



An open science rare diseases research initiative: the University of North Carolina Catalyst

Kaleb Naegeli^{1,2}, Tammy Havener^{1,2}, Wen Yih Aw^{1,2}, Ed Anderson^{1,2} & Dave Morris^{*,1,2}

¹UNC Catalyst for Rare Diseases, 120 Mason Farm Road, Chapel Hill, 27599, NC

²UNC Eshelman School of Pharmacy, 120 Mason Farm Road, Chapel Hill, 27599, NC

*Author for correspondence: dave.morris@unc.edu

First draft submitted: 8 January 2019; Accepted for publication: 31 January 2019; Published online: 19 March 2019

Keywords: drug discovery • drug re-purposing • functional genomics • gene editing • open science • rare disease

The past decade has seen tremendous advances in our ability to sequence and characterize human genetic code. Rare disease research has benefited significantly from this progress and the more recent availability of cellular and molecular tools that enable the editing of DNA has made it possible to explore the functional impact of human genetic mutations at a basic biology level. Furthermore, the availability of chemical probes and bioanalytical technologies has the potential to delineate more accurately the genotype to phenotype translation in cellular models. That knowledge is in turn informing how best to approach a therapeutic intervention strategy for rare diseases. This article describes the initial experiences and learning from the UNC Catalyst for Rare Diseases (NC, USA), a research-focused group tasked with this mission.

Background

The UNC Catalyst for Rare Disease research ('Catalyst') was funded in 2016 by the Eshelman Institute for Innovation following an internal competitive proposal submission and review process. The group was established as a precompetitive open access research group with a mission to understand the functional phenotypic manifestations of genetic mutations. To identify specific diseases where the Catalyst could have substantial scientific impact, partnerships were established with the Genetic Alliance, a rare disease patient advocacy organization based in Washington DC, USA [1], and the Structural Genomics Consortium (SGC) group at the University of North Carolina (UNC) in Chapel Hill (NC, USA). The SGC is a precompetitive private/public partnership that operates using an open science model of collaboration to accelerate drug discovery [2]. This way of working was adopted by the Catalyst with a similar aim of driving research and the discovery of therapeutics for rare diseases. Disease-specific target enabling packages ('TEPs') of biological and chemical tools are a primary deliverable from the more than 250 scientists now working at six SGC centers. The UNC-based SGC group is a team of chemical biologists with a substantial history of developing target class-specific modulators for therapeutic programs at a major Pharma company [3,4]. Recent work by this group has focused on the discovery of kinase-specific inhibitors together with the creation of a kinase chemogenomic set [4,5]. The latter is a collection of compounds, each with a primary kinase target and limited cross-reactivity with other members of the kinome. This set of molecules represents a unique and invaluable resource for investigators who are exploring the biology associated with a specific kinase target. Additionally, UNC at Chapel Hill has an outstanding clinical research faculty and over 20 biomedical science core groups that provide support for projects requiring help with cloning, protein production and scale up, gene sequencing, proteomics, metabolomics, imaging, stem cell procurement and disease models.

Information & evaluation

Using data provided by the team at Genetic Alliance, search terms used by NORD/GARD/NCATS and a Google/SCOPUS-driven analysis of the literature, we compiled a list of rare diseases for which there is a reasonable understanding of genotype and the associated human phenotype. We applied several evaluation frameworks to map the rare disease landscape at a basic science level [6]. One involved the characterization of diseases using

a cell, matrix or organellar alignment – for example, cilia, connective tissue proteins, intermediate filaments, lysosomal enzymes or mitochondrial proteins. Other frameworks reflected the drug discovery experience resident in the group with considerations of target class, signaling pathways, availability of phenotypic assays, chemical probes and data covering off-target activity/promiscuity. The resulting dossier contained information on the disease such as the specific target/pathway, documented mutations, key (seminal) references (if available), reagent availability (proteins, assays, antibodies, chemical tools), animal models/model organisms, publications, funded labs (NIH Reporter search), disease foundations, repurposing/repositioning efforts and modality options for cure or amelioration (e.g., alternate effector pathway target; protein replacement [peptide, gene therapy]). In a related project, an information repository will house summaries of this rare disease information and be available to the public.

Generation of rare disease TEPs

Since the inception of the Catalyst in 2016, several diseases that were evaluated using the aforementioned analysis have received research support from the staff and postdoctoral fellows working in the group. In all cases, each example of disease engagement has led to the generation of specific sets of research tools to enable experimental interrogation (TEPs).

The first example involves autosomal recessive spastic ataxia of Charlevoix Saguenay (ARSACS), which is a progressive disease affecting muscle function throughout the body. ARSACS was originally described in Quebec, where the incidence is 1 in 1900. ARSACS is caused by the functional loss of the saccin protein, which is highly expressed in neurons, with lower expression in skeletal muscle and skin. *Saccin* is one of the largest genes in the human genome, which complicates experiments aimed at understanding saccin's role in normal development and disease. However, homologous protein domains in saccin suggest the protein plays a role as a molecular chaperone and scaffold, and early studies in ARSACS patient fibroblasts have found deficits with cytoskeletal proteins, the proteasome and mitochondrial proteins [7]. In addition, saccin has been identified as a substrate of the ubiquitin ligase Ube3a, which is implicated in autism spectrum disorder, and is the causal gene in Angelman syndrome, suggesting saccin misexpression may contribute to other neurological disorders [8]. However, little is known about the role of saccin in postmitotic neurons, where its expression is highest. To that end, the Catalyst has generated modified neuronal cells in which saccin has been mutated through gene editing procedures. Experiments are currently in progress to analyze how loss of functional saccin protein in these cells impacts kinase signaling through the use of a proteomic analysis platform [9].

Another example is vascular Ehlers-Danlos Syndrome (vEDS), a rare disease resulting from mutations in fibrillar collagen *COL3A1*. Afflicted individuals exhibit thinner arterial walls that are prone to spontaneous aneurysm and dissection, resulting in markedly decreased lifespan. The standard of care is management of blood pressure, particularly with celiprolol, a cardioselective β -adrenergic receptor inhibitor shown to reduce incidence of major deleterious effects of this rare disease [10]. However, the particular unique mechanism of celiprolol and the underlying molecular mechanisms of vEDS are poorly understood. Furthermore, vEDS lacks established *in vitro* models necessary for elucidation of biomarkers and pathways as well as screening of potential therapeutics.

The Catalyst model afforded an opportunity to advance vEDS research by developing research methods and systems to discover vEDS biological mechanisms. We have utilized a screening platform in primary cells to identify celiprolol-regulated biomarkers that appear to be unique to celiprolol among β -adrenergic receptor inhibitors; this may provide insight into the mechanism by which this compound confers benefit to patients. Additionally, the postdoctoral scholars of the Catalyst for Rare Diseases have been able to begin building an *in vitro* model system for vEDS by identifying cell culture systems that can generate biomarker and physiological readouts representative of the disease. These basic science initiatives begun by the Catalyst will be shared with the entire vEDS research community to accelerate the pace of discovery and therapeutic development.

A final example covers a partnership with a clinical geneticist at the UNC Children's Hospital who specializes in the etiologies of rare otological disorders. A major aim of this collaboration is to characterize the phenotype variations in rare mutations underlying hereditary hearing loss [11]. The advent of next-generation sequencing technologies has revolutionized our ability to identify rare genetic variants underlying diseases. With the ever-increasing pace of genomic variants discovery, there is a paucity of data on understanding the relationship between genotypic changes and molecular/phenotype variations – this complicates the translation of genetics information to clinical practice. Understanding how certain mutations can exhibit phenotypic differences in hearing loss will be critical for the development of genetic counseling and testing guidelines for patients.

In terms of the functional analysis objective, the Catalyst scientists recognized that a number of model organisms including mouse, *Drosophila melanogaster*, *Caenorhabditis elegans* and zebrafish have been widely employed to study the phenotypes arising from disease mutations [12–14]. Thanks to rapid advances in CRISPR-mediated genome editing, the generation of organisms containing precise orthologs of human disease mutations is now an accessible approach to disease modeling. Mouse models can allow exploration of poorly characterized disease physiology and the testing of focused therapeutic hypothesis. Many genetic disorders affect genes conserved across invertebrate model organisms, which can be used for powerful genetic or small molecule screens that allow unbiased discovery of disease pathways and potential targets. Furthermore, screens in *Drosophila* and *C. elegans* can be based on highly relevant whole-organism phenotypes, such as locomotion in the case of neurodegenerative disorders. The relative merits of available model organisms should be carefully considered for each disease of interest and the Catalyst intends to establish collaborations with experts in these models to help extend findings beyond the *in vitro* approach.

Mutation impact

One of the goals of the UNC Catalyst is to try and understand in greater depth the impact of a gene mutation on intracellular signaling and cell function. We have noted with interest the use of this approach [15,16], as well as the development of platform-based methodologies to understand the functional consequences of mutations in genes linked with cancer [17]. More recently, CRISPR-Cas9 screening approaches have been used to identify cancer targets and genetic interaction networks [18]. These screens could be a very powerful method for identifying genetic suppressors of disease-causing mutations, which would be high-priority drug targets. Expression phenotyping has been used to characterize and determine if specific mutations alter resistance to treatment with specific kinase inhibitors [17]. Our interest in and pursuit of the ‘ripple effect’ of a gene mutation at the level of kinase signaling is driven by several factors. First, kinases represent a proven family of druggable targets; second, UNC’s proteomic core group has developed a profiling platform that utilizes bead-linked kinase inhibitors and mass spectrophotometric analysis to determine the identity of kinases functionally active in biological samples [9]. Finally, our SGC partners at UNC have built specific kinase inhibitors for therapeutic programs and shared sets of these compounds with investigators who have unique biological assays that can be used for molecule screening and profiling. In combination, these capabilities effectively represent the tools we need in order to detect and follow-up on changes in kinase signaling in cells in the presence and absence of a specific gene mutation. Importantly, the reagents, chemical tools and knowledge relating to these studies are freely available.

Repurposing/repositioning

A number of recent publications have elaborated on the value of drug repurposing in the context of rare diseases [19,20]. It is beyond the scope of this commentary to go into great depth on this topic and it is fair to say that advances in information science-based analyses have revealed putative relationships between molecules functional for one indication and biological pathway promiscuity, opening possibilities for alternate disease indications. Additionally, it is likely that many small molecule-focused companies harbor ‘shelf portfolios’ of molecules which, although safe, have not progressed to the clinic for portfolio alignment or strategic fit reasons. This begs the question as to how the value of these compounds can be realized. Some organizations and companies have taken the initiative to invite partnerships with external groups to help determine how these molecules could provide therapeutic value [21,22]. It will be important to continue these types of collaborative arrangements, especially when the biological assays that can define value are spread across a number of academic groups working in a specific disease area. The systematic screening of these nonprogressed compounds across a broad swath of biological targets would be a very useful experiment to run in the precompetitive discovery space.

Tools required/needed

Our early experiences with a small set of rare disease research projects have highlighted the need for a specific set of tools that can delineate the molecular and cellular mechanisms of a rare disease in order to accelerate biological discovery and therapeutic development. Such tools include the generation of cell lines and *in vitro* models (CRISPR, patient-derived samples), -omics technologies of gene and protein expression (RNA-seq, ChIP-seq, ATAC-seq, CoIP, etc.) and small molecules or therapeutics that can acutely target signaling pathways.

CRISPR-Cas9 screening approaches have been used to identify cancer targets and genetic interaction networks [18]. Precise implementation of these methods could be utilized to generate rare disease tools such as cell lines

and animal models while ‘blunt’ screening approaches could be employed to identify suppressors of disease-causing mutations, potentially opening therapeutic windows for rare diseases.

Tools that need to be developed further by the scientific community would include *in vitro* recapitulation of 3D tissue environments (heterogeneity of tissues, cell–ECM interactions, etc.), humanized mouse models (or other model systems) and assays to connect cell types and genetic variants to disease phenotypes. Some of these systems exist for big diseases such as cancer and neurodegenerative conditions but need to be improved and honed to be optimally applicable to rare diseases.

Another valuable resource would be patient-derived iPSCs and a standardized set of protocols or platforms for culturing and differentiating iPSCs, so that data from each iPSC disease modeling publication are comparable and replicable. Additionally, the ability to generate iPSCs with engineered recombinase sites around a locus of interest could be very useful. Recombinase-mediated cassette exchange would permit experiments with large allelic series in true isogenic background and preclude the costs and potential for suppressor mutations with CRISPR knock-ins. iPSCs can be used to model many different cell or tissue types that can be affected in a single disease.

Targets could be identified using genetic suppressor screens in cell culture. Such screening approaches were impossible in diploid cells until the development of CRISPR-based reverse genetic screens. Other useful tools for uncovering genetic modifiers of disease mutations include forward genetic screens using insertional mutagenesis in haploid human cells [23]. These screens will depend on the development of effective selection methods and could be particularly useful for clusters of diseases (mitochondrial, lysosomal storage, etc.)

Open science

Most rare diseases lack the robust research and clinical funding support of ‘big diseases’ such as cancer, Alzheimer’s and heart disease. To make the best use of the money that is raised for rare disease research, scientists need to work together and commit to sharing what they learn with the broader research communities.

The scientists working in the UNC Catalyst assert that research can progress fastest by openly sharing knowledge and materials, especially in rare diseases where resources are often limited. As one example, we could imagine that if all companies released the identity of all the molecules that were made for specific programs but did not progress because the chemical itself caused major toxicities and was declared unsafe, that would be useful information to share with other medicinal chemists. Likewise, if a target is invalidated, we should alert the appropriate community to that finding. Negative data or failed replications of biology (gathered in a rigorous manner) should be published. This would help correct the literature as well as saving time and money on trying to ‘optimize’ futile experiments. In addition to failed drug candidates, we add a plea for sharing information about models that do not prove useful – this could be *in vitro* screens where the hits failed to produce a phenotype in mouse models, or mouse models that ‘validate’ candidate molecules that are ineffective in clinical trials. This does happen over time and it could be even more effective if the knowledge was shared sooner and more deliberately through the use of prepublication sites such as bioRxiv and the SGC’s ‘open notebook’ concept.

A shift in emphasis from competition for grant dollars to cooperation on disease progress requires a level of trust between scientists to work collaboratively with their peers – in the ethos of helping advance rare disease research [24]. Within most labs, post-docs, grad students and associates will share data to improve how they study their (usually related) problems. We can envision a scientific community in which that kind of sharing extends beyond a single lab to unify all labs that study a given problem.

In this context, new publication models are available that aim to disseminate one-off observations [25]. This may be particularly useful in the rare disease space where smaller grants are enough to generate a few potentially useful observations but not enough to launch a continuous research program. Going forward, this could be an excellent approach for sharing results that could – ultimately – contribute to a more comprehensive story.

Conclusion

Rare disease drug discovery and development efforts are theoretically helped by a number of considerations. The first involves the power of human genetics; a genetically validated target is more likely to be successful in putting a molecule in the clinic [26,27]. Additionally, the advocacy of the groups/foundations that champion the research and coordinate engagement with individuals directly impacted by a specific disease provides a direct link to the appropriate clinical study population.

While Foundations can effectively promote and support advocacy, fundraising and patient registries (including biobanking), reducing testable hypotheses to experimental practice for a rare disease usually involves the aggregation

of disease-specific knowledge and the construction of biological and chemical tools. Funding mechanisms and collaborative partnerships can enable access to labs that have these capabilities to address key research questions and build the appropriate tools.

In terms of research directions, the convergence of some rare diseases at the level of organellar function offers the opportunity to create platform screening approaches to help identify molecules that lead to a desired functional phenotype [28]. Most drug companies and academic drug discovery centers have the capabilities and capacity to build the tools to enable this activity, which would be surely welcomed by rare disease communities. Additionally, information derived from rare disease research will inform broader disease understanding.

In summary, our early experiences indicate that the Catalyst's mission to openly share knowledge and materials is a way of working that has the potential to accelerate progress for rare disease research.

Financial & competing interests disclosure

The work described in this article was supported by a grant award from the Eshelman Institute for Innovation in the UNC Eshelman School of Pharmacy.

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. Terry SF, Terry PF, Rauven KA, Uitto J, Bercovitch LG. Advocacy groups as research organizations: the PXE International example. *Nat. Rev. Genet.* 8(2), 157–164 (2007).
2. Lee WH. Open access target validation is a more efficient way to accelerate drug discovery. *PLoS Biol.* 13(6), e1002164 (2015).
3. Drewry DH, Willson TM, Zuercher WJ. Seeding collaborations to advance kinase science with the GSK Published Kinase Inhibitor Set (PKIS). *Curr. Top. Med. Chem.* 14(3), 340–342 (2014).
4. Elkins JM, Fedele V, Szklarz M *et al.* Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat. Biotechnol.* 34(1), 95–103 (2016).
5. Drewry DH, Wells CI, Andrews DM *et al.* Progress towards a public chemogenomic set for protein kinases and a call for contributions. *PLoS ONE* 12(8), e0181585 (2017).
6. Strande NT, Riggs ER, Buchanan AH *et al.* Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the clinical genome resource. *Am. J. Hum. Genet.* 100(6), 895–906 (2017).
7. Girard M, Lariviere R, Parfitt DA *et al.* Mitochondrial dysfunction and Purkinje cell loss in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *Proc. Natl Acad. Sci. USA* 109(5), 1661–1666 (2012).
8. Greer PL, Hanayama R, Bloodgood BL *et al.* The Angelman syndrome protein Ube3A regulates synapse development by ubiquitinating arc. *Cell* 140(5), 704–716 (2010).
9. Arend KC, Lenarcic EM, Vincent HA *et al.* Kinome profiling identifies druggable targets for novel human cytomegalovirus (HCMV) antivirals. *Mol. Cell. Proteomics* 16(4 Suppl. 1), S263–S276 (2017).
10. Ong KT, Perdu, De Backer J *et al.* Effect of celioprolol on prevention of cardiovascular events in vascular Ehlers-Danlos syndrome: a prospective randomised, open, blinded-endpoints trial. *Lancet* 376(9751), 1476–1484 (2010).
11. Rost S, Bach E, Neuner C *et al.* Novel form of X-linked nonsyndromic hearing loss with cochlear malformation caused by a mutation in the type IV collagen gene COL4A6. *Eur. J. Hum. Genet.* 22(2), 208–225 (2014).
12. Davis EE, Frangakis S, Katsanis N. Interpreting human genetic variation with *in vivo* zebrafish assays. *Biochim. Biophys. Acta.* 1842(10), 1960–1970 (2014).
13. Giacomotto J, Segalat L. High-throughput screening and small animal models, where are we? *Br. J. Pharmacol.* 160(2), 204–216 (2010).
14. Bier E, McGinnis W. Model organisms in development and disease. In: *Molecular Basis of Inborn Errors of Development*. Epstein CJ, Erikson RP, Wynshaw-Boris A (Eds). Oxford University Press, Oxford, UK, 25–45 (2003).
15. Sahni N, Yi S, Taipale M *et al.* Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* 161(3), 647–660 (2015).
16. Yi S, Lin S, Li Y, Zhao W, Mills GB, Sahni N. Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat. Rev. Genet.* 18(7), 395–410 (2017).

17. Berger AH, Brooks AN, Wu X *et al.* High-throughput phenotyping of lung cancer somatic mutations. *Cancer Cell* 30(2), 214–228 (2016).
18. Kampmann M. CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chem. Biol.* 13(2), 406–416 (2018).
19. Pushpakom S, Iorio F, Eyers PA *et al.* Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* doi:10.1038/nrd.2018.168 (2018) (Epub ahead of print).
20. Baker NC, Ekins S, Williams AJ, Tropsha A. A bibliometric review of drug repurposing. *Drug Discov. Today.* 23(3), 661–672 (2018).
21. Boehringer Ingelheim Open Innovation Portal. (2019) <https://openme.com/>
22. National Center for Advancing Translational Sciences. NCATS Funding Available to Repurpose Existing Drugs (2018). <https://ncats.nih.gov/news/releases/2018/ntu-foa-2018>
23. Li Y, Shuai L. A versatile genetic tool: haploid cells. *Stem Cell Res. Ther.* 8(1), 197 (2017). doi:10.1186/s13287-017-0657-4
24. Edwards A, Morgan M, Al Chawaf A *et al.* A trust approach for sharing research reagents. *Sci. Transl. Med.* 9(392), eaai9055 (2017).
25. ScienceMatters. <https://sciencematters.io/>
26. Nelson MR, Tipney H, Painter JL *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* 47(8), 856–860 (2015).
27. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* 12(8), 581–594 (2013).
28. Wagner BK, Schreiber SL. The power of sophisticated phenotypic screening and modern mechanism-of-action methods. *Cell Chem. Biol.* 23(1), 3–9 (2016).