frontiers in Pharmacology

ORIGINAL RESEARCH
published: 26 June 2020
doi: 10.3389/fphar.2020.00927

Drug Repositioning for Noonan and LEOPARD Syndromes by Integrating Transcriptomics With a Structure-Based Approach

Liyuan Zhu[1], Ruth Roberts[1,2,3], Ruili Huang[4], Jinghua Zhao[4], Menghang Xia[4], Brian Delavan[1,5], Mike Mikailov[6], Weida Tong[1] and Zhichao Liu[1]**

[1] National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, United States, [2] Department of Drug Safety, ApconiX, Alderley Edge, United Kingdom, [3] Department of Biosciences, University of Birmingham, Birmingham, United Kingdom, [4] National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, United States, [5] Joint Bioinformatics Graduate Program, University of Arkansas at Little Rock and University of Arkansas for Medical Sciences, Little Rock, AR, United States, [6] Office of Science and Engineering Labs, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, United States

OPEN ACCESS

Edited by:
Thomas Hartung,
Bloomberg School of Public Health,
Johns Hopkins University,
United States

Reviewed by:
Paolo Versacci,
Sapienza University of Rome, Italy
Nicole C. Kleinstreuer,
National Institute of Environmental
Health Sciences (NIEHS),
United States

***Correspondence:**
Weida Tong
weida.tong@fda.hhs.gov
Zhichao Liu
zhichao.liu@fda.hhs.gov

Specialty section:
This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 30 September 2019
Accepted: 08 June 2020
Published: 26 June 2020

Citation:
Zhu L, Roberts R, Huang R, Zhao J, Xia M, Delavan B, Mikailov M, Tong W and Liu Z (2020) Drug Repositioning for Noonan and LEOPARD Syndromes by Integrating Transcriptomics With a Structure-Based Approach. Front. Pharmacol. 11:927. doi: 10.3389/fphar.2020.00927

Noonan and LEOPARD syndromes (NS and LS) belong to a group of related disorders called RASopathies characterized by abnormalities of multiple organs and systems including hypertrophic cardiomyopathy and dysmorphic facial features. There are no approved drugs for these two rare diseases, but it is known that a missense mutation in PTPN11 genes is associated with approximately 50% and 70% of NS and LS cases, respectively. In this study, we implemented a hybrid computational drug repositioning framework by integrating transcriptomic and structure-based approaches to explore potential treatment options for NS and LS. Specifically, disease signatures were derived from the transcriptomic profiles of human induced pluripotent stem cells (iPSCs) from NS and LS patients and reverse correlated to drug transcriptomic signatures from CMap and L1000 projects on the basis that if disease and drug transcriptomic signatures are reversely correlated, the drug has the potential to treat that disease. The compounds that were ranked top based on their transcriptomic profiles were docked to mutated and wild-type 3D structures of PTPN11 by an adjusted Induced Fit Docking (IFD) protocol. In addition, we prioritized repositioned candidates for NS and LS by a consensus ranking strategy. Network analysis and phenotypic anchoring of the transcriptomic data could discriminate the two diseases at the molecular level. Furthermore, the adjusted IFD protocol was able to recapitulate the binding specificity of potential drug candidates to mutated 3D structures, revealing the relevant amino acids. Importantly, a list of potential drug candidates for repositioning was identified including 61 for NS and 43 for LS and was further verified from literature reports and on-going clinical trials. Altogether, this hybrid computational drug repositioning approach has highlighted a number of drug candidates for NS and LS and could be applied to identifying drug candidates for other diseases as well.

Keywords: transcriptomic profiles, molecular docking, induced pluripotent stem cell, drug repositioning, rare diseases

INTRODUCTION

Noonan syndrome (NS) and LEOPARD syndrome [LS, also called Noonan syndrome with multiple lentigines (NSML)] are autosomal dominant disorders with overlapping phenotypic characteristics including hypertrophic cardiomyopathy, short stature, pectus deformity, and dysmorphic facial features (Tartaglia and Gelb, 2005). NS has an estimated prevalence of 1 in 1000~2500. The exact prevalence of LS is unknown and approximately 200 cases have been reported globally (https://ghr.nlm.nih.gov/condition/noonan-syndrome-with-multiple-lentigines#statistics) (Sarkozy et al., 2008). The missense mutation in PTPN11 genes, resulting in a gain-of-function of the non-receptor protein tyrosine phosphatase SHP-2 protein causes approximately 50% of NS cases. Causative mutations in other genes (e.g., SOS1, RIT1, and KRAS) involved in RAS-MAPK pathways have also been identified in a small portion of patients with NS (Schubbert et al., 2006; Pandit et al., 2007; Kouz et al., 2016). Mutations in PTPN11 also causes approximately 70% of LS. Considerable efforts have been made to uncover the molecular basis of the two diseases (Tartaglia et al., 2010). The current management for LS and NS mainly relies on treating the phenotypic presentation of a specific organ or system based on the individual patient (Romano et al., 2010). No drugs have been approved for treating NS or LS by the U.S. Food and Drug Administration (FDA).

Drug repositioning is a well-established approach to providing quicker, safer, and more affordable drugs to fulfill the unmet medical need for rare disease therapies (Ekins et al., 2011; Liu et al., 2016; Delavan et al., 2018). Advances in genomics and bioengineering provide unprecedented opportunities to uncover the underlying mechanisms of diseases where there may be accumulated data profiles that represent the different aspects of biological complexity. These data profiles also enable *in silico* drug repositioning as a parallel approach to exploring potential opportunities for rare disease treatment (Butte et al., 2015; Jia et al., 2018). To date, approximately 1,000 crystal structures for rare disease-associated proteins have been made available *via* The Research Collaboration for Structural Bioinformatics Protein Data Bank (RCSB PDB) (Yang et al., 2018), stimulating research in molecular docking-based drug repositioning. For example, Govindaraj et al. (2018) carried out molecular docking-based virtual screening with optimized drug-binding sites generated by the *e*MatchSite to systematically assess opportunities to repurpose approved drugs for 980 rare diseases.

Novel high throughput screening platforms such as LINCS1000 (Subramanian et al., 2017) and TempO-Seq™ (Liu et al., 2019), combined with the falling cost of next-generation sequencing (NGS), have resulted in transcriptomics data of more than 30,000 compounds being publicly available. Some initial efforts have been made to apply these transcriptomics data sets to seek repositioning candidates for rare diseases. For example, Dudley et al. (2011) comprehensively compared disease gene signatures to drug transcriptomic signatures from CMap to suggest the potential reuse of the anticonvulsant topiramate for inflammatory bowel disease (IBD) treatment. Also, some other approaches such as information retrieval or text mining (Chen and Altman, 2017), miRNA transcription factor feed-forward loops (Liu et al., 2014), and high throughput or high content screening assays (Bellomo et al., 2017) have also been applied to drug repositioning of rare diseases.

Due to our incomplete knowledge of rare diseases, the selection of "fit-for-purpose" *in silico* drug repositioning approaches varies among different rare diseases based on data availability (Delavan et al., 2018). Progress in structural chemistry makes mutated protein structures of NS and LS such as PTPN11 available, enabling the implementation of molecular docking approaches to look for potential repositioning candidates. Furthermore, human induced pluripotent stem cells (iPSCs) derived directly from patient cells hold great promise in mimicking the complex pathogenesis of diseases. Some transcriptomic profiles of NS and LS have been generated using human iPSCs, which provides further information in support of *in silico* drug repositioning. Transcriptomics-based drug repositioning can produce a comprehensive picture of gene perturbation by experimental compounds to assist in identifying different disease-associated biological processes. However, this approach cannot point directly to a therapeutic target for further investigation. Molecular docking aims to rank the binding affinity of tested molecules on the potential therapeutic target, offering a complementary strategy for transcriptomics-based drug repositioning. However, to the best of our knowledge, there are as yet no approaches described that integrate transcriptomics and molecular docking to explore treatment opportunities for NS and LS.

In this study, we implemented a hybrid approach that combines transcriptomics with molecular docking to explore potential repositioned candidates the treatment of NS and LS. First, transcriptomic profiles generated in human iPSC from NS and LS patients with PTPN11 mutations were employed to generate diseases signatures and were further compared using functional and network analysis. Second, the repositioned candidates for NS and LS were enriched by reverse correlation with drug signatures derived from LINCS1000 and CMap data sets. Third, the enriched repositioned candidates were further prioritized using molecular docking approaches. Finally, the promising repositioned candidates for NS and LS were verified from on-going clinical trials and literature reports.

RESULTS

Transcriptomic Signatures of Noonan and LEOPARD Syndrome

An outline of the study is illustrated in **Figure 1**, and a summary for transcriptomic data sets of Noonan syndrome (NS), LEOPARD syndrome (LS), and hypertrophic cardiomyopathy (HCM) are given in **Table 1**. Initially, the NS and LS transcriptomic signatures were generated. We used two microarray data sets generated from human iPSC lines for NS (e.g., GEO accession number: GSE54538) and LS (e.g., GEO accession number: GSE20473), respectively. The NS and LS

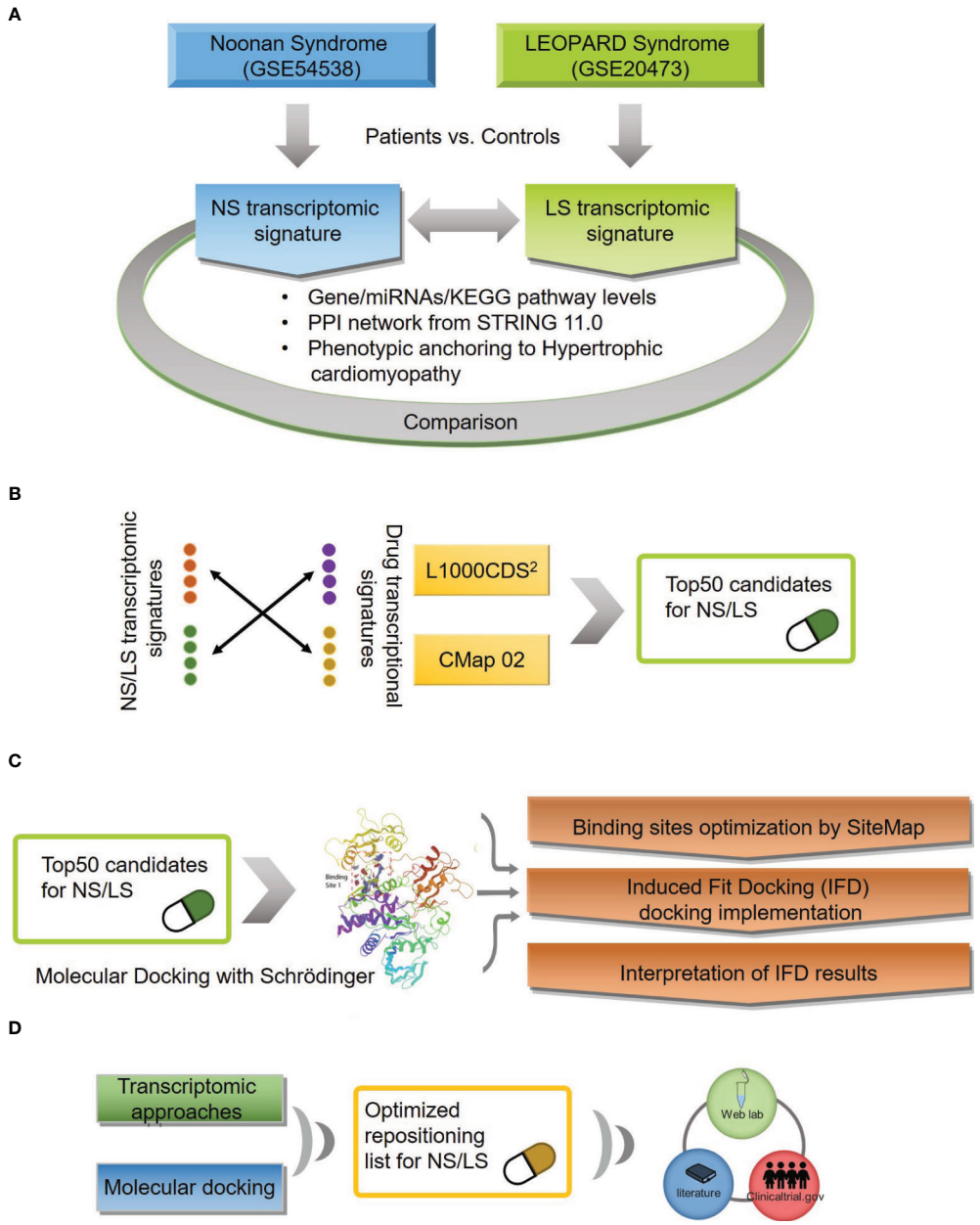

FIGURE 1 | An outline of workflow **(A)** Transcriptomic signatures of Noonan and LEOPARD syndrome. **(B)** Repositioning candidates enriched by transcriptomic approaches. **(C)** Molecular docking for repositioning candidates enriched by transcriptomic approaches. **(D)** Optimization of repositioning candidates by a hybrid strategy.

transcriptomic signatures were calculated using patient samples versus their matched controls. The top 250 genes based on ranked fold change values were extracted as NS and LS transcriptomic signatures (see **Supplementary Table S1**).

NS and LS have much in common regarding their phenotypic characters and etiology. To investigate whether the transcriptomic signatures could be used to distinguish between NS and LS, we compared NS and LS signatures at the gene and pathway levels. There was very little overlap (less than 0.1%) between NS and LS transcriptomic signatures, indicating the power of transcriptomic profiles to differentiate between NS from LS (**Figure 2A**). Two common up-regulated genes (*LGALS1* and *PHLDA1*) and six common down-regulated genes (*PRKCB*, *ZIC3*, *PGAM1*, *PMEL*, *FRAT2*, and *MT1G*) were found in the transcriptomic signatures of NS and LS. An analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways highlighted two over-represented KEGG for NS included Ribosome (KEGG id: hsa03010) and Biosynthesis of antibiotics (KEGG id: hsa01130). Meanwhile, a

TABLE 1 | Transcriptomic data information for Noonan syndrome, LEOPARD syndrome, and hypertrophic cardiomyopathy (HCM).

GEO accession numbers	Tissue type	Platform	Sample information
Noonan syndrome (NS)			
GSE54538	Human iPSC line	Illumina HumanHT-12 V4.0 expression beadchip	• 2 patient samples: 1 NS patient with E76D mutation in PTPN11; 2 biological duplicates from human iPSC line • 4 control samples: 1 human embryonic stem cell sample, and 3 induced pluripotent stem cell samples from healthy people
LEOPARD syndrome (LS)			
GSE20473	Human iPSC line	Affymetrix Human Exon 1.0 ST Array	• 2 patient samples: 2 LS patients with T468M mutation in PTPN11; each patient with 2 biological duplicates from human iPSC line. • 2 control samples: 1 HES2 human embryonic stem cell sample, and 1 induced pluripotent stem cell sample from healthy people
Hypertrophic cardiomyopathy (HCM)			
GSE89714	Human hypertrophic heart tissue	Illumina HiSeq 2000	• 5 patient samples: patients previously diagnosed with hypertrophic cardiomyopathy, undergoing septal myectomy surgery. • 4 Control samples: normal heart donor left ventricles
GSE68316	Human myocardial tissues	CapitalBio Human LncRNA Microarray v2.0	• 7 patient samples: 7 HCM patients • 5 Control samples: 5 disease-free individuals
GSE36961	Human surgical myectomy tissue	Illumina HumanHT-12 V3.0 expression beadchip	• 106 patient samples: 106 HCM patients • 39 Control samples: 39 healthy donors
GSE32453	Human cardiac myectomy tissue	Illumina humanRef-8 v2.0 expression beadchip	• 8 male patients: hypertrophic obstructive cardiomyopathy (HOCM) • 5 controls: people without cardiac disease

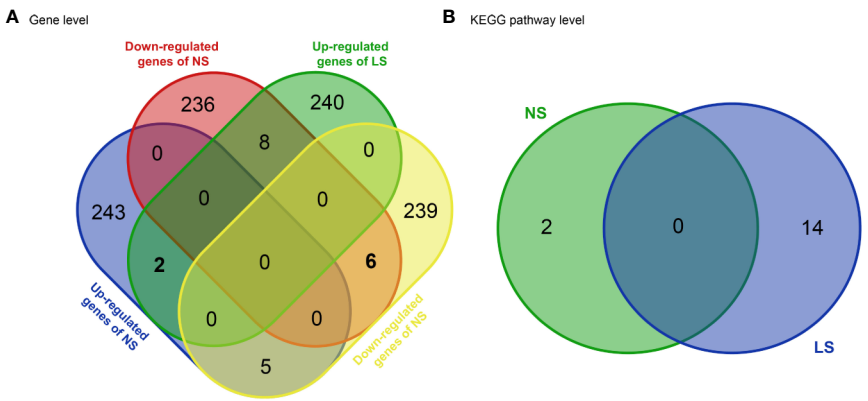

FIGURE 2 | Venn diagram of transcriptomic signatures and related KEGG pathways between Noonan syndrome (NS) and LEPOARD syndrome (LS). **(A)** common genes of up- or down- regulated genes of NS and LS; **(B)** common KEGG pathways enriched by NS and LS transcriptomic signatures.

total of 14 KEGG pathways were over-represented for LS, which consisted of several cardiovascular-related pathways such as Hypertrophic cardiomyopathy (KEGG id: hsa05410) and some cancer-related pathways such as PI3K-Akt signaling pathway (KEGG id: hsa04151) (**Figure 2B** and **Supplementary Table S2**). Similarly, no overlapping pathways were found between NS and LS.

The results of the KEGG Pathway analysis were further confirmed *via* a network analysis of NS and LS transcriptomic signatures and protein and protein interactions (PPIs) from the

STRING database. The top three clusters of PPIs were extracted using Cytoscape (plugin MCODE), and the over-represented KEGG pathways were enriched using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (**Figure 3** and **Supplementary Table S3**). For NS, the only enriched KEGG pathway was Ribosome (KEGG id: hsa03010). For LS, the enriched pathways were Hypertrophic cardiomyopathy (KEGG id: hsa05410) and PI3K-Akt signaling pathway (KEGG id: hsa04151). These results are consistent with the KEGG pathways enriched with the whole transcriptomic signatures.

The concordance between the NS and LS signatures was also compared at the miRNA level. Specifically, the related miRNAs for NS and LS signatures were obtained by using a consensus of 12 miRNA target prediction algorithms (**Supplementary Table S4**). The concordance between LS and NS was dramatically increased at the miRNA level (**Supplementary Figure S1**); there was a total of 118 and 81 up- and down-regulated miRNAs shared between NS and LS, respectively. This improved concordance indicates that the regulatory mechanisms of the two diseases are interrelated.

The cardiac system is one of the organs most affected by NS and LS. Cardiac disorders such as HCM were present in approximately 20% of NS and LS patients. The pathway analysis indicated that the LS patients had more characteristic features of HCM compared with NS patients. To further understand the relationship between HCM and NS and LS, we extracted HCM-related gene signatures from four transcriptomic profiling data sets (i.e., GSE89714, GSE68316, GSE36961, and GSE32453) and compared them with the NS and LS signatures. There was little overlap of genes among the different studies, possibly due to the high heterogenicity of HCM patients. The number of overlapping genes with HCM studies were 26 (10 up-regulated and 16 down-regulated) and 28 (16 up-regulated and 12 down-regulated) for NS and LS, respectively (**Supplementary Figure S2**).

Enrichment of Repositioning Candidates by Transcriptomics-Based Approaches

Repositioning candidates were enriched by using two large drug transcriptomic data sets including CMap02 and LINCS 1000 (L1000) (Lamb et al., 2006; Subramanian et al., 2017). The rationale behind transcriptomic-based drug repositioning is if the disease and drug transcriptomic signatures are reversely correlated, the drug has the potential to treat that disease (Iorio et al., 2013). We employed the L1000CDS[2] web-based tool to enrich repositioning candidates for NS and LS, respectively. The top 50 candidates with highest enrichment scores (e.g., cosine similarity) for each disease were extracted. We noticed that some repositioning candidates were enriched across multiple assay conditions. For example, CGP-60474, an inhibitor of a cyclin-dependent kinase, was identified as a potential treatment for NS based on its transcriptomic profiles in 12 different cell line/concentration/duration combinations. Repositioning candidates were also identified using CMap02

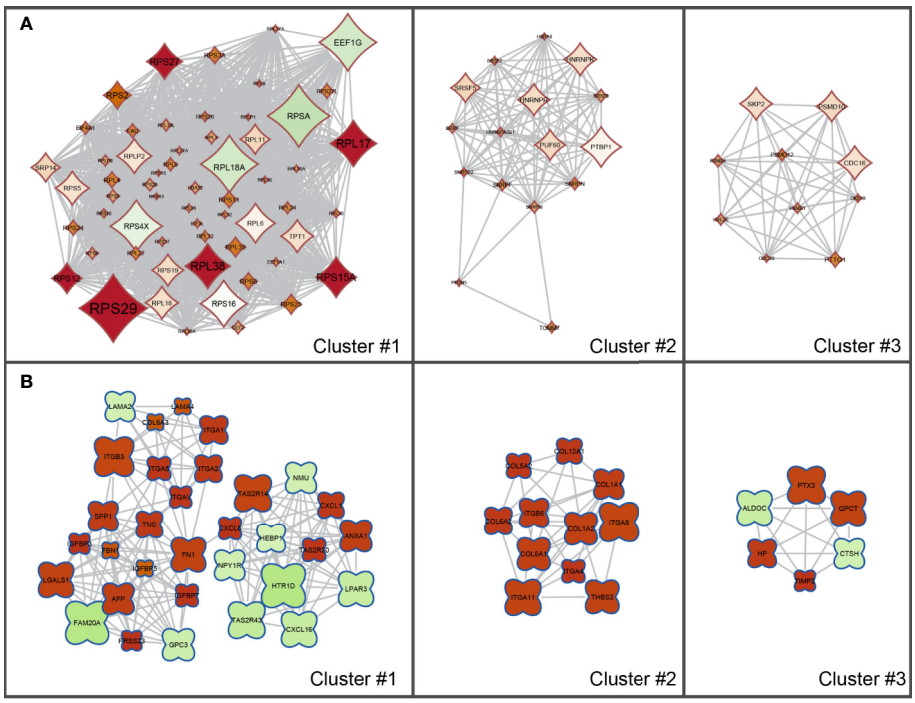

FIGURE 3 | Sub clusters of protein-protein interaction (PPI) networks of Noonan syndrome (NS) and LEPOARD syndrome (LS) transcriptomic signatures by Cytoscape MCODE plugin. **(A)** Top three clusters of NS-related PPI networks based on MCODE enrichment scores; **(B)** Top three clusters of LS-related PPI networks based on MCODE enrichment scores.

data sets. We collapsed these drug transcriptomic profiles generated under different assay conditions such as cell line/concentration/duration into one prototype ranking list (PRL) to generate its signature. The top 50 repositioning candidates were prioritized based on ranked GSEA scores for NS and LS, respectively. Together, we obtained 74 repositioning candidates for NS and LS from L1000 and CMap data sets, respectively (see **Supplementary Table S5**).

To investigate whether the enriched repositioning candidates were better by chance, we conducted a random test by using a pseudo disease signature to query the two drug transcriptomics datasets. The process was repeated 50,000 times. **Figure 4** illustrates the distribution of enrichment scores for real and pseudo disease signatures. The Cohen's d values with Hedges correction were calculated for assessing the effect sizes between the actual and pseudo distribution. All the Cohen's d values were more than 1.2, considered a very large effect size (Sawilowsky, 2009). Also, the unpaired student t-test was used, and all p values were less than 2.2×10^{-16}. Both statistical measures indicated that all the enriched repositioning candidates were not attributable to chance.

The therapeutic class distribution of enriched repositioning candidates is shown in **Figure 5**. Repositioning candidates covered more than ten different therapeutic categories as defined by the first level of the WHO Anatomical Therapeutic Chemical (ATC) Classification System. Of them, **L**-Antineoplastic and immunomodulating agents dominated all the therapeutic classes, corresponding to 22% and 32% of repositioning candidates for NS and LS, respectively. It was interesting that 11 repositioning candidates for NS were from **J**-Antiinfectives for systemic use and 10 for LS belonged to **N**-Nervous system, highlighting the divergence of the underlying mechanisms of the two diseases. It was notable that the 8 dermatological drugs were enriched for NS, perhaps correlating with facial deformaties in NS patients.

Molecular Docking for Enriched Repositioning Candidates

Binding Site Identification

Three 3D protein structures of PTPN11 including wild type (PDB ID: 4DGP), LS-Associated SHP2/T468M mutant (PDB ID: 4OHL), and NS-associated SHP2/E76D mutant (PDB ID: 6CMR) were used. The crystal structure of PTPN11 consisted of N- and C-terminal SH2 domains, and a PTP domain (**Figure 6A**). The mutated amino acids (i.e., E76D for NS and T468M for LS) were located at the PTP domain. Only the crystal structure of 6CMR contained the ligand SHP099. Thus, the binding sites of the other two crystal structures (i.e., 4DGP and 4OHL) can be considered as speculative. Herein, we employed SiteMap implemented in the Schrodinger suite to identify potential ligand binding sites (Halgren, 2007). The top five binding sites for each protein structure were predicted (see **Supplementary Figure S3**) and the binding sites with highest SiteMap scores were selected for further analysis. The predicted binding site of 6CMR (NS) covered the existing ligand SHP099, validating the performance of SiteMap (**Figure 6B**). The optimal binding sites for 4OHL (LS) and 4DGP (wild type) are illustrated in **Figures 6C, D**, respectively. It was observed that the

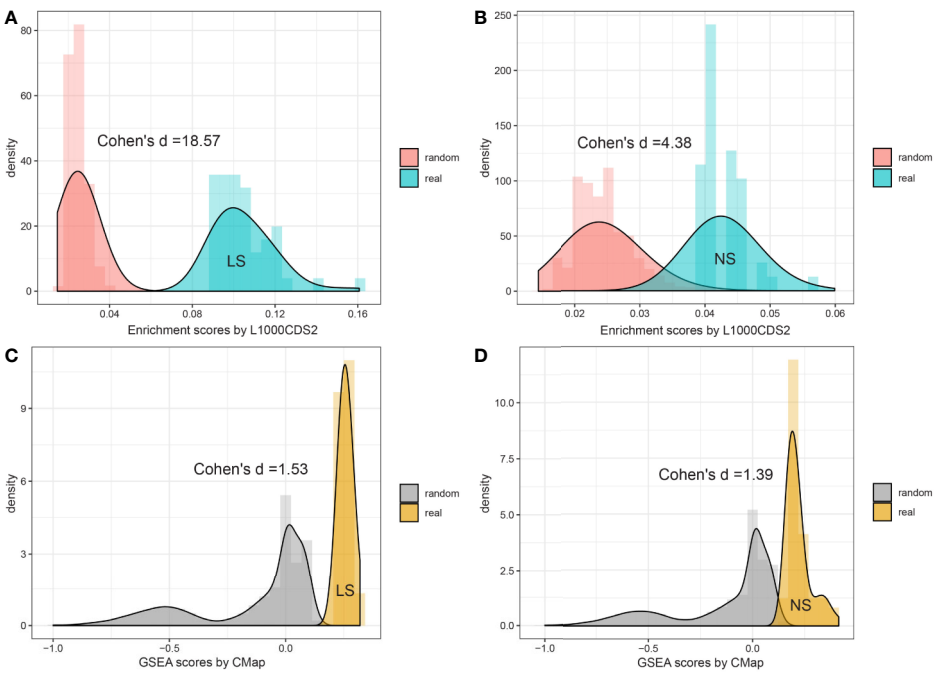

FIGURE 4 | Permutation test of enriched repositioning candidates by using transcriptomic approach. **(A, B)** distribution of enrichment scores generated using NS and LS transcriptomic signatures (shown in blue) and pseudo disease signatures (shown in red) against the L1000 dataset; **(C, D)** distribution of enrichment scores generated using NS and LS transcriptomic signatures (shown in yellow) and pseudo disease signatures (shown in gray) against CMap dataset.

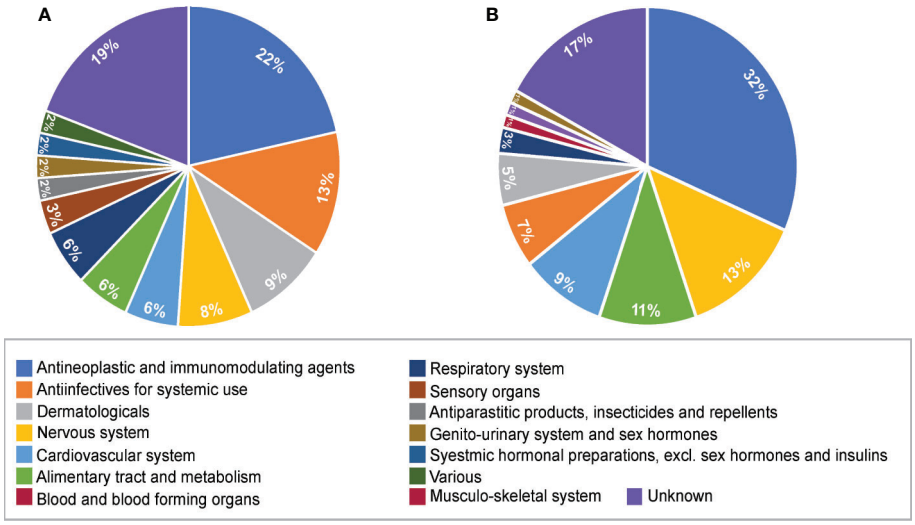

FIGURE 5 | Pie charts of therapeutic categories of enriched repositioning candidates by using transcriptomic approach. **(A)** Therapeutic categories of repositioning candidates for Noonan syndrome; **(B)** Therapeutic categories of repositioning candidates for LEPAROD syndrome.

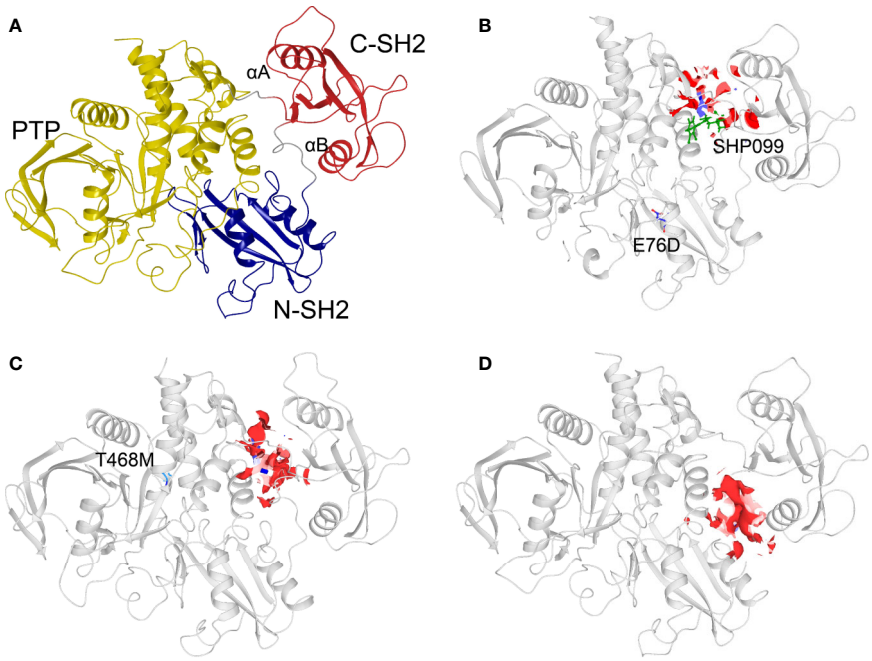

FIGURE 6 | Optimized binding sites of 3D SHP2 protein structures of Noonan syndrome (NS), LEPOARD syndrome (LS) and wild type by using Schrodinger SiteMap module. **(A)** Domain information of SHP2; **(B)** Predicted binding pocket for NS-associated SHP2/E76D mutant (PDB ID: 6CMR); **(C)** Predicted binding pocket for LS-Associated SHP2/T468M mutant (PDB ID: 4OHL); **(D)** Predicted binding pocket for SHP2 wild type (PDB ID: 4DGP).

predicted binding sites of mutated PTPN11 (i.e., 6CMR and 4OHL) shifted from Helix αB to Helix αA of C terminal-SH2 when compared to that of the wild type. The depth between the binding site of NS to the mutated site (E76D) was extended from wild type (17.1 Å) to the mutant state (28.0 Å). However, the distance between the binding site of LS and the mutated site (T468M) was shortened from 24.3 Å to 20.7 Å. Furthermore, hydrophobic areas within mutated structures (i.e., 40HL and 6CMR) were decreased from 95.0 Å to 62.2 Å and 62.1 Å compared to wild type, suggesting conformation changes from wild type to mutated states (**Table 2**). The key residual entropy of mutated structures such as 4OHL and 6CMR varied. For example,

TABLE 2 | Binding site optimization by SiteMap.

PDB ID	Volume of Pocket	Surface Area	Hydrophilic area (surface area %)	Hydrophobic area (surface area %)	Distance to mutated site E76D for LS	Distance to mutated site T468M of NS
Wild type of SHP2						
4DGP	222.6	1413.1	1069.4 (75.7%)	95.0 (6.7%)	17.1	24.3
LEOPARD Syndrome-Associated SHP2/T468M mutant						
4OHL	391.0	1442.5	1066.6 (73.9%)	62.2 (4.3%)	–	20.7
Closed structure of active SHP2 mutant E76D bound to SHP099 inhibitor for Noonan syndrome						
6CMR	358.1	1325.6	875.3 (66.0%)	62.1 (4.7%)	28.0	–

we observed that the hydrophobic area of 4OHL (LS) consisted of the residuals *PHE113*, *PRO215*, *LEU216*, *LEU233*, *LEU236*, *LEU254*, and *PRO491* and the hydrophobic area of 6CMR (NS) consisted of the residuals *PHE113*, *LEU117*, *LEU125*, *LEU233*, *LEU236* and *ALA237* (**Table 3**).

Adjusted IFD Model

The Schrödinger Induced Fit Docking (IFD) protocol was implemented to dock the repositioning candidates to the optimal binding site of 4DGP (wild type), 4OHL (LS), and 6CMR (NS), respectively. Some repositioning candidates (e.g., sirolimus, dactinomycin, atracurium besilate, rifabutin, tacrolimus, and asiaticoside for NS, and carbenoxolone, cyanocobalamin, and oxamic acid for LS) failed to dock into protein binding pockets. Building on this, compounds with positive IFD score values were also considered as docking failures. Overall, the success in the execution of IFD to generate repositioning candidates (e.g., 61 for NS and 43 for LS) resulted in multiple docking poses for each repositioning candidate. We selected the best docking pose based on the lowest IFD score for each repositioning candidate. We implemented an adjusted IFD score strategy by combining the IFD scores of ligands in both the mutated and wild type (see *Materials and Methods* section). We found that the IFD scores of LS were higher than that of their wild type, while the IFD scores of NS were lower than their wild type (see **Supplementary Table S6**) suggesting that the stability

TABLE 3 | Key residuals located at the optimal binding pocket.

PDB ID	Binding type of residuals				
	Hydrophobic	Polar	Negative charge	Positive charge	Not classified
Wild type of SHP2					
4DGP	PHE7	SER3	ASP106	ARG4	–
	PRO9	SER109	GLU110	ARG5	
	CYS104	THR191	GLU195	ARG111	
	PRO107	GLN214	GLU252	LYS198	
	TRP112	THR253		LYS199	
	PHE135	GLN255		LYS213	
	VAL194	GLN256		LYS260	
	PRO215	GLN257			
	CYS259				
LEOPARD Syndrome-Associated SHP2/T468M mutant					
4OHL	PHE113	THR108	GLU110	ARG111	GLY246
	PRO215	SER109	GLU232	ARG229	
	LEU216	HIS114	GLU249	LYS492	
	LEU233	HIS116	GLU250		
	LEU236	ASN217	GLU252		
	LEU254	THR218			
	PRO491	THR219			
		GLN245			
		THR253			
Closed structure of active SHP2 mutant E76D bound to SHP099 inhibitor for Noonan syndrome					
6CMR	PHE113	THR108	GLU139	ARG229	GLY115
	LEU117	HIS114	GLU232	ARG231	GLY246
	LEU125	HIS116	GLU238	LYS235	
	LEU233	THR218	GLU249		
	LEU236	THR219	GLU250		
	ALA237	SER234			
		GLN245			

of protein-ligand conformations for LS is lower than for NS. The adjusted IFD scores were used to reflect the specificity of repositioning candidates binding to the mutated protein structures. However, the prioritization of compounds based on adjusted IFD scores should generally conform to the original IFD approach. Therefore, we further compared the correlation among IFD scores for mutant structures and adjusted IFD scores (see **Supplementary Figure S4**). The Pearson's correlation coefficients between adjusted IFD scores and IFD scores for NS and LS were in the range of 0.883 to 0.982, suggesting the ranking of repositioning candidates was well preserved with only compounds with low binding specificity being excluded.

To further investigate the key residuals involving in the ligand binding process, we analyzed the ligand-residual binding interactions for 4OHL (LS) and 6CMR (NS) with hierarchical cluster analysis (HCA). For NS, the repositioning candidates were categorized into three clusters based on their binding residuals (**Figure 7A**). Dopamine receptors including BRD-K85818861, levomepromazine, quinpirole, and dopamine for

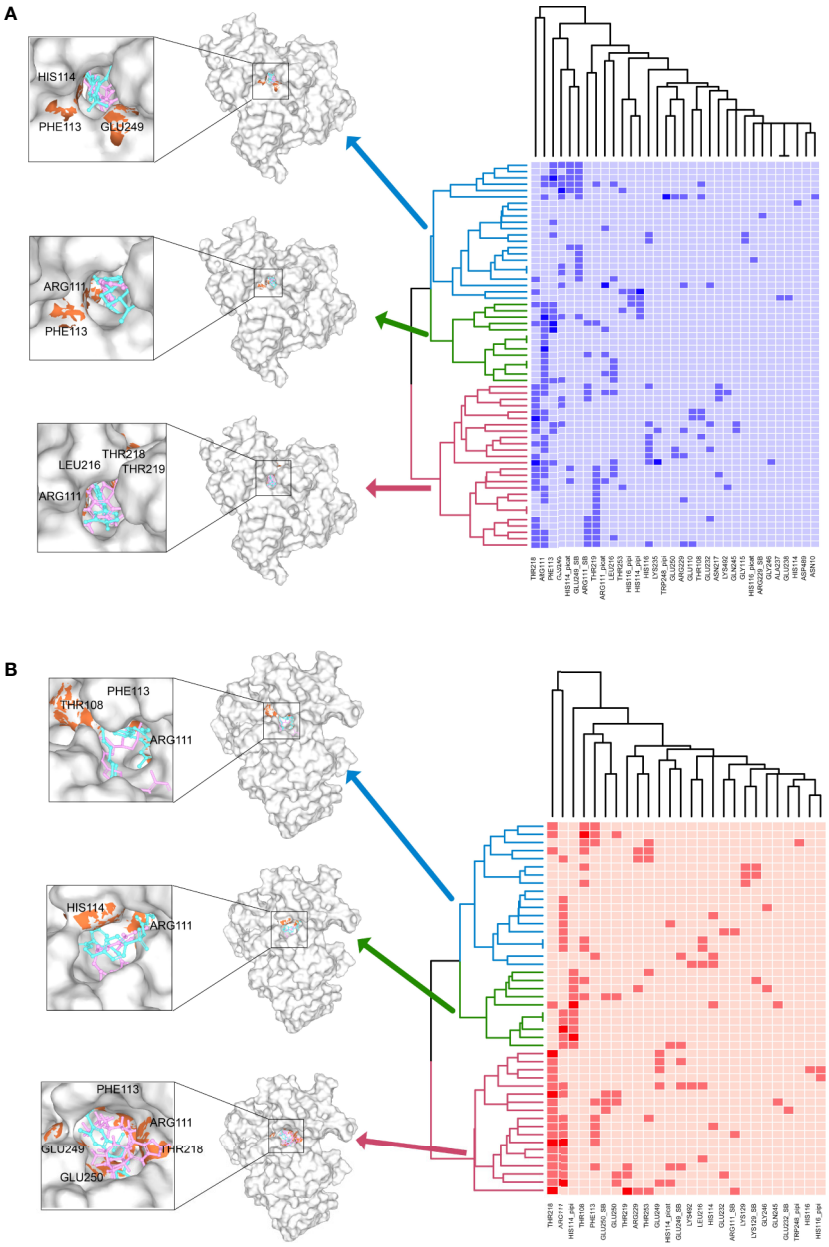

FIGURE 7 | Hierarchical cluster analysis (HCA) of repositioning candidates and binding amino acids for **(A)** NS and **(B)** LS.

NS were in the same cluster (show in blue). The repositioning candidates in this cluster preferentially bind with residuals such as *HIS114* (polar), *PHE113* (hydrophobic), and *GLU249* (negative charge). Repositioning candidates belonging to Phosphoinositide 3-kinases (PI3Ks) inhibitors (i.e., PI-103 and S1205) and non-selective, irreversible monoamine oxidase (MAOI) inhibitors (i.e., iopromide and iproniazid) were clustered together (shown in red), where more frequently binding residuals included *ARG111* (positive charge), *LEU216* (hydrophobic), *THR218* (polar), and *THR219* (polar). There was another group of repositioning candidates with diverse modes of action (shown in green). The average and standard deviation values of adjusted IFD scores for the three clusters (e.g., -1170.15 ± 4.382 for the blue cluster, -1172.15 ± 3.7695 for the green cluster, and -1171.97 ± 4.6227 for the red cluster) were not statistical significantly different. For LS, repositioning candidates were also divided into three clusters (**Figure 7B**). Most of the histone deacetylase (HDAC) inhibitors (i.e., S1030 and scriptaid) and the BCR-ABL inhibitor (i.e., dasatinib) were clustered together (shown in red). Key residuals such as *PHE113* (hydrophobic), *THR218* (polar), *GLU249/250* (negative charge), and *ARG111* (positive charge) were more frequently bound. Two heat shock protein 90 inhibitors (i.e., radicicol and BRD-K41859756) clustered together (shown in green) and key residuals including *ARG111* and *HIS114* were highlighted. Repositioning candidates that are MOA-dependent formed another cluster (shown in blue). Furthermore, the average adjusted IFD scores of the cluster including heat shock proteins (e.g., -1102.42 ± 14.65) was higher than that of the other two clusters (1111.96± 16.60 for the HDAC cluster and -1110.78 ± 15.18 for the MOA cluster).

A Combined Ranking Strategy

To take advantage of both transcriptomic and structure-based approaches for repositioning candidates prioritization, we developed a consensus ranking strategy (see *Materials and Methods* section). The prioritized repositioning list for NS and LS by proposed combined ranking strategy are listed in **Supplementary Table S5**. To further verify the prioritized repositioning list for their potential for LS or NS treatment, real-world evidence including the on-going clinical trials in clinicaltrial.gov and PubMed literature reports were collected. There were very few clinical trial studies focused on the two diseases, and only 1 clinical trial for LS and 13 clinical trials for NS were found. Many of the clinical trials were focused on repositioning candidates despite a lack of literature support. Moreover, the phenotype of NS patients involved in one study was not available in the original publication (Mulero-Navarro et al., 2015). However, LS patients with visible HCM symptoms were described in one initial report (Carvajal-Vergara et al., 2010) and further confirmed with our pathway and network analysis. Altogether, we have mainly focused on PubMed literature reports on the repositioning candidate list for LS to verify our results.

Table 4 shows the compiled list of selected repositioning candidates for LS with their original designated indications, mode of actions, boxed warnings (BW) status in drug labeling, and real-world repositioning evidence. Of 46 repositioning candidates for LS, we found 11 candidates (24%) with at least one literature report analyzing their potential utility for treating LS or LS-related cardiovascular disorders. Considering LS patients are mainly pediatric, we further evaluated safety profiles of repositioning candidates by assessing the boxed warnings in the label of the approved drug. Notably, 18 of 46 candidates (39.1%) were designated for oncology treatment, suggesting the safety profiles will need to be reevaluated when repositioning these candidates for pediatric uses.

DISCUSSION

There is still an extensive unmet medical need in the treatment of rare diseases such as NS and LS. Considering the limited knowledge and availability of data on rare diseases, we pursued a "fit-for-purpose" strategy for exploring treatment opportunities. In this study, we proposed a hybrid strategy to integrate transcriptomic and structure approaches to look for respective repositioning candidates for NS and LS. We found the transcriptomic signatures could discriminate the two diseases despite their high pathological similarity. Furthermore, an adjusted IFD model was able to take into consideration the binding behaviors of both mutated and wild types to reflect the binding specificity. Some key residuals involved in the ligand and mutated 3D PTPN11 binding were highlighted, which could be used for further repositioning candidate enrichment or the development of novel lead compounds. Also, a comparative ranking selection was developed to prioritize repositioning candidates. Finally, a list of repositioning candidates including 61 for NS and 43 for LS was proposed and verified in a literature survey.

Development of the tyrosine phosphatase SHP2 (PTPN11) inhibitors has been the focus of considerable research efforts in light of their diverse role in cancer and RASopathies disorders (Butterworth et al., 2014; Garcia Fortanet et al., 2016; Yu and Zhang, 2018). However, there is a remaining challenge around selectivity for specific phenotypic endpoints and cell permeability of potential inhibitors. The well-established approaches for PTPN11 inhibitor identification mainly rely on *in silico* and *in vitro* screening against large compound repositories. However, uncertainty remains regarding the potential adverse effects of the enriched lead compounds. The proposed hybrid approach is helpful to better understand the perturbation by of different biological process at the molecular level. Furthermore, the adjusted IFD model aims to prioritize the compounds based on their binding specificity to the mutated protein structure, which may increase the selectivity of compounds for treating more precise disease endpoints, further eliminating unexpected side effects.

A large proportion of enriched repositioning candidates for LS and NS has anticancer activity. In our previous study, we have proposed to reuse oncology drugs for rare disease treatment development based on the similarity of the two types of diseases at the molecular level (Liu et al., 2016). The mutation of the

TABLE 4 | Selected repositioning candidates for LEOPARD syndrome.

Candidate name	Market status	Boxed Warnings	Pharmacological class	Original indications	Notes	References
dasatinib	Approved	No	a BCR-ABL inhibitor	chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (ALL)	A low dose of dasatinib reversed the expression levels of molecular markers of cardiomyopathy and reduced cardiac fibrosis in NS and LS mice. Low-dose dasatinib may represent a unifying therapy for the treatment of PTPN11-related cardiomyopathies	PMID: 27942593
scriptaid	Investigational	–	a histone deacetylase (HDAC) inhibitor	Most early research on anti-cancer activity, cloning and regulation of metabolism	Scriptaid can further be used in the treatment of cardiac/vasculature diseases such as hypertrophy, hypertension, myocardial infarction, reperfusion, ischaemic heart disease, angina, arryhtmias, hypercholestremia, atherosclerosis and stroke	US20140235671A1
TWS119	Investigational	–	a glycogen synthase kinase 3 beta (GSK3β) inhibitor	Most early research on anti-cancer activity	Inhibition of GSK3 results in changes in the activities of transcription and translation factors in the heart and promotes hypertrophic responses, and it is generally assumed that signal transduction from hypertrophic stimuli to GSK3 passes primarily through protein kinase B/Akt (PKB/Akt)	PMID: 18204489
6-bromoindirubin-3-oxime	Investigational	–	a glycogen synthase kinase 3β (GSK3β) inhibitor	Most early research on anti-cancer activity	The hybrid hydrogel system can co-deliver 6-Bromoindirubin-3-oxime (BIO) and insulin-like growth factor 1 (IGF-1) to areas of myocardial infarction and thus improve cardiac function by promoting the proliferation of cardiomyocytes and revascularization.	PMID: 26251592
lansoprazole	Approved	No	a H+/K+-ATPase blocker	peptic ulcer disease, gastroesophageal reflux disease, and Zollinger–Ellison syndrome	Lansoprazole alleviates pressure overload-induced cardiac hypertrophy and heart failure in mice by blocking the activation of β-catenin	PMID: 30689763
resveratrol	Approved	No		Herpes labialis infections	there are potential beneficial effects of resveratrol in CVDs (atherosclerosis, hypertension, stroke, myocardial infarction, heart failure)	PMID: 27144581
nicardipine	Approved	No	a potent calcium channel blockader	marked vasodilator action	nicardipine may improve left ventricular diastolic function of patients with HCM without serious side effects	PMID: 2636085
bromocriptine	Approved	No	a D(2) dopamine receptor	Parkinsonian Syndrome	bromocriptine improves the hemodynamic profile in heart failure acutely	PMID: 6346844
LY-294002	Investigational		a strong inhibitor of phosphoinositide 3-kinases (PI3Ks); a BET inhibitor	Most early research on anti-cancer activity	PI3K inhibition by LY294002 or wortmannin abolished the myocardin-dependent SRF stimulation observed in the T468M LS cell line. Consequently, these data suggest that LS-induced PI3K upregulation can enhance SRF/myocardin activity in a reconstituted model.	PMID: 20308328
glycocholic acid	Investigational		a crystalline bile acid		bile acids (Bas) as newly recognized signaling molecules that modulate cardiovascular function	PMID: 21707953
glibenclamide	Approved	No	an ATP-sensitive potassium channels inhibitor	type 2 diabetes	Glibenclamide may increase the probability of spontaneous termination of ventricular fibrillation and facilitate the restoration of the myocardial function during regional ischaemia.	PMID: 1291084

PTPN11 gene is not the only causative link to NS and LS but is also one of the most somatic mutated genes (SMGs) in cancer pathogenesis (Kandoth et al., 2013). For example, RASopathies disorders such as NS and LS are correlated with specific cancer types such as Juvenile Myelomonocytic Leukemia (JMML), attributed to different mutation types of the PTPN11 gene (Kratz et al., 2005; Kiel and Serrano, 2014). One of the proposed repositioning candidates for LS was dasatinib, a BCR-ABL inhibitor, that was originally designed for treating chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (ALL). It was reported that a low dose of dasatinib reversed the expression levels of molecular markers of cardiomyopathy and reduced cardiac fibrosis in NS and LS mice with the PTPN11 mutation (Yi et al., 2016).

One of the most common phenotypic characteristics of NS and LS is hypertrophic cardiomyopathy (HCM). The LS patients involved in this study had a clear HCM phenotype. Multiple small molecule inhibitors of histone deacetylase (HDAC) have been shown to be efficacious in blocking pathological cardiac hypertrophy, and in improving cardiac function in various rodent animal models (Bush Erik and McKinsey Timothy, 2010; Ferguson et al., 2013). Several HDAC inhibitors including scriptaid, KM 00927, and S1030 emerged in our repositioning candidate list for LS treatment.

For example, one patent described the use of scriptaid in the treatment of cardiac/vasculature diseases such as hypertrophy, hypertension, myocardial infarction, and ischaemic heart disease (US20140235671A1).

Our data also suggested several heat shock protein 90 (HSP90) inhibitors including BRD-K41859756, geldanamycin, radicicol for potentially treating LS. HSP90 has been widely studied given its essential role in cancer etiology (Wu et al., 2017). Alongside this, the cardiac protection function of HSP90 has also been described. For example, induction of HSP90 by mild stress has a protective effect against severe heart pressure. Furthermore, over-expression of HSPs in cardiac cells in transgenic animal and virus models suggest a potential route to the development of treatments to protect heart function (Latchman, 2001).

Several areas would benefit from further verification of the strategies and results suggested from this study. First, due to the limited samples and data of NS (e.g., one sample with two replicates), the transcriptomic signatures may suffer a degree of bias leading to unspecific functional analysis results. Second, the adjusted IFD model is time-consuming due to the need for precise docking for both mutated and wild types, limiting its broader application for screening purposes. However, the adjusted IFD model could easily be migrated to the Glide docking framework to address this limitation. Finally, no single bioinformatics approach offers a one-stop solution for exploring rare disease treatment options. Each research question requires a "fit for purpose" approache based on biological understanding and data availability. More importantly, the prioritized repositioning list serves as a starting point for further treatment development, and wet lab experiment verification and real-world evidence are imperative (Delavan et al., 2018).

It is worthwhile to consider additional experimental studies to confirm the findings from our study. The current study set the focus on the novel bioinformatics approaches by integrating genomics and structure information to generate testable hypotheses and provide viable leads for other investigators to follow up on. However, the clinic validity of enriched candidates could be only established with further wet-lab experiments and clinical trials, which is out of the scope of the current study. Specifically, a more profound transcriptome analysis should be conducted to show that candidate drugs might influence the transcriptome in such a way that the NS and LS cells could come back to normal profile. Moreover, hiPS cells with NS and LS syndromes should be analyzed for cardiac defects. Then, the proposed drugs should be demonstrated to functionally rescue these defects. Furthermore, some experimental validation of the cellular activity of the proposed drugs against the disease targets (e.g., PTPN11) could be helpful to elucidate the pharmacological effects of enriched candidates. In our future work, we will further consider how to enhance biological validation for a better verification of proposed in silico approaches.

In conclusion, we have developed a computational drug repositioning framework that integrates transcriptomic and structural approaches to find novel potential NS and LS treatment options. The framework developed is robust and relatively straightforward to use as such it may be equally applicable to the search for treatments for other rare diseases.

MATERIALS AND METHODS

Transcriptomic Profiles of Noonan and LEOPARD Syndrome

The transcriptomic profiles of NS and LS were extracted from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) by using keyword "*Noonan syndrome*" AND ("*LEOPARD syndrome*" OR "*Noonan syndrome with multiple lentigines*"). Consequently, two GEO datasets GSE54538 (Mulero-Navarro et al., 2015) and GSE20473 (Carvajal-Vergara et al., 2010) were obtained for NS and LS, respectively. Interestingly, the transcriptomic profiles of the two studies were both generated based on samples derived from human iPSC (**Table 1**).

The samples of NS dataset (GSE54538) included NS patients, juvenile myelomonocytic leukemia (JMML) patients, NS/JMML patients, and healthy donors. Here, NS patient samples with E76D mutation in PTPN11 (i.e., one patient with two biological replicates) and four control samples from healthy donors were employed for further analysis.

The experiment samples of LS (GSE20473) consisted of two iPSC lines of two LS patients with T468M mutation in PTPN11, normal or patient Fibroblasts, and HES2 human ES cell line. In this study, two iPSC lines from each LS patient (i.e., four samples: L1-iPS1, L1-iPS13, L2-iPS6, and L2-iPS16) and two controls including the HES2 human ES cell line and a wild-type iPSC line and BJ-iPSB5 derived from a normal human fibroblast line (BJ) was used.

Microarray Data Normalization and Disease Signatures Generation

For NS dataset (GSE54538), Illumina HumanHT-12 V4.0 expression BeadChip was used, which targets more than 47,000 probes derived from the National Center for Biotechnology Information Reference Sequence (NCBI) RefSeq Release 38. The preprocessed data by Robust Multi-array Average (RMA) was downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54538. By comparing NS patient samples to the controls, the fold changes and p values were generated using R *limma* package. We further mapped the probe sets into gene symbols using R *biomaRt* packages. Then, the probes that could not be converted to gene symbols were excluded. Final, we ranked genes based on their fold change values and selected the top/down 250 genes as NS signatures (see **Supplementary Table S1**).

For LS dataset (GSE20473), the experiment was conducted using Affymetrix Human Exon 1.0 ST Array. The Raw CELL files were downloaded from https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE20473&format=file. First, the RAW CELL files were processed with alternative chip definition files (CDFs,

Version 19.0.1, ENTREZG) from Brainarray (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp) and RMA used to summarize the intensity ratios at probe set level to obtain expression values per gene. Like the procedure used for NS data, the fold change and *p* values were generated by using R limma and limited to gene levels. Finally, genes were ranked by their fold changes values, and the top/down 250 genes were extracted as the LS signatures (see **Supplementary Table S1**).

Pathway Analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was carried out with the obtained disease signatures including 500 genes (i.e., top/down 250 genes) for NS and LS by using the Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8) software (Huang et al., 2008). The pathways with a Benjamini-Hochberg (BH) adjusted *p*-value less than 0.05 were considered to be statistically significant.

Protein-Protein Interaction (PPI) Network Analysis

Protein-protein interactions (PPI) among NS and LS signatures were further determined by human PPI data extracted from the STRING v11.0 database (https://string-db.org/) (Szklarczyk et al., 2017). Only the interactions with high confidence interaction scores defined by the STRING database (e.g., ≥ 0.7) were considered. We further extracted the sub-networks by using the MCODE plug-in for Cytoscape (version 3.7.1), which is designed to expand the cluster from highly interconnected seed nodes by setting a certain threshold (Bader and Hogue, 2003).

miRNA Target Prediction

To further investigate the distribution of miRNA regulating NS and LS signatures, a consensus of twelve *in silico* miRNA prediction algorithms including miRWalk, miRDB, PITA, MicroT4, miRMap, RNA22, miRanda, miRNAMap, RNAhybrid, miRBridge, PICTAR2, and Targetscan were employed. We considered the mRNA-miRNA relationships that were predicted as positive by 10 of 12 algorithms. The calculation was carried out through miRNA 2.0 (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/miRretsys-self.html) (Dweep et al., 2011).

Phenotypic Anchoring

One of the most common phenotypes of NS and LS are cardiovascular disorders such as hypertrophic cardiomyopathy (HCM). We collected four transcriptomic profiles of HCM (i.e., GSE89714, GSE68316, GSE36961, and GSE32453) by searching against the GEO database (see **Table 1**). The differentially expressed genes (DEGs) were calculated for each dataset by comparing patient group to the control group. We then compared the NS and LS transcriptomic signatures to the DEGs of HCM, respectively. All the calculations were carried out with R *limma* packages.

Repositioning Candidate Enrichment by Using CMap and LINCS L1000

The hypothesis behind genome-based repositioning is that if the drug signature is reversely correlated with the disease signature, the drug could be potentially used to treat the diseases. Here, two drug-induced transcriptional profiles of human cell lines including NIH LINCS project (http://www.lincsproject.org/) (Subramanian et al., 2017) and Connectivity Map 02 (Lamb et al., 2006) were employed to enrich the repositioning candidates for NS and LS, respectively.

Specifically, LINCS L1000 characteristic direction signatures search engine (L1000CDS^2) was used to reversely compare the NS and LS transcriptomic signatures to the drug transcriptional signatures in LINCS project for repositioning candidate enrichment. The top 50 candidates based on cosine similarity scores were extracted as the repositioning candidates. For CMap 02 dataset, 6,100 experiments were conducted in six or seven different cancer cell lines using 1,309 compounds with two Affymetrix platforms. We followed the standard procedures to process and normalize the dataset (Lamb et al., 2006). To eliminate the divergence among cancer cell lines, we merged multiple experiments for the same compounds into a single Prototype Ranked List (PRL) following the processing described elsewhere (Iorio et al., 2010). Then, Gene Set Enrichment Analysis (GSEA) scores for measuring the reverse correlation degree between NS/LS transcriptomic signatures and drug PRL was calculated following the strategies described in previously (Sirota et al., 2011). The top 50 compounds by ranked GSEA score were used as repositioning candidates.

To further investigate whether the enriched compounds are by chance, an equal number of pseudo-NS and -LS transcriptomic signatures were generated by randomly selecting 500 genes (i.e., top/down 250 genes) for each drug-transcriptomic data set. Then, the pseudo enrichment scores were calculated. This process was repeated N= 50,000 to remove the potential bias in the gene selection process, and a null distribution of pseudo enrichment scores was obtained. Finally, the *p*-value for each drug was calculated based on the null distribution.

Molecular Docking

Repositioning candidates obtained from the transcriptomics-based approach were further prioritized by using molecular docking. Since the patients involved in the transcriptomic studies of NS and LS has PTPN11 mutants (i.e., T468M for NS and E76D for LS, respectively), the 3D structures of PTPN11 including wild type, T468M mutation, and E76D mutation were retrieved from RCSB PDB (Yang et al., 2018). The 3D crystal structures including PTPN11 wild type (PDB ID: 4DGP), mutant T468M (PDB ID: 4OHL), and mutant E76D (PDB ID: 6CMR) were downloaded. The structure of PTPN11 with E76D mutation (PDB ID: 6CMR) included a ligand SHP099 inhibitor.

Molecular docking was carried out by Schrödinger (Schrödinger, LLC, New York, NY, 2018). All protein structures were prepared with Protein Preparation Wizard in

the Schrödinger Suite. The structural integrity was checked, and missing site chains and loops segments were added using Prime. Hydrogen atoms were added without deleting the original ones. All ligands of repositioning candidates based on transcriptomics-based approach were prepared with LigPrep (LigPrep, Schrödinger, LLC, New York, NY, 2018) in the Schrödinger Suite with the default settings. The four poses with minimal energy were kept.

The potential binding sites of the 3D structure of mutated and wild type PTPN11 were predicted by the SiteMap program in Schrödinger software. The predicted binding sites for the PTPN11 with E76D mutations were compared with the pocket where the SHP099 inhibitor was located. The binding site with the highest score was selected as the most probable binding sites for further docking.

Induced fit docking (IFD) predicts optimal ligand-binding modes and concomitant structural movements in the receptor using Glide and Prime modules. In IFD, when a ligand binds to the receptor, it undergoes side-chain or backbone conformational changes or both in many proteins. These conformational changes permit better binding to the receptor according to the shape and binding mode of the ligand. Docking was performed using the Schrödinger Induced Fit Docking (IFD) protocol, consisting of GLIDE docking with the rigid protein followed by energy minimization of the side chain orientations with Prime of those residues within 7.5 Å of the ligand. In the first Glide docking round, the ring containing carbons 1, 2, 3, 4, 5, and 10, and the hydroxyl at C3 of the estradiol core were restricted to an RMSD of 1.0 Å from the crystal pose. A final round of GLIDE docking was then performed towards the optimized receptor structures, from which final Induced Fit Docking (IFD) scores for each ligand-receptor complex were calculated.

We developed an adjusted IFD score to reflect the specificity of compounds binding to the structure of mutated SHP-2 with the following equation:

$$adjusted\ \ IFD\ \ score_i = IFD\ \ score_{mutated,i} + c \tag{1}$$

$$c = (\max(IFDscore_{mutated,n}) - min(IFDscore_{mutated,n})) \times \frac{IFD\ \ score_{mutated,i} - IFD\ \ score_{wild\ \ type,i}}{\max(IFD\ \ score_{wild\ \ type,n}) - \min\ \ (IFDscore_{wild\ \ type,n}} \tag{2}$$

$$i, n\ \ \in 1, 2, 3, \ldots, 50$$

Where c, the adjusted parameter, is normalized the difference of IFD scores between mutated and wild type protein structures for compound i. The c was used to measure the specificity of compound binding to mutated PTPN11 structure.

Consensus Ranking Strategy

To combine the transcriptomic-based approach with molecular docking for generating a consensus ranking list, we developed a combined score as listed below,

$$Combined\ \ score_{i,CMap/L1000} = \frac{Enrichment\ \ score_{Cmap,i/L1000,i} + Normalized\ \ adjust\ \ IFD\ \ score_i}{2} \tag{3}$$

$$Normalized\ \ adjust\ \ IFD\ \ score_i \tag{4}$$

$$= \frac{|adjusted\ \ IFD\ \ score_i| - \min(|adjusted\ \ IFD\ \ score_n|)}{\max(|adjusted\ \ IFD\ \ score_n|) - \min(|adjusted\ \ IFD\ \ score_{n|})}$$

$$i, n\ \ \in 1, 2, 3, \ldots, 50$$

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: GSE54538; GSE20473;GSE89714;GSE68316; GSE36961; GSE32453.

AUTHOR CONTRIBUTIONS

ZL and WT conceived and designed the study. LZ and ZL performed data analysis and wrote the manuscript. RH, MX, and JZ designed and conducted the wet-lab experimental verification. LZ, RR, BD, RH, ZL, and WT revised the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

LZ is grateful to the National Center for Toxicological Research (NCTR) of the U.S. Food and Drug Administration (FDA) for postdoctoral support through the Oak Ridge Institute for Science and Education (ORISE).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.00927/full#supplementary-material

FIGURE S1 | Venn diagram of consensus predicted miRNAs for Noonan syndrome (NS) and LEPOARD syndrome (LS) transcriptomic signatures

FIGURE S2 | Venn diagram of Noonan syndrome (NS)/LEPOARD syndrome (LS) transcriptomic signatures and differentially expressed genes (DEGs) of four independent hypertrophic cardiomyopathy (HCM) patient-related whole genome gene expression data sets. **(A)** Commonality of up-regulated of genes among NS and four independent HCM patient-related transcriptomic studies; **(B)** Commonality of up-regulated of genes among LS and four independent HCM patient-related transcriptomic studies; **(C)** Commonality of down-regulated of genes among NS and four independent HCM patient-related transcriptomic studies; **(D)** Commonality

of down-regulated of genes among LS and four independent HCM patient-related transcriptomic studies.

FIGURE S3 | Top five binding sites of 3D SHP2 protein structures of Noonan syndrome (NS), LEPOARD syndrome (LS) and wild type by using Schrodinger SiteMap module. **(A)** NS; **(B)** LS; **(C)** wild type.

FIGURE S4 | Correlation between Induced Fit Docking (IFD) and adjusted IFD scores of repositioning candidates for Noonan syndrome (NS) and LEPOARD syndrome (LS). **(A)** Repositioning candidates for LS enriched from L1000 dataset; **(B)** Repositioning candidates for LS enriched from CMap dataset; **(C)** Repositioning candidates for NS enriched from L1000 dataset; **(D)** Repositioning candidates for NS enriched from CMap dataset.

REFERENCES

Bader, G. D., and Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* 4 (1), 2. doi: 10.1186/1471-2105-4-2

Bellomo, F., Medina, D. L., De Leo, E., Panarella, A., and Emma, F. (2017). High-content drug screening for rare diseases. *J. Inherited Metab. Dis.* 40 (4), 601–607. doi: 10.1007/s10545-017-0055-1

Bush Erik, W., and McKinsey Timothy, A. (2010). Protein Acetylation in the Cardiorenal Axis. *Circ. Res.* 106 (2), 272–284. doi: 10.1161/CIRCRESAHA.109.209338

Butte, A. J., Chen, B., Li, J., Swamidass, S. J., Zheng, S., and Lu, Z. (2015). A survey of current trends in computational drug repositioning. *Briefings Bioinf.* 17 (1), 2–12. doi: 10.1093/bib/bbv020

Butterworth, S., Overduin, M., and Barr, A. J. (2014). Targeting protein tyrosine phosphatase SHP2 for therapeutic intervention. *Future Med. Chem.* 6 (12), 1423–1437. doi: 10.4155/fmc.14.88

Carvajal-Vergara, X., Sevilla, A., D'Souza, S. L., Ang, Y.-S., Schaniel, C., Lee, D.-F., et al. (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature* 465, 808. doi: 10.1038/nature09005

Chen, B., and Altman, R. B. (2017). Opportunities for developing therapies for rare genetic diseases: focus on gain-of-function and allostery. *Orphanet J. rare Dis.* 12 (1), 61–61. doi: 10.1186/s13023-017-0614-4

Delavan, B., Roberts, R., Huang, R., Bao, W., Tong, W., and Liu, Z. (2018). Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discovery Today* 23 (2), 382–394. doi: 10.1016/j.drudis.2017.10.009

Dudley, J. T., Sirota, M., Shenoy, M., Pai, R. K., Roedder, S., Chiang, A. P., et al. (2011). Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease. *Sci. Trans. Med.* 3 (96), 96ra76–96ra76. doi: 10.1126/scitranslmed.3002648

Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWalk – Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J. Biomed. Inf.* 44 (5), 839–847. doi: 10.1016/j.jbi.2011.05.002

Ekins, S., Williams, A. J., Krasowski, M. D., and Freundlich, J. S. (2011). In silico repositioning of approved drugs for rare and neglected diseases. *Drug Discovery Today* 16 (7), 298–310. doi: 10.1016/j.drudis.2011.02.016

Ferguson, B. S., Harrison, B. C., Jeong, M. Y., Reid, B. G., Wempe, M. F., Wagner, F. F., et al. (2013). Signal-dependent repression of DUSP5 by class I HDACs controls nuclear ERK activity and cardiomyocyte hypertrophy. *Proc. Natl. Acad. Sci.* 110 (24), 9806–9811. doi: 10.1073/pnas.1301509110

Garcia Fortanet, J., Chen, C. H.-T., Chen, Y.-N. P., Chen, Z., Deng, Z., Firestone, B., et al. (2016). Allosteric Inhibition of SHP2: Identification of a Potent, Selective, and Orally Efficacious Phosphatase Inhibitor. *J. Med. Chem.* 59 (17), 7773–7782. doi: 10.1021/acs.jmedchem.6b00680

Govindaraj, R. G., Naderi, M., Singha, M., Lemoine, J., and Brylinski, M. (2018). Large-scale computational drug repositioning to find treatments for rare diseases. *NPJ Syst. Biol. Appl.* 4 (1), 13. doi: 10.1038/s41540-018-0050-7

Halgren, T. (2007). New Method for Fast and Accurate Binding-site Identification and Analysis. *Chem. Biol. Drug Des.* 69 (2), 146–148. doi: 10.1111/j.1747-0285.2007.00483.x

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44. doi: 10.1038/nprot.2008.211

Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., et al. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci.* 107 (33), 14621–14626. doi: 10.1073/pnas.1000138107

Iorio, F., Rittman, T., Ge, H., Menden, M., and Saez-Rodriguez, J. (2013). Transcriptional data: a new gateway to drug repositioning? *Drug Discovery Today* 18 (7-8), 350–357. doi: 10.1016/j.drudis.2012.07.014

Jia, J., An, Z., Ming, Y., Guo, Y., Li, W., Liang, Y., et al. (2018). eRAM: encyclopedia of rare disease annotations for precision medicine. *Nucleic Acids Res.* 46 (D1), D937–D943. doi: 10.1093/nar/gkx1062

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333. doi: 10.1038/nature12634

Kiel, C., and Serrano, L. (2014). Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations. *Mol. Syst. Biol.* 10 (5), 727. doi: 10.1002/msb.20145092

Kouz, K., Lissewski, C., Spranger, S., Mitter, D., Riess, A., Lopez-Gonzalez, V., et al. (2016). Genotype and phenotype in patients with Noonan syndrome and a RIT1 mutation. *Genet. Med.* 18, 1226. doi: 10.1038/gim.2016.32

Kratz, C. P., Niemeyer, C. M., Castleberry, R. P., Cetin, M., Bergsträsser, E., Emanuel, P. D., et al. (2005). The mutational spectrum of PTPN11 in juvenile myelomonocytic leukemia and Noonan syndrome/myeloproliferative disease. *Blood* 106 (6), 2183–2185. doi: 10.1182/blood-2005-02-0531

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., et al. (2006). The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* 313 (5795), 1929–1935. doi: 10.1126/science.1132939

Latchman, D. S. (2001). Heat shock proteins and cardiac protection. *Cardiovasc. Res.* 51 (4), 637–646. doi: 10.1016/s0008-6363(01)00354-6

Liu, Z., Borlak, J., and Tong, W. (2014). Deciphering miRNA transcription factor feed-forward loops to identify drug repurposing candidates for cystic fibrosis. *Genome Med.* 6 (12), 94. doi: 10.1186/s13073-014-0094-2

Liu, Z., Fang, H., Slikker, W., and Tong, W. (2016). Potential Reuse of Oncology Drugs in the Treatment of Rare Diseases. *Trends Pharmacol. Sci.* 37 (10), 843–857. doi: 10.1016/j.tips.2016.06.010

Liu, Z., Huang, R., Roberts, R., and Tong, W. (2019). Toxicogenomics: A 2020 Vision. *Trends Pharmacol. Sci.* 40 (2), 92–103. doi: 10.1016/j.tips.2018.12.001

Mulero-Navarro, S., Sevilla, A., Roman, A. C., Lee, D. F., D'Souza, S. L., Pardo, S., et al. (2015). Myeloid Dysregulation in a Human Induced Pluripotent Stem Cell Model of PTPN11-Associated Juvenile Myelomonocytic Leukemia. *Cell Rep.* 13 (3), 504–515. doi: 10.1016/j.celrep.2015.09.019

Pandit, B., Sarkozy, A., Pennacchio, L. A., Carta, C., Oishi, K., Martinelli, S., et al. (2007). Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. *Nat. Genet.* 39, 1007. doi: 10.1038/ng2073

Romano, A. A., Allanson, J. E., Dahlgren, J., Gelb, B. D., Hall, B., Pierpont, M. E., et al. (2010). Noonan Syndrome: Clinical Features, Diagnosis, and Management Guidelines. *Pediatrics* 126 (4), 746–759. doi: 10.1542/peds.2009-3207

Sarkozy, A., Digilio, M. C., and Dallapiccola, B. (2008). Leopard syndrome. *Orphanet J. Rare Dis.* 3 (1), 13. doi: 10.1186/1750-1172-3-13

Sawilowsky, S. (2009). New Effect Size Rules of Thumb. *J. Mod. Appl. Stat. Methods* 8, 597–599. doi: 10.22237/jmasm/1257035100

Schubbert, S., Zenker, M., Rowe, S. L., Böll, S., Klein, C., Bollag, G., et al. (2006). Germline KRAS mutations cause Noonan syndrome. *Nat. Genet.* 38, 331. doi: 10.1038/ng1748

Sirota, M., Dudley, J. T., Kim, J., Chiang, A. P., Morgan, A. A., Sweet-Cordero, A., et al. (2011). Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data. *Sci. Trans. Med.* 3 (96), 96ra77–96ra77. doi: 10.1126/scitranslmed.3001318

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 171 (6), 1437–1452.e1417. doi: 10.1016/j.cell.2017.10.049

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Tartaglia, M., and Gelb, B. D. (2005). Noonan Syndrome and Related Disorders: Genetics and Pathogenesis. *Annu. Rev. Genomics Hum. Genet.* 6 (1), 45–68. doi: 10.1146/annurev.genom.6.080604.162305

Tartaglia, M., Zampino, G., and Gelb, B. D. (2010). Noonan syndrome: clinical aspects and molecular pathogenesis. *Mol. Syndromol.* 1 (1), 2–26. doi: 10.1159/000276766

Wu, J., Liu, T., Rios, Z., Mei, Q., Lin, X., and Cao, S. (2017). Heat Shock Proteins and Cancer. *Trends Pharmacol. Sci.* 38 (3), 226–256. doi: 10.1016/j.tips.2016.11.009

Yang, H., Tan, L., Sala, R., Burley, S. K., Hudson, B. P., Bhikadiya, C., et al. (2018). RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* 47 (D1), D464–D474. doi: 10.1093/nar/gky1004

Yi, J.-S., Huang, Y., Kwaczala, A. T., Kuo, I. Y., Ehrlich, B. E., Campbell, S. G., et al. (2016). Low-dose dasatinib rescues cardiac function in Noonan syndrome. *JCI Insight* 1 (20), e90220–e90220. doi: 10.1172/jci.insight.90220

Yu, Z.-H., and Zhang, Z.-Y. (2018). Regulatory Mechanisms and Novel Therapeutic Targeting Strategies for Protein Tyrosine Phosphatases. *Chem. Rev.* 118 (3), 1069–1091. doi: 10.1021/acs.chemrev.7b00105

Disclaimer: The views presented in this article do not necessarily reflect current or future opinion or policy of the U.S. Food and Drug Administration and/or the U.S. National Institute of Health. Any mention of commercial products is for clarification and not intended as an endorsement.

Conflict of Interest: RR is co-founder and co-director of ApconiX, an integrated toxicology and ion channel company that provides expert advice on non-clinical aspects of drug discovery and drug development to academia, industry and not-for-profit organizations.

Copyright © 2020 Zhu, Roberts, Huang, Zhao, Xia, Delavan, Mikailov, Tong and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.