyurea treatment or mouse cortical cultures (Figure 2A). To determine the validation rate of the HU-elicited *FOX M1*-associated DEGs, human U87 glioblastoma cells were treated for 0, 4, and 8 h with 250 μM HU. Quantitative RT-PCR revealed HU-mediated increases (minimum significance of $p < 0.05$) of seven of the eight *FOX M1*-network genes at 4 and/or 8 h (Figure 2C–E; Supplementary File, Figure S1). The level of the upstream regulator gene *FOX M1* was not significantly affected by 4 or 8 h

of HU treatment (Figure 2B). These results confirmed the efficiency of identifying DE genes from URA network analysis, while serving to highlight that such results are not contingent on the individual upstream gene (in this case *FOXM1*).

The *PPARD* network, activated by DEX treatment (z -score = 3.126 and p -value = 0.0458), was next used to assess the in vivo validation of DEGs identified by URA networks. The activity of 1 mg/kg DEX oral treatments was confirmed by the upregulation of the DEX-responsive gene *Fkbp5* in the cortex of C57BL6 mice shown by qRT-PCR [33,34] (Supplementary File, Figure S2A). Subsequently, qRT-PCR analysis of the *PPARD* network in cortical tissue from DEX-treated mice was undertaken. Seven of the ten genes in the *PPARD* network were studied by qRT-PCR. Three genes (*Bcl2l1*, *Ilk*, *Mfsd2a*) showed statistically significant upregulation in DEX-treated mice (Figure 3B–D), while *Pdk4* and *Kyat3* showed a trend toward upregulation in DEX-treated mice (Figure 3E; Supplementary File, Figure S2B). Only two of the seven genes tested (*Mertk* and *Lrp5*) showed no increase in response to DEX treatment in the mouse cortex (Supplementary File, Figure S2C,D).

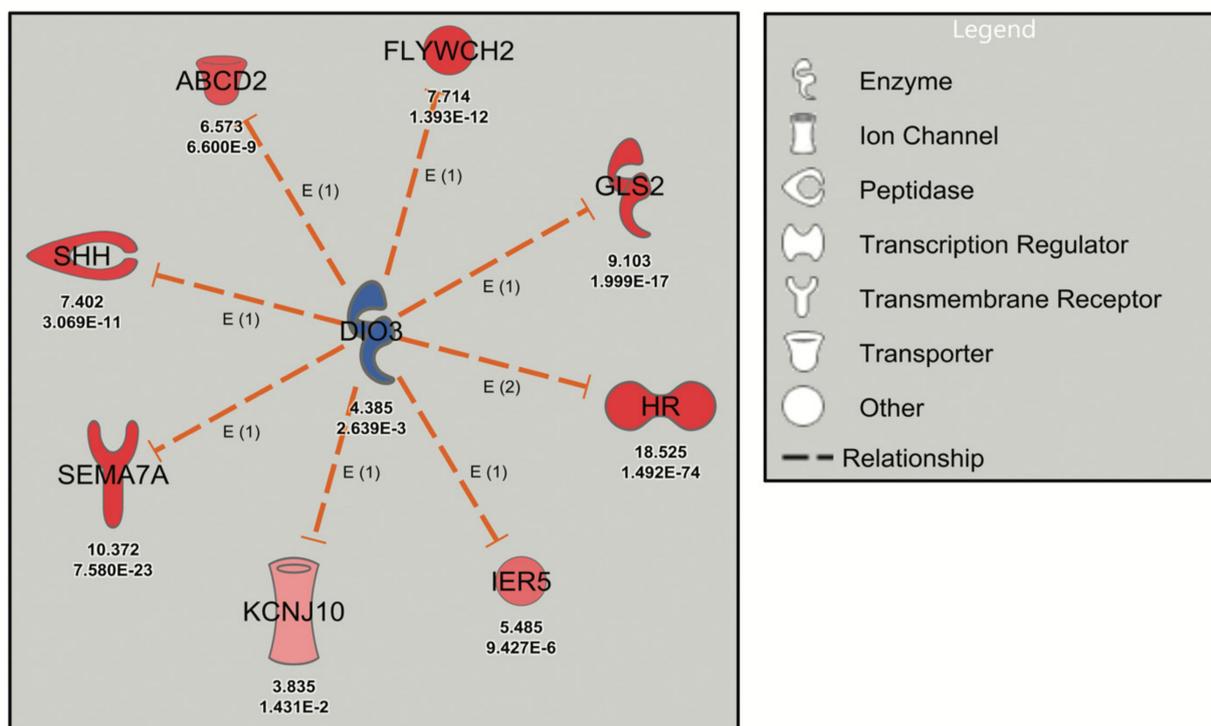


Figure 1. Robust URA network induced by levothyroxine treatment of mouse cerebrocortical cultures. Dashed orange lines signify activation (loss of inhibition) and red symbols signify upregulated genes with the Neuron Screen Z-score and adjusted p -value (p -adj) (<0.05) appearing directly below each symbol. The upstream regulator *DIO3* is blue, indicating downregulation based on the C4R drug screen data.

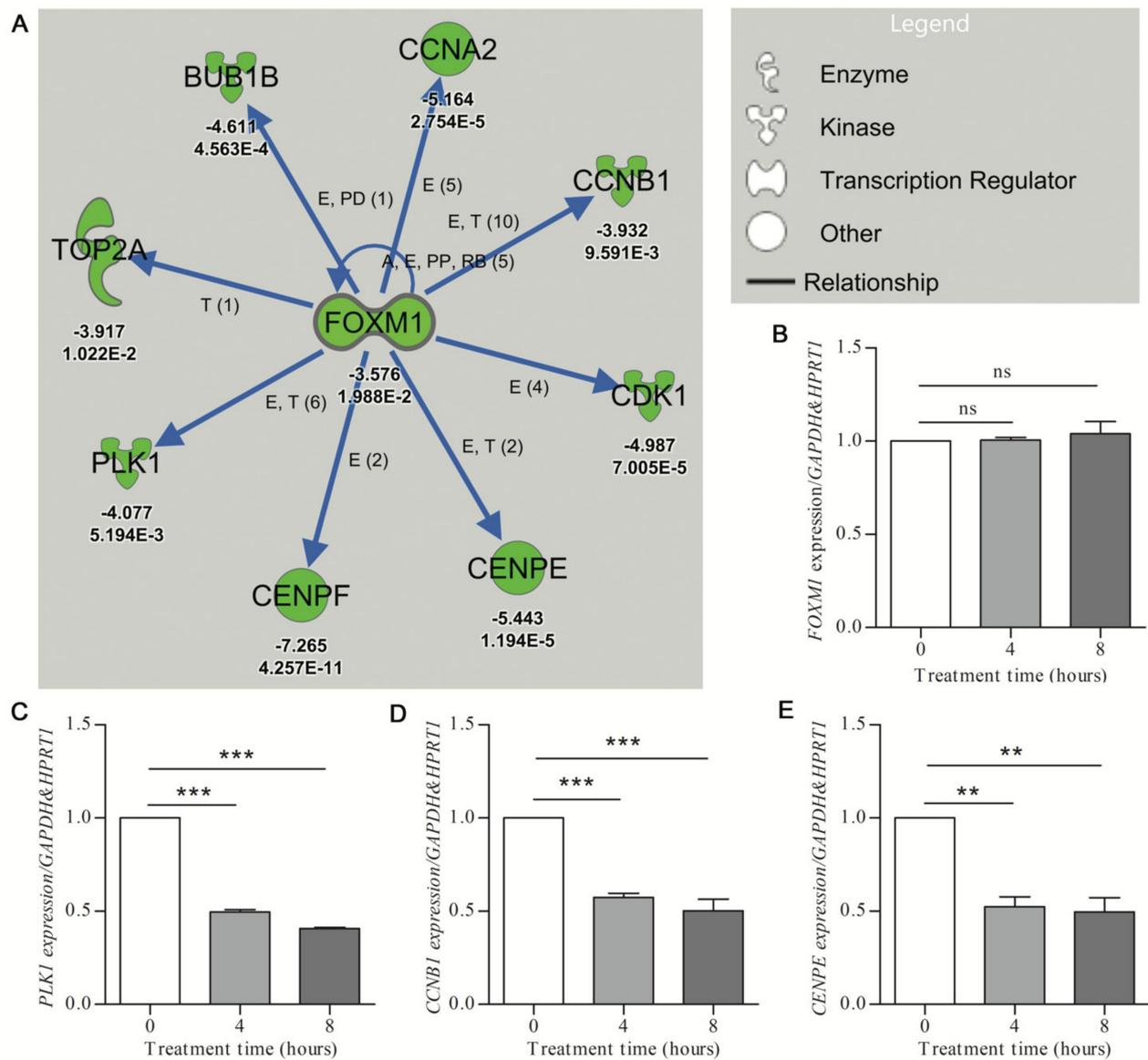


Figure 2. Validation of downregulated gene expression prioritized by IPA analysis. (A) The URA network shows downregulation (green symbols) of the putative upstream regulator (*FOXM1*) and the downstream molecules (with associated Z-score and *p*-value given under each gene, for *p* values, E-n symbolizes E⁻ⁿ) (Care4Rare Neuron Screen). Blue arrows indicate inhibition of the physiological activating relationship between *FOXM1* and related genes. Only genes for which *p*-adj < 0.05 are included. (B–E) U87 glioblastoma cells were treated with 250 μ M hydroxyurea for 0, 4, and 8 h. qRT-PCR (*n* = 3) was employed to determine target gene expression with geometric normalization against *GAPDH* and *HPRT1*. (B) Fork-head box M1 (*FOXM1*). (C) Polo-like kinase 1 (*PLK1*). (D) Cyclin B1 (*CCNB1*). (E) Centromere protein E (*CENPE*). Statistical significance was measured by one-way ANOVA (nonparametric) with Tukey post hoc analysis (** *p* < 0.01, *** *p* < 0.001, ns = nonsignificant).

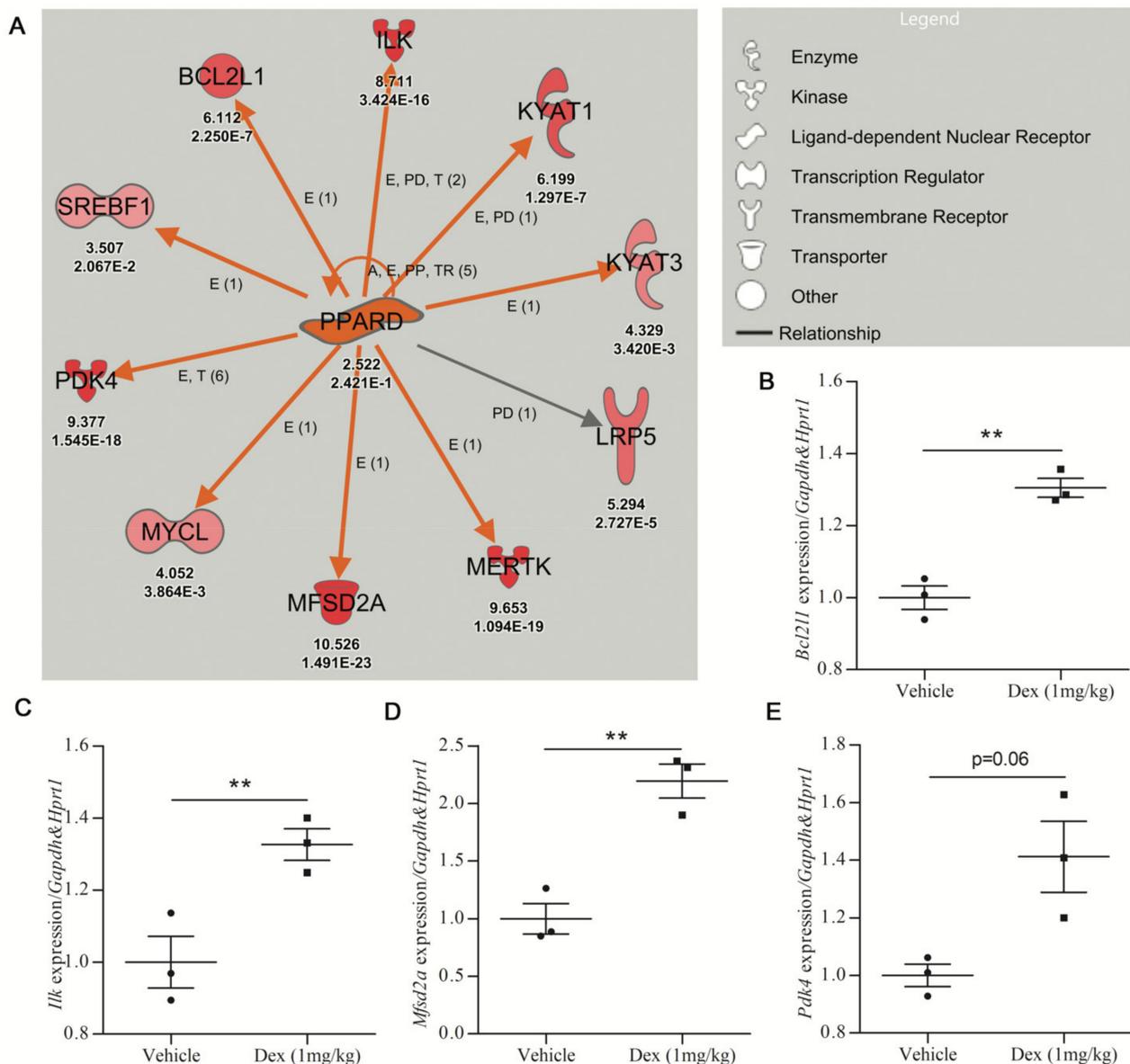


Figure 3. Validation efficiency for upregulated gene expression prioritized by IPA analysis. (A) The *PPARD* upstream regulator identified by URA analysis is upregulated in dexamethasone-treated mouse cerebrocortical cultures (Care4Rare Neuron Screen). Orange arrows signify an activating relationship with downstream genes and red symbols signify upregulated genes with the Neuron Screen Z-score and *p*-adj appearing directly below each symbol. Only genes for which *p*-adj < 0.05 are included. (B–E) Adult male C57BL6 mice were treated p.o. with vehicle or DEX (1 mg/kg) for 5 days, and qRT-PCR was employed to determine gene expression of 7 of the 10 targets of *PPARD* (*n* = 3). The four genes with the most robust upregulation are shown. (B) Bcl-2-like protein 1 (*Bcl2l1*). (C) Integrin-linked kinase (*Ilk*). (D) Major facilitator superfamily domain containing 2A (*Mfsd2a*). (E) Pyruvate dehydrogenase kinase 4 (*Pdk4*). For all hits, statistical significance was measured by Student's paired two-tailed *t*-test (** *p* < 0.01).

3. Discussion

Recent technologies have permitted the massively parallel interrogation of biologic systems generating many orders of magnitude of data points. The challenge is how to extract reliable information of biological utility from such massive sequencing and gene-array datasets. In this regard, although multiple biological replicates are necessary for traditional statistical calculations, it is not always feasible (high cost of RNA-seq) or possible (personalized medicine) to use biological replicates to truly validate a given result.

This problem is offset to some extent when a subset of data is analyzed, as the statistical strength of such profiles is greater than that achieved with the measurement of a single data point. This can be seen in rare disease drug repurposing studies utilizing genome-wide DE analyses that involve the orthogonal signature technique, pioneered by Lamb et al. (2006) [19,35,36]. In contrast, our lab has adopted a more targeted approach, mining DE datasets to identify drug-based modulation of single genes with potential therapeutic benefit for monogenic diseases [12,20–22]. However, the inherent variability of single-gene instances within system-wide datasets is borne out by in vitro and in vivo transcriptional validation rates of 20% or lower documented in these studies.

In particular, in the analysis of our recent RNA-seq drug screen using primary cerebrocortical cultures, the attempted validation of 32 DEGs taken from 60 rare neurogenetic disease-associated genes originally catalogued by Mears et al. (2017) was successful in only six cases, a validation rate of less than 20% [12,20].

One technique that has been used to solve the lack of biological replicates is hypergeometric distribution analysis, which has been proposed for individualized therapy development for glioblastoma multiforme [9]. In an alternative approach, we have in the present study explored the use of IPA-based URA to assign a reliability probability for single-gene instances found within system-wide RNA-seq datasets. The URA technique is well suited given the lack of replicates, as the addition of a minimum of four biologically related genes to serve as biological replicates of a gene of interest dramatically increases the efficiency of the individual gene validation. Importantly, our technique of reliably identifying DEGs does not depend on the degree of gene expression of the upstream regulator. For example, validation of the *FOXM1* network revealed a high rate of validation of downstream genes (e.g., *PLK1*, *CCNB1*), yet a lack of validation of the *FOXM1* gene itself (despite being downregulated in the RNA-seq data in response to HU, it was not downregulated in qRT-PCR validation). With regard to the *STAT4* network, gene expression results for *STAT4* were not even included in the original RNA-seq data (that were input into the IPA software). However, the network achieved statistical significance due to the number of transcriptional targets of *STAT4* that were upregulated by DEX.

Our robust networks include a minimum of five DEGs per URA network both for optimal statistical strength and given that roughly a fourth of protein-coding genes have been linked to a rare genetic disease [37,38] (with the general expectation that this proportion will increase with the ongoing delineation of novel ultrarare diseases). Each network would thus have a reasonable probability of including at least one rare disease gene; for example, rare disease-causing genes are seen in both the HU-responsive *FOXM1* network (*CENPE*, *CENPF*) [39,40] and the DEX-responsive *PPARD* network (*MERTK*, *MFSD2A*) [41,42]. Obviously, a robust network could comprise a different number of genes depending on the research question being posed.

Several other studies using URA-based analysis of networks to interpret transcriptomic data, while not quantifying the validation rate, have reported good reproducibility [43–46]. However, in contrast to the present report, the goal of these studies was to validate the upstream regulators, not to identify and prioritize DEGs for validation as we have done. Although the sample size was small, our network-directed prioritization of DEGs resulted in 100% validation of mouse cerebrocortical culture hits in human immortalized cultures (*FOXM1* DEGs). Moreover, a roughly 40% validation from tissue culture to whole animals was documented (*PPARD* DEGs). Collectively, the results suggest that for drug-conferred DEGs that are also constituents of a robust transcriptional network response, there is a greater likelihood of validation in the original system as well as in other in vitro and in vivo models, even extending to different species, thus representing true physiological relationships that transcend the boundaries of species and experimental conditions.

Finally, novel network analyses, such as weighted correlation network analysis (WGCNA), could be used to search for clusters of highly correlated genes [47]. WGCNA algorithms have been used in dyslipidemias [48], cancer [49,50], and autism spectrum disorder [51].

4. Materials and Methods

4.1. Animals and Treatments

All animal protocols were approved by the Animal Care and Veterinary Services (ACVS) and Ethics Board of the University of Ottawa. For in vivo drug studies, male C57BL6 mice (8 weeks old) were obtained from Charles River, housed in triplicate, and given food and water *ad libitum*. Mice were treated once daily with 1 mg/kg DEX (Sigma-Aldrich, Oakville, ON, Canada) by oral gavage for 5 days. Roughly four hours after the last dose, the mice were anesthetized by isoflurane (Sigma-Aldrich, ON, Canada) and euthanized by cervical dislocation. Cerebral cortices were then collected from each animal and flash-frozen in liquid nitrogen. Total RNA extraction from the cortex was performed by QIAzol lysis reagent according to the manufacturer's recommendations (Qiagen, Montreal, QC, Canada). The RNA was then purified over Qiagen RNeasy Mini spin columns (Qiagen, Montreal, QC, Canada) and frozen at -80°C .

4.2. Human Glioblastoma Cell Culture

Human U87 glioblastoma cells (ATCC) were cultured in DMEM high-glucose supplemented with 10% fetal calf serum (Life Technologies, Burlington, ON, Canada) and 2 mM L-glutamine (Life Technologies, ON, Canada). For the validation of *FOXM1* targets, U87 cells were plated in 10 cm dishes at 120,000 cells per dish in a volume of 10 mL. After an overnight settling period, U87 cells were treated for 0, 4, or 8 h with 0.25 mM HU (Sigma-Aldrich, ON, Canada). After treatment end-points, cells were washed with $1 \times$ phosphate-buffered saline (PBS) (Sigma-Aldrich, ON, Canada), trypsinized, and pelleted by centrifugation at $300 \times g$ for 5 min. Pellets were rinsed in 10 mL of $1 \times$ PBS and then stored at -8°C . RNA extraction was performed by Qiagen RNeasy Mini column-based extraction (Qiagen, Montreal, QC, Canada) and purification (Qiagen, Montreal, QC, Canada) according to the manufacturer's instructions.

4.3. Dataset Used and Ingenuity Pathway Analysis

Primary murine neuronal cultures were incubated with 219 drugs mostly derived from the Screen-Well Food and Drug Administration approved drug library v2 (Enzo Life Sciences). Drugs were selected for blood-brain barrier penetrance, comparative safety consistent with long-term use, and oral bioavailability representing 80 therapeutic classes, using concentrations approximating those attained in patients (more details are available at <http://bigbear.med.uottawa.ca:1000/>, accessed on 7 June 2021). cDNA libraries, constructed from the poly-A fraction of total RNA extracted from drug- and control-treated cultures, were sequenced and analyzed, establishing a transcriptome-wide DE dataset representing all 227 sequenced libraries. For each sample, gene expression for 14,000 genes with $\text{cpm} > 1$ was normalized against the average of all conditions. Given that the screen was performed as a single replicate, the expression of a given gene across all samples served as the background to represent unaffected expression distributions. Robust z-scores and *p*-adj for each potential drug-gene interaction were generated, with DE being defined as *p*-adj < 0.05 . The transcriptome-wide DE datasets of 50 of the most transcriptionally active drugs from the Neuron Screen were entered into the IPA software. Default settings were used for the analysis with a statistical limit of z-score > 3 for all data points. The two metrics of differential expression that were entered were "z-score" and *p*-adj. Each dataset was individually analyzed by IPA, and the URA function was used to identify upstream regulator pathways that were "activated" (z-score > 2) or "inhibited" (z-score < -2). IPA also calculated the *p*-value of overlap that was considered significant if *p* < 0.05 . Upstream regulator networks chosen for validation were curated by removing genes that had *p*-adj > 0.05 . Networks identified as "robust" contained ≥ 5 statistically significant gene-targets that were modulated in a direction consistent with the IPA prediction.

4.4. Quantitative qRT-PCR

DEGs from robust URA networks were investigated by qPCR in human U87 cells and C57BL6 mouse cortex. Gene-specific primers were designed for each gene of interest, and *Gapdh* and *Hprt1* were used as housekeeping genes (Table S1). Primers for qPCR were designed using NCBI Primer-BLAST in accordance with the MIQE guidelines. Reverse transcription of purified RNA using the Bio-Rad iScript advanced RT kit (Bio-rad, Mississauga, ON, Canada) on a T100 Thermal Cycler (Bio-rad, Mississauga, ON, Canada) and qPCR using the iQ SYBR Green Master Mix (Bio-Rad, Mississauga, ON, Canada) on a CFX96 Touch Real-Time PCR Detection System (Bio-rad, Mississauga, ON, Canada) were performed as previously described [12]. Bio-Rad CFX software (Bio-rad, Mississauga, ON, Canada) enabled geometric mean normalization of target gene expression to *Gapdh* and *Hprt1*.

4.5. Statistical Analysis

Statistical tests were performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA). All data are expressed as mean and standard deviation unless otherwise indicated. Statistical significance was measured by one-way ANOVA with Tukey post hoc analysis (for in vitro validation) and by Student's paired two-tailed *t*-test (for in vivo validation).

5. Conclusions

In conclusion, we present a method to prioritize the investigation of DEGs identified in transcriptome-wide studies in neurogenetic disease. We believe our approach, combining the statistical rigour of gene enrichment and pathway analysis while utilizing accessible bioinformatics tools, may serve as a useful low-cost rapid filter to prioritize single DEGs worthy of further analysis. Although IPA was used for gene prioritization in our study, it is likely that other roughly equivalent tools (iPathwayGuide, PathVisio, GeneGo) would allow for similar gene prioritization. As an ever-expanding number of organisms are being added to the list of sequenced genomes, pathway tools such as IPA may be developed that can integrate gene expression results of nonmodel organisms. The URA-directed prioritization technique could then be broadly applicable in the context of transcriptomic data with impacts in fields as diverse as personalized drug discovery, the effect of ocean acidification on marine species, or elucidating mechanisms of antibiotic resistance.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22126295/s1>, Figure S1: Validation of the effect of HU on four *FOXM1* downstream genes. U87 glioblastoma cells were treated for 0, 4, and 8 h, and qRT-PCR was employed to determine gene expression of (A) cyclin-dependent kinase 1 (*CDK1*), (B) centromere protein F (*CENPF*), (C) BUB1 mitotic checkpoint serine/threonine kinase (*BUB1*), and (D) cyclin A2 (*CCNA2*). Statistical significance was measured by one-way ANOVA (nonparametric) with Tukey post hoc analysis (** $p < 0.01$, *** $p < 0.001$, ns = not significant), Figure S2: Validation of the effect of dexamethasone on three *PPARD* targets in vivo. Male C57BL6 mice were treated orally for 5 days with vehicle ($n = 3$) or DEX (1 mg/kg) ($n = 3$). Gene expression of *PPARD* targets was determined by qRT-PCR. (A) FKBP prolyl isomerase 5 (*Fkbp5*), (B) kynurenine aminotransferase 3 (*Kyat3*), (C) C-mer proto-oncogene tyrosine kinase (*Mertk*), (D) LDL receptor-related protein 5 (*Lrp5*), Table S1: Mouse and human primers.

Author Contributions: J.H., S.S., and A.M. designed research; J.H. performed research; J.H., S.S., F.F., A.M., and J.P.-D. wrote the manuscript. All authors assisted in editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was performed under the Care4Rare Canada Consortium funded by Genome Canada, the Canadian Institutes of Health Research, the Ontario Genomics Institute (OGI-049), Ontario Research Fund, Genome Quebec, Genome British Columbia, and CHEO Foundation (3 July 2014).

Institutional Review Board Statement: Protocol number CHEO-1810, approved on 4 November 2014. All animal protocols were approved by the Animal Care and Veterinary Services (ACVS) and Ethics

Board of the University of Ottawa. The ACVS is certified by the Canadian Council on Animal Care, licensed under the Ontario Animals for Research Act, and adheres to all Canadian, provincial, and international regulations and standards pertaining to laboratory research with animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Care4Rare Neuron Screen database that was used to identify the transcriptional drug–gene hits can be accessed from <http://bigbear.med.uottawa.ca:1000>, accessed on 7 June 2021. The original fastq files generated and analyzed to create the database are available from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) (accession number GSE110256, accessed on 7 June 2021).

Acknowledgments: The authors would like to thank Kate Daniels for her technical help with mouse brain dissections. Julio Plaza-Diaz is part of the “UGR Plan Propio de Investigación 2016” and the “Excellence actions: Unit of Excellence on Exercise and Health (UCEES), University of Granada”. Julio Plaza-Diaz is supported by a grant awarded to postdoctoral researchers at foreign universities and research centers from the “Fundación Ramón Areces”, Madrid, Spain.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wakap, S.N.; Lambert, D.M.; Olry, A.; Rodwell, C.; Gueydan, C.; Lanneau, V.; Murphy, D.; Le Cam, Y.; Rath, A. Estimating cumulative point prevalence of rare diseases: Analysis of the Orphanet database. *Eur. J. Hum. Genet.* **2020**, *28*, 165–173. [[CrossRef](#)] [[PubMed](#)]
2. Gonzaludo, N.; Belmont, J.W.; Gainullin, V.G.; Taft, R.J. Estimating the burden and economic impact of pediatric genetic disease. *Genet. Med.* **2019**, *21*, 1781–1789. [[CrossRef](#)] [[PubMed](#)]
3. Walker, C.E.; Mahede, T.; Davis, G.; Miller, L.J.; Girschik, J.; Brameld, K.; Sun, W.; Rath, A.; Aymé, S.; Zubrick, S.R. The collective impact of rare diseases in Western Australia: An estimate using a population-based cohort. *Genet. Med.* **2017**, *19*, 546–552. [[CrossRef](#)] [[PubMed](#)]
4. Mazzucato, M.; Dalla Pozza, L.V.; Manea, S.; Minichiello, C.; Facchin, P. A population-based registry as a source of health indicators for rare diseases: The ten-year experience of the Veneto Region’s rare diseases registry. *Orphanet J. Rare Dis.* **2014**, *9*, 37. [[CrossRef](#)] [[PubMed](#)]
5. Chen, Z.; Farrell, A.P.; Matala, A.; Narum, S.R. Mechanisms of thermal adaptation and evolutionary potential of conspecific populations to changing environments. *Mol. Ecol.* **2018**, *27*, 659–674. [[CrossRef](#)]
6. Goncalves, P.; Jones, D.B.; Thompson, E.L.; Parker, L.M.; Ross, P.M.; Raftos, D.A. Transcriptomic profiling of adaptive responses to ocean acidification. *Mol. Ecol.* **2017**, *26*, 5974–5988. [[CrossRef](#)]
7. Suzuki, S.; Horinouchi, T.; Furusawa, C. Prediction of antibiotic resistance by gene expression profiles. *Nat. Commun.* **2014**, *5*, 5792. [[CrossRef](#)]
8. Li, L.; Dai, X.; Wang, Y.; Yang, Y.; Zhao, X.; Wang, L.; Zeng, M. RNA-seq-based analysis of drug-resistant *Salmonella enterica* serovar Typhimurium selected in vivo and in vitro. *PLoS ONE* **2017**, *12*, e0175234. [[CrossRef](#)]
9. Stathias, V.; Pastori, C.; Griffin, T.Z.; Komotar, R.; Clarke, J.; Zhang, M.; Ayad, N.G. Identifying glioblastoma gene networks based on hypergeometric test analysis. *PLoS ONE* **2014**, *9*, e115842. [[CrossRef](#)] [[PubMed](#)]
10. Gagan, J.; Van Allen, E.M. Next-generation sequencing to guide cancer therapy. *Genome Med.* **2015**, *7*, 80. [[CrossRef](#)]
11. Kremer, L.S.; Bader, D.M.; Mertes, C.; Kopajtich, R.; Pichler, G.; Iuso, A.; Haack, T.B.; Graf, E.; Schwarzmayr, T.; Terrile, C. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **2017**, *8*, 15824. [[CrossRef](#)]
12. Hadwen, J.; Schock, S.; Mears, A.; Yang, R.; Charron, P.; Zhang, L.; Xi, H.S.; MacKenzie, A. Transcriptomic RNAseq drug screen in cerebrocortical cultures: Toward novel neurogenetic disease therapies. *Hum. Mol. Genet.* **2018**, *27*, 3206–3217. [[CrossRef](#)] [[PubMed](#)]
13. Poirier, M.B.; Hadwen, J.; MacKenzie, A. Pharmacologic normalization of pathogenic dosage underlying genetic diseases: An overview of the literature and path forward. *Emerg. Top. Life Sci.* **2019**, *3*, 53–62. [[PubMed](#)]
14. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [[CrossRef](#)] [[PubMed](#)]
15. Williams, C.R.; Baccarella, A.; Parrish, J.Z.; Kim, C.C. Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. *BMC Bioinform.* **2017**, *18*, 38. [[CrossRef](#)]
16. Noble, W.S. How does multiple testing correction work? *Nat. Biotechnol.* **2009**, *27*, 1135–1137. [[CrossRef](#)]
17. Li, J.; Witten, D.M.; Johnstone, I.M.; Tibshirani, R. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* **2012**, *13*, 523–538. [[CrossRef](#)]
18. Kunkel, S.D.; Suneja, M.; Ebert, S.M.; Bongers, K.S.; Fox, D.K.; Malmberg, S.E.; Alipour, F.; Shields, R.K.; Adams, C.M. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. *Cell Metab.* **2011**, *13*, 627–638. [[CrossRef](#)]

19. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.-P.; Subramanian, A.; Ross, K.N. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935. [[CrossRef](#)] [[PubMed](#)]
20. Mears, A.; Schock, S.; Hadwen, J.; Putos, S.; Dymont, D.; Boycott, K.; MacKenzie, A. Mining the transcriptome for rare disease therapies: A comparison of the efficiencies of two data mining approaches and a targeted cell-based drug screen. *NPJ Genom. Med.* **2017**, *2*, 14. [[CrossRef](#)] [[PubMed](#)]
21. Witherspoon, L.; O'Reilly, S.; Hadwen, J.; Tasnim, N.; MacKenzie, A.; Farooq, F. Sodium channel inhibitors reduce DMPK mRNA and protein. *Clin. Transl. Sci.* **2015**, *8*, 298–304. [[CrossRef](#)]
22. Farooq, F.; Abadía-Molina, F.; MacKenzie, D.; Hadwen, J.; Shamim, F.; O'Reilly, S.; Holcik, M.; MacKenzie, A. Celecoxib increases SMN and survival in a severe spinal muscular atrophy mouse model via p38 pathway activation. *Hum. Mol. Genet.* **2013**, *22*, 3415–3424. [[CrossRef](#)] [[PubMed](#)]
23. Hadwen, J.; MacKenzie, D.; Shamim, F.; Mongeon, K.; Holcik, M.; MacKenzie, A.; Farooq, F. VPAC2 receptor agonist BAY 55-9837 increases SMN protein levels and moderates disease phenotype in severe spinal muscular atrophy mouse models. *Orphanet J. Rare Dis.* **2014**, *9*, 4. [[CrossRef](#)]
24. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2009**, *37*, 1–13. [[CrossRef](#)]
25. Krämer, A.; Green, J.; Pollard, J., Jr.; Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **2014**, *30*, 523–530. [[CrossRef](#)]
26. Ahsan, S.; Drăghici, S. Identifying significantly impacted pathways and putative mechanisms with iPathwayGuide. *Curr. Protoc. Bioinform.* **2017**, *57*, 7–15. [[CrossRef](#)] [[PubMed](#)]
27. Van Iersel, M.P.; Kelder, T.; Pico, A.R.; Hanspers, K.; Coort, S.; Conklin, B.R.; Evelo, C. Presenting and exploring biological pathways with PathVisio. *BMC Bioinform.* **2008**, *9*, 399. [[CrossRef](#)]
28. Lv, Y.-W.; Wang, J.; Sun, L.; Zhang, J.-M.; Cao, L.; Ding, Y.-Y.; Chen, Y.; Dou, J.-J.; Huang, J.; Tang, Y.-F. Understanding the pathogenesis of Kawasaki disease by network and pathway analysis. *Comput. Math. Methods Med.* **2013**, *2013*, 989307. [[CrossRef](#)]
29. Abatangelo, L.; Maglietta, R.; Distaso, A.; D'Addabbo, A.; Creanza, T.M.; Mukherjee, S.; Ancona, N. Comparative study of gene set enrichment methods. *BMC Bioinform.* **2009**, *10*, 275. [[CrossRef](#)] [[PubMed](#)]
30. Huang, S.A. Physiology and pathophysiology of type 3 deiodinase in humans. *Thyroid* **2005**, *15*, 875–881. [[CrossRef](#)]
31. Gil-Ibañez, P.; García-García, F.; Dopazo, J.; Bernal, J.; Morte, B. Global transcriptome analysis of primary cerebrocortical cells: Identification of genes regulated by triiodothyronine in specific cell types. *Cereb. Cortex* **2017**, *27*, 706–717. [[CrossRef](#)] [[PubMed](#)]
32. Wierstra, I. The transcription factor FOXM1 (Forkhead box M1): Proliferation-specific expression, transcription factor function, target genes, mouse models, and normal biological roles. *Adv. Cancer Res.* **2013**, *118*, 97–398.
33. Vermeer, H.; Hendriks-Stegeman, B.I.; van der Burg, B.; van Buul-Offers, S.C.; Jansen, M. Glucocorticoid-induced increase in lymphocytic FKBP51 messenger ribonucleic acid expression: A potential marker for glucocorticoid sensitivity, potency, and bioavailability. *J. Clin. Endocrinol. Metab.* **2003**, *88*, 277–284. [[CrossRef](#)] [[PubMed](#)]
34. Frahm, K.A.; Peffer, M.E.; Zhang, J.Y.; Luthra, S.; Chakka, A.B.; Couger, M.B.; Chandran, U.R.; Monaghan, A.P.; DeFranco, D.B. Research resource: The dexamethasone transcriptome in hypothalamic embryonic neural stem cells. *Mol. Endocrinol.* **2016**, *30*, 144–154. [[CrossRef](#)] [[PubMed](#)]
35. Stuart, J.M.; Segal, E.; Koller, D.; Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **2003**, *302*, 249–255. [[CrossRef](#)]
36. Delahaye-Duriez, A.; Srivastava, P.; Shkura, K.; Langley, S.R.; Laaniste, L.; Moreno-Moral, A.; Danis, B.; Mazzuferi, M.; Foerch, P.; Gazina, E.V. Rare and common epilepsies converge on a shared gene regulatory network providing opportunities for novel antiepileptic drug discovery. *Genome Biol.* **2016**, *17*, 245. [[CrossRef](#)]
37. Boycott, K.M.; Vanstone, M.R.; Bulman, D.E.; MacKenzie, A.E. Rare-disease genetics in the era of next-generation sequencing: Discovery to translation. *Nat. Rev. Genet.* **2013**, *14*, 681–691. [[CrossRef](#)]
38. Boycott, K.M.; Rath, A.; Chong, J.X.; Hartley, T.; Alkuraya, F.S.; Baynam, G.; Brookes, A.J.; Brudno, M.; Carracedo, A.; den Dunnen, J.T. International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* **2017**, *100*, 695–705. [[CrossRef](#)] [[PubMed](#)]
39. Mirzaa, G.M.; Vitre, B.; Carpenter, G.; Abramowicz, I.; Gleeson, J.G.; Paciorkowski, A.R.; Cleveland, D.W.; Dobyns, W.B.; O'Driscoll, M. Mutations in CENPE define a novel kinetochore-centromeric mechanism for microcephalic primordial dwarfism. *Hum. Genet.* **2014**, *133*, 1023–1039. [[CrossRef](#)]
40. Waters, A.M.; Asfahani, R.; Carroll, P.; Bicknell, L.; Lescai, F.; Bright, A.; Chanudet, E.; Brooks, A.; Christou-Savina, S.; Osman, G. The kinetochore protein, CENPF, is mutated in human ciliopathy and microcephaly phenotypes. *J. Med. Genet.* **2015**, *52*, 147–156. [[CrossRef](#)]
41. Gal, A.; Li, Y.; Thompson, D.A.; Weir, J.; Orth, U.; Jacobson, S.G.; Apfelstedt-Sylla, E.; Vollrath, D. Mutations in MERTK, the human orthologue of the RCS rat retinal dystrophy gene, cause retinitis pigmentosa. *Nat. Genet.* **2000**, *26*, 270–271. [[CrossRef](#)] [[PubMed](#)]
42. Alakbarzade, V.; Hameed, A.; Quek, D.Q.; Chioza, B.A.; Baple, E.L.; Cazenave-Gassiot, A.; Nguyen, L.N.; Wenk, M.R.; Ahmad, A.Q.; Sreekantan-Nair, A. A partially inactivating mutation in the sodium-dependent lysophosphatidylcholine transporter MFSD2A causes a non-lethal microcephaly syndrome. *Nat. Genet.* **2015**, *47*, 814–817. [[CrossRef](#)]

43. Sugimura, S.; Kobayashi, N.; Okae, H.; Yamanouchi, T.; Matsuda, H.; Kojima, T.; Yajima, A.; Hashiyada, Y.; Kaneda, M.; Sato, K. Transcriptomic signature of the follicular somatic compartment surrounding an oocyte with high developmental competence. *Sci. Rep.* **2017**, *7*, 6815. [[CrossRef](#)]
44. Oghumu, S.; Nori, U.; Bracewell, A.; Zhang, J.; Bott, C.; Nadasdy, G.M.; Brodsky, S.V.; Pelletier, R.; Satoskar, A.R.; Nadasdy, T. Differential gene expression pattern in biopsies with renal allograft pyelonephritis and allograft rejection. *Clin. Transplant.* **2016**, *30*, 1115–1133. [[CrossRef](#)]
45. Hurem, S.; Martín, L.M.; Brede, D.A.; Skjerve, E.; Nourizadeh-Lillabadi, R.; Lind, O.C.; Christensen, T.; Berg, V.; Teien, H.-C.; Salbu, B. Dose-dependent effects of gamma radiation on the early zebrafish development and gene expression. *PLoS ONE* **2017**, *12*, e0179259. [[CrossRef](#)]
46. Boucher, J.G.; Gagné, R.; Rowan-Carroll, A.; Boudreau, A.; Yauk, C.L.; Atlas, E. Bisphenol A and bisphenol S induce distinct transcriptional profiles in differentiating human primary preadipocytes. *PLoS ONE* **2016**, *11*, e0163318. [[CrossRef](#)] [[PubMed](#)]
47. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)] [[PubMed](#)]
48. Benincasa, G.; de Candia, P.; Costa, D.; Faenza, M.; Mansueto, G.; Ambrosio, G.; Napoli, C. Network Medicine Approach in Prevention and Personalized Treatment of Dyslipidemias. *Lipids* **2020**. [[CrossRef](#)]
49. Luo, R.; Song, J.; Zhang, W.; Ran, L. Identification of MFI2-AS1, a Novel Pivotal lncRNA for Prognosis of Stage III/IV Colorectal Cancer. *Dig. Dis. Sci.* **2020**, *65*, 3538–3550. [[CrossRef](#)]
50. Weng, W.; Zhang, Z.; Huang, W.; Xu, X.; Wu, B.; Ye, T.; Shan, Y.; Shi, K.; Lin, Z. Identification of a competing endogenous RNA network associated with prognosis of pancreatic adenocarcinoma. *Cancer Cell Int.* **2020**, *20*, 231. [[CrossRef](#)]
51. Kopp, N.; Climer, S.; Dougherty, J.D. Moving from capstones toward cornerstones: Successes and challenges in applying systems biology to identify mechanisms of autism spectrum disorders. *Front. Genet.* **2015**, *6*, 301. [[CrossRef](#)] [[PubMed](#)]