ORPHANET JOURNAL
OF RARE DISEASES

REVIEW   Open Access

A generalizable pre-clinical research approach for orphan disease therapy

Chandree L Beaulieu[1], Mark E Samuels[2], Sean Ekins[3], Christopher R McMaster[4], Aled M Edwards[5], Adrian R Krainer[6], Geoffrey G Hicks[7], Brendan J Frey[5], Kym M Boycott[1,8*] and Alex E MacKenzie[1,8*]

Abstract

With the advent of next-generation DNA sequencing, the pace of inherited orphan disease gene identification has increased dramatically, a situation that will continue for at least the next several years. At present, the numbers of such identified disease genes significantly outstrips the number of laboratories available to investigate a given disorder, an asymmetry that will only increase over time. The hope for any genetic disorder is, where possible and in addition to accurate diagnostic test formulation, the development of therapeutic approaches. To this end, we propose here the development of a strategic toolbox and preclinical research pathway for inherited orphan disease. Taking much of what has been learned from rare genetic disease research over the past two decades, we propose generalizable methods utilizing transcriptomic, system-wide chemical biology datasets combined with chemical informatics and, where possible, repurposing of FDA approved drugs for pre-clinical orphan disease therapies. It is hoped that this approach may be of utility for the broader orphan disease research community and provide funding organizations and patient advocacy groups with suggestions for the optimal path forward. In addition to enabling academic pre-clinical research, strategies such as this may also aid in seeding startup companies, as well as further engaging the pharmaceutical industry in the treatment of rare genetic disease.

Keywords: Orphan disease therapy, Preclinical drug development, Generalizable screening methods, Translational toolbox

Introduction

Single gene disorders, which typically result from mutations having severe effects on gene function, are of particular importance in pediatrics. A significant fraction of pediatric hospital admissions involve genetic conditions [1,2]. The curated Online Inheritance in Man (OMIM) human genetics database currently lists over 3300 genes for which DNA sequence variants have been associated with human disease [3]. The Human Gene Mutation Database (HGMD), using slightly different criteria, lists approximately 3200 such genes [4]. Apart from a fraction of curated variants arising from genome-wide association studies with uncertain functional significance, the majority of these annotations reflect medical conditions with either congenital or childhood onset. Optimal management of patients suffering from these disorders entails both rapid and accurate molecular diagnosis, and, where possible, treatment. The advent of new high throughput, low cost DNA sequencing technologies (so-called next generation systems) has already and will continue to increase the efficiency of new causal gene identification that will facilitate their molecular diagnoses [5]. However, given the tremendous variety of such disorders, their treatment remains an intractable and difficult-to-generalize problem.

Genetic disorders of high penetrance are typically caused by mutations that result in i) loss-of-function (LOF), i.e., a reduction in the level and/or activity of a given protein, usually seen in recessively inherited disorders or ii) gain-of-function (GOF), i.e., an increase in protein level and/or activity with the introduction of a novel pathological function often associated with activation of a pathway, usually seen in dominantly inherited disorders [4]. In broad terms, therapy in either case can be directed at normalizing the pathogenic imbalance; that is to

* Correspondence: kboycott@cheo.on.ca; mackenzie@cheo.on.ca
[1]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada
[8]Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8 L1, Canada
Full list of author information is available at the end of the article

enhance mRNA, protein or protein activity in disorders caused by LOF mutations and moderate the mRNA, protein, protein function, or pathway activity excess observed in GOF mutations (Figure 1). Disorders with a molecular test available for early diagnosis and a presymptomatic window or likelihood of clinical reversibility are potential candidates for therapeutic intervention by this strategy.
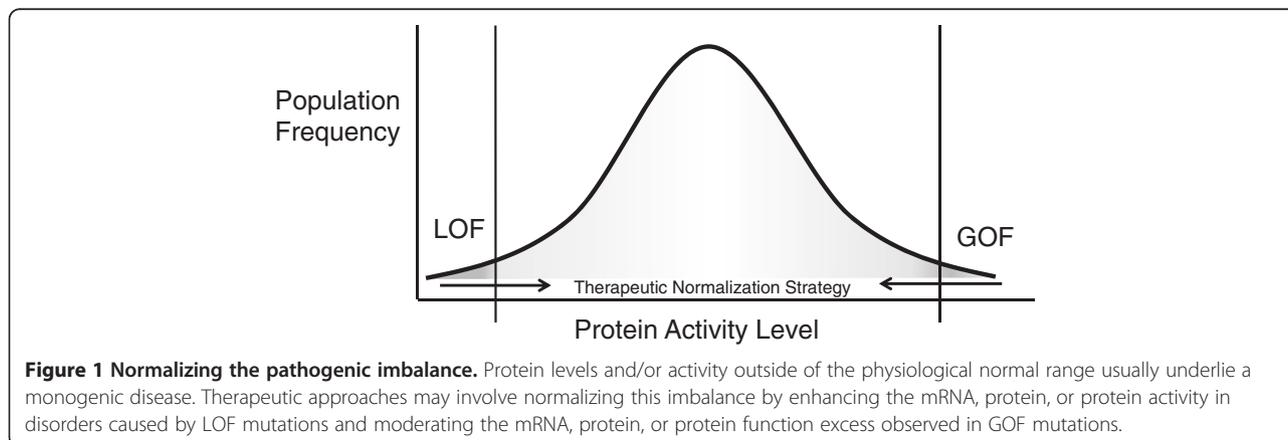
In the case of LOF mutations, a functional agonist may be considered, for example via DNA (e.g., gene therapy) or protein replacement (e.g. enzyme replacement therapy). For partial loss of enzymatic function, exogenously increasing levels of the biochemical substrate may be feasible. For some LOF mutations, it may also be possible, with antisense oligonucleotides (ASO) or small molecules, to increase the rate of transcription, correct RNA mis-splicing, allow read through of a premature stop codon, stabilize the transcript, or stabilize or enhance the activity of the mutated protein itself. Finally, for this class of mutation, there may be the option of activating a wild-type homologous gene, thereby recapitulating the function lost in the primary gene mutation, an approach which we term activating the "rescuing paralog". Conversely, in the case of GOF mutations, a functional antagonist may have clinical value. The therapeutic reduction of the supraphysiologic levels of pathogenic RNA, protein and/or function may be achievable directly by transcriptional or translational inhibition, transcript or protein destabilization, or by direct inhibition of the activity of the protein itself. Alternatively, instead of targeting the gene and its products, the modulation of pathogenically dysregulated pathways by substrate-reduction therapy or product replacement therapy are possible treatments for many LOF and GOF orphan disorders.

Therapeutic approaches for loss-of-function (LOF) mutations
LOF: Pharmacologic suppression of premature termination codons (PTC)
The nature of the specific LOF mutation may suggest specific therapeutic approaches, for example a number of drugs are currently being investigated which result in read-through of premature termination codons (PTCs) [6]. These agents reduce ribosome termination at the PTC, resulting in the insertion of a random amino acid and the translation of the remainder of the correct full-length protein. There are a number of limitations to this approach including reduction of the available transcript because of nonsense-mediated mRNA decay, efficient translation and the activity of the protein containing the random amino acid [6,7]. Nonetheless, read through inducing agents show promise for Duchenne muscular dystrophy (DMD) and cystic fibrosis (CF) [8-10] with the potential of eventually treating a wider range of genetic disorders. The current exemplar of nonsense codon suppression is the small molecule PTC124. PTC124 suppresses nonsense mutations at nanomolar concentrations in mammalian cells (but does not appear to alter global protein or mRNA profiles) and is associated with some phenotypic improvement in mouse models with Duchenne muscular dystrophy and cystic fibrosis [6,10,11]. Both diseases have had PTC124 Phase I trials conducted; Phase II clinical trials of PTC124 administered to cystic fibrosis patients that carry nonsense mutations have also been completed, with restoration of measurable CFTR function in half of treated patients and no appreciable impact on CFTR in a second study [12,13]. More recently, PTC124 treatment improved total chloride transport and showed trends toward improvements in pulmonary function and CF-related coughing in a subset of patients with classic CF phenotype and at least one CFTR nonsense mutation allele [14].

In principle, any nonsense mutation that does not trigger significant nonsense-mediated mRNA decay is a candidate for this approach [15]. A challenge in this field is whether there will be a general *in silico* approach to identify small molecules allowing read-through for a given PTC sequence, or whether painstaking and expensive library screening will be needed to identify an agent for each particular disease-causing PTC. Drugs, such as

Figure 1 Normalizing the pathogenic imbalance. Protein levels and/or activity outside of the physiological normal range usually underlie a monogenic disease. Therapeutic approaches may involve normalizing this imbalance by enhancing the mRNA, protein, or protein activity in disorders caused by LOF mutations and moderating the mRNA, protein, or protein function excess observed in GOF mutations.

PTC124, that permit read-through of a wide spectrum of stop codons clearly carry a risk of off-target side effects. The expectation in the field is that there may exist unifying principles for the targeting, combined with minimal chemical modification and appropriate dosing, to impart sufficient specificity to minimize such side effects.

LOF: Therapeutic oligonucleotide based modulation of splicing

Another class of therapeutic targets comprises the large number of disease-causing mutations occurring in splice site motifs [16]. Such mutations may be intronic or exonic and typically result either in exon skipping and/or use of nearby alternative splice donor or acceptor sites. This results in the deletion of potentially important parts of the encoded protein or, more frequently, exon skipping or mis-splicing resulting in frame shifting errors and resulting aberrant polypeptide with partial or total loss of protein function. Chemically modified oligonucleotides can sterically mask pathogenic splice signals activated by mutation resulting in close to normal levels of mRNA encoding functional protein [17]. Alternatively, in some cases they can be designed to mediate forced splicing exclusion of functionally dispensable in-frame exons containing a pathogenic mutation, resulting in the production of near physiologic levels of internally deleted but partially functional proteins [18]. The application of oligonucleotide-based methods to redirect and modulate pre-mRNA splicing was first demonstrated for β-thalassaemia and now shows promise for both spinal muscular atrophy (SMA) [19] and DMD [20,21]. The design of effective oligonucleotides for this process is undergoing continuous improvement; there now exists both a considerable published body of experience as well as a 'splicing code' that relates genomic features to splicing levels in diverse normal and disease tissues, including predictive analyses of how genomic mutations impact splicing [22,23]. There has been recent progress in utilizing the splicing code to identify disease associated mutations in HGM (unpublished data, Frey); it is hoped that the *in silico* identification of disease-causing splice mutations is a realistic goal for the future. Given the increased predictability of the oligonucleotide action on splicing mutations, a central issue is to determine to what extent the absorption, distribution, metabolism, and excretion- toxicity (ADME-tox) characteristics of a given oligonucleotide shall also be predictable; is oligonucleotide ADME-tox a general drug class effect for a given type of chemical modification, minimizing expensive preclinical assessment or is it sequence-dependent, varying from molecule to molecule [24]? Further understanding of this cause and effect

relationship will emerge as this approach is applied to more disorders.

LOF: Pharmacologic modulation of gene activity, mRNA stability and protein function

Therapeutic upregulation in LOF mutations may be feasible when i) the mutated protein retains some degree of residual function; ii) the disease is caused by a haploinsufficiency; iii) there exists a paralogous normal protein that functionally recapitulates the mutated protein; or iv) there is a protein that mediates a bypass salvage pathway. The general goal in such cases is to increase effective gene function, either directly by modulating the residual expression level or enzymatic activity of the mutated protein itself, or indirectly by upregulating other genes in the genome which themselves may be paralogs of the mutated gene or else active in related biochemical pathways. The introduction of a drug into the complex intracellular and extracellular topography of a human will have, in addition to the impact on its known target, off-target and, from the perspective of orphan genetic disease, potentially beneficial effects. These effects may include the up or down regulation of multiple non-targeted genes, mRNAs or proteins. Indeed, it has been recognized in recent years that small-molecular weight molecules can affect a substantial proportion of the metazoan transcriptome [25,26]. Although such off-target gene modulating effects are currently impossible to predict, the advent of systems biology has begun to permit these effects to be identified with greater effectiveness.

An excellent opportunity to identify beneficial off-target effects comes with novel microarray based databases containing system-wide transcriptome profiles from cell lines grown in the presence of clinically used compounds. This approach has been utilized to catalogue system-wide gene expression profiles elicited by different drugs and drug classes [26-29]. In addition to identifying genomic signatures composed of many transcripts to elucidate drug effects, with these databases one can also ascertain the impact of hundreds of FDA approved drugs on the level of almost every individual transcript in the transcriptome. It is the modulation of individual transcripts that is of particular value for monogenic disorders. For example, we have mined the Johnson and Johnson and Connectivity Map databases [26] for agents that modulate SMN2 mRNA levels, the rescuing paralog for spinal muscular atrophy (SMA; OMIM [253300]), rapidly identifying a role for p38 kinase in the modulation of SMN levels [30]. This approach not only identifies agents that impact transcriptional activity but also those that impact transcript stability as well, such as is seen with p38 activation on SMN2 [30]. The same *in silico* screening approaches can be used to identify agents that upregulate mRNA encoding mutated proteins with residual function, such as seen in milder

variants of a recessive disease; given the low levels of protein frequently observed in these disorders, even a modest increase in activity might be anticipated to have a clinically appreciable effect. Similarly, for disorders caused by haploinsufficiency, upregulation of the remaining normal allele may also yield a clinically beneficial effect. It is likely that industry has additional system-wide transcriptional databases that catalogue the impact of large numbers of preclinical and/or clinically used agents. Similarly there exist hundreds of genes which if significantly modulated would have a good likelihood of impacting rare genetic disease. Thus, one goal is to enable improved access to both public and private datasets to expand the set of genes that are both pharmacologically responsive and clinically relevant (Figure 2).
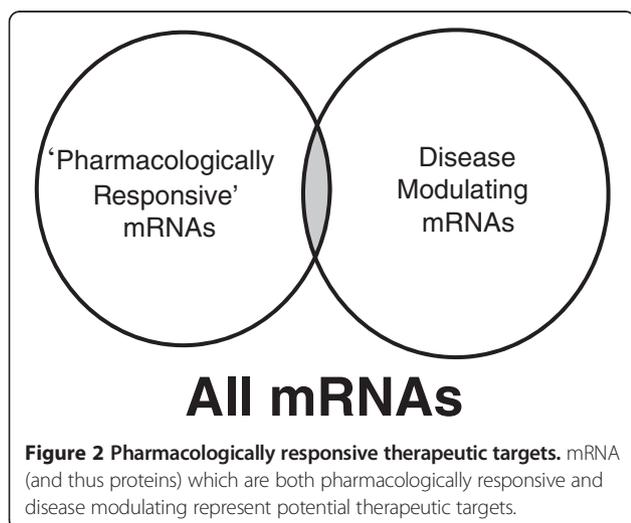
An alternative method of identifying compounds to modify clinically relevant genes is to perform a computational screen for binding sites of pharmacologically inducible transcription factors upstream of the gene of interest. Returning to the example of spinal muscular atrophy, the complementing SMN2 paralog contains in its 5' region a number of STAT5 kinase binding sites; it has been recently shown that the STAT5 activating hormone prolactin both upregulates SMN protein derived from SMN2 and confers significant survival benefit in a mouse model of SMA [31]. The increasing annotation of system-wide transcription factor binding sites via large CHIP-Seq datasets, wedded with knowledge of agents activating these transcription factors augur well for the wider applicability of this approach.

Limitations do exist to the pharmacologic upregulation of gene activity and mRNA level. The potential strength of systems-wide transcriptional profiling comes with the large number of addresses that are sampled; however, the reproducibility of data for a single microarray address is famously unreliable and it is a case of caveat emptor when

utilizing such data. Clearly even before any demonstration of clinical utility, the more independent verification of a given observation from different datasets, the more likely it is to be reliable. Protein modulation is the goal in most rare disease therapeutic approaches and there are not yet systems-wide databases documenting protein levels/activity. Regardless, mRNA serves as a valid proxy for protein level more often than not [32]. In addition, even if a protein:RNA correlation is observed *in vitro*, a given transcript response detected in cell culture may not hold true for a whole organism. It also has to be hoped, when upregulating mutated mRNA and protein, that both are stable and not degraded as can be seen in inherited disorders. Moreover, this approach will not work if the mutated protein has any dominant negative effect. Finally those small molecules which significantly perturb a large proportion of the transcriptome may have a greater incidence of side effects or toxicity than do molecules with milder perturbing effects. For example, histone deacetylase inhibitors such as valproic acid have a significant impact on a substantial proportion of the transcriptome and have simultaneously been suggested for use with a large number of diverse disorders, yet have well documented toxic side effects. The goal of identifying an agent which modulates a specific target in an entirely safe fashion continues to represent a significant challenge.

LOF: In silico cheminformatic and bioinformatic methods for target proteins

The complex topology of proteins, and the lack of a comprehensive algorithmic mapping of primary amino acid sequences to three-dimensional structures, makes generalizable approaches to protein functional modulation problematic. Nevertheless, proteins represent a far more frequent therapeutic target for orphan diseases than mRNA. Diverse *in silico* cheminformatic and bioinformatic methods, including target/ligand-based strategies and systems biology methods, are a means of screening large datasets in the hopes of predicting the binding of molecules to proteins [33-40]. Data integration platforms for systems biology, using both ligand and binding site similarity [36,41,42] can help chaperonin identification, possibly involving drug repositioning. In addition, 2D ligand and 3D protein based approaches employing algorithms and networks, have been used to link molecular structure and biological activity [43-45]. The continual development of machine learning methods and databases for drug repurposing also have promise [46-50]. By connecting data on drugs, proteins and diseases, these various computational methods may enable repurposing of molecules and possibly enable *in vitro* screening efforts for orphan disease therapy [51-55]. Implicit in using these molecular, protein and gene expression databases is the requirement

Figure 2 Pharmacologically responsive therapeutic targets. mRNA (and thus proteins) which are both pharmacologically responsive and disease modulating represent potential therapeutic targets.

that the data are sufficiently free from error so as not to hamper the quality of the prediction [56].

Perhaps the best categorized area of protein modulation is pharmacologic chaperone therapy: small molecules that specifically bind to target proteins modulating folding and/ or stability [57-61]. This approach is currently either being investigated or used for various orphan conditions, particularly lysosomal storage disorders [60,62]. Pharmacological chaperones stabilize the folding of mutant proteins and allow for correct trafficking of the enzyme, often resulting in increased egress from the Golgi to the cytosol or cell surface. Molecules that would be predicted to inhibit the lysosomal enzyme counterintuitively often serve as protein stabilizers and de facto enzyme activators with therapeutic potential; presumably the negative effect of competitive inhibition for substrate binding is more than offset by the positive effect of increased correct protein folding on total enzyme produced [61]. Such observations serve as a key starting point for drug development for other enzymes besides those involved in lysosomal storage diseases; a considerable proportion of molecularly characterized monogenic genetic disorders involve enzymes and present with aberrant, often life-threatening metabolomic profiles (e.g. phenylketonuria).

High content cellular screening for compounds leading to correction of mutated trafficking anomalies, may be one robust route to active pharmacologic chaperones although given the cost of these approaches the preferable scenario would include a relatively common or recurrent causal mutation (which are often seen in local and even large founder populations) and a good *in silico* lead into a compound class.

In this context, RNAi approaches with cellular screens for protein localization may identify currently unappreciated genes that would have a broad impact on folding of mutated polypeptides, irrespective of specific mutation or even particular protein involved. Once identified, altered regulation of such genes by the approaches described above might be clinically beneficial for a wide range of genetic disorders involving missense mutations and aberrant protein folding. Finally, modulation of the specific function of a mutated protein can be obtained, as has recently been documented in a successful clinical trial of CFTR conductance modulators for cystic fibrosis [63].

LOF and GOF: Pathogenic pathway modulation
Regardless of the underlying genetic lesion, there is usually a dysregulation of a biochemical/metabolic/intracellular pathway that is the primary pathogenic driver of a given disorder. Identifying the pathway underlying disease pathogenesis, along with an understanding of the nature of the dysregulation, brings with it the opportunity for therapeutic pathway modulation. The intervention is clearly contingent upon the specific pathway involved and thus is less open to generalizable methods, although the approach may allow for a common intervention to be used for different disease genes that are within the same pathway. Nonetheless, dysregulated pathways may for many disorders, represent the most accessible and relevant therapeutic targets. The pathogenic effect may be as simple as a build-up of a toxic substrate or dearth of a critical product in a metabolic pathway; in such cases, means of reducing the former and increasing the latter are obvious therapeutic approaches. For example, the zinc deficiency observed in acrodermatitis enteropathica can be addressed by dietary supplementation [64,65], while the accumulation of a toxic metabolite may be addressed by dietary means, such as the phenylalanine restriction used to treat phenylketonuria [62]. Where dietary modification is either not possible or insufficient, modulation of the activity of an alternate pathway may be employed as a possible adjunctive therapy. One example is the recessive Tay Sachs disease (TSD), resulting from mutation in the *HEXA* gene encoding the α subunit of the lysosomal enzyme, β-hexosaminidase [66]. The enzyme mediates breakdown of GM2 ganglioside; its loss results in GM2 accumulation and progressive neurodegenerative disease. There exists a metabolic bypass, mediated by Neu1/PPCA or Neu3 sialidase enzymes, which can also break down GM2 ganglioside; thus, the pharmacologic induction of these genes might be therapeutic in TSD [67].

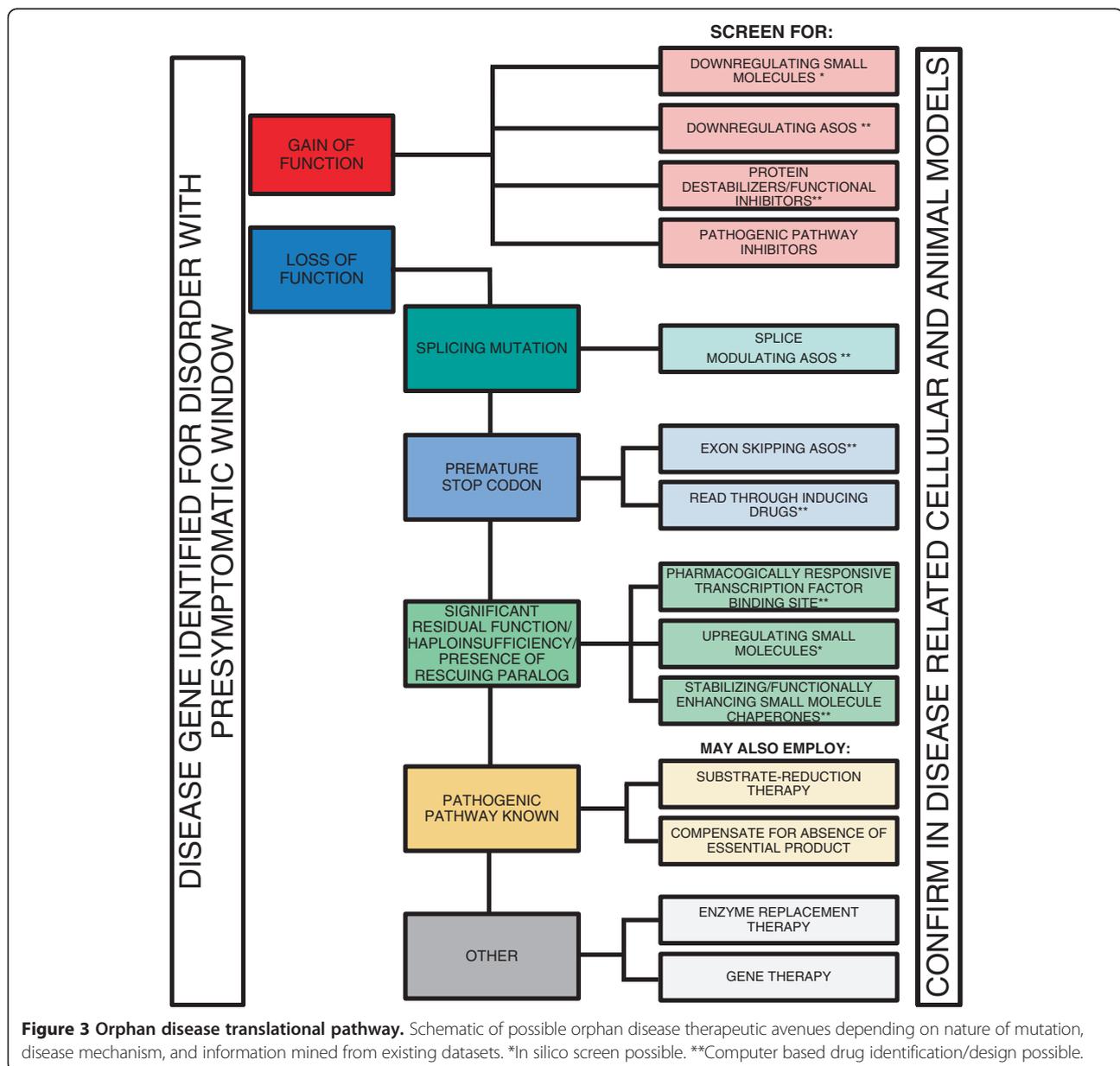### LOF: Other translational approaches
Completing the list of promising orphan disease treatments for LOF mutations are enzyme replacement and gene replacement therapies. Enzyme replacement is a proven approach (although neurologic correction remains elusive due to the impermeant blood brain barrier) and gene therapy after a period of disappointments appears to be coming of age. Unfortunately these approaches remain research intensive and likely prohibitively expensive for the treatment of a large number of very rare disorders. Extensive research is ongoing within these fields and the hope is that they will eventually become more viable options, but they are not the focus of our current translational path.

Therapeutic approaches for gain-of-function (GOF) mutations
XGain-of-function mutations, usually dominantly inherited, are caused by a supra- and thereby pathophysiologic level of mRNA/protein and/or a novel pathologic gain-of-function. There are a number of methods to moderate the mRNA, protein, novel protein function, or pathway activity excess observed in GOF mutations. Oligonucleotides generated using novel second- and third-generation backbone-modifications can reduce RNA and thus protein by at least 11 distinct mechanisms, including the direct sequence-specific steric block by hybridization to pre-mRNA (RNAi), the alteration of

pre-mRNA splicing, hybridization to processed mRNAs thereby preventing ribosome recruitment, and the blockage of protein translation [68]. Oligonucleotides can also inhibit the actions of non-coding RNAs such as microRNAs. In addition to antisense oligonucleotides (ASOs), one can also screen for pharmacologic (in this case down regulatory) modulation of the transcriptome such as outlined above for LOF mutations or screen for pharmacologic direct inhibitors of protein function also using the *in silico* approaches outlined in the previous section. Screens for small molecules binding to the GOF protein itself could be done both through high-throughput screens or *in silico* modeling.

If there is a pathogenic activation of a given pathway, then other means of inhibiting that particular cascade can be attempted. In general terms, this concept was described above in "pathogenic pathway modulation". For GOF disorders, the dominantly inherited kinasopathies represent an instructive disease class in this regard; the majority of these involve a gain of pathogenic function frequently leading to a constitutive activation of a signaling pathway. This has resulted in a rare disease drug repurposing opportunity, as some of these pathways have also been implicated in onco-genesis and are targets for kinase inhibitor programs at pharmaceutical companies. One example is the autosomal dominantly inherited human developmental syndromes

Figure 3 Orphan disease translational pathway. Schematic of possible orphan disease therapeutic avenues depending on nature of mutation, disease mechanism, and information mined from existing datasets. *In silico screen possible. **Computer based drug identification/design possible.

termed the RASopathies caused by (usually) activating mutations of the RAS-RAF-MEK-ERK MAPK pathway [69,70]. The most common of these is Noonan syndrome, characterized by proportional short stature, facial dysmorphia, and cardiovascular abnormalities. A clinical trial with the MEK1/2 inhibitor (also being trialed for several malignancies) for treatment of cardiomyopathy in adults with Noonan syndrome has recently been launched (ClinicalTrials.gov Identifier: NCT01556568); the hope is that other kinase inhibitors will find their way from cancer trials into orphan disease trials.

A schematic outlining the approaches and the proposed translational research path for both LOF and GOF classes of mutation is summarized in Figures 3 and 4.
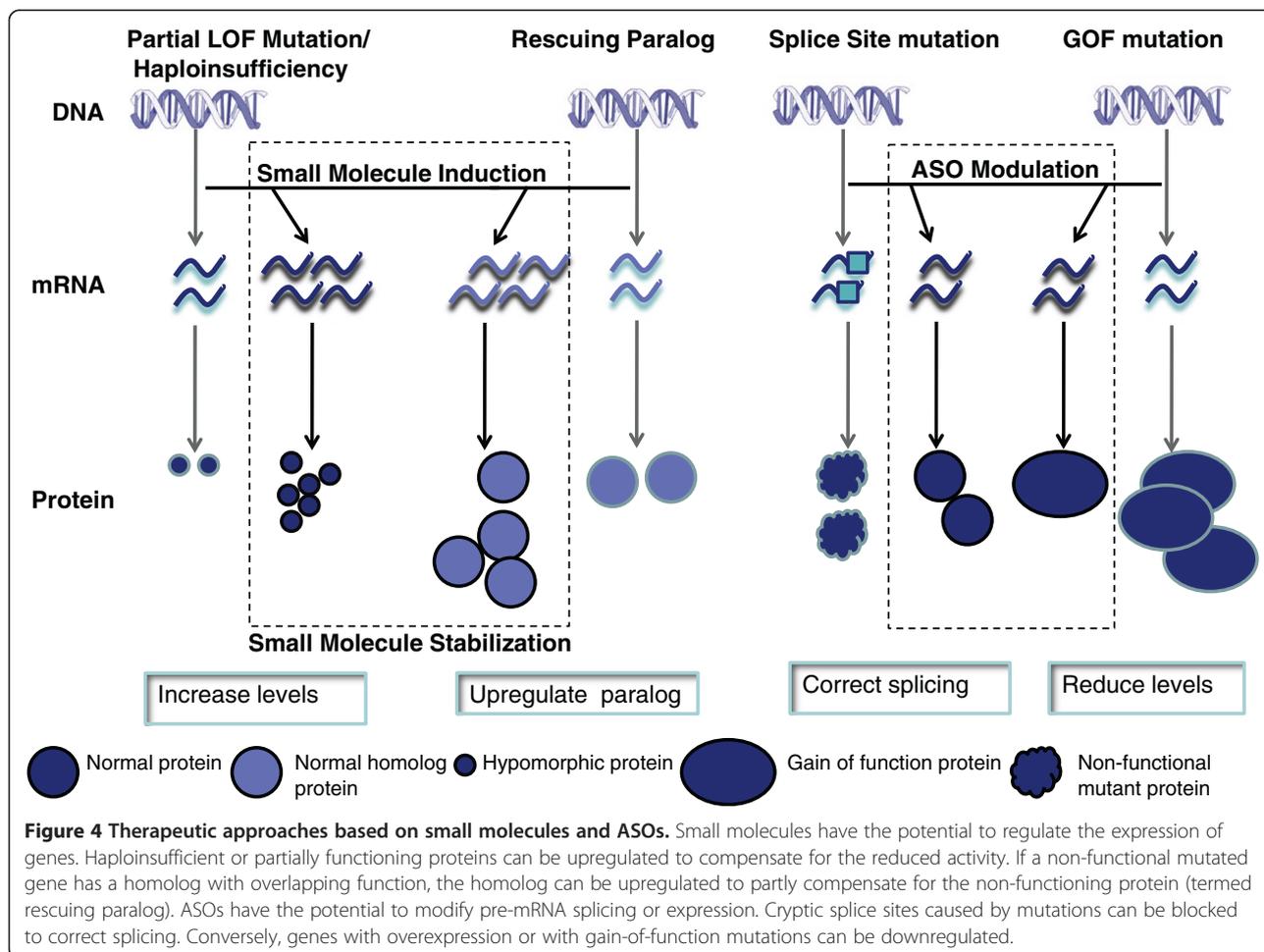
The path following potential therapeutic agent identification

Following the *in silico* identification of a putative therapeutic agent, the confirmation of the proposed effect both in cell cultures and *in vivo* will be required. We have found that this assessment can be done both in wild-type and patient cell lines (using either immortalized fibroblasts or lymphocytes, or with new induced pluripotent cell (iPS) technology) as well as wild-type animals in addition to animal models of a given disorder. For example, an increase of the protein level and/or activity in LOF mutations (or diminution of these parameters for GOF mutations) in wild-type cells and animals can be suggestive of a true effect. The presence of a scorable phenotype in patient cell cultures as well as a biomarker in the animal models (and naturally the human disease), help further in the assessment of the proposed therapy. The summary outlined in Table 1 highlights the various resources that will be of assistance for this phase of disease pathogenesis and preclinical research.

Conclusion

Translational research for rare diseases is clearly a resource intensive undertaking, both in terms of time and real cost. Therefore, as much as possible, we have attempted to define an approach that relies on databases and computational analyses prior to the more expensive

Figure 4 Therapeutic approaches based on small molecules and ASOs. Small molecules have the potential to regulate the expression of genes. Haploinsufficient or partially functioning proteins can be upregulated to compensate for the reduced activity. If a non-functional mutated gene has a homolog with overlapping function, the homolog can be upregulated to partly compensate for the non-functioning protein (termed rescuing paralog). ASOs have the potential to modify pre-mRNA splicing or expression. Cryptic splice sites caused by mutations can be blocked to correct splicing. Conversely, genes with overexpression or with gain-of-function mutations can be downregulated.

Table 1 Translational toolbox: Conditions and resources that will enable effective orphan disease translational research

Critical Information	Known Gene
	Known Inheritance (dominant or recessive)
	Known Mechanism (LOF or GOF)
Helpful Information	Existence of a presymptomatic window or likeliness of clinical reversibility
	Dysregulated pathway known
	Diagnostic assay associated with primary defect available
	Purified protein available
	Protein crystal structure known
	Antibody directed against protein available
	Previous screen related to disease gene exists (screen for pharmacologic modulation, ASO screening, etc)
	Gene under control of transcription factors responsive to drugs
	Scorable cell culture phenotype exists
	Animal or other model available
	Scorable biomarker reflecting disease state exists (metabolomic, transcriptomic marker, etc)
	Disease management protocol in use

experimental validation of potential therapies (examples of potential new therapies for rare diseases identified using these approaches are listed in Table 2). This is especially important when a disease is so rare as to render the prospects of any commercial profit from a therapy, regardless of its effectiveness, unlikely. Thus, an important issue is the question of generalization of approach and minimization of costs. Key steps in this direction include access to system-wide datasets, compounds and reagents for the orphan disorder research community, advances in both systems biology and computational prediction of small molecule-macromolecule interaction, the identification of additional generalizable therapeutic approaches and ultimately more collaboration.

Accessibility to a greater pharmacophoric compound library space, data currently existent in pharmaceutical companies and readily available reagents

There currently exists, in major pharmaceutical companies, extensive databases describing the system-wide transcriptional response to thousands of compounds, the vast majority of which have not seen clinical use. Although most of these compounds are designed for proteases, kinases and G-protein coupled receptors, mining these datasets and subsequent access to those compounds that modulate orphan disease related transcripts might serve as a source of rare disorder therapeutic leads. In addition there does not yet exist a definitive gold standard transcriptomal database to mine computationally for FDA approved drugs; the configuration of such an entity would be of significant value. Finally, but of no less importance, is the key set of reagents that needs to be generated for as many orphan diseases as possible to encourage research in that area by the community. Purified protein, crystal structures, antibodies as well as cellular and animal models, will both help advance the prospects of meaningful translational research as well as enlist new research teams into the field of rare genetic diseases.

Computational prediction of small molecule-macromolecule interaction

Computational advances are also needed in the prediction of small molecule-RNA/protein interactions including the identification of agents that suppress PTCs, bind and then stabilize, destabilize or antagonize protein activity. Improved prediction of the splicing impact of relevant genomic mutations and of optimal antisense oligonucleotides that may reverse the effects of such mutations would also be of value.

Additional generalizable therapeutic approaches

A key aspect of any broadly successful approach for this class of disease will be to conceive and employ

Table 2 Examples of computational technologies used for rare disease drug discovery

Strategy	Rare disease	Computational Technology	Drug	Reference
Small Molecule Upregulation	Spinal Muscular Atrophy	Connectivity Map	Anisomycin	[30]
Small Molecule Upregulation	Spinal Muscular Atrophy	Transcription factor binding site identification	Prolactin	[31]
Chaperone: drug safety predictions	Gaucher disease	Leadscope	Core structures of aminoquinoline, sulfonamide, and triazine	[71]
Chaperone: identify binding sites and compounds	Huntington disease	AutoDock, Patch Dock Server, CastP	Metoprolol, minocyclines, and 18 F fluorodeoxyglucose	[72]
Drug similarity predictions	Neurodegenerative disorders due to protein misfolding	Mode of Action by Network Analysis, MANTRA	Fasudil	[27]
Prediction of which mutations respond to treatment	Fabry disease	Position specific substitution matrix	1-deoxy-galactonojirimycin	[73,74]

generalizable methodologies wherever possible. Some examples of these include the possibility of generalizing ADME-toxicology for oligonucleotides, identifying novel broad spectrum PTC suppressor compound classes or possibly identifying a druggable pathway that allows persistence of higher levels of non-toxic, partially functional misfolded proteins.

While the approaches we propose here are credible and feasible, the prospect of a rapid configuration of numerous effective orphan disease therapies should be viewed in perspective. Two of the most "common" rare diseases, CF and DMD, have been the subject of many years and millions of dollars of pre-clinical and clinical assessment; indeed many of the approaches outlined above have been pioneered in the analysis of these disorders, yet there is today still not an effective therapy in routine clinical use for either disorder. Nonetheless, we hope that among the many genetic disorders that have been or will shortly be molecularly characterized; there will be some that are tractable to the approaches reviewed here. At a minimum, the generation of a standardized toolbox will help to move a larger number of disorders closer to the day of effective therapy.

Abbreviations
ADME-tox: Absorption distribution, metabolism, and excretion – toxicity; ASO: Antisense oligonucleotide; CF: Cystic Fibrosis; DMD: Duchenne muscular dystrophy; GOF: Gain-of-function; LOF: Loss-of-function; PTC: Premature termination codon; SMA: Spinal muscular atrophy; TSD: Tay Sachs disease.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed to the literature review and the writing of the article. All authors read and approved the final manuscript.

Author details
[1]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON, Canada. [2]Centre de Recherche du CHU-Ste-Justine and Department of Medicine, University of Montreal, Montreal, Canada. [3]Collaborations in Chemistry, Fuquay Varina, NC, USA. [4]Atlantic Research Centre, Dalhousie University, Halifax, NS, Canada. [5]Structural Genomics Consortium, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada. [6]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [7]Manitoba Institute of Cell Biology, University of Manitoba, Manitoba, MB, Canada. [8]Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, ON K1H 8 L1, Canada.

References
1. McCandless SE, Brunger JW, Cassidy SB: **The burden of genetic disease on inpatient care in a children's hospital.** *Am J Hum Genet* 2004, **74**:121–127.
2. Scriver CR, Neal JL, Saginur R, Clow A: **The frequency of genetic disease and congenital malformation among patients in a pediatric hospital.** *Can Med Assoc J* 1973, **108**:1111–1115.
3. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders.** *Nucl Acids Res* 2005, **33**:D514–D517.
4. Cooper DN, Chen JM, Ball EV, Howells K, Mort M, Phillips AD, Chuzhanova N, Krawczak M, Kehrer-Sawatzki H, Stenson PD: **Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics.** *Hum Mutat* 2010, **31**:631–655.
5. Kingsmore SF, Dinwiddie DL, Miller NA, Soden SE, Saunders CJ: **Adopting orphans: comprehensive genetic testing of Mendelian diseases of childhood by next-generation sequencing.** *Expert Rev Mol Diagn* 2011, **11**:855–868.
6. Keeling KM, Bedwell DM: **Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases.** *Wiley Interdisciplinary Reviews: RNA* 2011, **2**:837–852.
7. Dietz HC: **New therapeutic approaches to mendelian disorders.** *N Engl J Med* 2010, **363**:852–863.
8. Finkel RS: **Read-through strategies for suppression of nonsense mutations in Duchenne/Becker muscular dystrophy: aminoglycosides and ataluren (PTC124).** *J Child Neurol* 2010, **25**:1158–1164.
9. Sermet-Gaudelus I, Boeck KD, Casimir GJ, Vermeulen F, Leal T, Mogenet A, Roussel D, Fritsch J, Hanssens L, Hirawat S: **Ataluren (PTC124) induces cystic fibrosis transmembrane conductance regulator protein expression and activity in children with nonsense mutation cystic fibrosis.** *Am J Respir Crit Care Med* 2010, **182**:1262–1272.
10. Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, Paushkin S, Patel M, Trotta CR, Hwang S: **PTC124 targets genetic disorders caused by nonsense mutations.** *Nature* 2007, **447**:87–91.
11. Du M, Liu X, Welch EM, Hirawat S, Peltz SW, Bedwell DM: **PTC124 is an orally bioavailable compound that promotes suppression of the human CFTR-G542X nonsense allele in a CF mouse model.** *PNAS* 2008, **105**:2064–2069.
12. Clancy J, Konstan M, Rowe S, Accurso F, Zeitlin P, Moss R, Bebok Z, Northcutt V, Elfring G, Miller L: **A phase II study of PTC124 in CF patients harboring premature stop mutations [abstract].** *Ped Pulmonol Suppl* 2006, **41**:A269.
13. Kerem E, Hirawat S, Armoni S, Yaakov Y, Shoseyov D, Cohen M, Nissim-Rafinia M, Blau H, Rivlin J, Aviram M: **Effectiveness of PTC124 treatment of cystic fibrosis caused by nonsense mutations: a prospective phase II trial.** *Lancet* 2008, **372**:719–727.
14. Wilschanski M, Miller L, Shoseyov D, Blau H, Rivlin J, Aviram M, Cohen M, Armoni S, Yaakov Y, Pugatch T: **Chronic ataluren (PTC124) treatment of nonsense mutation cystic fibrosis.** *Eur Respir J* 2011, **38**:59–69.
15. Frischmeyer PA, Dietz HC: **Nonsense-mediated mRNA decay in health and disease.** *Hum Mol Genet* 1999, **8**:1893–1900.
16. Wang GS, Cooper TA: **Splicing in disease: disruption of the splicing code and the decoding machinery.** *Nat Rev Genet* 2007, **8**:749–761.
17. Wilton SD, Fletcher S: **RNA splicing manipulation: strategies to modify gene expression for a variety of therapeutic outcomes.** *Curr Gene Ther* 2011, **11**:259–275.
18. Wilton SD, Fletcher S: **RNA splicing manipulation: strategies to modify gene expression for a variety of therapeutic outcomes.** *Curr Gene Ther* 2005, **5**:467–483.
19. Hua Y, Sahashi K, Rigo F, Hung G, Horev G, Bennett CF, Krainer AR: **Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model.** *Nature* 2011, **478**:123–126.
20. Cirak S, Arechavala-Gomeza V, Guglieri M, Feng L, Torelli S, Anthony K, Abbs S, Garralda ME, Bourke J, Wells DJ, et al: **Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study.** *Lancet* 2011, **378**:595–605.
21. Aartsma-Rus A, Fokkema I, Verschuuren J, Ginjaar I, van Deutekom J, van Ommen GJ, den Dunnen JT: **Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations.** *Hum Mutat* 2009, **30**:293–299.

22. Barash Y, Blencowe BJ, Frey BJ: **Model-based detection of alternative splicing signals.** *Bioinformatics* 2010, **26**:i325–i333.
23. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ: **Deciphering the splicing code.** *Nature* 2010, **465**:53–59.
24. Jason TLH, Koropatnick J, Berg RW: **Toxicology of antisense therapeutics.** *Toxicol Appl Pharmacol* 2004, **201**:66–83.
25. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD: **Functional discovery via a compendium of expression profiles.** *Cell* 2000, **102**:109–126.
26. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN: **The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.** *Science* 2006, **313**:1929–1935.
27. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, Murino L, Tagliaferri R, Brunetti-Pierri N, Isacchi A, di Bernardo D: **Discovery of drug mode of action and drug repositioning from transcriptional responses.** *PNAS* 2010, **107**:14621–14626.
28. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, Sage J, Butte AJ: **Discovery and preclinical validation of drug indications using compendia of public gene expression data.** *Sci Transl Med* 2011, **3**(96ra):77.
29. Lamb J: **The connectivity map: a new tool for biomedical research.** *Nat Rev Cancer* 2007, **7**:54–60.
30. Farooq F, Balabanian S, Liu X, Holcik M, MacKenzie A: **p38 Mitogen-activated protein kinase stabilizes SMN mRNA through RNA binding protein HuR.** *Hum Mol Genet* 2009, **18**:4035–4045.
31. Farooq F, Molina FA, Hadwen J, MacKenzie D, Witherspoon L, Osmond M, Holcik M, MacKenzie A: **Prolactin increases SMN expression and survival in a mouse model of severe spinal muscular atrophy via the STAT5 pathway.** *J Clin Invest* 2011, **121**:3042–3050.
32. Maier T, Guell M, Serrano L: **Correlation of mRNA and protein in complex biological samples.** *FEBS Lett* 2009, **583**:3966–3973.
33. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL: **Global mapping of pharmacological space.** *Nat Biotechnol* 2006, **24**:805–815.
34. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M: **DrugBank: a knowledgebase for drugs, drug actions and drug targets.** *Nucleic Acids Res* 2008, **36**:D901–D906.
35. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M: **Drug-target network.** *Nat Biotechnol* 2007, **25**:1119–1126.
36. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK: **Relating protein pharmacology by ligand chemistry.** *Nat Biotechnol* 2007, **25**:197–206.
37. Metz JT, Hajduk PJ: **Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks.** *Curr Opin Chem Biol* 2010, **14**:498–504.
38. Cases M, Mestres J: **A chemogenomic approach to drug discovery: focus on cardiovascular diseases.** *Drug Discov Today* 2009, **14**:479–485.
39. Berg EL, Yang J, Melrose J, Nguyen D, Privat S, Rosler E, Kunkel EJ, Ekins S: **Chemical target and pathway toxicity mechanisms defined in primary human cell systems.** *J Pharmacol Toxicol Methods* 2010, **61**:3–15.
40. Li YY, An J, Jones SJ: **A computational approach to finding novel targets for existing drugs.** *PLoS Comput Biol* 2011, **7**:e1002139.
41. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kuijer MB, Matos RC, Tran TB, *et al*: **Predicting new molecular targets for known drugs.** *Nature* 2009, **462**:175–181.
42. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE: **Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis.** *PLoS Comput Biol* 2009, **5**:e1000423.
43. Fliri AF, Loging WT, Thadeio PF, Volkmann RA: **Biological spectra analysis: linking biological activity profiles to molecular structure.** *PNAS* 2005, **102**:261–266.
44. Fliri AF, Loging WT, Thadeio PF, Volkmann RA: **Biospectra analysis: model proteome characterizations for linking molecular structure and biological response.** *J Med Chem* 2005, **48**:6918–6925.
45. Fliri AF, Loging WT, Thadeio PF, Volkmann RA: **Analysis of drug-induced effect patterns to link structure and side effects of medicines.** *Nat Chem Biol* 2005, **1**:389–397.
46. Gurulingappa H, Kolarik C, Hofmann-Apitius M, Fluck J: **Concept-based semi-automatic classification of drugs.** *J Chem Inf Model* 2009, **49**:1986–1992.
47. Chiang AP, Butte AJ: **Systematic evaluation of drug-disease relationships to identify leads for novel drug uses.** *Clin Pharmacol Ther* 2009, **86**:507–510.

48. von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R: **PROMISCUOUS: a database for network-based drug-repositioning.** *Nucleic Acids Res* 2011, **39**:D1060–1066.
49. Ha S, Seo YJ, Kwon MS, Chang BH, Han CK, Yoon JH: **IDMap: facilitating the detection of potential leads with therapeutic targets.** *Bioinformatics* 2008, **24**:1413–1415.
50. Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, Bourne PE: **A machine learning-based method to improve docking scoring functions and its application to drug repurposing.** *J Chem Inf Model* 2011, **51**:408–419.
51. Chong CR, Sullivan DJ Jr: **New uses for old drugs.** *Nature* 2007, **448**:645–646.
52. Jensen NH, Roth BL: **Massively parallel screening of the receptorome.** *Comb Chem High Throughput Screen* 2008, **11**:420–426.
53. Strachan RT, Ferrara G, Roth BL: **Screening the receptorome: an efficient approach for drug discovery and target validation.** *Drug Discov Today* 2006, **11**:708–716.
54. O'Connor KA, Roth BL: **Finding new tricks for old drugs: an efficient route for public-sector drug discovery.** *Nat Rev Drug Discov* 2005, **4**:1005–1014.
55. Roth BL, Lopez E, Beischel S, Westkaemper RB, Evans JM: **Screening the receptorome to discover the molecular targets for plant-derived psychoactive compounds: a novel approach for CNS drug discovery.** *Pharmacol Ther* 2004, **102**:99–110.
56. Williams AJ, Ekins S: **A quality alert and call for improved curation of public chemistry databases.** *Drug Disc Today* 2011, **16**:747–750.
57. Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ: **Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc.** *J Am Chem Soc* 2009, **131**:7390–7401.
58. Ulloa-Aguirre A, Janovick JA, Brothers SP, Conn PM: **Pharmacologic rescue of conformationally-defective proteins: implications for the treatment of human disease.** *Traffic* 2004, **5**:821–837.
59. Dunker AK, Silman I, Uversky VN, Sussman JL: **Function and structure of inherently disordered proteins.** *Curr Opin Struct Biol* 2008, **18**:756–764.
60. Valenzano KJ, Khanna R, Powe AC, Boyd R, Lee G, Flanagan JJ, Benjamin ER: **Identification and characterization of pharmacological chaperones to correct enzyme deficiencies in lysosomal storage disorders.** *Assay Drug Dev Technol* 2011, **9**:213–235.
61. Fan JQ: **A counterintuitive approach to treat enzyme deficiencies: use of enzyme inhibitors for restoring mutant enzyme activity.** *Biol Chem* 2008, **389**:1–11.
62. Feldhammer M, Durand S, Pshezhetsky AV: **Protein misfolding as an underlying molecular defect in mucopolysaccharidosis III type C.** *PLoS One* 2009, **4**:e7434.
63. Accurso FJ, Rowe SM, Clancy JP, Boyle MP, Dunitz JM, Durie PR, Sagel SD, Hornick DB, Konstan MW, Donaldson SH, *et al*: **Effect of VX-770 in persons with cystic fibrosis and the G551D-CFTR mutation.** *N Engl J Med* 2010, **363**:1991–2003.
64. Kury S, Dreno B, Bezieau S, Giraudet S, Kharfi M, Kamoun R, Moisan JP: **Identification of SLC39A4, a gene involved in acrodermatitis enteropathica.** *Nat Genet* 2002, **31**:239–240.
65. Neldner KH, Hambidge KM: **Zinc therapy of acrodermatitis enteropathica.** *N Engl J Med* 1975, **292**:879–882.
66. Gravel RA, Kaback M, Proia RL, Sandhoff K, Suzuki K, Suzuki K: *(Eds): The Metabolic and Molecular Bases of Inherited Disease.* New York: McGraw-Hill; 2001.
67. Igdoura SA, Mertineit C, Trasler JM, Gravel RA: **Sialidase-mediated depletion of GM2 ganglioside in Tay-Sachs neuroglia cells.** *Hum Mol Genet* 1999, **8**:1111–1116.
68. Sepp-Lorenzino L, Ruddy M: **Challenges and opportunities for local and systemic delivery of siRNA and antisense oligonucleotides.** *Clin Pharmacol Ther* 2008, **84**:628–632.
69. Tidyman WE, Rauen KA: **The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation.** *Curr Opin Genet Dev* 2009, **19**:230–236.
70. Rauen KA, Schoyer L, McCormick F, Lin AE, Allanson JE, Stevenson DA, Gripp KW, Neri G, Carey JC, Legius E: **Proceedings from the 2009 genetic syndromes of the Ras/MAPK pathway: From bedside to bench and back.** *Am J Med Genet A* 2010, **152**:4–24.
71. Zheng W, Padia J, Urban DJ, Jadhav A, Goker-Alpan O, Simeonov A, Goldin E, Auld D, LaMarca ME, Inglese J, *et al*: **Three classes of glucocerebrosidase inhibitors identified by quantitative**

high-throughput screening are chaperone leads for Gaucher disease. *PNAS* 2007, **104**:13192–13197.

72. Gupta S, Misra G, Pant MC, Seth PK: Prediction of a new surface binding pocket and evaluation of inhibitors against huntingtin interacting protein 14: an insight using docking studies. *J Mol Model* 2011, **17**:3047–3056.

73. Andreotti G, Citro V, De Crescenzo A, Orlando P, Cammisa M, Correra A, Cubellis MV: Therapy of Fabry disease with pharmacological chaperones: from in silico predictions to in vitro tests. *Orphanet J Rare Dis* 2011, **6**:66.

74. Andreotti G, Guarracino MR, Cammisa M, Correra A, Cubellis MV: Prediction of the responsiveness to pharmacological chaperones: lysosomal human alpha-galactosidase, a case of study. *Orphanet J Rare Dis* 2010, **5**:36.