



# **Undiagnosed Diseases Network**

## **Manual of Operations**

May 2026

## Table of Contents

I. Network Overview and Operating Procedures .....	3
II. Becoming a UDN Site .....	9
III. Institutional Review Board Communications.....	13
IV. Publications and Research .....	16
V. Communications .....	20
VI. Technology .....	23
VII. Data Sharing .....	25
VIII. Clinical Protocols.....	29
IX. Clinical Research Support Core Subawards.....	37
X. Environmental Exposures.....	39
XI. Sequencing .....	41
XII. Biospecimens .....	63
XIII. Central Biorepository .....	73
XIV. Model Organism Screening .....	76
XIV. Appendices .....	87
APPENDIX 1: Research Concept Sheet, Manuscript Drafts, and Accepted Conference Abstract Submission.....	87
APPENDIX 2: Diagnosis Coding Tool v1.0 .....	91
APPENDIX 3: Onboarding Workflow .....	94

## **I. Network Overview and Operating Procedures**

### **A. Mission, Vision, and Values**

Mission: Our mission is to equitably provide diagnoses for patients with the most challenging conditions through multidisciplinary collaboration and exploration of new diagnostic approaches, advancing scientific discovery and sharing advances with the broader community.

Vision: We envision a world in which an evidence-based, collaborative approach ends the diagnostic odyssey for all with undiagnosed conditions, advances the science of disease identification, and delivers exceptional care for patients and families.

Values: Equity, Collaboration, Innovation, Partnership, Patient-centeredness, Discovery, Inclusive access, Community, Continuous improvement, Dedication, Integrity

### **B. National Institutes of Health (NIH) Goals for Phase III**

The NIH's goals for the Undiagnosed Diseases Network (UDN) in Phase III include:

1. Continuing the proven success of the UDN in improving the clinical evaluation of difficult-to-diagnose patients using a collaborative team approach and investigating and validating promising new diagnostic technologies.
2. Facilitating research into the etiology of undiagnosed diseases by collecting and sharing standardized, high-quality clinical and laboratory data including genotyping, phenotyping, and documentation of environmental exposures.
3. Promoting a broader integrated and collaborative community across multiple Clinical Sites and among laboratory and clinical investigators prepared to investigate the genetic, pathophysiologic, cell biologic, and molecular mechanisms underpinning these difficult-to-diagnose conditions.

### **C. Network Components**

The UDN consists of Clinical Sites (CSs), a Data Management Coordinating Center (DMCC), a Sequencing Core (SC), Model Organisms Screening Centers (MOSCs), a Metabolomics Consultation Group (MC), and a Central Biorepository. Actively enrolling CSs must be selected by the NIH as part of either the U01 (<https://grants.nih.gov/grants/guide/pa-files/PAR-23-289.html>) or X01 (<https://grants.nih.gov/grants/guide/pa-files/PAR-23-171.html>) mechanism.

For a current list of UDN sites and PIs, see the UDN protocol document or the UDN PI and Site Coordinator Contact List. Both documents are stored in Dropbox and are linked on the UDN Investigator Resource page on the UDN website: <https://undiagnosed.hms.harvard.edu/udn-investigator-resource-page>.

### **D. Cooperative Agreement Responsibilities**

The administrative and funding instrument used for the UDN is the cooperative agreement, an "assistance" mechanism (rather than an "acquisition" mechanism), in which substantial NIH programmatic involvement with the awardees is anticipated during the performance of the activities. Under the cooperative agreement, the NIH purpose is to support and stimulate the

recipients' activities by involvement in and otherwise working jointly with the award recipients in a partnership role; it is not to assume direction, prime responsibility, or a dominant role in the activities. Consistent with this concept, the dominant role and prime responsibility reside with the awardees for the project as a whole, although specific tasks and activities may be shared among the awardees and the NIH as defined below.

**NIH staff have substantial programmatic involvement that is above and beyond the normal stewardship role as described below:**

The NIH Project Scientist(s) have substantial scientific and programmatic involvement during the conduct of this activity through technical assistance, advice, and coordination. However, the role of NIH staff is to facilitate and not to direct the activities. It is anticipated that decisions in all activities are reached by consensus of the UDN and that NIH staff are given the opportunity to offer input to this process. The Project Scientist(s) will participate as members of the Steering Committee and will have one vote. The Project Scientist(s) have the following substantial involvement:

1. Participating with the other Steering Committee members in addressing issues that arise with UDN planning, operation, and analysis. The Project Scientist(s) assist and facilitate the group process and do not direct it.
2. Serving as a liaison, helping to coordinate activities, including acting as a liaison to other NIH Institutes/Centers, and as an information resource for the awardees. The Project Scientist(s) also help coordinate the efforts of the UDN with other groups conducting similar efforts.
3. Attending all Steering Committee meetings as a voting member and all working group meetings, assisting in developing operating guidelines, quality control procedures, and consistent policies for dealing with situations that require coordinated action. The Project Scientist(s) are responsible for working with the grantee(s) as needed to manage the logistic aspects of the resource.
4. Reporting periodically on Network progress to the NIH UDN Working Group (a trans-NIH working group made up of staff from multiple NIH Institutes and Centers).
5. Serving on subcommittees of the Steering Committee and Working Groups as appropriate.
6. Assisting awardees in the development, if needed, of policies for dealing with situations that require coordinated action.
7. Providing advice in the management and technical performance of the award.
8. Assisting in promoting the availability of the data and related resources developed in the course of this program to the scientific community at large.
9. Participating in data analyses, interpretations, and, where warranted, co-authorship of the publication of results of studies conducted through the program.
10. Other NIH UDN Working Group staff may assist the awardee as designated by the UDN NIH Program Official (The NIH official responsible for the programmatic, scientific, and/or technical aspects of the grant).

**Collaborative responsibilities**

Close interaction among the participating investigators is required, as well as significant involvement from the NIH, to develop and operate the UDN. Principal investigators participate in in-person meetings on an annual basis; during months in which there are not in-person meetings, there are monthly conference calls as needed to share information on data resources, methodologies, analytical tools, as well as data and preliminary results. Key co-investigators

and pre- and post-doctoral trainees, especially those who are members of under-represented minority groups or those from different but related disciplines, are also eligible to attend these meetings.

All Awardees agree to work collaboratively to:

1. Assist in refining a common approach to patients with undiagnosed diseases.
2. Work collaboratively with other UDN investigators to provide for secure, accurate and timely data submission.
3. Participate in presenting and publishing new processes and substantive findings.
4. Participate in the governance of the UDN as a member of the Steering Committee by attending meetings and contributing to discussions.
5. Interact with other relevant and NIH activities, as needed, to promote synergy and consistency among similar projects.

The CS Awardees agree to fulfill responsibilities outlined in their NIH funding announcements.

## E. Steering Committee Policies

**Guideline:** The Network Steering Committee is composed of the following voting members:

- The contact PI of NIH-awarded CSs (U01 and X01) and the UDP PI
- The contact PI of the DMCC
- A representative from the participating undiagnosed diseases patient advocacy group(s). The patient advocacy groups will have one collective vote.
- NIH Program Official(s)/Projects Scientist(s) from participating Institutes. NIH will have one collective vote.
- As the Network evolves, other key stakeholders may be invited to serve on the Steering Committee as appropriate (e.g., outside funding or resource partners)

**Policies:**

1. The Steering Committee is responsible for policy decisions regarding the Network, and for the discussion and resolution of procedural issues that affect the operation and status of the Network as a whole.
2. The UDN Steering Committee will be the operational group through which the NIH UDN Working Group interacts with the UDN.
3. The Steering Committee will have at least monthly conference calls.
4. There will be at least one in-person UDN meeting per year.
5. The minutes for all Steering Committee discussion will be documented and posted on Dropbox.
6. The Steering Committee may add additional members, and other NIH Program staff may attend the Steering Committee meetings as desired.

## F. Election of Steering Committee Co-Chairs

**Guideline:** The position of Chairperson of the Steering Committee of the UDN is filled by 2-3 co-chairs who serve overlapping terms. Co-chairs are nominated by the UDN Executive Committee, approved by the UDN Steering Committee, and serve as chairs of the Executive Committee simultaneously.

**Principles:**

1. The term of the position of Chair is 2 years in duration unless an extension is agreed to by the Steering Committee.
2. The individual holding the position of Chair must be a current member of the Steering Committee and Executive Committee.

## **G. Executive Committee**

The Executive Committee consists of 7 members approved by the Steering Committee, including the 2-3 co-chairs of the Steering Committee. The group will consist of 1 representative from: U01 sites, X01 sites, UDP, DMCC, patient advocacy groups, NIH, and resource partners (MOSC, sequencing core). The representatives will serve 2-year terms.

The Executive Committee meets weekly to review and monitor UDN progress. Members may send a proxy to attend the meeting if they are unable to attend. Each representative is responsible for reaching out to their group to discuss issues that need to be raised with the Executive Committee.

## **H. Other Network Committees**

**Guideline:** The Steering Committee may establish working groups as needed to address particular issues, which will include representatives from the network, the NIH, and possibly other experts. The Steering Committee has the overall responsibility of assessing and prioritizing the progress of the various working groups and other needed subcommittees of the working groups.

### **Working Group Governance**

1. Any individual or group proposing a new working group will present their idea to the Steering Committee. A formal vote of the Steering Committee is needed to create a new working group.
2. Volunteers for chair or co-chairs of the new working group will be solicited when the new working group is proposed. A formal vote of the Steering Committee is needed to confirm the chair or co-chairs.
3. Co-chairs are not required for all working groups but may be recommended by the Steering Committee.
4. Working group co-chairs may come from the same site.
5. If there are no volunteers, or only one, the Steering Committee may recommend a site or type of site that may be a good fit for the working group and one of the Steering Committee co-chairs will solicit the site(s) for a recommended chair.
6. Any working group proposing to close will present their idea to the Steering Committee. A formal vote of the Steering Committee is needed to close a working group.

### **Active Committees and Groups**

All active committees and groups are open to new members, except for the Executive and Publications and Research Committees. UDN members can request to join groups by completing the following form: <http://redcap.link/udnprotocolrequest>.

Group	Purpose
Central Consultation Committee	Review difficult cases. Investigators must be on the UDN IRB protocol to attend.
Clinical Protocols Working Group	Create and modify protocols related to the UDN evaluation process
Diagnostic Best Practices Paper Writing Group	Paper writing group
Ethics Paper Writing Group	Paper writing group
Executive Committee	Invitation only meeting for Executive Committee members
Grand Rounds	Present and discuss UDN cases, open to the public unless otherwise specified
Model Organisms Working Group (MOWG)	Discuss cases sent to Model Organisms Screening Centers
Site Operations	Discuss challenges and potential solutions related to site operations. Mandatory for site coordinators.
Steering Committee	Discuss collaborations with external groups and network-wide changes. At least one PI from each site should attend or send a proxy.
Tool Building Coalition	Discuss analysis of UDN data

### Inactive Committees and Groups

1. Environmental Trigger Working Group
2. Genetic Counseling & Testing Working Group
3. Metabolomics Working Group
4. Sustainability Working Group
5. Therapeutic Matching Committee

## I. Implementing and Revising the Manual of Operations

1. Working groups are responsible for development and maintenance of related chapters for the Manual of Operations.
2. Chapters of the Manual of Operations are ratified by the Steering Committee.
3. Working groups have the authority to make decisions regarding implementation of ratified chapters of the Manual of Operations that are assigned to the working group for implementation.
4. If a working group cannot resolve an implementation decision internally, the Steering Committee is consulted.
5. Working groups consult with other relevant working groups on implementation decisions that involve multiple areas of expertise. A cross-working group liaison may be assigned to facilitate these interactions.
6. All working groups make their agendas and minutes available to other working groups via Dropbox.

7. Working groups that would like to recommend a change to a ratified Manual of Operations chapter that affects network-wide operations or a new chapter should recommend the change to the Steering Committee for ratification.
8. Groups that would like to recommend a change to the network-wide Institutional Review Board (IRB) protocol or consents should recommend the change to the Steering Committee.

## **J. Milestones and Metrics**

One of the core functions of the NIH Program and the DMCC is to monitor each component of the network (DMCC, CSs, SCs, other Cores) as well as the network as a whole. The rationale for this monitoring is two-fold: (1) to encourage understanding and continuous improvement of components of the network and (2) to codify the expectations of our funders in order to ensure that our efforts are aligned with expectations.

The DMCC calculates a range of metrics to track network progress. These metrics are available on the public UDN website:

1. Facts and Figures: <https://undiagnosed.hms.harvard.edu/about-us/facts-and-figures/>
2. Quarterly Reports: <https://undiagnosed.hms.harvard.edu/about-us/>

## II. Becoming a UDN Site

### A. Clinical Site Responsibilities and Resources

UDN Clinical Sites (CSs) play a crucial role in the network, leveraging their expertise, resources, and collaborative efforts to advance rare disease research and improve patient care. These CSs are typically at academic medical centers and include members with expertise in various medical specialties, genetics, and rare diseases. The onboarding process is described in this section and depicted in Appendix 3.

The expectations of CSs, as set forth by the NIH, include:

1. Evaluation of patients with undiagnosed conditions: CSs are expected to assess patients referred to them who have complex medical conditions that have eluded diagnosis through conventional means. Evaluations involve a thorough review of medical records, comprehensive multi-disciplinary clinical workups, and specialized testing. Sites should reference their NIH grant documentation for enrollment metrics.
2. Collaboration with other network members: CSs work closely with other UDN sites, as well as with the DMCC and NIH, to share data, insights, and expertise. Collaboration and data sharing are crucial for solving challenging cases and advancing research into rare diseases.
3. Contribution to research efforts: In addition to providing clinical care to patients, CSs contribute valuable data and samples to research studies aimed at understanding the genetic and molecular basis of undiagnosed conditions. CSs should have the capability to apply innovative data analysis approaches.
4. Education and outreach: CSs engage in educational and outreach activities to raise awareness about undiagnosed conditions among healthcare professionals, researchers, and the general public. This may include disseminating information about the UDN's mission, services, and resources.

All CSs have access to a range of centralized resources to support their efforts, including:

Resource	Time period				
	Apr '23- Mar '24	Apr '24- Mar '25	Apr '25- Mar '26	Apr '26- Mar '27	Apr '27- Mar '28
<b>Data sharing</b>					
Mosaic					
External data sharing coordination					
Gateway					
<b>Diagnostic tools and technologies<sup>^</sup></b>					
Genome sequencing (short and long read)	Sept '23- Jan '24	Jul '24- Mar '25	Jan '26- Mar '26	?	?
RNA sequencing	Sept '23- Jan '24	Jul '24- Mar '25	Jan '26- Mar '26	?	?
Reanalysis		Jul '24- Mar '25		?	?
Optical genome mapping			Jan '26- Mar '26	?	?
<b>Participant financial support</b>					

National Organization for Rare Disorders*					?
Warren Alpert Foundation (only travel)			Apr '25- Dec '25		
<b>Research tools and technologies</b>					
Model organisms	Sept '23- Mar '24	Jul '24- Mar '25	Jan '26- Mar '26	?	?
Clinical Research Support Core Subawards**			Jan '26- Mar '26	Jun '26- Mar '27	?

**Support provided**

? Support is not determined at this time

^ The NIH issued Administrative Supplements to the DMCC to support clinical sequencing through the UDN Sequencing Core. Additional details about the Administrative Supplements are shared directly with sites.

\* Details in Clinical Protocols MOO chapter

\*\*Details in Clinical Research Support Core Subawards MOO chapter

## B. Agreements

For a site to join the network, the site must be added to the UDN Institutional Review Board (IRB) protocol and sign the UDN Data Sharing and Use Agreement (DSUA). The first step in this process is for the DMCC to complete a reliance agreement request form with input from the site. The form is submitted to the NIH by the DMCC. With this information, the NIH drafts a reliance agreement for the site to review and sign. This signed agreement is returned to the NIH for execution.

Once the reliance agreement is executed, the DMCC must submit a modification to the NIH IRB. Sites enrolling participants must also create consent forms using the template forms (stored in the UDN Site Coordinators Dropbox folder, quick link available on the UDN Investigator Resource webpage). The DMCC informs the site when the modification is approved.

The site must also sign the UDN DSUA. The DMCC will draft a signature page for the site. Upon receipt of the signed signature page, UDN DSUA Exhibits 1 and 2 will be amended to reflect the new site and PI(s) and made available to the parties.

## C. Site Operations

*Team Structure:* CSs should designate at least 1 site coordinator to serve as the primary contact for other Network sites. Site coordinators are responsible for managing the participant workflow, ensuring data is entered into the central database, the “Gateway,” in a timely manner, and maintaining compliance with network-wide agreements and protocols. The site coordinator should be a project manager, nurse practitioner, genetic counselor, or research coordinator familiar with the institution and investigators. The site coordinator is expected to participate in the UDN Site Operations group and other working groups as needed.

Sites should also consider establishing teams to address the various needs of the UDN. An example team structure from a CS is outlined below.

Leadership team: represents the site at network meetings

- PI(s)
- 2-3 additional clinicians

Clinical research team: day-to-day study coordination and visit planning

- Site coordinator(s)
- Genetic counselor(s)
- Nurse practitioner(s)
- Research assistant(s)

Bioinformatics team: research analysis of sequencing data

- Bioinformatician
- Molecular laboratory director (if available)

*Meetings:* In addition to network-wide meetings, sites should anticipate having several local meetings. An example meeting schedule from a CS is outlined below.

Weekly: team meeting (90 minutes)

- Discuss new cases and assign them to reviewers
- Assign tiers for reviewed cases
- Review upcoming participant evaluations
- Review recently completed evaluations

Weekly: clinical research staff (45-60 minutes)

- Discuss logistical issues of travel, scheduling, sample collection, and paperwork

As needed: pediatric or adult case planning meetings (~30 minutes/case)

*Internal communications:* In addition to data stored in the Gateway or shared UDN folder in Dropbox, sites are expected to track research progress and data locally. Several sites use REDCap and SharePoint for this purpose. Please reach out to your local IT department for institutional recommendations on secure platforms to store PHI. Components to be tracked locally include, but are not limited to, medical records, communication with participant/family, and progress through the case submission and evaluation processes.

*Consultants:* Several CSs developed relationships with more than 50 consultants to achieve the goals of the multidisciplinary UDN evaluation. Based on these experiences, the sites developed this list of recommended consultants.

- Neurologist (adult and pediatric)
- Neuromuscular specialist
- Developmental specialist
- Radiologist
- Ophthalmologist
- Others as needed, such as an immunologist, surgeon (for muscle biopsy), rheumatologist, dermatologist, nephrologist, cardiologist, case-relevant pathologist

*Space:* Several sites choose to use clinical research centers for UDN evaluations, tests, and procedures. This provides flexibility and convenience not typically available in inpatient or

outpatient settings. Of note, this approach worked extremely well for some sites and did not work at all for other sites. It is recommended that sites consider this approach before exploring alternatives. Several sites find it helpful for team members' offices to be located close to one another.

*Website:* The DMCC maintains the primary UDN website at <https://undiagnosed.hms.harvard.edu/>. Sites may find it useful to create a site-specific UDN website to share contact information, publications, and other content related to their specific UDN site. Site-specific websites are designed, maintained, and hosted by individual sites.

*Participant triaging:* It is highly recommended to utilize a shared email address for all participant and consultant communications. Shared electronic fax numbers have been helpful for sites to allow participants to fax documents to the site, which are received by a shared email inbox. Having a shared phone number and voicemail has also worked well for some sites.

*Helpdesk:* The Helpdesk is staffed by members of the DMCC that respond to inquiries from participants and family members, referring providers, UDN sites, and the general public. Participants who speak languages other than English or Spanish can speak with the Helpdesk staff via an on-demand telephone interpreter. Members of the Helpdesk team are available to answer questions from 9 a.m. – 5 p.m. ET, Monday – Friday. Contact information for the Helpdesk is as follows: [UDN@hms.harvard.edu](mailto:UDN@hms.harvard.edu), 1.844.746.4836 (1 844 Ring UDN).

### III. Institutional Review Board Communications

NIH IRB policies and template documents can be found on the NIH Office of Human Subjects Research Protections website: <https://irbo.nih.gov/>.

#### A. Consent Form Development

1. The UDN PI, Central IRB (CIRB), and DMCC CIRB Liaison will develop model informed consent and assent form (ICF) templates for the UDN, noting sections of the template that must be customized by each CS.
2. The DMCC CIRB Liaison will make the template ICFs available to the CS Site Coordinators.
3. The CS Site Coordinators will customize only the areas of the ICFs specified in the template, including:
  - a. Placing the consent form on the institutional letterhead
  - b. Adding standardized language as required by the CS (due to local policy requirements)
  - c. Incorporating HIPAA authorization for use and disclosure of PII if HIPAA is not available as a separate document, as per with the CS institution's standard approach. If HIPAA is provided as a separate document, it does not need to be submitted.
4. The CS Site Coordinators will send the completed site specific ICFs to the DMCC CIRB Liaison.
5. The DMCC CIRB Liaison will review the ICFs and send them to the UDN PI to submit to the CIRB.
6. The CIRB will review the site specific ICFs with all of the other submitted site materials provided for site approval.
7. The CIRB will communicate the results of the review to the UDN PI and the DMCC CIRB Liaison.
8. The DMCC CIRB Liaison will communicate the results of the review to the CS Site Coordinators.
9. The CIRB will provide to the DMCC CIRB Liaison the approved ICFs for each CS. ICFs will have a version control reference in the header and on the last page of the form.
10. The DMCC CIRB Liaison will make the CIRB-approved ICFs available to the CS Site Coordinators and will store centrally for all CSs to access.

#### B. Reportable New Information (RNI)

Please see UDN IRB protocol and [3014-801 - Reporting Research Events](#) for more information regarding reportable new information (RNI) and adverse events (section: "Collection, monitoring, analysis and reporting of adverse events").

Unanticipated problems, non-compliance, major protocol deviations, and deaths possibly related to research activities are reported to the IRB **as soon as possible but not more than 7 days** after any study team member first learns of the event. If also an unanticipated problem, deaths are reported to the IRB **within 24 hours** after the principal investigator first learns of the event. Adverse events that occur during this study, including serious adverse events that are not unanticipated problems and minor protocol deviations, are sent to the DMCC and reported to the IRB at continuing review. Adverse events that are unrelated to research do not need to be

sent to the DMCC or reported at continuing review if more than 1 week has elapsed since the last invasive research procedure (e.g., blood draw, skin biopsy) was performed.

If unsure if an event needs to be reported, the CS and Core PIs will contact the DMCC CIRB Liaison. The DMCC CIRB Liaison will ensure CS and Core PIs have access to the CIRB portal.  
RNI Reporting:

1. The CS and Core PIs will report all unanticipated problems, non-compliance, major protocol deviations, and deaths possibly related to research activities as soon as possible to the DMCC.
2. The DMCC will facilitate reporting to the CIRB.
3. The CS and Core PIs will facilitate follow up activities deemed necessary by the CIRB.
4. The CIRB will communicate the results of the review to the CS and Core PIs, UDN PI, and the DMCC CIRB Liaison, and in some cases, the Institutional Official related to unanticipated problems or serious and/or continuing noncompliance.
5. The DMCC CIRB Liaison will communicate outcomes to the CSs and Cores.

### **C. Continuing Review**

1. Three months prior to the continuing renewal deadline, the CIRB will notify the DMCC CIRB Liaison regarding information required for Continuing Review (CR) and provide the forms that all CSs and Cores, including the intramural site, must complete.
2. The DMCC CIRB Liaison will notify each CS and Core Site Coordinator regarding information required for CR, and any other forms required by the CIRB.
3. The CS and Core Site Coordinators will submit their responsive information for CR to the DMCC CIRB Liaison within 1 month of the CR deadline.
4. The DMCC CIRB Liaison will review the UDN site forms for accuracy and completeness.
5. The DMCC CIRB Liaison will provide the individual site-specific CR forms as well as submit a single CR Application to the UDN PI.
6. The UDN PI will review the applications and submit all documents to the CIRB.
7. The CIRB will conduct CR of all submitted materials.
8. The CIRB will communicate the results of the review to the UDN PI and the DMCC CIRB Liaison.
9. The DMCC CIRB Liaison will communicate the results of the review to the CS and Core Site Coordinators.

### **D. Modifications**

1. Study-wide protocol modifications will be reviewed with the Steering Committee, as needed, before submission to the CIRB. Site-specific modifications can be submitted directly to the DMCC CIRB Liaison with supporting documentation.
2. The DMCC CIRB Liaison will finalize modifications and submit to the UDN PI.
3. The UDN PI will submit modifications to the CIRB, including tracked and clean copies of all modified documents with updates to the version control of each document.
4. If the modification requires changes to the ICF:
  - a. The DMCC CIRB Liaison will modify the ICF model template.
  - b. The DMCC CIRB will send tracked and clean copies of the modified ICF model template to the UDN PI to submit for CIRB review.
  - c. Once the model template ICFs are approved by the CIRB, the DMCC CIRB Liaison will work with the CS Site Coordinators to modify each site-specific document.

5. The CIRB will communicate the results of the review to the UDN PI and DMCC IRB Liaison.
6. The DMCC CIRB Liaison will communicate the results of the review to the CS and Core Site Coordinators.
7. The DMCC CIRB Liaison will make available to the CSs and Cores the approved documents.
8. If there are updated CS ICFs, the DMCC CIRB Liaison will make the ICFs available to the CS Site Coordinators and will store centrally for all CSs to access.

## E. Study Staff Changes

1. NIH UDP CS:
  - a. The NIH IRB CS Site Coordinators will submit completed site-specific key study personnel forms and training certificates to the DMCC CIRB Liaison.
  - b. The DMCC CIRB Liaison will review the documentation and make personnel changes.
  - c. NIH IRB CS Site Coordinators will instruct investigators being added to complete this survey: <http://redcap.link/udnprotocolrequest>.
  - d. NIH IRB CS Site Coordinators will complete this form to remove someone: <https://forms.gle/2drX4s141hvgStoN6>.
2. All other groups:
  - a. Site Coordinators provide the investigator being added with a completed key study personnel form and instruct them to upload the form along with their human subjects research trainings via this survey: <http://redcap.link/udnprotocolrequest>
  - b. Site Coordinators complete this form to remove investigators: <https://forms.gle/2drX4s141hvgStoN6>
  - c. Site Coordinators monitor the list of those currently on the IRB protocol linked on UDN Investigator Resource Page.

## F. Affiliated Studies

1. CSs and Cores may propose studies involving UDN participants that are not network wide.
2. These studies require permission from the Publications and Research Committee. The UDN site carrying out an affiliated study will provide information about the study, which may include a brief summary of the study, sites involved, nature and characteristics of participant involvement and consent required, and lead UDN investigator.
3. If the aims of the study are distinct from the aims of the broader UDN study, a separate IRB protocol through the site proposing the study will be submitted. If the aims of the study are consistent with the aims of the broader UDN study, a modification to the CIRB detailing the site-specific research will be submitted.
4. The DMCC CIRB Liaison keeps track of these studies.
5. The DMCC CIRB Liaison informs the UDN PI and the CIRB of affiliated studies, as needed.

## IV. Publications and Research

One dimension of UDN success is the number and quality of its publications and presentations. The purpose of this document is to establish a framework that facilitates and streamlines collaborative manuscript submission, as well as associated work, including meeting abstracts and presentations. The Publications and Research Committee (PRC) oversees the activities set out herein on behalf of the UDN Steering Committee, and report to it. Changes to the policy described herein, which are expected from time to time, must be approved by the Steering Committee. The publication policy applies to a proposed publication if the results are the product of research conducted under the UDN study protocol.

### A. Scope

1. To facilitate manuscript submission.
2. To provide input on abstract submission and scientific presentation, when requested.
3. To help the DMCC with content for the website and, if required, social media.
4. To maintain an up-to-date list of all presentations, abstracts, publications, and proposals. The DMCC assists in tracking projects.
5. Notwithstanding anything to the contrary in this document, the scope of PRC does not include evaluation of the scientific merit of any publication produced as a result of UDN participation.

### B. Membership

The PRC is made up of one representative from each UDN CS and Core, the DMCC, and the NIH Program.

### C. Concept Sheets

Concept sheets are required for research projects, data analysis projects, grant applications, and UDN-led manuscripts. This includes but is not limited to: single-site projects, case reports, analysis of Gateway data from participants evaluated at other sites, and biorepository specimen requests from external groups. Concept sheets are encouraged but not required for manuscripts led by external groups. Manuscripts led by external groups should follow all guidelines outlined in section IV.D. Manuscript Review.

Concept sheets must be submitted online at [https://hms.az1.qualtrics.com/jfe/form/SV\\_3CyZOKOuiEvnCx7](https://hms.az1.qualtrics.com/jfe/form/SV_3CyZOKOuiEvnCx7) (see Appendix 8: Research Concept Sheet). Once a concept sheet is submitted, the DMCC forwards the concept sheet to the PRC for review. The PRC reviews all concept sheets within 2 weeks of circulation to members. The DMCC notifies submitters of approval at the end of the review period, including any feedback received from the PRC.

Concept sheets must be reviewed and approved by the PRC before substantial work begins. In addition to PRC review, investigators should note the following requirements for research projects and grant applications:

1. Research projects involving human subjects research require IRB review and approval.
2. Grant applications and research projects led by external investigators require the involvement of a UDN member.
3. Grant applications and research projects involving external investigators and network-wide data and/or samples require a Steering Committee presentation and vote.

Note that UDN investigators do not need to contact each site individually to obtain approval for analysis of Gateway data from participants evaluated at other sites. Only concept sheet submission is required.

Concept sheets are not required when UDN investigators request biorepository specimens from participants evaluated at their own site.

## **D. Manuscript Review**

### **Review process**

All manuscripts must be reviewed and approved by the PRC prior to publication. Final manuscripts should be submitted to the PRC and journal of interest simultaneously.

Manuscripts must be submitted online at [https://hms.az1.qualtrics.com/jfe/form/SV\\_3CyZOKOuiEvnCx7](https://hms.az1.qualtrics.com/jfe/form/SV_3CyZOKOuiEvnCx7). Once a manuscript is circulated, the PRC reviews all manuscripts within 2 weeks. The DMCC notifies submitters of approval at the end of the review period, including any feedback received from the PRC.

The DMCC notifies submitters of approval at the end of the review period, including any feedback received by PRC members.

### **Corporate authorship**

The UDN should be acknowledged as a corporate author in the author list as “Undiagnosed Diseases Network”.

### **Member list**

The UDN member list should be submitted to the journal of interest in accordance with the journal’s requirements for consortia. The UDN member list is updated on a quarterly basis and includes members from the sites as determined by the site PIs. When preparing manuscripts, authors should go Dropbox to download the current version of the member list: <https://www.dropbox.com/sh/4jz2btq03dw5n1a/AAOxKT4B-dlf9LhJuTPz0aga?dl=0>

### **Author indexing**

When possible, authors included in the UDN member list should be indexed in PubMed.

### **Funding acknowledgement**

All UDN papers (network-wide and local) where research was conducted with NIH prime or sub-prime awards (past or current) must include the following funding acknowledgement:

Research reported in this publication was supported by the National Institute Of Neurological Disorders And Stroke of the National Institutes of Health under Award Number(s) [xxxxx]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The funding acknowledgement statement and award numbers are available at <https://www.dropbox.com/scl/fi/6je77gccjuu1s0skcf0oi/UDN-Funding-Acknowledgement-Statement.docx?dl=0&rlkey=pfokv50ddekxnrj7osmdqn0oz> .

### **Notification of acceptance**

Investigators should notify the DMCC when manuscripts are accepted and provide the expected publication date.

### **Notification of publication**

Investigators should notify the DMCC when manuscripts are published.

### **Authorship guidelines**

Authors (first, middle, and senior) are determined by the type, scope and site of project. Given the nature of the UDN's work, shared first or last authors should be remembered as an option. Generally, it is expected that authors would make contributions to any or all of the following including but not limited to the concept, design, acquisition and analyses of data, drafting of manuscript, editing, and revision of manuscript. The PRC resolves all authorship disagreements.

### **Scope of manuscript review**

PRC approval shall be for purposes of satisfaction of the points in this section alone and, for purposes of clarity, this means that the PRC shall not withhold approval of a manuscript on the basis of its scientific merit.

## **E. Accepted Conference Abstracts and Presentations**

Accepted conference abstracts and presentations must be submitted online: [https://hms.az1.qualtrics.com/jfe/form/SV\\_3CyZOKOuiEvnCx7](https://hms.az1.qualtrics.com/jfe/form/SV_3CyZOKOuiEvnCx7)

Accepted conference abstracts and presentations sharing data from a single site do not require PRC review, but submission is needed to track UDN contributions.

Accepted conference abstracts and presentations including UDN-wide data or data from more than one UDN site require PRC and DMCC review to ensure accurate reporting of UDN-wide data.

Abstracts and presentations should reference UDN grant number(s) and include "Undiagnosed Diseases Network" in the author list. Abstracts and presentations should include the UDN logo.

If there is an NIH co-author, final versions of the abstract must be submitted to the Program Officer for review and approval.

## **G. Start of UDN Publication**

The PRC developed a manuscript, which described, defined, and introduced the network to the medical and scientific community:

Ramoni RB, Mulvihill JJ, Adams DR, et al. The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease. *Am J Hum Genet.* 2017;100(2):185-192.  
doi:10.1016/j.ajhg.2017.01.006

## **H. End of Funding Cycle Publications**

In collaboration with the DMCC and SC leadership, the PRC is involved in the preparation of manuscripts that describe the UDN experience towards the end of each funding cycle.

## V. Communications

The purpose of this section is to describe the public-facing UDN websites and outline how the UDN engages with the public. It addresses both official and unofficial UDN social media communications. It provides guidance for UDN investigators interested in sharing information about the UDN on personal social media accounts and profiles. Guidelines for responding to media inquiries are also provided.

### A. Website

Information about the UDN is made available on public-facing websites maintained by the DMCC, NIH, and CSs.

- DMCC: <https://undiagnosed.hms.harvard.edu/> (alias: udnconnect.org)
- NIH: <https://www.ninds.nih.gov/current-research/focus-disorders/focus-undiagnosed-diseases-network>

The DMCC website is considered the primary source of up-to-date information about the UDN, as it contains the link to the Gateway, content related to the case submission process, CSs and core laboratories, current research, and press coverage. The DMCC website also includes contact information for the Helpdesk located at the DMCC. The NIH and CS websites are primarily for informational purposes.

### B. Social Media

#### Official UDN social media communications

Since the launch of the Network, the DMCC has worked to cultivate a social media presence to 1) engage with potential participants, referring providers, and collaborators, and (2) share the work of the UDN with interested stakeholders (e.g., general public, advocacy organizations, scientific community). A secondary goal is to share the UDN participant pages to identify individuals with similar conditions.

When representing the UDN on social media, the DMCC:

- Carefully reviews all content for accuracy and relevance prior to posting on any social media platform
- Deletes spam and comments that are unrelated to the topic in question
- Replies to questions and comments when appropriate
- Avoids making political statements
- Notifies the Executive Committee and CS, if applicable, of comments that rise above the level of general questions or concerns

Official social media accounts managed by the DMCC include Facebook, LinkedIn, and Instagram, which can be accessed via the following links:

- Facebook: <https://www.facebook.com/udnconnect>
- LinkedIn: <https://www.linkedin.com/company/undiagnosed-diseases-network/>
- Instagram: <https://www.instagram.com/udnconnect>

## **Unofficial UDN social media communications**

Unofficial UDN social media communications are defined as communications about the UDN that have not been reviewed by a UDN investigator or NIH Program staff member. Individuals and groups engaging in unofficial UDN social media communications are advised to include a disclaimer such as “Views are my own and do not necessarily reflect the views of the UDN or the NIH” on their social media profiles. They are also asked to route inquiries that may warrant or benefit from an official UDN or NIH response to the DMCC.

## **Institutional social media communications**

The DMCC engages with institutional social media accounts belonging to the NIH and UDN sites as appropriate.

## **Best practices for personal social media accounts and profiles**

UDN investigators are encouraged to share the work of the UDN on their personal social media profiles and engage with the official UDN social media accounts with the following guidelines in mind:

- Do not share sensitive or confidential information about UDN research activities or participants
- Avoid expressions that could be misinterpreted and should not engage in discourse that may reflect poorly on the UDN or the NIH
  - Example: politically charged posts related to UDN funding
- Avoid making statements about unpublished UDN metrics and outcomes without appropriate context
  - Example: public Facebook posts about the high number of diagnoses made at your site

If the DMCC observes public social media posts made by a UDN investigator violate any of the above guidelines, the DMCC contacts the individual to suggest that the post be removed or made private. The DMCC notifies the UDN Executive Committee of such posts when appropriate, as determined by internal review at the DMCC.

## **Social media content suggestions**

All members of the UDN are invited to send content suggestions for the UDN websites and social media accounts to [UDN@hms.harvard.edu](mailto:UDN@hms.harvard.edu).

# **C. Media Inquiries**

## **General guidelines**

In general, UDN sites are welcome to respond to media inquiries directed at their site. If a site determines that the inquiry may benefit from additional perspectives, the site may opt to contact other UDN sites or the NIH. Comment requests related to NIH funding or other administrative issues should be directed to the DMCC for routing to NIH staff, as appropriate. Note that any comments from NIH staff require approval by one or more NIH communications offices. In some

circumstances, the DMCC may engage experts at the UDN sites (e.g., individual with experience in broadcast television) when responding to media inquiries.

These guidelines do not apply to press releases and content developed by the UDN sites. In addition, they do not override institutional policies at the individual UDN sites.

### **Requests for interviews with participants**

The CSs coordinate requests for interviews with participants. Cores and other sites that do not interact with participants should direct interview requests to the DMCC and include the following in writing:

1. An outline of the proposed interview
2. Criteria for selecting interview participants
3. Expected publication date

Once this information is received, the DMCC works with the CSs to identify participants or families appropriate for the interview.

### **Unpublished metrics**

Each UDN site is welcome to comment on unpublished metrics for their site. However, the only network-wide metrics that should be shared with the media are those displayed on the UDN public website.

## VI. Technology

The UDN Gateway (<https://gateway.undiagnosed.hms.harvard.edu/>) is the centralized web-based patient and investigator portal for the UDN. The Gateway was built by developers at the DMCC to meet the unique needs of the network. The Gateway is administered by the DMCC in a manner that is consistent with Federal Information Security Management Act (FISMA) Moderate regulations.

Of note, each CS and Core uses local systems to track study data that falls outside the scope of the Gateway, such as research plan notes. The DMCC does not dictate which systems a given CS or Core chooses to use as long as that decision does not impair or threaten the overall security and compliance posture of the UDN data and technology network. Examples of local tracking systems include electronic medical record (EMR) systems, laboratory information management systems (LIMS), and REDCap. Data entered into these systems are stored securely under the oversight and policies of the individual CS and Cores.

### A. UDN Gateway

#### Capabilities

The Gateway has 4 distinct interfaces: participant, parent/guardian, licensed healthcare provider, and UDN researcher. The participant, parent/guardian, and licensed healthcare provider interfaces are used to submit cases, including study recommendation letters, medical records, and personal narratives, for consideration by the UDN. When a case is created, it is assigned a universally unique identifier (UUID) and a 6-digit UDN ID (e.g., UDN123456).

The UDN researcher interface is organized into CS- and Core-specific dashboards. This structure allows for seamless, secure sharing of clinical and research information between UDN investigators.

#### Data entry

Upon addition to the UDN IRB protocol, new investigators receive instructions to sign up for an account on the Gateway using their institutional email address. Site coordinators are expected to enter data in the Gateway in real time as participants progress through the UDN case submission and evaluation processes. Timelines for data entry are described in the UDN Gateway Guide available on Dropbox. Reports are available for download in the Gateway and can help identify missing data.

#### Bugs

UDN investigators should contact the UDN Gateway Support Desk at [udngateway@hms-dbmi.atlassian.net](mailto:udngateway@hms-dbmi.atlassian.net) if they suspect the Gateway is not working as expected.

#### Downloading sequencing files

Sequencing files may be downloaded **one time across all users at a given CS**. Sites are strongly encouraged to implement file storage workflows that will make any files downloaded available to local team members.

If an investigator is blocked from downloading a file because it was already downloaded, they should first follow up with their local team to determine if the file is available locally. If unable to locate the file, they should follow the steps provided in the Gateway to request an additional download.

### **Feature requests**

The Gateway continues to evolve to meet the needs of the network. Investigators may submit requests for new features or improvements to the UDN Gateway Support Desk at [udngateway@hms-dbmi.atlassian.net](mailto:udngateway@hms-dbmi.atlassian.net).

## **B. Data Security**

### **Physical security**

For physical records (paper, photographs, etc.), appropriate physical access controls (e.g., locked cabinet in a locked room) are used at all UDN sites.

### **Electronic security**

In accordance with the UDN DSUA, all transfers of data between UDN sites utilize encrypted and authenticated data transfer methods. Each site is required to implement technical safeguards to ensure the security and confidentiality of electronic records, such as authenticated and authorized access, end-to-end encryption at rest and in transit, and audit trails.

### **Auditing**

The Gateway is assessed periodically via a risk-based approach to determine security, privacy, and compliance requirements. Local systems should have up-to-date virus scanners, disk encryption, workstation-level authentication, and lockouts on all machines accessing the Gateway.

### **Incidents**

If information were to be released from the Gateway, the DMCC would notify PIs and site coordinators, report the incident to the NIH IRB, and follow local procedures for incident reporting (<https://privsec.harvard.edu/get-help>).

## **VII. Data Sharing**

### **A. UDN Data Sharing Principles**

Data are expected to be shared so as to maximize the value of the significant public investment in the creation and operation of the network. In addition, given the special nature of the project, every effort should be made to share data in a timely and comprehensive manner. Individuals with undiagnosed conditions have often gone for extended periods of time without a diagnosis. Sharing results as rapidly and as broadly as possible holds the promise of accelerating diagnoses and identifying potential treatments for those with these often very serious conditions.

#### **NIH policies**

The UDN data sharing principles are consistent with the goals of the NIH Data Sharing Policy<sup>1</sup> which states that “Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.”<sup>2</sup> The UDN data sharing practices are aligned with the goals and requirements of the NIH Genomic Data Sharing Policy<sup>3</sup>. The Policy expects genomic data generated by NIH-funded research studies to be submitted to a relevant database no later than 3 months after data quality and control have been completed, and expects data to be released up to 6 months after data submission is initiated or at the time of acceptance of initial publication, whichever occurs first.

#### **Expectations of research participants**

Participants expect that their data will be shared as early and as broadly as possible in order to maximize the chances of reaching a diagnosis. Several participants have made their data available on the UDN website in the form of participant pages (<https://undiagnosed.hms.harvard.edu/participants/>).

#### **Commitment of investigators**

Investigators agree to further the mission of the UDN by creating new knowledge regarding the biochemistry, physiology, and mechanisms of undiagnosed diseases and improve diagnostic and management options for patients afflicted with them.

Investigators further agree to acknowledge that in pursuit of this mission, common UDN goals may supersede individual goals. Specifically, in the interest of rapid progress, investigators commit to:

- Model a collaborative, open, interdisciplinary spirit, characterized by mutual trust and respect across disciplines, individuals, areas of expertise, institutions, and by demonstrating interest and engagement beyond their own specific domains.
- Ensure that data generated at individual sites are comparable and additive by adhering to UDN data standards.
- Make data available to the Gateway and external databases in a timely and effective manner.

The following outlines the minimum requirements for sharing the data that are collected in the course of participation in the UDN.

## B. UDN Data Sharing Procedures

### Data Sharing and Use Agreement

Policies and procedures for data sharing within and outside of the UDN are described in the UDN Data Sharing and Use Agreement.

### Informed consent and data security

At the time of case submission, individuals consent to have their data shared within the UDN. Each participant is assigned a UUID and UDN ID by the Gateway. These identifiers are used for all data associated with that participant. Role-based access and physical security controls that are fully aligned with the sensitivity of the data at each point of use and access are employed by the Gateway. De-identified data shared outside of the UDN does not reveal individual identifying information, consistent with the HIPAA Privacy Rule<sup>4</sup>. Individuals assigned to tier 2-4 evaluations give consent to have their data shared more broadly as part of a network-wide informed consent process.

### Sharing identified data

*With UDN investigators:* All clinical, biospecimen, and sequencing data that are generated by the UDN effort are shared in a secure and compliant manner within the UDN. UDN data include data generated both in human subjects and laboratory research. All UDN investigators have access to the fully identified UDN data. UDN investigators who acquire UDN data must use the data responsibly and must monitor the use of the data by study personnel at their site.

*With disease experts:* In aid of a possible diagnosis, UDN data may be shared with disease experts who are not currently part of the UDN. In this case, the UDN CS that evaluated the participant must be notified that an outside expert has been consulted.

### Sharing de-identified data

#### DMCC

The UDN DMCC shares de-identified data in the following publicly accessible databases on behalf of UDN CSs.

- dbGaP: dbGaP<sup>5</sup> (Database of Genotypes and Phenotypes) is an NIH-sponsored controlled access repository that was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in humans. For each participant, phenotypic data entered into the Gateway and genotypic data, in the form of BAM and VCF files, are deposited in dbGaP.
- PhenomeCentral: PhenomeCentral<sup>6</sup> is a global repository and collaboration platform for secure data sharing in the rare and undiagnosed diseases community. PhenoTips<sup>7</sup> is a web-based tool for collecting phenotypic data that was fully integrated with the Gateway from the launch of the UDN on September 16, 2015 through June 30, 2024. During this period, phenotypic data entered through the Gateway and genotypic data, in the form of candidate genes, were deposited in PhenomeCentral, a node of Matchmaker Exchange<sup>8</sup>. Although records are no longer submitted to PhenomeCentral, existing

records remain in PhenomeCentral and the DMCC continues to triage match inquiries as they are received.

- ClinVar: ClinVar<sup>9</sup> is a freely accessible, public archive of the relationships between specific human variations and phenotype. For each participant diagnosed with a genetic condition, variant level data will be shared in ClinVar. All data are submitted in accordance with ClinVar requirements. Currently, this includes: gene symbol, reference sequence, descriptions of DNA and protein level sequence variants using Human Genome Variation Society (HGVS) nomenclature, variant type, transcript, cytogenetic and genomic positions, clinical significance and supporting evidence, demographic information, and phenotypic information. This information is retrieved from the CS and the Gateway.

### *Individual sites and Cores*

Matchmaker Exchange: As of June 30, 2024, the DMCC is no longer sharing UDN participant data in Matchmaker Exchange via PhenomeCentral. CSs are encouraged to share case level data via the Matchmaker Exchange node of their choice, The current list of linked Matchmaker Exchange nodes is available at <https://www.matchmakerexchange.org/>.

Additional databases: As data are generated by additional research cores, relevant databases for deposit of UDN data will be identified. These might include model organism databases, such as Flybase, Zebrafish Information Network, and the Mouse Genome Database, and metabolomics databases, such as the Human Metabolome Database and MetaboLights. In addition, the network will share data publicly on the UDN website. The UDN website<sup>10</sup> is a resource that provides extensive information about the goals, activities, and outcomes of the network. Phenotypic data and candidate gene names are displayed. Sharing data on the UDN website could lead to the identification of additional patients with similar phenotypic and genotypic profiles and could be useful for advancing scientific investigation and discoveries about these rare conditions.

### **Data sharing timelines**

De-identified participant data must be ready for deposit in dbGaP no later than 3 months after the clinical evaluation is complete and the initial sequencing report is issued by the clinical laboratory. This means that phenotyping needs to be complete and, if relevant, candidate genes need to have been identified and entered into the Gateway.

De-identified participant data must be ready for deposit in ClinVar and for display on the UDN website no later than 6 months after the clinical evaluation is complete and the initial sequencing report is issued by the clinical laboratory.

Although broad and timely data sharing are expected, UDN investigators have the option of delaying data sharing in certain unusual cases. A rationale for the delay and an estimated data release date must be submitted to the DMCC. The DMCC contacts the investigator shortly after the estimated data release date to assess progress and to determine if the data are ready to be shared.

### **Data release at publication**

There is a separate publications section within this document that describes these policies. If broad data release is required as a condition of publication by the authors or the publisher, the

DMCC should be contacted as soon as possible prior to making any commitments to ensure that the data release is feasible.

### Role of the DMCC

The DMCC facilitates, monitors, and reports on the effective and timely sharing of data within the UDN and beyond.

## C. References

1. NIH Data Sharing Policy. [http://grants.nih.gov/grants/policy/data\\_sharing/](http://grants.nih.gov/grants/policy/data_sharing/)
2. Final NIH Statement on Sharing Research Data, February 26, 2003. (<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>)
3. NIH Genomic Data Sharing Policy. <https://www.federalregister.gov/documents/2014/08/28/2014-20385/final-nih-genomic-data-sharing-policy>
4. HHS - Office for Civil Rights - HIPAA. <http://www.hhs.gov/ocr/hipaa/>
5. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, Feolo M. NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D975-9.
6. Buske OJ, Girdea M, Dumitriu S, Gallinger B, Hartley T, Trang H, Misyura A, Friedman T, Beaulieu C, Bone WP, Links AE, Washington NL, Haendel MA, Robinson PN, Boerkoel CF, Adams D, Gahl WA, Boycott KM, Brudno M. PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum Mutat.* 2015 Oct;36(10):931-40.
7. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, Chitayat D, Faghfoury H, Meyn MS, Ray PN, So J, Stavropoulos DJ, Brudno M. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat.* 2013 Aug;34(8):1057-65.
8. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, Brunner HG, Buske OJ, Carey K, Doll C, Dumitriu S, Dyke SOM, den Dunnen JT, Firth HV, Gibbs RA, Girdea M, Gonzalez M, Haendel MA, Hamosh A, Holm IA, Huang L, Hurles ME, Hutton B, Krier JB, Misyura A, Mungall CJ, Paschall J, Paten B, Robinson PN, Schiettecatte F, Sobreira NL, Swaminathan GJ, Taschner PE, Terry SF, Washington NL, Züchner S, Boycott KM, Rehm HL. 2015. [The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery](#). *Human Mutation*, **36**: 915–921. doi:[10.1002/humu.22858](https://doi.org/10.1002/humu.22858)
9. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D862-8.
10. The Undiagnosed Diseases Network. <https://undiagnosed.hms.harvard.edu/>

## **VIII. Clinical Protocols**

### **A. Introduction**

The clinical protocols component of the Manual of Operations “provides preliminary protocols and operating guidelines that will define an initial framework for common approaches to participant selection, data collection, laboratory investigation, and diagnosis, and serve as a base for further refinement by UDN investigators.” (From [RFA-RM-12-020](#)).

These approaches are outlined primarily in the UDN IRB protocol, which can be found here: <https://www.dropbox.com/sh/gcumcc9s7m8h15p/AABi0kMohQFGnCo2bdzz73gia?dl=0>. See [ClinicalTrials.gov record](#) for a publicly available summary of the protocol.

Key operational elements that are not included in the protocol are documented below.

#### **UDN Clinical Protocols Working Group**

The Clinical Protocols Working Group developed the clinical protocols as part of the Manual of Operations and, with input from the Steering Committee, continues to refine it. Should there be a need to vote on matters within the working group, each CS, the NIH Program, and the DMCC cast a single vote.

### **B. Detailed UDN Clinical Protocol**

#### **DMCC Helpdesk communications**

The Helpdesk email and phone number is staffed by personnel at the DMCC. When the Helpdesk receives a participant inquiry that rises above the level of general questions or concerns, the CS involved is notified. Such inquiries will also be brought to the attention of the Executive Committee and CIRB, as needed. If a participant expresses intent to contact Congress or any portion of the NIH directly, the Program Officer will also be notified. A general log of participant inquiries is kept at the DMCC.

#### **Identification of a UDN participant**

As general practice, no identifiable information should be shared with individuals not on the IRB protocol unless they have been directly contacted by a participant. If identifiable information needs to be shared with individuals not on the IRB protocol, sites should contact [UDN@hms.harvard.edu](mailto:UDN@hms.harvard.edu).

#### **Case submission process**

Cases are submitted through the UDN Gateway by the participant, parent guardian, or licensed healthcare provider. The UDN Helpdesk also submits cases on behalf of participants that complete the paper case submission form, which is available in English and Spanish on the UDN website. Additional case submission materials, including example study recommendation letters, are also available on the website: <https://undiagnosed.hms.harvard.edu/about-us/faqs>.

#### **Stages of enrollment**

- Case Submission Stage – from time of first inquiry or case submission start in the UDN Gateway through submission of a case to the UDN
- Review Stage – from submission of a case through medical records review and assignment by the CS to tiers 2-4 and referred to the consent stage
- Enrolled – participants assigned to tiers 2-4 in the UDN that have completed the stage 2 consent

Participants that have completed the review stage may be informed that their case has completed the first level of review. Participants should also be informed that they are not fully enrolled in the UDN study until stage 2 consent is completed, if they are assigned to tiers 2-4 evaluation. These stages of enrollment are also outlined on the UDN website: <https://undiagnosed.hms.harvard.edu/about-us/faqs>.

### **Tier evaluations**

The definitions of the tiers of UDN evaluations are as follows:

- Tier 1 – initial screening/record review
- Tier 2 – telemedicine visits and remote evaluations (may include short or long read DNA or RNA sequencing, clinical testing, clinical or research reanalysis of existing sequencing data)
- Tier 3 – comprehensive on-site evaluations (may include short or long read DNA or RNA sequencing, clinical testing, clinical or research reanalysis of existing sequencing data)
- Tier 4 – research activities (may include model organisms studies, research functional studies, cell-based assays)

### **Case review**

A minimal medical record review includes:

1. A reasonable effort to obtain all relevant medical records (potentially assigning the case to tier 1 if critical existing records cannot be obtained).
2. Sufficient review of records to:
  - a. Understand the significant components of the presenting illness,
  - b. Understand the differential diagnoses entertained during prior evaluations and evaluate the evidence used to endorse or discard each potential diagnosis,
  - c. Develop a summary differential diagnosis and evaluate the available evidence to endorse or discard each possibility, and
  - d. Make recommendations if further evaluation is needed to complete item c if the case is assigned to tier 1.

Medical information, such as medical records and the recommendation letter, are also analyzed and summarized by a language model (e.g., local language model, GPT-4) as a tool to assist UDN investigators in the review process. The language model is trained using UDN participant data. Language models are run by the DMCC team on local, secure devices or in a cloud-based, HIPAA-compliant, secure enclave only accessible by the UDN team.

The CS reviews the language model summary, the study recommendation letter, and medical records to determine the appropriate tier of UDN evaluation for each participant.

When the CS completes the record review, the following information is included:

- Category of primary condition
- Narrative summary of the participant's condition, which may include: brief summary, list of HPO term findings, prior workups, family history, medications, genetics, differential diagnosis, and next steps
- General measurements (height, weight, head circumference)
- Prior genetic testing requested and whether insurance denied coverage
- Affected family members
- Plan for UDN evaluation
- Total personnel time spent reviewing records and summary
- Tier assignment

The tier assignment is made primarily by individual CSs, except those for whom the CS is uncertain, has questions about, or thinks may be better served at another CS. The CS is encouraged to submit questionable cases to the UDN Consultation Service.

It is expected that about half of cases are pediatric, half adults. No more than half should be patients known to a CS, half completely new to any CS. "Patients known to a CS" are defined as any patient that is recommended by a health care provider from any of the institutions that are on that CS's award. A patient recommended by an outside provider is not considered "known to a CS" even if they have been seen previously at an institution on that CS's award.

CSs may also identify when they believe that a case assigned to tier 2-4 falls outside of the inclusion and exclusion criteria specified in the protocol. For the inclusion and exclusion criteria, please see NIH IRB Protocol (section: "Description of study population"). In this scenario, the case is submitted to the UDN Consultation Service for recommendations. Cases identified as falling outside of the inclusion and exclusion criteria may be sent to the Executive Committee, as needed.

### **Participants assigned to tier 1 evaluation (medical record review)**

In addition to the record review form, which must be completed for all participants, cases assigned to tier 1 evaluation during the review stage require the following additional information to be entered in the UDN Gateway:

- Tier assignment – Tier 1
- Reason Assigned to Tier 1 (select one)
  - Unlikely to find a diagnosis
  - Participant should seek consultation or specific testing
  - Agree with existing diagnosis or diagnoses
  - Diagnosis made by external agent since initial application
  - New diagnoses suggested
  - Insufficient records
  - Other (free text)
- Recommendations Made (select all that apply)
  - Consultation with clinician
  - Single gene testing
  - Panel testing
  - Genome-scale sequencing
  - Non-sequencing genome wide diagnostic assay
  - Other laboratory test
  - Imaging

- Treatment/medication
- Other research study
- Support group
- Other
- Tier 1 assignment letter – should be uploaded into the Gateway when sent to the participant and referring provider

### **Case submission outcomes**

1. Participant and site are appropriate for tier 2, 3, or 4 evaluation
2. Participant appropriate for tier 2, 3, or 4 evaluation but reassigned to a different site
3. Participant requires further testing or evaluation and may be reconsidered following receipt of the results
4. Participant appropriate for tier 1 evaluation (medical record review) only
  - a. A diagnosis was identified based on the review.
  - b. A potentially beneficial referral was identified based on review. Further evaluation at the UDN may not be necessary to make a diagnosis.
  - c. The participant is more appropriate for an expert site outside of UDN such as a research protocol at the NIH or elsewhere, the participant is more appropriate for an expert at an academic medical center or elsewhere, or the participant is not appropriate for further evaluation in the UDN and no alternative can be identified.

In all cases the referring provider and the participant are informed of the decision, generally within 60 days after receipt of all the requested information. The CSs send out the disposition letters for all tier assignments. If assigned to tiers 2-4, the CSs send out the Letter to Participant(s) to the participant's healthcare provider(s) and the participant (see Appendix 3: Template Letters for Patients and Health Care Providers). When individuals are assigned to tier 1, their case information is stored securely and indefinitely in the Gateway.

The DMCC administers follow-up surveys for participants assigned to tier 1 to collect feedback on the information received and participant-reported diagnostic outcomes.

### **UDN Consultation Service**

The UDN Consultation Service is intended to support investigators of UDN CSs and unaffiliated clinical providers seeking consultation for prospective and enrolled participants in the UDN who do not yet have a diagnosis.

Purpose:

- Facilitate access to expert advice for UDN members seeking guidance regarding diagnostic strategies for participants assigned to tiers 2-4 in the UDN.
- Provide consultation from specialists including, but not limited to, medical genetics, bioinformatics, designing and conducting gene function studies in a variety of model systems (e.g., mouse, Drosophila, zebrafish, C. elegans, human cell-based models), and non-routine clinical genomics or metabolomics investigation.
- Support affiliated and non-affiliated UDN providers in assessing a patient's eligibility for the UDN
- Aid in the transfer of applications and participants between UDN sites

Instructions for submission:

1. Please use the following REDCap survey to request a consultation:  
<https://redcap.link/bfdh4q8h>
  - a. Note: please be as specific as possible when formulating your question(s) for the consultation team. This will help us provide a prompt and productive service.
2. Upon submitting the survey, the UDN Consultation Service team will reach out by email within 7 business days of submission to schedule a 30-minute time slot for consultation. Consultation sessions will be conducted on Zoom.
3. Attend consultation session at the scheduled time.
4. After consultation session is complete, a summary of the recommendations made during the session will be provided. The summary will be sent by email to the address provided in the initial survey submission.

The team will attempt to address the request at the time of the session but may not be able to answer questions in full. In these cases, recommendations for further actions and a follow-up session to continue discussion of the case may be provided.

If there are issues with scheduling or with signing up for a consultation, please reach out to the UDN Consultation Service team. The team will attempt to accommodate scheduling requests.

### Appeal and reconsideration process

The UDN will not accept appeals related to enrollment decisions. Language that there is no appeal process and that the UDN tier assignment is binding should be communicated to participants during the case submission process and in the FAQs online.

If new medical information becomes available, participants may request sites reconsider their tier assignment. For these requests to be considered, participants must provide an updated letter of recommendation from their referring provider. They may also provide an updated patient narrative if desired. These documents should detail any new information that may be relevant (i.e., new test results, new symptoms, new family member with symptoms, etc.). Sites may choose to re-review cases at their discretion. Sites are not obligated to reconsider cases, especially if the participant continues to approach the site over time. At any point, the site may choose to submit the case to the UDN Consultation Service for a secondary review.

### Medically actionable findings

The SC reports medically actionable findings in variants in the genes recommended for such reporting by the American College of Medical Genetics and Genomics. In addition, the SC reports other medically actionable findings beyond the currently recommended genes, provided these additional findings meet the threshold of having a defined medical treatment or specific management guidelines for disease surveillance. Stringent criteria for interpretation of variants in these medically actionable genes is applied; reported medically actionable variants have either been previously reported as pathogenic or expected to be pathogenic based on the usual molecular mechanism associated with the gene.

EXAMPLES OF MEDICALLY ACTIONABLE CONDITIONS NOT INCLUDED ON THE ACMG LIST	
Syndrome/Disorder	Gene
Progressive Familial Heart Block, Type 1B	<i>TRPM4</i>
Long QT type 4	<i>ANK2</i>

Familial Gastric Cancer	<i>CDH1</i>
Von Willebrand disease	<i>VWF</i>

Carrier status for autosomal recessive conditions may also be identified. Examples include cystic fibrosis and Tay-Sachs disease. Individuals who are carriers of a specific genetic disorder have one normal, working copy of the gene, and one changed copy of the gene. Because they have one working copy of the gene, carriers do not usually have symptoms of the disorder. Findings without medical treatment or specific management guidelines are not reported by the SC. Examples of these types of disorders include Alzheimer disease and late onset skeletal dysplasias (e.g., spondyloepiphyseal dysplasia - *TRAPPC2* gene).

During the pre-test genetic counseling/informed consent session, probands (or legal guardians) are given the option of receiving genetic results that are unrelated to the indication for testing, including results related to: (1) medical conditions with treatment or management options, known as medically actionable findings, and (2) carrier status findings (only applicable for adult probands). The proband's genetic testing results are shared with family members only if the proband (or legal guardian) provides permission. During the pre-test genetic counseling/informed consent session, family members are given the choice of receiving medically actionable findings if they are identified. A search for medically actionable findings will not routinely be performed in family members. However, medically actionable findings may be discovered by chance during the genetic testing process.

If a medically actionable finding is detected in a proband and is also present in a family member who consented to receive medically actionable findings, the finding will be returned on the family member's report. If a medically actionable finding is present in a family member but not in the proband, it will not be detected or reported by the SC. If a medically actionable finding is detected in a proband and a family member who is not enrolled in the UDN is interested in learning if they also have this finding, the family member will be referred to a clinical genetics program for genetic counseling and testing. As this counseling and testing is not done as part of the UDN, cost associated with this clinical follow-up is billed to insurance. It is important to note that the lack of reportable medically actionable findings does not rule out any disease-causing genetic changes. If a medical concern is raised regarding a specific condition, further genetic testing may be warranted.

### **Withdrawal procedures**

During the UDN study, participants may withdraw or be withdrawn by UDN investigators. Withdrawal scenarios include, but are not limited to, the following:

1. Participant elects to withdraw
2. Participant is not compliant with study procedures or does not follow instructions given by UDN investigators
3. Participant passes away before stage 2 consent
4. Participant is assigned to tier 2-4 evaluation, but is unable to be evaluated

If the participant chooses to withdraw from the study, the participant sends a written withdrawal to the PI at the CS (if possible) and their status is changed to withdrawn in the UDN Gateway. No new information is collected in accordance with the UDN consent form. If the participant asks to have their clinical and research information and specimens destroyed, the site coordinator destroys the participant's research records and specimens at the CS and contacts

the DMCC to request deletion of information from the participant's Gateway record. The site coordinator also contacts the biorepository to destroy the participant's research specimens and collaborators to destroy research records or specimens, if possible. If the individual was assigned to tier 2-4 evaluation and signed the UDN consent form, the site coordinator saves the signed consent form with the written withdrawal (if available).

### **Billing procedures**

The RFA stated that the CSs could bill participants' health insurance for clinically indicated evaluations, procedures and tests, and use grant funds for underinsured or uninsured participants. The CSs were also required to provide participant transportation and lodging/meals during the one-week stay at the CS. These practices would ensure that participants did not incur out-of-pocket expenses and enable all subjects to have access to the UDN, irrespective of their health insurance status. This would also allow all participants the same experience as at the NIH UDP with no out-of-pocket expenses. However, while establishing billing procedures at the CSs outside of the NIH UDP, it became evident that there were several challenges to implementing these practices. All the CSs were told by institutional representatives that insurance co-pays and deductibles could not be waived or reimbursed by the grant or institutional funds, due to a federal anti-inducement law that is framed for Medicare and Medicaid but is often applied to other insurance policies {42 U.S.C. § 1320a-7a(i)(6)}. Two sites were told by institutional representatives that they could not both bill insurance and pay for participant travel/lodging due to a federal anti-kickback law {42 U.S.C. § 1320a-7b}.

This led to the formation of a Billing Working Group to resolve the issues so as to not place the CSs and the participants at an undue disadvantage. After considering the legalities and the available choices, two billing options were created. The first option utilizes grant funds solely to cover all the participant evaluations, made feasible by institutional discounts (~80%) for participant care performed as part of NIH-funded studies. The second option bills the participant's insurance companies for the clinical evaluations and cover underinsured/uninsured participants or tests not reimbursed by insurance with grant or institutional funds. Each CS can choose which option is best based on their institutional policies. In addition, each CS can re-evaluate and change to the other option based on their institutional policies.

### **NORD fund**

To enable payment of participation-related costs, the UDN is collaborating with the National Organization for Rare Disorders (NORD, <https://rarediseases.org/>). NORD Patient Assistance funds may be used for pre-evaluation tests and co-pays/deductibles for UDN tests and procedures billed to insurance. It may also be used to support sequencing costs or travel costs. If you have questions about what the funds may be used for, please contact the email below.

The UDN site must refer the participant to NORD via their RareAccess portal (<https://rareaccess.org/>). Typically, NORD reaches out to the participant to create their profile within 1 business day. Invoices may be submitted once the participant's profile has been created. UDN participants and site coordinators may call or email NORD for more information on the Patient Assistance Fund. NORD will provide specific instructions on how to apply for the funds.

Phone number: 844-858-0451  
Email: [undiagnosed@rarediseases.org](mailto:undiagnosed@rarediseases.org)

Workflow for using NORD funds for sequencing performed at the SC:

1. CS determines if the participant meets criteria for accessing NORD funds
2. CS checks if their funding will support the cost of the sequencing
3. CS refers the participant to NORD through the rareaccess.org referral portal.
  - a) If sequencing is covered by insurance - choose copay program
  - b) If sequencing is not covered by insurance - choose medical assistance program
4. NORD contacts the participant to complete the enrollment process
5. UDN site submits sequencing request and selects "Site will submit invoice to NORD for payment" as payment option
6. SC issues a family level invoice to the CS
7. CS pursues payment of the invoice through NORD

### **Mercy Medical Angels**

Mercy Medical Angels (<http://mercymedical.org/>) may allow for provision of commercial air travel expenses for participants who are financially stressed.

### **Congressional inquiries**

If a congressional communication related to a participant is received, the UDN Program Director should be notified as soon as possible so that a coordinated response can be initiated. Most congressional offices will reach out to the NIH directly as the federal agency involved; however, if a request for information is received by a CS or the DMCC, general guidelines on replies are detailed below.

Timeline for response: Technically, there is no timeline to respond to a letter from a personal office, but a response should be made as soon as possible. For an email inquiry, a response email should be sent immediately acknowledging receipt and noting that the requested information will be provided as soon as possible. If the request was received via email, an identical postal mail can be sent if desired. If the request was received over the phone, a call to acknowledge receipt is appropriate.

Individuals to be contacted: The CS PI, DMCC, and UDN Program Officer should be contacted regarding any communications with a congressional office.

Official response: The NIH will notify the CS or DMCC when an official response can be distributed.

### **C. References**

1. Napoles AM, Gregorich SE, Santoyo-Olsson J, O'Brien H, Stewart AL. Interpersonal processes of care and patient satisfaction: do associations differ by race, ethnicity, and language? *Health Serv Res.* 2009;44(4):1326-1344. doi: 1310.1111/j.1475-6773.2009.00965.x. Epub 02009 Mar 00931.
2. Stewart AL, Napoles-Springer A, Perez-Stable EJ. Interpersonal processes of care in diverse populations. *Milbank Q.* 1999;77(3):305-339
3. Stewart AL, Napoles-Springer AM, Gregorich SE, Santoyo-Olsson J. Interpersonal processes of care survey: patient-reported measures for diverse groups. *Health Serv Res.* 2007;42(3 Pt 1):1235-1256.

## **IX. Clinical Research Support Core Subawards**

### **A. Background**

In collaboration with the Steering Committee and NIH Program staff, the Clinical Research Support Core developed and implemented a versatile subawards' administration platform that permits evaluation, management, unbiased prioritization, and tracking of subawards. Subawards will be funded to support local site coordination or data submission (if not supported through an NIH U01 award or Administrative Supplement), to confirm a diagnosis with metabolomics or gene function studies, and to obtain pilot data to facilitate future, larger research studies. UDN participants regardless of participation phase can be included in subaward projects. There will be \$50-\$100,000 total (direct plus indirect) costs available per subaward (average \$50,000/subaward) with the potential of larger allocations if approved by the Steering Committee. All subawards must be completed within the grant year of submission (04/01-03/31), will be of 12 months' duration, and PIs must have all invoices submitted by 03/01 of the grant year. No carry over funding will be permitted. The Subawards Review Committee will consider a limited number of phased applications each year with milestones that will determine funding for subsequent phases and may extend beyond 1 year.

### **B. Standing Subaward Review Committee (SRC)**

Permanent members with expertise in CS coordination/operations, clinical genetics, metabolomics, model organisms, bioinformatics, data submission to the Gateway, genomics, NIH Program staff, and ad hoc members with expertise required for consideration of specific subawards will be identified.

SRC members must complete the Confidentiality and Conflicts of Interest Agreement before each review. This agreement includes guidelines about disclosing and reviewing conflicts of interest as well as situations where members cannot be part of the review process.

### **C. Subaward submission dates and format**

1. Subaward recipients must be an X01 or U01 funded PD/PI
2. Distributed funding across CSs and project areas will be taken into consideration. Sites should consider only submitting 1-2 proposals and making proposal prioritization known to the SRC.
3. Subawards will not support more advanced research proposals (e.g., characterizing a gene or genomic variant already known to cause the participant's disease) since these types of proposals can be submitted to the NIH using standard grant mechanisms and will not support projects funded through U01 or Administrative Supplements.
4. Subaward submission dates will be February 1 and will be 12-month awards.
5. Brief summaries (~1/2 page) that describe achieved vs. planned project progress and assurance that funds will be spent by the end of the grant year will be required at 6-month intervals for all awards; summary reports will be required at the completion of all subawards; no cost extensions will not be permitted.
6. Format of subawards:
  - i. CS operations and data submission

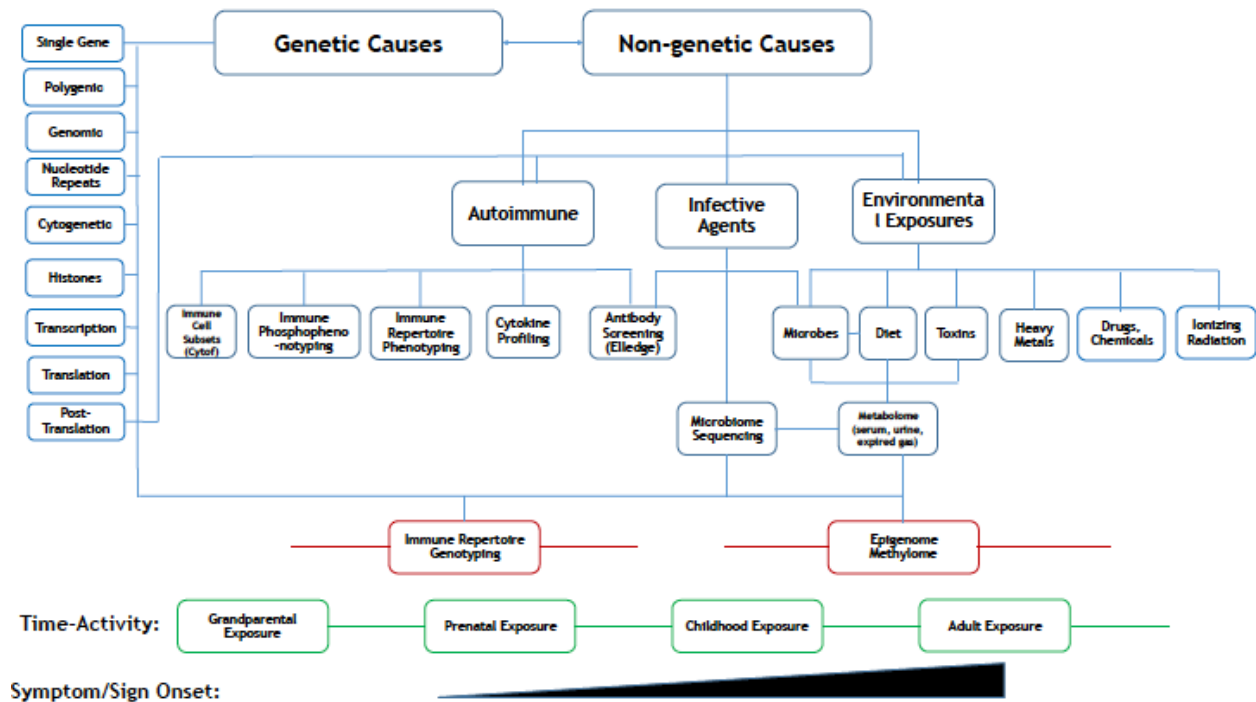
- a) Plan (total 2 pages): overview of current plans for and progress of site-led efforts to increase efficiency of participant evaluation, throughput, and follow up and/or to increase diversity of UDN participants; Scope of Work; description of how activities/personnel funded as part of the subaward proposal contribute to existing or new activities but are not fully financially supported by other means (e.g., U01 grants, administrative supplements, institutional support)
- ii. Gene function studies and other research
  - a) Research plan (total 5 pages): Title, PI/Program Director, key personnel, Specific Aims, Research Strategy, Scope of Work and Deliverables and Milestones
  - b) For studies that propose use of model organisms or other pathogenicity models (e.g., cell culture), subaward applications should provide evidence of variant pathogenicity and include a bioinformatics review by the Model Organism Screening Core that supports the application. Requests for the bioinformatics review must be submitted to Dustin Baldrige and Michael Wangler no later than January 1 prior to the February 1 application due date.
  - c) For applications that propose use of other strategies for functional confirmation, availability of relevant resources (e.g., primary cells or cell lines, proposed metabolomics signature or environmental exposure) should be provided.
- iii. Additional materials required for subaward applications: R&R Detailed Budget, Budget Justification, Letter of Intent, Performance Location Sites, Resource and Data Sharing Plan, Facilities and Other Resources, Key Personnel Profiles, NIH Biosketches and Other Support for all key personnel, Equipment (if applicable), IACUC approval (if applicable), IRB approval (if applicable), documentation of executed material transfer agreements and financial subcontracts required for the proposed project (if applicable)
  - a) The Subawards Review Committee will be willing to consider a limited number of 2 year phased applications. Continuation of phased funding will be contingent on achieving initial milestones and can extend no longer than 2 grant years.

## X. Environmental Exposures

### A. Background

The UDN fully recognizes that all diseases are likely determined by many factors, including “minor modifying” genes and environment, through molecular mechanisms such as by epigenetics, transcriptomics, post-translational modification of the proteome, the microbiome, infections, and immune responses. One strategy used for identifying possible environmental factors, outlined in the figure below, is intended to systematically screen participants for such factors, expedite the efforts of clinical investigators, and perhaps improve the participants’ health by intervening against current environmental hazards.

Figure: A holistic view of the multifactorial, ecogenetic hypothesis of disease etiology.



### B. Environmental Exposures Questionnaire

Current and previous versions of the Environmental Exposures Questionnaire are available on Dropbox: <https://www.dropbox.com/sh/ikfdp6ta5vp92y8/AAD6jlgYvfUVJ8nCXNrS2tFua?dl=0>. As of February 26, 2024 the survey is no longer administered electronically to all consented probands. Existing survey responses remain visible in the UDN Gateway. Sites interested in continuing to administer the questionnaire may do so using the paper version.

To create the questionnaire, existing questionnaires were leveraged including the American Academy of Clinical Toxicology Exposure History Form,<sup>1</sup> CDC Environmental Exposure History,<sup>2</sup> PhenX project,<sup>3</sup> and NHANES,<sup>4</sup> in addition to questions developed with the assistance of various experts in multiple environmental exposures.

The areas addressed by the questionnaire include:

- a) occupational exposures to dust and chemicals,
- b) residential exposures to chemicals and pollution,
- c) use of alcohol, tobacco, and other drugs (prescribed and recreational),
- d) exposures from hobbies and travel,
- e) exposures to infectious disease vectors and sexually transmitted diseases,
- f) unusual dietary histories, and
- g) parental environmental and occupational exposures before and during pregnancy that resulted in the birth of the participant.

This questionnaire was intended to be used as a clinical screening tool to identify participants who would potentially benefit from a more intensive environmental exposure assessment at the UDN CSs.

As of November 2018, the Environmental Exposures Working Group suspended review of all questionnaires, pending design and implementation of improved procedures for efficiently identifying participants who report highly likely significant environmental exposures. If administering the paper version of the questionnaire, CSs are urged to review responses and contact Working Group members for expert opinion in assessing alleged exposures of concern.

## C. References

1. American Academy of Clinical Toxicology Exposure History Form  
<http://www.clintox.org/sections/occupationalenvironmental>
2. Agency for Toxic Substances and Disease Registry. Case Studies in Environmental Medicine (CSEM): Taking an Exposure History.  
[https://www.atsdr.cdc.gov/csem/exphistory/docs/exposure\\_history.pdf](https://www.atsdr.cdc.gov/csem/exphistory/docs/exposure_history.pdf)
3. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, Hammond JA, Huggins W, Jackman D, Pan H, Nettles DS, Beaty TH, Farrer LA, Kraft P, Marazita ML, Ordovas JM, Pato CN, Spitz MR, Wagener D, Williams M, Junkins HA, Harlan WR, Ramos EM, Haines J. The PhenX Toolkit: Get the most from your measures. *Am J Epidemiol.* 2011;174(3):253-60. PMID 21749974..
4. Centers for Disease Control and Prevention, National Center for Health Statistics, National Health and Nutrition Examination Survey.  
[http://www.cdc.gov/nchs/nhanes/nhanes\\_questionnaires.htm](http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm).

## XI. Sequencing

### A. Flow of Samples to Sequencing Core

#### I. Sample collection and Specimen Requirements

1. The clinical sites (CSs) arrange for sample collection before or during the clinical evaluation. Test kits may be shipped to a CS or directly to a patient's home. To request kits (blood in EDTA or buccal swabs):
  - a. Email [logistics@baylorgenetics.com](mailto:logistics@baylorgenetics.com) and specify buccal or blood kit
  - b. Include name of one individual, address and phone number, and number of kits requested.
    - i. Example: Could we please request that you send out a kit for buccal collection for TWO UDN family members? [Name, phone number, address.]
  - c. Website: <https://www.baylorgenetics.com/supplies/>
  - d. Test codes: use UDN codes
  - e. Specify blood or buccal, and number of kits
  - f. Buccal kits come with instructions for the patient and family. They will need to label each sample collection individually. Only one name in the email is required for shipping purposes.
2. Specimen Requirements:
  - a. Blood in EDTA (purple top)
    - i. Draw blood in an EDTA (purple top) tube(s) and send 3-5 cc (adults and children) and 3 cc (Infants under age 2).
    - ii. Ship at room temperature in an insulated container by overnight courier. Do not heat or freeze. Sample must arrive within 72 hours of collection.
  - b. Buccal swab
    - i. Collect with ORAcollect Dx (OCD-100) self-collection kit provided by Baylor Genetics with instructions. It is highly recommended the sample be collected by a healthcare professional.
    - ii. Ship at room temperature in an insulated container by overnight courier. Do not heat or freeze. Sample must arrive within 72 hours of collection.
  - c. Skin biopsy
    - i. Note that there will be an additional charge for cell culture.
    - ii. Collect 5 cubic millimeters of skin from a central location (e.g., buttock or upper thigh) rather than a distal location (e.g., foot) to enhance cell viability. Place the sample in a separate sterile container with RPMI media. If RPMI media is unavailable, use a sterile container with a small amount of sterile saline.
    - iii. Ship at ambient temperature (18-25°C/64-77°F). Protect from excessive heat and ship in a cooled container during summer months. The sample must arrive within 72 hours of collection.
  - d. Cultured fibroblasts
    - i. Send 2 T25 flasks at 80-100% confluence.
    - ii. Ship at room temperature in an insulated container by overnight courier. Do not heat or freeze.
  - e. Extracted DNA
    - i. Send 15-20ug of extracted DNA (minimal concentration of 50ng/uL and A260/A280 of ~1.7).

- ii. NOTE: We do not need exactly 20 ug, but if sending less, please specify how much you are planning to send when submitting the sequencing request in the UDN Gateway.
  - iii. Ship at room temperature in an insulated container by overnight courier. Do not heat or freeze.
3. Additional sample collection considerations:
  - a. If the individual has had a bone marrow transplant or recent blood transfusion, DNA from fibroblasts is preferred.
  - b. Samples from deceased individuals should be discussed with the sequencing core (SC) on a case-by-case basis.

## II. Sample shipping guidelines

1. CS prepares DNA samples for shipment to the SC. DNA samples should be sent as complete families (including all family members that will be included in the analysis) excepting clinically urgent samples that warrant prioritized sequencing. Urgency is at the discretion of the CS.
2. CS completes a sequencing request form in the Gateway for each DNA sample being sent for sequencing.
3. CS enters and releases updated phenotype information in the Gateway for use by the SC in their analyses.
4. CS enters shipping information (date DNA sent and tracking number) in the Gateway and ships samples.
  - a. Please note that the Gateway provides alerts for shipping of UDN samples, but shipment tracking needs to occur at the CS/SC level.
5. Gateway sends an automated email to alert the SC of sample shipment and available phenotypic data.
6. SC acknowledges receipt of samples by entering date DNA received in the Gateway.
7. If a submitted DNA sample does not pass QC at the SC or is otherwise deemed unacceptable, the SC contacts the CS site directly via phone or email to request a replacement.
8. Sample labeling discrepancies are addressed on a case-by-case basis at the discretion of each SC.
9. Please ship samples for exome sequencing, genome sequencing, and targeted variant testing to the address below. For RNA-Seq, skip to section XI.F.

Baylor Genetics  
2450 Holcombe Blvd. Ste. 2210  
Houston, TX 77021-2024  
(800) 411-4363

## B. Flow of Clinical Information to Sequencing Core

1. The CS are responsible for sample collection. The DMCC and SC will not know about or track the DNA specimens until they are shipped to the SC. The CSs are encouraged to collect all specimens for a family before sending them, but additional family members may be added later if necessary. For sites without an option for DNA extraction from a CLIA-certified laboratory, arrangements can be made with the SC for DNA extraction. Sequencing will commence when samples from all family members has been received. The 2-week TAT will commence after all samples have been extracted.

2. Typically, CSs send an aliquot of DNA extracted in a CLIA-certified lab and keep the remainder of the DNA for future procedures.
3. Samples may be sent either before or after the evaluation of the proband. In either case clinical information and a pedigree, including the relationships of all submitted family members to the proband, should be added to the Gateway as soon as possible for samples submitted for sequencing. This information is used by the SC for their analysis.
4. DNA samples submitted for sequencing must be labeled with participant name, date of birth, and "UDN".
  - a. Local identifiers may also be included at the discretion of the CS. Local identifiers will appear with participant name on sequencing reports (i.e., John Doe, UDP\_12345).
5. Other required information for sample submission includes:
  - a. Gateway consent form
  - b. Gateway sequencing form
  - c. Affection status of family members

## C. Exome Sequencing

This section covers sample intake, library preparation, whole exome capture, and sequencing at Baylor Genetics. It describes the sample flow from DNA sample receipt to production of WES data, including appropriate quality control and assurance procedures.

### Sample intake

A visual inspection of the sample tubes is conducted when received by the SC. Sample tube label is compared with information entered in the Gateway to ensure consistency and completeness of the Gateway data, consent forms, proper sample labeling, and sample tube integrity. Samples are accepted if no discrepancies are found, sample labels match, and no tube damage is observed. If any of the above criteria is not met, Baylor notifies the referring CS.

Once accepted, samples are accessioned into the Laboratory Information Management System (LIMS) system. Sample information in the Gateway is entered into the WGL LIMS database. Each sample is assigned an internal lab number, as well as a family number in LIMS. 1D bar code labels with participant specific information (unique identifier) including participant name, date of birth (DOB), lab number and family number are attached to the stock DNA tube. Subsequently samples are aliquoted from the stock tubes into 2D barcode tubes. The samples in 2D bar code tubes are processed for exome sequencing. Before sample transfer, the record for a sample is first opened in LIMS, then the 1D barcode label on the stock tube of the sample is scanned and the LIMS automatically verifies if the sample ID in LIMS matches that on the label. Then, the 2D bar code on the aliquot tube is also scanned to link the two bar codes in LIMS before sample transferring occurs. These steps are to ensure the chain of custody remains intact during sample transfer.

### DNA sample QC

DNA samples are screened to quantify DNA as well as determine DNA quality. To determine DNA concentration and purity, the samples are evaluated using the Lunatic Spectrophotometer. Passing criteria include:

1. Sample A260/A280 ratio between 1.8-2.0 and A260/A230 greater than 1.5.

2. Sample concentration greater than 30ng/ul and contains more than 1ug of DNA

To verify DNA integrity and relative size, the same dilution of sample is loaded on a 0.8% E-Gel. Passing criteria include:

1. Gel image is clear and shows no DNA degradation

If a sample does not meet the criteria above, the CS is notified.

### **Pre-capture library preparation**

In order to meet UDN sequencing objectives, we use our quick whole exome sequencing (QWES) protocol. QWES is an optimized version of the standard Illumina (ILM) library preparation workflow that reduces library construction time to 3-4 hours.

Library construction either manually or through a completely automated process on the Tecan liquid handler platform with an incorporated LIMS tracking system. Before starting library preparation, all primers and adapters lots are validated, and the appropriate dilutions are prepared. Negative (H<sub>2</sub>O) and reagent blank controls are included. Robot operator closely monitors each transfer step. Pre-capture library preparation involves the following steps:

#### Normalization and shearing

DNA samples are normalized to 750 ng total. Samples are loaded into Covaris microtubes in 50 ul aliquots and sheared to approximately 250-500 bp using the Covaris E220 ultrasonicator. Shearing efficiency is assessed using a 2.2% flash gel. Fragments should range from 100-600bp with average of 250-500bp. If most sheared fragments are larger than 800 bp, the sample is re-sheared.

#### End repair and 3' Adenylation

Fragmented DNA samples are treated with the KAPA HyperPrep Kit (Roche KK8504). After the master mix has been added to fragmented DNA, the samples go through incubations at 20°C for 30min followed 65°C for 30min to complete both End Repair and A-tailing in one reaction.

#### Ligation

Post dA-Tailing, DNA Samples are ligated with Illumina multiplexing paired-end (PE) adapters by using the ligase and buffer provided in the KAPA HyperPrep kit, incubate at 20°C for 20 minutes. SPRI beads and 80% ethanol is used for cleaning up the post ligation products. SPRI beads are removed post 40 ul elution buffer added.

#### Enrichment

Ligated DNA samples are enriched for total 6 cycles with 2X KAPA HiFi HotStart Ready Mix PCR kit (catalog#: KK2612) and Illumina PE PCR primers. AB GeneAmp PCR System 9700/Veriti are used for amplification Enrichment.

#### Post-enrichment QC

The enrichment PCR efficiency analysis is performed on 2.2% FlashGel by checking the product intensity. The FlashGel analysis is performed after 6 cycles and is re-run if the band is too weak. Additional PCR cycles can be added for samples with low yield.

1. No more than 9 cycles total can be run for samples
2. If the amount of the post-PCR product is insufficient after a total of 9 cycles, the whole process needs to be repeated.
3. The negative (H<sub>2</sub>O) and reagents blank control should give no product

To check size distribution and quantify the final library, the sample is run on an Agilent Bioanalyzer 2100 DNA 7500 Chip.

1. The library sizes should range from 200 – 750 bp (Majority are 250-550bp) with the peak ranging from 250 - 350 bp.
2. The yield of library should be more than 1 ug. If pre-capture library yield is lower than 1ug, the library preparation is repeated. No adaptor dimer and free primers are visible.
3. See Figure 1 for an example of a passing pre-capture library.

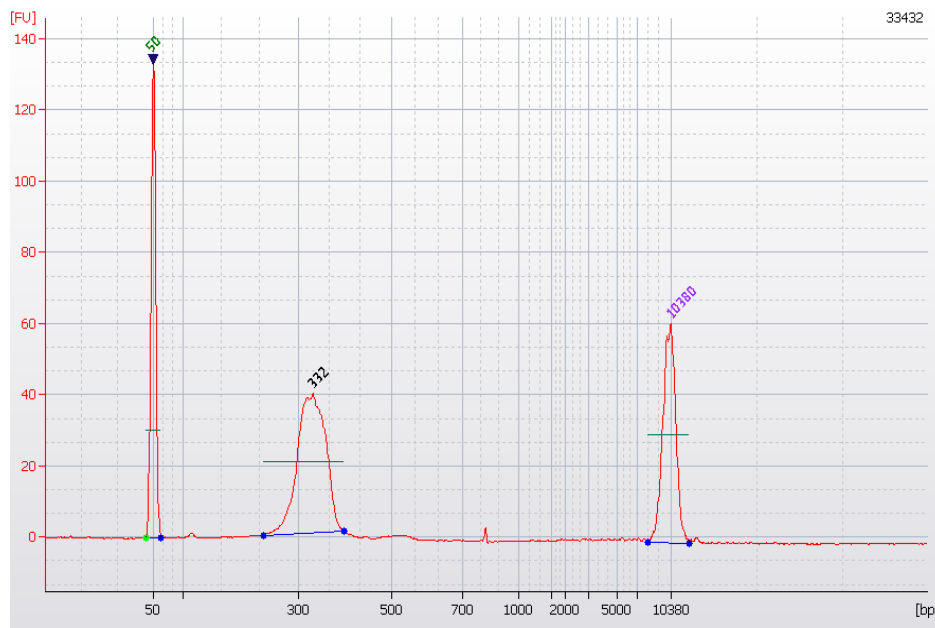


Figure 1: Illumina paired end DNA library on Agilent Bioanalyzer 2100

### Whole exome capture

The whole exome capture utilizes the Twist Comprehensive Exome probe, that targets approximately 36.8 Mbp of genomic DNA based on the most accurate curated subset—CCDS database. Additionally, the single nucleotide polymorphism (SNP) probe is spiked in to assist with contamination calculation.

Solution capture is initiated by combining 300ng-1ug of each of the pre-capture library (up to 12 samples & 3.6 ug total in a pool), using the enhancing oligos and wash buffers from the KAPA HyperCapture Reagent kit (Roche, 9075917001)

This mixture is concentrated using SPRI beads and re-suspended in Hybridization buffer and Formamide. The mixture is denatured for 10 minutes and Twist Exome probe is added. The mixture is incubated at 56°C for 16-18 hours. The following day captured DNA is washed and recovered.

Post-Capture PCR amplification is performed using KAPA HiFi HotStart DNA polymerase with total 12 cycles.

### Final library QC

#### FlashGel

Capture efficiency is checked using a 2.2% FlashGel. If the intensity is too low, the capture process needs to be repeated. Only primer bands should be seen for negative and reagent blank controls.

#### Bioanalyzer

To assess size distribution and quantify the final post-capture product, libraries are loaded onto DNA 7500 Chip for assessment on the Agilent 2100 Bioanalyzer. The majority size should be around 300-400 bp with the concentration above 20 nmol/L. See Figure 2 for an example of a passing final library.

Post-capture libraries must pass all QC checks to proceed to cluster generation and sequencing. For each sequencing run, individual WES pool containing up to 12 barcoded libraries are normalized to 10nM and pooled into one pool with other sample pools to fit in the capacity of one lane of the sequencing flowcell

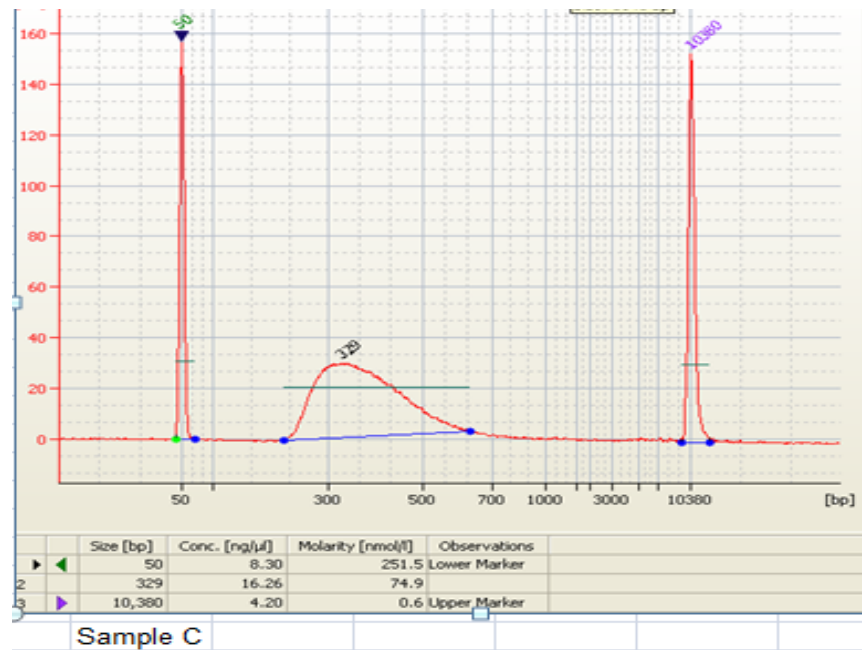


Figure 2: Final post-capture library on Agilent Bioanalyzer 2100

## Sequencing on NovaSeq X plus

The NovaSeq X plus are employed for sequencing (19 to 24-hr cycle time) to generate 100bp or 150bp paired-end reads in a format of 3 samples per lane to generate 10-12 GB per sample. Target coverage for proband and parental samples is >100x. The WGS LIMS system is utilized to track the run set up, status and quality metrics. Pertinent metrics and passing thresholds are provided in the tables that follow.

The performance of the run is monitored, and the metrics below are recorded to assess quality at a particular step of the sequencing run, evaluate library quality and concentration, detect any potential sequencing reagents and/or optical issues.

Passing filter rate	>60%
Q30	>85%

The capture analysis is incorporated in the analysis pipeline and provides metrics to gauge the overall quality of the capture process. This pipeline reports:

1. Proportion of the aligned reads that map to the targeted region, which is relative to the effective enrichment of the capture
2. Distribution of coverage across the targeted bases; specifically, the fraction of targeted bases covered at 1x; 10x, 20x, 40x

The complexity of the capture library is assessed by calculating the number of alignment reads that occur from PCR duplicates. If needed, these reads can be removed from the analysis.

Key metrics that have been developed and are reviewed before moving to case review/reporting steps following the table below.

Reads Aligned to target	>50%
Target bases covered at >20x	>90%
Target bases covered at >40x	>80%
Mean coverage of target bases	>100x

As an additional quality control measure, samples are also analyzed by SNP on NGS WES data to calculate the contamination and use molecular tracer to ensure correct sample identification and to assess sequencing quality. The data is analyzed using an automated pipeline that produces concordance and contamination scores.

Tracer	match
Contamination	<4%

## D. Genome Sequencing

This section covers sample intake, library preparation, and whole genome sequencing at Baylor Genetics. This section describes the sample flow from DNA sample receipt to production of WGS data, including appropriate quality control (QC) and quality assurance procedures. The

protocol described above for accessioning, DNA extraction, and DNA quality QC are to be followed for genome sequencing.

### **Sample intake**

A visual inspection of the sample tubes is conducted when received by the SC. Sample tube label is compared with information entered in the Gateway to ensure consistency and completeness of the Gateway data, consent forms, proper sample labeling, and sample tube integrity. Samples are accepted if no discrepancies are found, sample labels match, and no tube damage is observed. If any of the above criteria is not met, Baylor notifies the referring CS.

Once accepted, samples are accessioned into the Laboratory Information Management System (LIMS) system. Sample information in the Gateway is entered into the WGL LIMS database. Each sample is assigned an internal lab number, as well as a family number in LIMS. 1D bar code labels with participant specific information (unique identifier) including participant name, date of birth (DOB), lab number and family number are attached to the stock DNA tube. Subsequently samples are aliquoted from the stock tubes into 2D barcode tubes. The samples in 2D bar code tubes are processed for exome sequencing. Before sample transfer, the record for a sample is first opened in LIMS, then the 1D barcode label on the stock tube of the sample is scanned and the LIMS automatically verifies if the sample ID in LIMS matches that on the label. Then, the 2D bar code on the aliquot tube is also scanned to link the two bar codes in LIMS before sample transferring occurs. These steps are to ensure the chain of custody remains intact during sample transfer.

### **DNA sample QC**

DNA samples are screened to quantify DNA as well as determine DNA quality. To determine DNA concentration and purity, the samples are evaluated using the Lunatic Spectrophotometer. Passing criteria include:

1. Sample A260/A280 ratio between 1.8-2.0 and A260/A230 greater than 1.5.
2. Sample concentration greater than 30ng/ul and contains more than 1ug of DNA

If a sample does not meet the criteria above, the CS is notified.

### **Library preparation**

The purpose of this SOP is to provide instructions on how to prepare a PCR free sample for whole genome sequencing by using Illumina DNA PCR-Free prep. Tagmentation kit (Illumina, 20041795).

DNA samples are normalized to 20ng/ul in a final volume of 25ul with RSB from the kit. All gDNA samples will be fragmented into  $475 \pm 75$  bp through Tagmentation reaction. This longer insert size improves overall library performance and allows the longer sequencing read lengths on the Illumina NovaSeq 6000 or X plus platform (150bp) to be efficiently used without producing a significant number of over-lapping reads. Library QC is performed with single strand Qubit (use 450bp as average library size and 660 g/mol as the DNA mass), expected final library yield above 3nM in 20ul..

After the library preparation is complete, samples are stored in low-bond tubes and kept at -20°C before sequencing.

### Sequencing on NovaSeq X Plus

Kit: NovaSeq X Plus Series

Kit	Source	Catalog #
NovaSeq X Series 10B Reagent Kit (300Cycle)	Illumina	20085594 (300 cycles)
NovaSeq X Series 25B Reagent Kit (300Cycle)	Illumina	20104706 (300 cycles)
NovaSeq X Series 1.5B Reagent Kit (300Cycle)	Illumina	20104705 (300 cycles)

Equipment	Reference Source	Catalog/Model #
Novaseq X Plus	Illumina	X Plus
Pipet-Lite LTS pipette 100-1000 µl	Rainin	L-1000
Pipet-Lite LTS pipette 20-200 µl	Rainin	L-200
Pipet-Lite LTS pipette 2-20 µl	Rainin/Intergra	L-20
Minifuge	Any	N/A
Vortexer	Any	N/A
BioShaker XP Mixer	BullDog Bio Inc.	1808-0505

Once a library passes QC, the production sequencing on the NovaSeq X Plus is performed. PhiX must be loaded with the libraries to monitor the sequencing metrics. Record the quality metrics of the run at any time after the 26<sup>th</sup> cycle.

The metrics are:

Quality Metric	10B (2x150)
Cluster Passing Filter	60%
Estimated Yield	3TB
Q30 Score	Above 85%

Clustering and sequencing are performed as per standard Illumina protocols for NovaSeq sequencing. UDN samples will be sequenced on NovaSeq X plus10B flow cells which will yield a minimum of 40X coverage for each sample.

## E. Analysis

### Overview

The steps in the analysis of WES or WGS data can generally be divided into 4 phases: primary analysis, secondary analysis, tertiary analysis, and interpretation. Secondary analysis can be further subdivided into read mapping and variant calling phases. Best-of-breed standards in

analysis of WES or WGS sequence data are followed (as defined in this manual and agreed to across the SC, CSs, and the CC).

Annual review of current methodologies is performed by the SC with an aim of identifying and potentially incorporating advances of note in analytical approaches supporting interpretation of sequencing data. Any alterations that are considered for inclusion are shared with the Sequencing Working Group. Those prioritized are implemented leading to revision of the analysis steps outlined in this section. Each CS may conduct analyses on UDN cases sequence data as they see fit, but the SC undertakes primary, secondary, and tertiary analysis of the sequencing data. The purpose of this is to provide consistency in the format and quality of the data provided and to create maximal utility for the widest range of consumers of these data. For example, this method ensures that sites without existing clinically certified variant annotation and prioritization pipelines have access to richly annotated data.

### **Sequencing files in the Gateway**

Output files from various stages in the analysis process derived from various applications are uploaded to the Gateway by the SC. Files may also be stored locally according to clinical data retention policies in place at the SC. As currently defined these are:

1. Standard compliant format FASTQ files
2. Standard compliant BAM files
3. Standard compliant VCF files
4. Annotated variant files (tab delimited text file format)
5. An interpreted clinical report is also be provided by the SC.
  - a. The format of this report follows the existing industry standards for clinical sequencing reports. Clinical reports include the following report sections:
    - i. Lab contact information and general test information
    - ii. Participant name and date of birth
    - iii. Indication for testing
    - iv. Primary findings (pathogenic, likely pathogenic, and variants of unknown clinical significance) in tabular format
    - v. Medically actionable findings in tabular format
    - vi. Interpretation of findings – textual discussion of the relevance of the findings given the clinical presentation of the proband
    - vii. Specific recommendations
    - viii. A description of the methods used
    - ix. Limitations for both the sequencing technology and analytical processes
    - x. References
6. Each family member receives a report containing one of three possible results:
  - a. Medically actionable finding identified
  - b. No medically actionable finding identified
  - c. Family member opted out of receiving medically actionable findings

### **Analysis, variant calling and annotation**

For analysis, we will utilize the Dragen system (<http://www.edicogenome.com/dragen/dragen-genome-pipeline/>) to generate BAM and VCF analyses. Dragen processors enable NGS data analysis to be performed at extreme speed and with very high accuracy. Dragen uses field-programmable gate arrays (FPGA) - a reconfigurable processor to provide hardware-accelerated implementations of genome analysis algorithms. The main applications of the

Dragen FPGA are Map/Align, sorting, duplicate marking haplotype variant calling and compression. Map/Align acceleration is about 2 orders of magnitude faster when compared to modern CPU analysis speed.

The output data from the Illumina HiSeq or NovaSeq are converted from BCL files to FastQ files according to each sample's specific adapter sequence using Illumina's recommended procedure. FastQ data are aligned to the human reference genome using the Edico Dragen BioIT Platform. The output of the alignment is a BAM file; QC metrics of the map-align process are recorded for quality review. QC statistics include coverage for exome genes and genes known to be implicated in human disease, mate-pair alignment information as well as number of total and duplicate reads. Variant calling on the BAM file is performed using the Edico Dragen haplotype-based variant calling system and the output is a VCF file. For indels and SNV, the variant calling step generates a "raw" VCF file containing a list of detected variants, which are then annotated using a locally installed annotation system. Edico Dragen output also includes gVCF genotype calls on a dense collection of approximately 500,000 exonic SNPs derived from the 1000 Genomes Project, and these SNPs are used for identity analysis as well as for analysis of copy number variation (CNV) and segments of homozygosity or uniparental inheritance. Our pipeline also includes a joint calling step to enable variant discovery leveraging data from multiple samples, allowing us to detect variants with greater sensitivity and genotype samples more accurately. BAM files are first generated per sample followed by creation of gVCF file. The last step, and the only step that involves all samples, is to process all previously generated gVCFs together, creating the joint calling of variants. All information is integrated to a single VCF that contains all family variants information.

### **Primary analysis – demultiplexing**

Primary analysis (demultiplexing) is performed on the HiSeq or NovaSeq instrument workstation according to Illumina guidelines. The Dragen BCL conversion pipeline enables fast BCL to FASTQ conversion of NGS data including both HiSeq and NovaSeq chemistries. FASTQ files are written out to directory structure that identifies the run, flow cell, and each filename will contain the lab number of the sample.

### **Secondary analysis – read mapping**

SC performs clinically appropriate validation of all datasets and algorithms/software applications in use within its clinically validated analysis pipelines. Significant pipeline component changes undergo re-validation at the discretion of the SC.

Secondary analysis step 1 (mapping and realignment; FastQs serve as input to the Dragen software which performs the mapping/alignment, calculates coverage, performs variant calling, and joint analysis.

Data analysis and interpretation are performed by the Baylor Genetics analytics pipeline. Data are aligned to the human reference genome build GRCh38 using the Illumina Dragen BioIT Platform. Variant calling is performed using the Illumina Dragen haplotype-based variant calling system and Illumina Dragen genome wide depth based CNV (Copy Number Variants) caller with custom modifications from Baylor Genetics. Short tandem repeat calling is performed using the Illumina Expansion Hunter tool.

The automated genetic interpretation platform from Emedgene is implemented for variant annotation and interpretation with proprietary algorithms and open-source data sets such as gnomAD, EVS (Exome Variant Server), ClinVar, and professional resources such as HGMD (Human Gene Mutation Database) Pro. The variants were interpreted according to ACMG (American College of Medical Genetics and Genomics) guidelines (reference 1) and patient phenotypes.

Variants are confirmed by Sanger sequencing, specific PCR, Multiplex Ligation-dependent Probe Amplification (MLPA), or another methodology, if sequencing or copy number variants detected by NGS do not meet internal quality standards or are within highly homologous regions.

### Tertiary analysis – variant annotation

The variant annotation is performed by a locally installed program that leverages the GenomOncology Knowledge Management System API for phase and provides annotations using open source data sets such as gnomAD, ExAC, EVS, and ClinVar and professional resources such as HGMD Pro, and later replaced by EMG/Illumina platform. The GenomOncology Knowledge Management System also provides HGVS nomenclature built using the Biocommons open-source suite of tools: HGVS python library, the UTA transcript repository, and SeqRepo sequence database. Annotation process reports zygosity as well as inference of mutation types including nonsense, missense, synonymous, splicing and frameshift, and others. Annotation is highly parallelized and operates on several thousand lines per second.

Variant types being analyzed include sequencing variants and CNV in nuclear genes and mitochondrial DNA (mtDNA), and short tandem repeats (STR). Uniparental disomy (UPD) analysis is done only in a trio setting. Regions of homozygosity (ROH) are reported when greater than 5 Mb. Sensitivity of detection is decreased for mtDNA sequencing variants and CNVs at heteroplasmy levels lower than 5% and 10%, respectively. Pathogenic events of short tandem repeats can be detected within the genomic regions of following genes: *AFF2*, *AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *ATXN10*, *ATNX8OS*, *C9orf72*, *CACNA1A*, *CNBP*, *CSTB*, *DIP2B*, *DMPK*, *FMR1*, *FXN*, *GLS*, *HTT*, *JPH3*, *NOP56*, *NOTCH2NLC*, *PABPN1*, *PHOX2B*, *PPP2R2B*, *RFC1*, *TBP*, and *TCF4*. Sensitivity of detection is reduced for borderline or incomplete penetrance alleles.

The specific variant annotations and the tools used to produce them are outlined in Table A.

Table A. Depiction of the set of variant annotations and tools used by SC.

<b>Annotation</b>	<b>BCM</b>
Alamut	yes
splice sites	yes
near splice site	yes
protein coding flag	yes
syn change flag	yes
non-syn change flag	yes
AA change	yes
Sift prediction	yes
Polyphen2 HVAR prediction	yes
Polyphen2 HDIV prediction	yes

Mutation Taster prediction	yes
MutationAssessor prediction	yes
stop gain flag	yes
stop loss flag	yes
frameshift flag insertion/deletion/indel	yes
non-frameshift flag insertion/deletion/indel	yes
location: intron/exon	yes
location: 5'UTR/3'UTR/Intergenic/Promotor	yes
HGNC appropriate Gene Symbol	yes
Transcript ID	yes
COSMIC	yes
HGMD ID	yes
HGMD variant level association	yes
HGMD gene level association	yes
OMIM ID	yes
OMIM variant level association	yes
OMIM gene level association	yes
ClinVar ID	yes
ClinVar metadata (various; to be clarified)	yes
dbSNP ID	yes
dbSNP AF	yes
1000 Genomes AF	yes
ESP EVS AF	yes
Mappability score	yes

### Coverage analysis

Depth of coverage is calculated for the WES regions of interest by Dragen. For WES, the regions used in the coverage calculation are defined by the unpadded target file. The direct input for the coverage analysis is the BAM file produced in the preceding alignment step. The output is a genome coverage bed file, which is a tabular histogram of the coverage depths of the target regions. The resulting coverage bed is used as an input to the DART QC Scoring step for calculating various coverage statistics for each individual sample to be used during QC review.

### Interpretation

The SC also provides an interpreted clinical report. A systematic process is followed in accordance with published American College of Medical Genetics and Genomics (ACMG) guidelines to determine the clinical significance of each variant considered for reporting. WGS clinical report consists of reporting of the digital exome using the same exome BED file as the Baylor clinical exome. In addition, 5' and 3' UTR will be analyzed as well as deep intronic changes that were previously described. . The whole genome clinical report consists of reporting of the above-described digital exome regions and elements, as well as deep intronic small variants that are predicted by computational algorithms including spliceAI to significantly affect splicing. The whole genome clinical report also consists copy number variant analysis of genes and regions known to be causative to Mendelian diseases. Extended regions of absence of heterozygosity and uniparental disomies will also be reported. Further development of a

clinical genome report including balanced structural variants is underway. The MOO will be updated when these additional elements are launched clinically.

### **Analysis output delivery turnaround times**

Initial analysis (to end of tertiary analysis phase) is completed within a 2-week turnaround time (TAT). Preliminary clinical reports are typically available 4-6 weeks after raw data is uploaded to the Gateway. All variants included in clinical reports are confirmed by Sanger sequencing. CSs may request additional Sanger sequencing of variants identified during their analysis. Sanger sequencing is performed by the SC. The SC can Sanger confirm up to 8 variants per case. The turnaround time for final clinical reports depends on the timing of CS data analysis but will be 4-6 weeks from receipt of the CS's additional variants. Please note that TAT may sometimes be delayed depending on the complexity of Sanger primer design for certain genes, particularly those not encountered before in the sequencing core.

### **Requesting release of sequencing data**

The SC follows existing institutional policies for fulfilling raw data requests. Raw data is only released after a final report has been returned to the CS and test results have been communicated to the participant. Raw data release forms are available on dropbox.com.

Variants that will appear on the clinical report include:

1. Pathogenic or likely pathogenic variants (SNVs, CNVs or SVs) related to clinical phenotype
2. Variants of unknown clinical significance (VUS) related to clinical phenotype
3. Medically actionable pathogenic variants unrelated to clinical phenotype, if opted for reporting
4. Pathogenic variant indicating carrier status for recessive Mendelian disorders, if opted for reporting

The final step in the interpretation process is a weighing of the evidence regarding each candidate variant. The clinical interpretation of variants and the preparation of the clinical report is conducted in three independent reviews, the first by a PhD-level scientist, the second by an ABMGG-certified clinical molecular geneticist, and the third by a clinical geneticist or genetic counselor. These three reviews are correlated to build a final report, with differing interpretations vetted by an advisory group of laboratory diagnosticians and clinicians who may have expertise in the gene of interest. The final weighing of evidence for pathogenicity of each candidate variant is a manual process that requires critical review of the primary literature, information from databases and in silico prediction programs, and correlation with the patient's phenotype and familial inheritance patterns. In addition to the categories of results listed above, during UDN Phase I we developed a collaborative interpretation process between the sequencing core lab and the CSs to confer on individual cases as needed. An extra table known as the "Research Variants" table is communicated to the CSs which includes variants of interest that do not meet criteria for clinical reporting, such as novel genes. After the CS performs their independent analysis of the sequence data, they add additional variants of interest and indicate which of these they would like to have Sanger validation. Sanger is performed and a final report issued.

The clinical WES/WGS report will be issued for the proband to include variants belonging to sections 1-4 above and the allele transmissions of those variants in the parents. The clinical WES/WGS report will be issued for the unaffected parents for incidental findings only according

to current UDN protocol. The clinical WES/WGS report includes patient phenotype, information at variant level, gene level, disease level, as well as result interpretation and recommendations for the patient and family. . The timeline for providing a preliminary report is 8 weeks. A final report which contains additional Sanger sequencing requested by the CS after their analysis will be issued 4-6 weeks after receiving the request.

Research analysis is also performed at UDN CSs. Computational workflows are documented in the following manuscript: Kobren, S.N., Baldrige, D., Velinder, M. et al. Commonalities across computational workflows for uncovering explanatory variants in undiagnosed cases. *Genet Med* (2021). <https://doi.org/10.1038/s41436-020-01084-8>.

## F. Transcriptome Sequencing

This section covers sample intake, library preparation, and whole transcriptome sequencing at Baylor College of Medicine. This section describes the sample flow from RNA sample receipt to production of transcriptome data, including appropriate quality control (QC) and quality assurance procedures. The protocol described above for accessioning, DNA extraction, and DNA quality QC are to be followed for exome and genome sequencing.

### Intended use

RNA-Seq utilizes next generation sequencing to analyze the transcriptome. It has the capability to quantify gene expression and can also facilitate the discovery of novel transcripts, identification of alternatively spliced genes, and detection of allele-specific expression. In addition to poly A RNA(mRNA transcripts), other populations of RNA such as noncoding RNA and microRNA can be investigated. Following genomic sequencing, RNA-Seq can be used to provide evidence to further classify variants as disease causing or provide evidence for causality of new disease genes. This is primarily the role of RNA-Seq in the UDN. While WES or WGS analysis can detect SNVs, CNVs and SVs that are disease-causing, RNA-Seq may point to additional disease etiologies by identifying alternative splicing, differential expression of genes or alleles and gene fusions that cannot be readily seen at the DNA level. RNA-Seq provides a dynamic, in-depth view of the transcriptome by detecting and quantifying allele-specific expression arising from coding or non-coding variants and novel alternatively spliced RNA transcripts that may be the result of SNVs, CNVs and SVs as well as revealing the perturbative impact of pathogenic variation on gene expression networks and the larger transcriptome. Importantly a recent study describes a 35% increase in diagnoses using RNA-Seq in patients with a non-diagnostic WES or WGS. The purpose of this protocol is to generate RNA-Seq data for integration of functional omics data for variant prioritization and interpretation, which will be incorporated into the UDN workflow for selected cases. In addition, the analysis may help evaluate the utility of multi-omics data in diagnosing genetic disorders as well as establish procedures for integrating multi-omics into clinical practice.

In Phase 1, an RNA-Seq pilot study was undertaken with three labs providing the technical component. In addition, the groups agreed upon a harmonized analytical pipeline. Insofar as possible within the constraints of our laboratory, we will replicate these procedures that were developed during the pilot including sample types, type of RNA-Seq, cycles, number of reads per sample and analytics.

## Test order

We will follow the workflow for RNA-Seq requests that was established during the UDN RNA-Seq pilot study:

1. Clinical Site (CS) completes a Transcriptome sequencing request in the Gateway for each sample being sent.
2. CS enters and releases updated phenotype information in the Gateway for use by the MGML in their analyses.
3. CS ensures that DNA sequencing data is available for MGML to facilitate interpretation of the RNA-seq data. Proband DNA sequencing data must either already be stored in the Gateway, or uploaded to the gateway if it originates from prior exome or genome sequencing performed at a laboratory other than Baylor. The uploaded data files should pertain to the proband only (family member data is not required). FASTQ files are preferred, but BAM and VCF files in hg38 are acceptable alternatives if FASTQ files are not available.
4. CS enters shipping information (date RNA sent and tracking number) in the Gateway and ships samples. Please note that Gateway provides alerts for shipping of UDN samples, but shipment tracking needs to occur at the CS/MGML level.
5. Gateway sends an automated email to alert the MGML of sample shipment.

## Sample intake

A visual inspection of the sample tubes is conducted when received by the SC. Sample tube label is compared with information entered in the Gateway to ensure consistency and completeness of the Gateway data, consent forms, proper sample labeling, and sample tube integrity. Samples are accepted if no discrepancies are found, sample labels match, and no tube damage is observed. If any of the above criteria is not met, Baylor notifies the referring CS.

Once accepted, samples are accessioned into the Laboratory Information Management System (LIMS) system. Sample information in the Gateway is entered into the LIMS database. Each sample is assigned an internal lab number, as well as a family number in LIMS. 1D bar code labels with participant specific information (unique identifier) including participant name, date of birth (DOB), lab number and family number are attached to the stock DNA tube. Subsequently samples are aliquoted from the stock tubes into 2D barcode tubes. The samples in 2D bar code tubes are processed for exome sequencing. Before sample transfer, the record for a sample is first opened in LIMS, then the 1D barcode label on the stock tube of the sample is scanned and the LIMS automatically verifies if the sample ID in LIMS matches that on the label. Then, the 2D bar code on the aliquot tube is also scanned to link the two bar codes in LIMS before sample transferring occurs. These steps are to ensure the chain of custody remains intact during sample transfer.

## Assay description

RNA-Seq consists of isolating RNA, converting it to complementary DNA (cDNA), enriching for polyadenylated transcripts or ribo-depletion to remove ribosomal RNAs, preparing the sequencing library and sequencing on an NGS platform. Additional considerations include the use of biological and technical replicates, depth of sequencing, and desired coverage across the transcriptome.

## RNA-Seq library prep and sequencing

The transcriptome data will be generated by RNA-Seq from RNA isolated from peripheral whole-blood samples or cultured fibroblasts available from the UDN CSs or the Biorepository. Extracted RNA shipped in dry ice will also be accepted. Sample collection and RNA purification will be performed according to PaxGene protocols. All RNA samples will be kept at -80C in 2-3 aliquots. Library construction will be carried out by using the Illumina Stranded Total RNA Prep with Ribo-Zero Plus kit for the blood samples. RNA from fibroblast and other sample types will be prepared using the Illumina Stranded mRNA Prep kit. This kit works with high-quality RNA with polyA tails. The quantity, library concentration and size distribution of the final library is assessed with the bioanalyzer. The expected size distribution has an average size of 400bp with a range between 200-800bp. Sequencing is performed on the NovaSeq X with 150 cycles of paired-end reads. Only the Q<sub>30</sub> score >85 data will be used. We aim at 100M reads per sample.

## RNA-Seq analysis

The SC have been uploading output files generated from the following pipeline:

[https://github.com/furbelows/udn\\_rnaseq/blob/master/UCLA\\_UDN\\_DNAnexus.md](https://github.com/furbelows/udn_rnaseq/blob/master/UCLA_UDN_DNAnexus.md)

These uploaded files consist of:

FASTQs

Alignment files in BAM

BAM index

After clinical validation, the pipeline will be updated and the RNA-Seq data will be analyzed following the V10 version of the GTEx consortium as outlined in the github link below.

<https://github.com/broadinstitute/gtex-pipeline/blob/master/rnaseq/README.md>

Output files generated from this updated clinical pipeline will consists of:

FASTQs

CRAM sequences aligned to the genome

Gene-Level Quantification Files (RPKM, TPM)

Transcript-Level Quantification Files

Junction Count Files

Chimeric junction

Exon Quantification Files

QC matrix and reports

The turnaround time for RNA-Seq datafile deposition after receiving blood or fibroblasts is approximately 4 weeks.

## Sample requirements

Below are requirements for submitting samples to the Medical Genetics Multi-omic Laboratory (MGML) at Baylor College of Medicine. Please follow them accordingly to prevent delays and rejection of the submitted samples. They are only able to accept blood or cultured fibroblast samples or total RNA extracted from these two sample types. Please contact before sending other sample types.

The quality of the data is directly related to the quality of the submitted RNA. It is imperative that the quality and quantity of the submitted RNA is within the sample requirements discussed below. Upon sample arrival, sample quantity and quality check will be performed using Qubit to measure the concentration and Tapestation for measuring RNA quality. A **RIN (RNA Integrity Number) score of 7** is required to proceed with RNA-seq library construction.

### Total RNA sample requirements

1. For extracted total RNA, please submit two (2) aliquots of each sample.
  - A 5  $\mu$ L aliquot for quality control, and
  - A separate aliquot for processing
2. The recommended minimum concentration of RNA sample is 50 ng/ $\mu$ L. The recommended RNA amount is 1.5  $\mu$ g (not including the 5  $\mu$ L aliquot submitted for QC) and minimum volume  $\geq$  20  $\mu$ L.
3. Do not dilute the sample if concentration is more than 50 ng/ $\mu$ L.
4. The A260/280 ratio (Nanodrop) should be 1.7 - 2.2 and A260/230 should be 1.5 – 2.2.
5. RNA must be treated with DNase.

### In summary

<b>Application</b>	Poly A RNA-seq or Total RNA-seq
<b>Sample Type</b>	Total RNA, DNase Treated
<b>Recommended Amount</b>	$\geq$ 1.5 $\mu$ g, Volume: $\geq$ 20 $\mu$ L
<b>Minimum Amount</b>	100 ng, Volume: $\geq$ 20 $\mu$ L
<b>Recommended Buffer</b>	Nuclease free H <sub>2</sub> O, Low TE (<0.1 mM EDTA)
<b>Purity (Nanodrop)</b>	A260/280: 1.7 – 2.2, A260/230: 1.4 – 2.2
<b>Quality or RIN Number</b>	$\geq$ 7.0

For samples that do not meet the criteria above and/or with poor QC, the principal investigator will be consulted beforehand to discuss the possibility of either replacing them or processing with their consent.

### Sample labeling

1. Submit samples **only** in 1.5 mL DNase and RNase free micro centrifuge tubes. PCR tubes or another format will not be accepted.
2. Each tube must be clearly labelled. Use a label printer if possible (freeze proof labels) or use a waterproof marker.
3. Samples submitted for transcriptome sequencing must be labeled with participant first and last name, date of birth, and the UDN number.

### Sample labeling for PAXgene blood RNA tubes

1. Each PAXgene Blood RNA Tube must be clearly labeled using freeze proof labels (recommended) or waterproof markers (if sending in for RNA extraction)
2. Samples submitted for transcriptome sequencing must be labeled with participant first and last name, date of birth, and the UDN number.

### Sample ordering guidelines

1. CS completes a Transcriptome sequencing request form in the Gateway for each RNA sample being sent.
2. CS enters and releases updated phenotype information in the Gateway for use by the MGML in their analyses.
3. CS enters shipping information (date RNA sent and tracking number) in the Gateway and ships samples.
  - a. Please note that Gateway provides alerts for shipping of UDN samples, but shipment tracking needs to occur at the CS/MGML level.
4. Gateway sends an automated email to alert the MGML of sample shipment and available phenotypic data.

### Sample packing and shipping

1. Submit RNA samples only in 1.5 mL DNase and RNase free micro centrifuge tubes. Please use parafilm to seal each tube before packaging.
2. Total RNA samples should be shipped in dry ice only.
3. The PAXgene Blood RNA Tubes should be frozen at  $-20^{\circ}\text{C}$  and  $-70^{\circ}\text{C}$  in a wire rack, not in an expanded polystyrene foam tray. Please follow FEDEX or UPS guidelines for shipping biohazard specimens. They should be shipped in a Styrofoam container covered by an outer cardboard box along with absorbent materials.
4. Select a reliable courier and choose priority option (overnight preferred) for shipment. We prefer FedEx or UPS. Samples should be shipped Monday-Wednesday to avoid delivery during the weekend.
5. Please ship samples for RNA-Seq to:

Medical Genetics Multi-omics Laboratory  
Baylor College of Medicine  
One Baylor Plaza  
Room 223E  
Houston, TX 77030  
(713) 798-7725

#### **Skin biopsy** samples should be sent to:

Baylor Genetics  
2450 Holcombe Blvd.  
Ste. 2210  
Houston, TX 77021-2024  
(800) 411-4363

6. Contact Information:  
[MGML@bcm.edu](mailto:MGML@bcm.edu)  
713-798-7725
7. After arriving at Baylor site, samples will be stored in  $-80^{\circ}\text{C}$  freezer. The MGML lab will provide timely feedback on the progress of your project.

## Blood collection using PAXgene blood RNA tube

MGML prefers using PAXgene Blood RNA tubes for blood collection, but RNA extracted with any method is acceptable if it passes QC.

## Collection of whole blood using PAXgene blood RNA tube

Required Items:

- |                           |                  |               |
|---------------------------|------------------|---------------|
| • PAXgene Blood RNA Tubes | Catalog # 762165 | BD Vacutainer |
| • Blood Collection Set    | Catalog # 367281 | BD Vacutainer |
| • Discard Tube*           | catalog # 368975 | BD Vacutainer |

Instructions on Blood Collection:

1. PAXgene Blood RNA Tube must be at room temperature prior to use and properly labeled with first and last name, date of birth and the UDN number.
2. If PAXgene Blood RNA is the only tube to be drawn, a small amount of blood should be into a discard tube (\* any serum collection vacutainer tube can be used). Otherwise PAXgene Blood RNA Tube should be the last tube drawn during phlebotomy.
3. Using your institutions recommended standard procedure for venipuncture, collect blood in to PAXgene Blood RNA Tube
  - A blood collection set is required to be used with the PAXgene Blood RNA Tube to eliminate the possibility of patient contact with the reagent in the tube, due to backflow of blood, when used in accordance with the instructions for use.
  - DO NOT PIPETTE blood from an EDTA tube into the PAXgene Blood RNA Tube
  - DO NOT INJECT blood from a syringe into the PAXgene Blood RNA Tube
4. Allow at least 10 seconds for a complete blood draw to take place. Make sure that blood has stopped flowing into the tube before removing the tube from the holder.
5. Gently invert the PAXgene Blood RNA Tube 8 to 10 times
6. Store the PAXgene Blood RNA Tube at room temperature for at least 2 hours after collection to ensure complete lysis of blood cells.
7. Store PAXgene Blood Tubes at -20°C for 24 hours and then transfer them to -70°C before shipping. It is very important that PAXgene Blood RNA Tubes should be stored in **wire rack** and **not** in an expanded polystyrene foam tray when in freezer. Freezing the PAXgene tubes in a foam container may result in cracked tubes because the foam will prevent the tube from expanding during freezing.
8. Keep PAXgene RNA Blood Tubes upright all the time.

## Cell culture instructions for RNA-Seq

Method to Harvest Cultured Cells:

1. Cells growing in suspension
  - Do not use more than  $1 \times 10^7$  cells.
  - Pellet the appropriate number of cells by centrifuging for 5 minutes at 300 x g in a micro centrifuge tube. Carefully remove **all** supernatant by aspiration.
  - Incomplete removal will inhibit lysis and dilute lysate, affecting the conditions for DNA removal and RNA purification, both effects may reduce RNA quality and yield.
  - Loosen the cell pellets by flicking the tube. Add appropriate volume of lysis buffer.
  - Vortex or pipet to mix until the cell pellet is completely dispersed, and the cells appear lysed.

- If needed (cell lysate is viscous or processing  $\geq 1 \times 10^7$  cells) homogenize by transferring the lysate to homogenizer (Cat # 12183-026, Thermofisher Scientific) inserted in a collection tube and centrifuge at 12,000 X g for 2 minutes. Remove the homogenizer when done, proceed to protocol for binding, washing and elution of RNA from cell culture.
2. Cells growing in a monolayer
- Harvest a maximum of  $1 \times 10^7$  cells as a cell pellet.
  - Determine the number of cells and aspirate the medium and wash the cells with PBS.
  - Aspirate the PBS and add appropriate amount of 0.05% Trypsin-EDTA (1X) (see table below)

Flask Size	Trypsin/EDTA Solution (mL)	Stop Medium (mL)
T25	1.5 mL	3.5 mL
T75	2.5 mL	5.5 mL

**Stop media is the same as growth medium.**

- After the cells detached from the flask, add appropriate volume of stop media, transfer the cells into a 1.5- or 2.0-mL RNase free micro centrifuge tube and centrifuge at 300 x g for 5 minutes.
- Carefully remove **all** supernatant by aspiration. Incomplete removal will inhibit lysis and dilute lysate, affecting the conditions for DNA removal and RNA purification, both effects may reduce RNA quality and yield.
- Loosen the cell pellets by flicking the tube. Add appropriate volume of lysis buffer.
- Vortex or pipet to mix until the cell pellet is completely dispersed, and the cells appear lysed.
- If needed (cell lysate is viscous or processing  $\geq 1 \times 10^7$  cells) homogenize by transferring the lysate to homogenizer (Cat # 12183-026, Thermofisher Scientific) inserted in a collection tube and centrifuge at 12,000 X g for 2 minutes. Remove the homogenizer when done, proceed to protocol for binding, washing and elution of RNA from cell culture.
- Cell Pellets can be stored at  $-80^{\circ}\text{C}$  for later use. Determine the number of cells before freezing. Frozen pellets should be thawed slightly so that they can be dislodged by flicking the tube. Homogenized cell lysates can also be stored at  $-80^{\circ}\text{C}$  for several months and can be thawed at  $37^{\circ}\text{C}$  in a water bath until completely thawed and salts are dissolved. Avoid prolonged incubation, which may compromise RNA integrity.

## G. Reanalysis of Exome and Genome Data

Reanalysis of previously sequenced exomes and genomes is a standard approach for clinical sequencing laboratories. The power from reanalysis comes from multiple activities within the field, changes in analytical pipelines and changes in clinical status of the patients/participants within the CSs. Reanalysis may also be helpful for those cases in which one diagnosis may not completely explain the observed phenotype (e.g., J. Posey, Resolution of disease phenotypes resulting from multilocus genomic variation, *New Engl J Med* 376:21-31, 2017).

### Approach

There are two different analytic approaches proposed: reanalysis and reinterpretation, as described below. Reanalysis and reinterpretation of individual cases is not currently performed on a UDN-wide basis but rather is performed by the clinical sites, using varying protocols.

### **Reanalysis**

Definition: Start with raw data and rerun the entire analysis and interpretation. Reanalysis was available through Baylor for a subset of Phase 1 cases at the request of the CSs. Upon reanalysis, an amended report was issued which communicated results, new findings, or no changes to original interpretation.

CSs may request reanalysis for exome and genome cases on a fee-for-service basis through Baylor Genetics. Additionally, supplemental funds are available for a limited number of reanalyses to be completed by March 2025. The test requisitions to order reanalysis can be obtained at 1) EXOME: and 2) GENOME: <https://www.baylorgenetics.com/wp-content/uploads/2022/12/baylor-genetics-genome-reanalysis-req.pdf>. Upon reanalysis, an amended report will be issued which communicates results, new findings, or no changes to original interpretation.

## **XII. Biospecimens**

### **A. Recommendations for Research Specimen Collection and Processing**

#### **Overview**

The following specimens may be collected on all probands and affected family members at the discretion of the CS unless they are refused by the participant or their collection would compromise participant safety. Clinical sample collection is understood to be the priority and blood draw volume limits are guided by institutional policies. Blood volume limits in pediatric probands are addressed below. Specimens may also be collected from unaffected family members at the discretion of the CS. When blood is drawn for research purposes, DNA collection and plasma for metabolomics should be given the highest priority. It is then up to the individual CS to prioritize the remaining available blood for serum, plasma, and peripheral blood mononuclear cells (PBMCs) based on planned studies.

*Serum:* 3 ml serum in 0.5 ml aliquots

*Plasma:* 3 ml plasma in 6 x 0.5 ml aliquots (for general use) and 3 x 0.1 ml (for untargeted metabolomics analyses)

*DNA:* at least 20 micrograms (with goal of 50 micrograms) at a target concentration of 100-200 ng/ul (with a minimum of 50 ng/ul)

*PBMCs:* stored in 4 aliquots of  $2 \times 10^6$  cells followed by multiple aliquots of  $5 \times 10^6$  cells each

*Urine:* 5 ml urine in 5 x 1.0 ml aliquots (for general use) and 3 x 0.2 ml (for untargeted metabolomics analyses)

In the instructions below, aliquots for the Metabolomics Consultation Group must be placed in 0.5ml Sarstedt Biosphere® SC Micro Tubes (Sarstedt Catalog # 72.730.217), snap frozen, and shipped to the lab within 3-6 months for analysis. Collection of samples for metabolomics analysis is at the discretion of the CS.

#### **General sample collection guidelines**

Blood samples for serum and plasma should be obtained in the fasting state, defined as an overnight fast for adults and at least 3 hours of fasting for children. If a subject is unable to fast, samples should still be obtained. The CS should record whether blood samples were collected as fasting (including duration) or non-fasting.

Blood for PBMCs should be collected in CPT Vacutainer® tubes with citrate (one 8 ml tube for pediatric subjects and as total blood draw volume allows, and two 8 ml tubes for adult subjects; if sample volumes are limited by proband body weight use the smaller 4 ml CPT tube). If the CS has expired 8 ml CPT tubes or wants to be cost-efficient and keep only 4 ml CPT tubes in stock, the CS uses as many 4 ml CPT tubes as needed.

Blood for DNA should be collected in one 10 ml purple top EDTA Vacutainer® tube and sent to a local CLIA laboratory for DNA extraction and quantification. If DNA from blood cannot be obtained, an alternative source of DNA such as skin fibroblasts should be considered.

### **Adult probands (and older pediatric probands)**

Older pediatric probands, that allow for larger volume of blood draw, may follow the adult protocol or the adult protocol with slight modifications (ex. Reduced collection volume for PBMCs)

1. 10 ml blood for DNA (EDTA)
2. 10 ml blood for plasma (EDTA)
3. 10 ml blood for serum (SST)
4. 2 x 8 ml blood for PBMCs (Citrate CPT tube)
  - a. 1 x 8 ml blood for PBMCs (Citrate CPT tube) probands with limits on blood vol

### **Pediatric probands (and other participants with limited blood collections)**

The volume of blood drawn should be consistent with the allowable blood collection based on participant body weight. In cases where the blood volume that can be obtained is the limiting factor, samples should be obtained in the following order:

1. 3 ml blood for DNA (EDTA)
2. 3 ml blood for plasma (EDTA)
3. 3 ml blood for serum (SST)
4. 4 ml blood for PBMCs (Citrate CPT tube)

### **Sample tracking and storage**

Please note that the central biorepository is no longer accepting new samples but will continue to maintain previously submitted samples.

Labels of research grade samples (not CLIA) to be stored at the local CS should include non-identifiable information: the UDN ID, sample type, and date of sample collection.

Samples should be stored as follows:

- Serum: -80°C or colder (e.g., liquid nitrogen)
- Plasma: -80°C or colder (e.g., liquid nitrogen)
- Urine: -80°C or colder (e.g., liquid nitrogen); stored at local biorepository only
- DNA: -20°C or colder
- PBMC: cryopreserved in liquid nitrogen

Note that aliquots for metabolomics analysis should not be subjected to freeze thawing.

## Sample collection and processing protocols (I-VII)

### I. Blood sample collection and processing

#### 1. Overview

- a. Processing of blood for plasma and serum, as well as urine, should be performed within two hours of sample collection. PBMCs should be processed within 24 hours of sample collection. All steps should be performed as quickly as possible and on wet ice or otherwise at 4°C, as appropriate, to minimize artifacts in metabolomics data.
- b. The standard large serum, plasma, urine, DNA, and PBMC aliquots should be stored in screw-cap cryovials appropriate for ultra-cold storage (e.g., Nalgene NUNC 1.8 ml Cryovials, Fisher Scientific Catalog #: 12-565-170N).
- c. Aliquots for untargeted metabolomics need to be stored in Sarstedt Biosphere® SC Micro 0.5ml Tubes, Sarstedt Catalog # 72.730.217.

#### 2. Serum sample collection and processing

- a. Blood for serum should be collected in one 10 ml red top Vacutainer® Serum Separator Tubes (SST) with clot activator
- a. After obtaining the SST sample, allow sample to clot 30 minutes in a vertical position
- b. Centrifuge vacutainer tube at 2500 RPM (1-1.3 x 10<sup>3</sup> g) for 10 minutes either at ambient temperature or at 4°C.
- c. Carefully remove rubber stopper from vacutainer tube and remove caps from cryovial tubes
- d. Aliquot 0.5 ml of serum into 6 cryovial storage tubes (3 ml total), and store samples in a -80°C or liquid nitrogen freezer with appropriate labels (UDN ID, sample type, and collection date). Pediatric samples generate fewer tubes.

#### 3. Plasma sample collection and processing

- a. Blood for plasma should be collected in one 10 ml purple top EDTA Vacutainer® tube
- b. Centrifuge vacutainer tube at 2500 RPM (1-1.3 x 10<sup>3</sup> g) for 10 minutes at 4°C
- c. Aliquot 0.5 ml of plasma into 6 cryovial storage tubes (3 ml total), and store samples in a -80°C or liquid nitrogen freezer with appropriate labels (UDN ID, sample type, and collection date). Pediatric samples generate fewer tubes. Note: if metabolomic aliquots are needed (section d below) you may need to preserve fewer aliquots in this step.
- d. Aliquot plasma to storage tubes for untargeted metabolomic studies (Sarstedt Biosphere® SC Micro Tube 0.5ml, Sarstedt Catalog # 72.730.217).
  - i. For untargeted metabolomics analyses: 3 aliquots of plasma (0.1 ml each). Otherwise, see the Metabolomics section for details regarding other analyses.
  - ii. For optional targeted analyses (available at the Metabolomics Consultation Group): 1 aliquot of plasma (1 ml). Please see the Metabolomics section for details regarding other analyses.
- e. Flash freeze all samples in liquid nitrogen or quick freeze in dry ice/ethanol prior to storage in a -80°C freezer, or in liquid nitrogen, with appropriate labels (UDN ID, sample type, and collection date).

#### 4. DNA extraction at CLIA laboratory

- a. Multiple acceptable DNA extraction protocols for the EDTA Vacutainer tube blood samples can be used (Examples of suitable Extraction kits: Qiagen Genra

Puregene Blood Kit Catalog # 158445, or Qiagen DNeasy Blood and Tissue Kit Catalog# 69504). DNA should be stored in TE (Tris-EDTA) buffer).

- b. CLIA DNA stored at the CS designated CLIA laboratory should be labeled with UDN ID &/or patient name, date of birth, sample type, and collection date.
- c. CLIA DNA aliquots sent to the UDN Sequencing Core must be labeled with participant name, date of birth, and "UDN" (do NOT include UDN ID).
- d. DNA quantification should be performed with PicoGreen (not NanoDrop), and DNA concentration should be between 100 to 200 ng/ul (with a minimum of 50 ng/ul).

## II. PBMC isolation and cryopreservation

### 1. PBMC isolation

- a. CPT tubes are the recommended cell separating device (refer to specific CPT tube manufacturer instructions for complete details for steps 1-4 and those below are offered as suggestions). Other cell separating devices may be utilized at individual sites for local biorepository.
- b. Centrifuge blood collected in CPT tubes at room temperature at 1500 to 1800 x g in a swing bucket rotor for 20-30 minutes with no brake following specific instructions from the CPT tube documentation. Visually inspect the CPT gel plug in addition to other guidance in the manufacturer's instructions.
- c. Use an aspirating pipette to remove the PBMC layer located at the gel interface in the CPT tube.
- d. Place the PBMCs in a new 50 ml conical tube.
- e. Wash cells by gently resuspending the cell pellet in 10 ml sterile 4°C or ambient temperature PBS (or other physiologic buffer) followed by centrifugation at 250-400 x g. Repeat once.
- f. Count the PBMCs on a hemocytometer, cellometer, or other standardized cell counting device.
- g. Separate into 4 aliquots of  $2 \times 10^6$  cells followed by multiple aliquots of  $5 \times 10^6$  cells each in separate polypropylene tubes and centrifuge at 250-400 x g to create the final cell pellet. The maximum number of aliquots should be made to allow for as many separate PBMC samples as possible to be saved from any one donor.

### 2. Preparation of PBMCs for storage in a cryorepository

- a. To prevent contamination, all processing shall be completed in a sterile biological safety cabinet by wiping all inside surfaces with 70% alcohol and performing UV light treatment for at least 5 minutes.
- b. Remove as much of wash buffer as possible from each aliquoted PBMC pellet. You may gently flick pellet to loosen prior to next step (resuspension).
- c. Resuspend each washed PBMC pellet containing 2 or 5 million cells in 1 ml of cold (2-8°C) CryoStor CS10 Freeze Media (BioLife Solutions).
- d. Mix cells by gently tapping the tube; do not use a pipette.
- e. Incubate resuspended cells at 2-8°C for 10 minutes.
- f. Pipet gently to minimize shear force and transfer into a labeled cryopreservation vial.
- g. For cryopreservation, transfer vials to a Controlled Rate Freezer to decrease the temperature in a controlled fashion (or if that is unavailable into an isopropanol containing cryopreservation system followed by transfer into a -80°C freezer for a minimum of 12 hours).

- h. Once cryopreservation vials have been appropriately cooled and contents frozen, transfer to designated receptacle within the liquid nitrogen storage unit for long-term storage with appropriate labels (UDN ID, sample type, and collection date).

### III. Urine sample collection and processing

1. Urine samples should be the first morning void urine collected in a polypropylene container. A 24-hour urine sample is not required, but may be elected by the CS.
2. Centrifuge urine at 1000 x g for 5 minutes at 4°C to remove any cells and particulates.
3. Transfer supernatant to storage tubes
  - a. Five or ten 1.0 ml aliquots in 1.8 ml screw-cap cryovials
    1. Please note the Central Biorepository no longer receives aliquots of urine. You can keep all 10 aliquots or only collect 5 to save at your site.
  - b. Three 0.2 ml aliquots in Sarstedt Biosphere® SC Micro 0.5ml Tubes (for untargeted metabolomics analyses). For other metabolomics services, see the Metabolomics section for details.
4. Flash freeze all samples in liquid nitrogen or quick freeze in dry ice/ethanol prior to storage in a -80°C or liquid nitrogen freezer with appropriate labels (UDN ID, sample type, and collection date).

### IV. Sample shipment to the Metabolomics Consultation Group

1. See the Metabolomics section for shipping details to the Metabolomics Consultation Group at Mayo Clinic.

## B. Recommendations for Optional Research Biospecimen Collection and Processing

### I. CSF collection and processing

CSF may be collected from neurological cases for clinical laboratory studies and for research use, including metabolomics, lipidomics, glycomics testing, and inflammome studies. All samples should be collected by lumbar puncture in the L3/L4 or the L4/L5 interspace. If neurotransmitters are to be analyzed, they should be collected first and will preempt the possibility of obtaining an opening pressure. Neurotransmitters are volatile and have a cephalo-caudal concentration.

1. Microtubes
  - a. Microtubes 1-5 (4.5 ml total): for shipment to Medical Neurogenetics (Note: Medical Neurogenetics and Baylor both supply 5 pre-marked CSF microtubes (total volume if filled only to the lines is 4-4.5 ml) - the studies offered by each lab varies though the methods are the same.
  - b. CSF must be collected from the 1<sup>st</sup> drop into the designated tubes in the order indicated in the following table. Do not collect the CSF in one large tube and aliquot into the tube set.
  - c. Fill each tube to the marked line with the following volumes, indicated in the table below. The total CSF volume required is 4-4.5 ml.

Tube Number	Required Volume
1	0.5 ml
2	1.0 ml
3	1.0 ml
4	1.0 ml
5	1.0 ml

- d. If the samples are not blood contaminated, place the tubes on ice (or dry ice if available), and then transfer to a -80°C freezer. If the samples are blood contaminated, then centrifuge immediately (prior to freezing) and transfer the clear CSF to new similarly labeled tubes then freeze and store at -80°C.
  - e. Microtubes are collected in 1 ml aliquots and placed in wet ice until they can be aliquoted: 5 x 1 ml (e.g., for energetics), 3 x 0.2 ml (for metabolomics, glycoproteomics, and lipidomics analyses), and the remainder in a final aliquot (e.g., Sarstedt Polypropylene 2ml tube #72.694.107). The Metabolomics Consultation Group requires that its 3 x 0.2 ml aliquots be in Sarstedt Biosphere® SC Micro Tube 0.5ml, Sarstedt Catalog # 72.730.217.
  - f. Polystyrene tubes 1-3 (0.5 ml each minimum): for measurement of glucose, cell count, protein, sterile fluid culture (Note: a standard lumbar puncture tray contains 4 polystyrene tubes with caps).
  - g. Polystyrene tube 4: for IgG index, Oligoclonal bands, etc. (Note: use 4<sup>th</sup> tube from lumbar puncture tray here).
2. CSF handling
- a. CSF is collected by lumbar puncture, place the CSF on wet ice immediately and transport to the laboratory. If dry ice is available and the CSF is not bloody the microtubules for neurotransmitters can be placed in the dry ice immediately.
  - b. If CSF is bloody, excessive blood may interfere with neurotransmitter and metabolomics testing
  - c. Flash freeze all samples in liquid nitrogen or quick freeze in dry ice/ethanol prior to storage in the -80oC freezer with appropriate labels (UDN ID, sample type, and collection date)
  - d. Samples should not be subjected to freeze-thawing

## II. Optional blood collection and processing

1. General guidelines
  - a. As noted above, blood is collected from all probands for DNA, plasma, serum, and PBMCs. Various other optional samples may be considered as well, including: PaxRNA, buffy coat, platelets, oxylipids, and blood spot cards.
  - b. Blood for RNA is collected in PAXgene blood RNA Vacutainer® tubes (VWR 77776-026).
  - c. Blood for additional PBMC or other buffy coat collection is collected in additional CPT (citrate) tubes or other site-specific collection tubes.
  - d. If platelets are to be collected, draw 31.5 ml blood in 7 tubes of light blue top sodium citrate tubes (BLU).
  - e. If oxylipids are to be analyzed, then blood should be collected in a chilled EDTA tube and placed on ice.

- f. Blood spot cards may also be obtained and stored at room temperature.
2. RNA processing
  - a. Use PAXgene Blood RNA Kit (Qiagen 762164)
  - b. Aliquot in 80ul aliquots and store in 500ul sterile, RNase- DNase-free tubes at -20°C until needed
3. RBC isolation
  - a. After removal of plasma (see procedures above), discard the remaining supernatant above the porous barrier using a plastic Pasteur pipette (wide orifice)
  - b. Using a glass Pasteur pipette (narrow orifice), transfer the erythrocyte (RBC) pellet to a 50 ml conical tube
  - c. Fill 50 ml conical tube to 40 ml with Phosphate Buffered Saline (PBS), pH 7.4 and invert several times to mix
  - d. Centrifuge for 5min at 1811 x g at 4°C
  - e. Remove saline layer and discard
  - f. Repeat wash with PBS pH 7.4 until PBS is clear (minimum of 3 times)
  - g. Aliquot 1 ml of the erythrocyte (RBC) pellet to clean cryovials. Store in -80°C freezer with appropriate labels (UDN ID, sample type, and collection date). *Note: this fraction also contains granulocytes*
4. Platelet Isolation
  - a. Blood has been collected in multiple light blue top sodium citrate tubes
    - a. 40-45 ml blood yields 1-3 x 10<sup>9</sup> platelets
  - b. Spin at 200xg for 20 min (no brake)
  - c. Remove two thirds of the top (platelet rich plasma) layer to a new tube
  - d. Add 1 volume HEP buffer + PGE1 (prostaglandin E1)
  - e. Mix very gently by inverting the tube slowly
  - f. Spin at 100 x g for 15-20 min at room temperature (with no brake applied) to pellet contaminating red and white blood cells
  - g. Transfer the supernatant into new plastic tube using a transfer pipette (wide orifice)
  - h. Pellet platelets by centrifugation at 800 x g for 15-20 min at room temperature (with no brake applied). Discard the supernatant.
  - i. Rinse the platelet pellet two times with platelet wash buffer by gently adding wash buffer and removing it slowly with a pipette. (DO NOT RESUSPEND! to avoid platelet activation)
  - j. Store the dry platelet pellet at -80°C freezer with appropriate labels (UDN ID, sample type, and collection date). *Note: Freezing the pellet disrupts the platelet granules. This pellet is only to be used for determination of glycomics, lipidomics and proteomics that does not include the platelet granules.*
5. Lipidomics sample processing
  - a. Centrifuge chilled blood for 10 min at 1200 x g at 4°C
  - b. Aliquot 1 mL plasma into a single storage vial
  - c. Keep on ice and flash freeze in liquid nitrogen or quick freeze in dry ice/ethanol the samples prior to storage at -80 °C within 4 hours (optimal is to freeze immediately and store at -80°C) and do not subject to freeze thaw cycles. If that is not possible then the time should be adjusted to be the same for all samples. Do not store at -20 °C. Redraw the sample if it is extensively hemolyzed.

### III. Skin biopsy collection and processing

Skin biopsies are used for culturing fibroblasts. These cell lines are then be used for various research purposes as well as glycomics testing. For subjects who are unable to provide PBMCs, skin fibroblasts provide an alternate source of living cells for future research.

1. Reagents

- a. DMEM High Glucose (Invitrogen #11965-118)
- b. Fetal Bovine Serum (FBS), Certified, Heat-Inactivated, US Origin (Invitrogen #10082-147)
- c. 100X Antibiotic-Antimycotic (Invitrogen #15240-062)
- d. 0.25% Trypsin-EDTA (Invitrogen #25200-056)
- e. 1 x PBS pH 7.4 w/o Calcium or Magnesium (Invitrogen #10010-023)
- f. Dimethyl sulfoxide (Sigma #D8418-100ML)

2. Collection protocol

- a. 3-5mm punch full thickness skin biopsy obtained according to standard medical procedure
- b. Place biopsy in sterile tissue culture medium (DMEM, 10% FBS, 1% antimycotic, antibiotic) contained in a 15 ml conical tube
- c. Store and transport the biopsy at ambient temperature
- d. Deliver the biopsy to the laboratory within 24 hours (Up to 96 hours is acceptable if shipped)

3. Initiation of skin fibroblast culture

- a. Spray hood and scalpels with Ethanol and wipe with Kimwipe
- b. Clean the biopsy tube by spraying well with Ethanol before placing in the hood
- c. Label 6-well tissue culture (TC) plate and place in TC hood
- d. Aspirate medium from biopsy sample using a 2 ml aspirating pipette
- e. Remove biopsy sample and place in one well of the 6-well dish
- f. Using the scalpels, cut the biopsy sample into 6 pieces (Try to attach the biopsy to the plate with the scratches made by the scalpel)
- g. Add 1 ml of pre-warmed culture medium (DMEM, 10% FBS, 1% antimycotic, antibiotic) to each well, being careful not to dislodge the biopsy
- h. Gently swirl the 6-well dish to coat the wells with culture medium
- i. Place in the 37°C, 5% CO<sub>2</sub> TC incubator for 4-5 days to allow the biopsy to attach to the well
- j. Gently add 2 ml of fresh, pre-warmed culture medium to each well being careful to not dislodge biopsy sample
- k. Allow the sample to remain in the 37°C incubator until a monolayer of cells is present in the wells, feeding cells with fresh culture medium every 3-4 days
- l. Once adequate cells have grown out of the biopsy fragment, remove the cells from each of the 6-wells by washing with 2 ml PBS, and following removal of the PBS, by adding 800ul trypsin (0.25% Trypsin EDTA, Invitrogen). Incubate at 37°C until cells are released from substrate. Then add 2 ml of tissue culture medium and transfer cells to one T75 culture flask.
- m. After 2 days, aspirate all but 1 ml of medium from flask. Scrape the bottom of the flask using a cell scraper, and wash cells with the 1 ml of medium remaining in the flask. Remove the 1 ml of culture, and place in a 1.5 ml eppendorf tube. Perform mycoplasma testing using ATCC Universal Mycoplasma Testing Kit (see section IX.B.III.7 below) on this 1 ml aliquot and record results.
- n. Add 10 ml of fresh DMEM to flask and return to 37°C tissue culture incubator to allow the remaining cells to proliferate.

- o. Allow samples to reach confluency
  - p. Remove medium and wash cells with 10 ml PBS
  - q. Detach cells as described below (Detaching and Passaging Cells). Add 8 ml of tissue culture medium to collect cell suspension, use 1 ml of the culture to start another T75 culture flask. As described below (Freezing Cells), count the cells remaining in suspension, centrifuge at 1000 RPM for 10 minutes, add 3 ml of freezing medium, and freeze the remaining culture into 3x 1 ml cryovials at -80°C in a cool cell for 3 days with appropriate labels (UDN ID, sample type, collection date, and passage number). Transfer to -150°C for permanent storage (Passage 1 cells).
  - r. Allow the second T75 flask to reach confluency, feeding with fresh DMEM every 2-3 days. Once confluent, trypsinize the flask, add 9 ml of DMEM count the cells, centrifuge at 1000 RPM for 10 minutes, aspirate supernatant, add 3 ml of freezing media, and freeze at -80°C in a cool cell for 3 days with appropriate labels (UDN ID, sample type, collection date, and passage number). Transfer to -150°C for permanent storage (Passage 2 cells).
4. Freezing fibroblast cells
- a. Prepare 3-4 ml of fresh freezing medium (10% DMSO, 90% FBS) per T75 flask and warm to 37°C.
  - b. Place the 9 ml fibroblast culture into a sterile 15 ml conical tube.
  - c. Count cells using the cell counter: Add 10ul of culture to BioRad cell counting slide and insert into the BioRad TC20 Automated Cell Counter. Multiply this number of cells by 9 (volume of total culture) and divide by the number of frozen culture aliquots you are making to determine the amount of cells frozen per tube.
  - d. Centrifuge culture at 1000 RPM for 5 minutes.
  - e. Aspirate off supernatant.
  - f. Add prepared freezing medium and mix by pipetting up and down.
  - g. Aliquot 1 ml into labeled cryovial.
  - h. Place cryovial into Cool-Cell that is labeled with your name and date.
  - i. Place Cool-Cell into -80°C freezer for 3 days (36 hours) with appropriate labels (UDN ID, sample type, collection date, amount of cells, and passage number) and then transfer into -150°C storage.
5. Tissue culture for frozen skin fibroblasts
- a. Once the tissue culture medium has warmed, remove cell vial from -150°C and immediately place at 37°C.
  - b. Prepare a T75 tissue culture flask or 10cm petri dish by adding 10 ml of pre-warmed tissue culture medium.
  - c. Remove cells from vial using a sterile 1 ml pipette tip or 1 ml pipette.
  - d. Add cells to flask and gently mix.
  - e. Place inoculated flask in a 37°C, 5% CO<sub>2</sub> tissue culture incubator for 24 hours to allow cells to attach to the dish.
  - f. After 24 hours, remove medium using a 2 ml aspirating pipette and replace with 10 ml of fresh tissue culture medium. Place at 37°C, 5% CO<sub>2</sub> in tissue culture incubator.
  - g. Feed cells every 2-3 days by aspirating off old tissue culture medium and replacing with fresh, pre-warmed, tissue culture medium.
  - h. Once cells have reached confluency, cells must be passaged and split into new T75 culture dishes.

6. Detaching and passaging cells
  - a. Warm 0.25% Trypsin EDTA to 37°C.
  - b. Warm DMEM, 10% FBS, and 1% Anti-Anti to 37°C.
  - c. Once reagents have warmed, aspirate medium from flask containing cells using a 2 ml aspirating pipette.
  - d. Rinse cells with 10 ml of 1X PBS pH 7.4 without Calcium or Magnesium.
  - e. Aspirate 1X PBS using 2 ml aspirating pipette.
  - f. Add 1 ml of pre-warmed 0.25% Trypsin EDTA to TC flask and spread across the attachment area by swirling the flask.
  - g. Incubate flask at 37°C for ~ 5 minutes (or until cells are rounding), then gently tap the flask to release the cells.
  - h. Add 8 ml of pre-warmed tissue culture medium and wash cells to the bottom of the flask.
  - i. Add 3 ml of culture to fresh T75 flask containing 7 ml of tissue culture medium. Place at 37°C, 5% CO<sub>2</sub> in tissue culture incubator.
  
7. Mycoplasma testing
  - a. After 2-3 days of cell growth, aspirate all but ~1 ml of medium from flask.
  - b. Using a cell scraper, scrape the cells from the bottom of the flask only.
  - c. Using a 1 ml serological pipette, wash cells using the 1 ml of media in the flask.
  - d. Remove the 1 ml aliquot and place in a sterile 1.5 ml eppendorf tube.
  - e. Add fresh pre-warmed DMEM to flask and place in 37°C, 5% CO<sub>2</sub> tissue culture incubator for future use.
  - f. Perform mycoplasma testing using the Universal ATCC Mycoplasma Testing Kit (ATCC #30-1012K) according to the protocol on the 1 ml aliquot.
  - g. Record mycoplasma results and upload gel image into LIMS. If mycoplasma free, you can continue to passage. If mycoplasma positive, cells must be treated with Plasmocin (InvivoGen #ant-mpt) according to the protocol for 2 weeks and retested.
  
8. Growth and isolation of fibroblasts for metabolomics analysis
  - a. Grow passaged fibroblasts (passage 1 or 2) to a density of ≥ 1 million cells, in replicate (n = 3 – 5)
  - b. Remove cell culture media, and place 1 mL of spent media into a storage tube. Also place 1 mL of fresh media into a storage tube (Sarstedt Biosphere® SC Micro Tube 0.5 ml, Sarstedt Catalog # 72.730.217)
  - c. Quench cell metabolism using a wash of ice-cold PBS
  - d. Scrape cells from culture plate into storage tube (Sarstedt Biosphere® SC Micro Tube 0.5 ml, Sarstedt Catalog # 72.730.217). If needed, then add a small amount of ice-cold PBS to facilitate scraping and collecting of cells, followed by centrifugation to pellet cells (remove PBS and discard).
  - e. Flash freeze all samples in liquid nitrogen or quick freeze in dry ice/ethanol prior to storage in a -80°C freezer, or in liquid nitrogen, with appropriate labels (UDN ID, sample type, and collection date)

### **XIII. Central Biorepository**

The UDN Central Biorepository (UDNCB) is no longer accepting new samples from the CS in Phase III. However, samples submitted to the UDNCB in Phases I and II continue to be maintained and may be requested by the CS and their collaborators in Phase III. These samples include DNA, urine, serum, plasma, PBMCs, and fibroblasts. Please refer to the Gateway for details on UDNCB inventory.

#### **A. Flow of samples to UDNCB**

CS are able to view samples available, request samples, and view contact information for the UDNCB in the Gateway.

##### **Submitting samples**

The UDNCB is no longer accepting new samples.

##### **Viewing samples available**

Samples available can be viewed two ways in Gateway:

1. Within the individual participant record, navigate to the Samples tab and click “View Inventory” to view all samples available for that participant.
2. From anywhere in the Gateway, navigate to the Biorepository tab and click “Specimen Inventory” to view all samples currently available from the UDNCB.

##### **Requesting samples**

Samples stored by the UDNCB are available to UDN investigators and their collaborators. An investigator wishing to obtain samples from the UDNCB should check the availability of the samples of interest through the Gateway. If the investigator wishes to obtain samples from the UDNCB that were deposited by their own site, no approval by the Publications and Research Committee (PRC) is required.

Since quantities are limited, sample requests for samples from a site other than the investigator’s require prior approval from the UDN. If the investigator wishes to obtain samples from the UDNCB that were deposited by another site, they must first obtain approval from the UDN Publications and Research Committee (PRC). After confirming the availability of samples, the investigator must submit a research concept sheet through Qualtrics ([https://hms.az1.qualtrics.com/jfe/form/SV\\_3CyZOKOuiEvnCx7](https://hms.az1.qualtrics.com/jfe/form/SV_3CyZOKOuiEvnCx7)). The concept sheet should describe the proposed studies and the samples needed from the UDNCB. The UDN Publications and Research Committee (PRC) reviews the proposal and the Director of the UDNCB, who is part of the PRC, reconfirms sample availability.

Once approval has been given by the UDN PRC, the investigator may contact the UDNCB to request the biomaterials. Depending on the type of request, an MTA may need to be executed prior to material transfer. Transfer of samples within the network does not require an MTA.

The UDNCB contacts the investigator to arrange a time to ship samples. The investigator is asked to notify the UDNCB when samples arrive. The UDNCB notifies the relevant CS whenever samples are released to a UDN investigator or third party.

## **B. UDNCB Procedures**

### **Storing and archiving of biological specimens**

Mailed samples are opened in a clean “no amplified DNA” laboratory. The frozen samples, pre-aliquoted in screw cap cryotubes and labeled by the sender, are placed on dry ice while the labels are checked against the Sample Submission Form accompanying the package for confirmation. This form is downloaded from the Gateway and must be included with all samples shipped to the biorepository. The Sample Submission Form contains the participant UDN ID (also printed on the sample tubes). All tubes have participant UDN ID, sample type, and date of collection typed on the label. Receipt of the samples is recorded in the laboratory sample intake book and their condition noted (if dry ice is gone, samples are partially thawed, any tubes are cracked, etc.). Samples are not required to have barcodes, but the UDNCB has the ability to read 1D and 2D barcodes. If barcodes are included on the labels by the CS, the information outlined in the Biospecimen section must still be typed legibly on the sample labels.

#### Storage of samples

The biological samples are placed into liquid nitrogen cryotanks (PBMCs and possibly fibroblasts) and -80°C freezers (DNA, serum, plasma, urine) for long-term storage in the locations assigned by the Progeny Laboratory Information Management System (LIMS) database. Samples with multiple tubes are divided into 2 separate freezers/cryotanks. All UDNCB equipment is on the Vanderbilt Delta alarm system with temperature and nitrogen fluctuation notification automatically going to the Director’s and the senior Research Assistant’s cell phone/pager.

#### Documentation

The sender is notified by email of shipment arrival and any problems that may have occurred with the shipment (late arrival, partially thawed tubes, broken tubes, etc.). Any problems with the shipment are also recorded into the Progeny LIMS database. Documentation into the Gateway Sample Submission Form is completed by the CS submitting the samples, and a printed copy of the CSV file containing the details of the submission is included with the samples when they are shipped. No participant names or identifying information is recorded in the Progeny LIMS, the sample intake notebook, or requested on the Sample Submission Forms.

### **Retrieval and shipping of biological specimens**

The UDNCB retrieves biological samples from liquid nitrogen and/or -80°C freezers, packages samples in dry ice, and ships to UDN investigators and collaborators. Sample information in Progeny LIMS is used to track quantities and distribution of biological samples.

#### Locating biological specimens in storage

The Progeny LIMS database contains participant UDN ID, date of birth, sample type, date of collection, and sample location. In addition to sample and location information, the Progeny LIMS database keeps track of original and current quantities of the biological samples and record the distribution of samples to investigators.

#### Sample retrieval and transfer

After completion of a Sample Submission Form through the Gateway, the UDNCB contacts the CS or CS-designated investigator requesting the samples by email to pre-arrange a date for

shipment. The samples are then located using the Progeny LIMS database, retrieved, and placed on dry ice to prepare for transfer to the investigator requesting the sample.

#### Documentation of retrieval

The type of sample, amount transferred, date of retrieval, and the CS-designated investigator receiving the sample are recorded in the Progeny LIMS database. The Biorepository sample inventory on the Gateway is updated so that all UDN investigators can see which samples and the amount(s) of each remain in the system.

#### Packaging and shipping

The biological samples, already labeled and in screw cap tubes, are packaged and shipped per International Air Transportation Association (IATA) requirements that apply to all dangerous goods (such as dry ice) by air. Samples must be triple packed which includes a leak proof bag with absorbent material. UDNCB ships frozen samples in EPS foam containers (1.5 inch minimum thickness) with corrugated cartons, 10 lbs dry ice by FedEx priority overnight (Monday–Wednesday). Average dry ice sublimation in a 1-1/2 inch thick wall EPS container with corrugated container is 5 pounds over 24 hours (< 10 lbs in 48 hours).

Delays can arise with FedEx and the extra dry ice is a safeguard to protect the samples in case of delays in delivery. A list of sample content is included with shipment and the CS and CS-designated investigator(s) are notified by email that the sample has shipped, given the FedEx tracking number, and an electronic copy of the sample sheet. The email requests that the UDNCB be notified upon receipt of the shipment and of any problems with the samples (e.g., tubes thawed or damaged, etc.).

#### Packaging and shipping budget

Shipping samples to the UDNCB is at the CS expense and can be in batches to reduce costs. Shipping samples out from the UDNCB to investigators is at the expense of the UDNCB. The budget accommodates 50 shipments per year. If sample approvals by the UDN exceed the number of shipments allotted per year, then additional funds will need to be made available for the UDNCB or, alternatively, the CS-designated investigator(s) receiving samples will be required to pay the shipping costs.

#### Quality control

The UDNCB keeps records on the number of samples received, their condition, date shipped/date arrived, etc. The UDNCB also keeps records on the number of samples shipped out, date shipped/date received, condition upon arrival, any problems reported by the recipient, etc.

#### **Updating UDNCB inventory in the Gateway**

Samples collected by the CS and entered into the UDNCB Sample Submission Form prior to shipping are automatically entered into the sample inventory and can be viewed in the Gateway. When samples are requested and shipped out through the UDNCB, the UDNCB edits/updates the sample inventory on the Gateway.

## **XIV. Model Organism Screening**

### **A. Goals**

The goal of the Model Organisms Screening Center (MOSC) is to provide compelling data based on studies in worms (*Caenorhabditis elegans*), flies (*Drosophila melanogaster*), or zebrafish (*Danio rerio*) that refute or support the disease causality of specific genetic variants identified in participants by the Clinical Sites (CSs) and the Baylor Sequencing Core after genome or exome sequencing.

### **B. General Considerations**

The MOSC began analyzing variants in Phase I in early 2016 with the *Drosophila* Core at Baylor College of Medicine and the Zebrafish Core at University of Oregon (BCM-UO). The MOSC had numerous discussions with the CSs about how to optimize the process of best diagnosing disease-associated variants in genes submitted by each site. MOSC investigators also had numerous interactions with human geneticists that helped shape approaches and logistics. The overall philosophy of the MOSC is that of a collaborative center for clinicians, human geneticists, and model organism researchers who work together to solve cases. This chapter of the Manual of Operations represents a summary of the general operations of the MOSC.

With the start of Phase II of the UDN in Fall 2018, there was an expansion of the MOSC to include a new *C. elegans* Core and a second Zebrafish Core at Washington University in St. Louis (WashU). The two groups (WashU and BCM-UO) function as a single multi-institutional core with the following components:

1. Initial Variant Review Informatics Team - Dustin Baldrige (WashU) and Michael Wangler (BCM)
2. *C. elegans* (Worm) Core - Tim Schedl and Stephen Pak (WashU)
3. *Drosophila* (Fly) Core - Hugo Bellen, Shinya Yamamoto, and Michael Wangler (BCM)
4. Zebrafish (Fish) Core - Monte Westerfield and John Postlethwait (U Oregon) and Lilianna Solnica-Krezel (WashU)
5. Administrative components - Hugo Bellen, Michael Wangler (BCM)

### **C. Variant Submissions and Initial Variant Review**

#### **Gene variant submissions**

Each CS submits cases to the MOSC for analysis, typically after sequencing is complete. Within the Gateway, CSs can enter clinical information about the participant, submit candidate variants for the MOSC, and explain the rationale for prioritizing candidate variants that may contribute to disease phenotypes. The Gateway tracks MOSC submissions and allows communication about prioritization and progress back to the site. The sites are encouraged to submit 1-5 variants per case.

Variants that will be considered higher priority have the following characteristics:

1. Come from an unsolved case.
2. Are in novel or candidate disease-causing genes.

3. Are novel variants (i.e. VUS) in known disease-causing genes, but with unique phenotypes (i.e. phenotype expansions).

Work that is outside the scope of the MOSC includes:

1. Studying complex multi-gene interactions.
2. Studying environmental triggers.
3. Solved cases: Pathogenic variants or variants of uncertain significance (VUS) in known genes when the phenotype matches the known disease phenotype.
4. Treatment/Drug screens (although the models generated in the MOSC will be made available for such studies).
5. Generating models for known disorders.

Each variant submitted by the CS through the Gateway will be analyzed in parallel by the Initial Variant Review Informatics Team at WashU and BCM. After parallel analysis, the two teams review the variants via online calls. The procedure is as follows:

1. Variant submissions are submitted from Phase II UDN CSs using an updated version of the current Gateway MOSC request form developed in Phase I.
2. Variant submissions are then independently assessed at WashU and BCM in a process called “Initial Review”, primarily from a medical genetics perspective (Figure 1). Initial review includes assessment of key human data, including confirmation of genomic coordinates, assessment of allele frequencies, search for additional cases with variants in the same gene, determination of whether there are existing disease-gene associations for each variant, and future filtering/prioritization based on non-human primate data sets. This process is a quality control step to ensure that the site has provided the correct information on good candidate variants to be further analyzed in the Cores through the secondary informatics analysis pipeline described below.

### **BCM review process**

Using information submitted through the Gateway, the BCM site reviews clinical phenotypes of participants as documented by the CS during the evaluation, reviews the list of submitted variants and assesses which variants are suited for functional analyses in model organisms. This process is an important step because it shapes all downstream procedures. The MOSC carefully reviews each of the genes and variants that the CSs submit and explores the vast amounts of information that are available in public human and model organism databases. The MOSC performs expansive bioinformatic analyses to assess submitted genes and their variants. The MOSC developed a systematic Internet-based database search engine that integrates many of the steps that the MOSC first performed manually: MARRVEL (Model organism Aggregated Resources for Rare Variant Exploration; <http://marrvel.org/>). This search tool is publicly available and greatly reduces time spent to aggregate information used to prioritize variants (Luo et al., 2017).

Genes and variants that are being considered are available for review within the Gateway, providing all members of the UDN access to this information. Collaborative analyses are pursued as described (Wangler et al., 2017), involving the MOSC variants and those being analyzed by the Canadian Rare Disease Models and Mechanisms Network (RDMM), the Canadian equivalent of the UDN (<http://www.rare-diseases-catalyst-network.ca/>). In addition, the MOSC is currently considering expansion of this approach with the Japanese (Initiative on Rare and Undiagnosed Diseases) and Oceania (UDP Australia and New Zealand) consortia that attempt to diagnose rare or unknown diseases primarily to find matching cases. This communication avoids duplication of work and fosters collaborations to identify additional

participants who have genotypes and symptoms similar to the UDN case. Similarly, the MOSC collaborates with the Baylor-Hopkins Center for Mendelian Genomics (CMG) and with Baylor Genetics (a clinical diagnostic sequencing facility) to attempt to identify additional patients who carry mutations in the same gene. A comparison of phenotypes of the additional patients may then lower or elevate the probability that the candidate variant is causative, and hence the priority with which the MOSC continues the characterization of gene and variant function *in vivo*. These collaborative searches and inquiries can exclude some variants or genes from functional analysis. In some cases, additional patients with similar phenotypes may be found that together lead to definitive diagnosis without the need of model organism data (Schoch et al., 2017). In other cases, the discovery of one or more additional patients with similar phenotypes in the collaborative search can significantly elevate the priority of the gene/variant to be studied in the MOSC.

### **WashU review process**

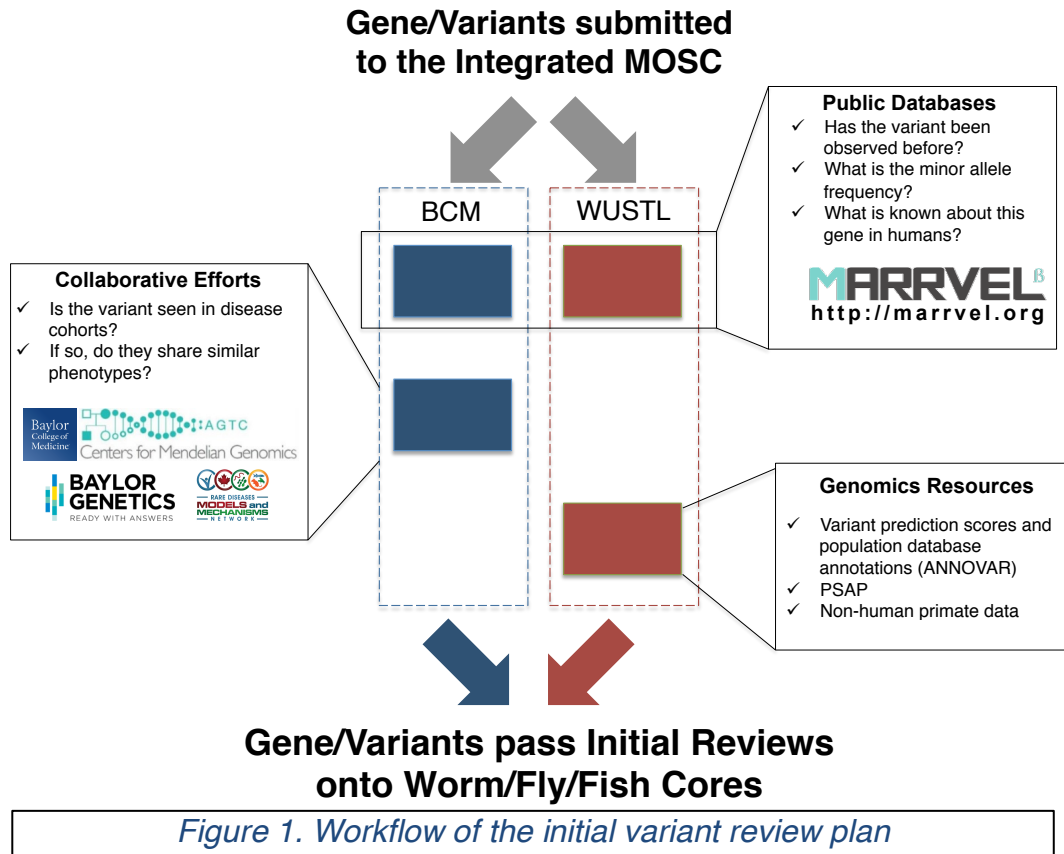
The WashU MOSC analyses submitted genes and variants in a parallel process that is similar to the above BCM review process. One reason for duplicate analysis is that the methods used are complementary to one another, such as RefSeq versus Ensembl for transcript annotation, allowing for a more comprehensive evaluation of submissions. Some tools are in common between both review processes, such as MARRVEL, as discussed above.

Variant nomenclature is assessed by use of the MutationTaster tool (<http://www.mutationtaster.org>), which is orthogonal to other methods used.

Variants are annotated by use of the ANNOVAR tool (<http://annovar.openbioinformatics.org/>): 1) genomic context and predicted effect on coding sequences (RefSeq), 2) population allele frequencies from multiple databases (gnomAD, ExAC, 1000 genomes), 3) multiple *in silico* nucleotide and protein functional predictions scores (e.g., SIFT, PolyPhen2, LRT, Mutation Taster, GERP++, PhyloP, CADD, REVEL, M-CAP), 4) splicing prediction (dbSNV), and 5) any previously observed clinical impact of variants (ClinVar, OMIM). Outputs from these analyses are placed onto a shared Google sheet to provide the “raw” data from which pathogenicity interpretations are made.

Literature searches are conducted to identify any human cases with disease-causing variants in the gene of interest, in order to compare to the phenotype of the UDN case.

Additional analysis modules are currently in development by WashU. One module involves the use of Population Sampling Probabilities (PSAP) (Wilfert et al., 2016), which leverages gene-specific models of pathogenicity derived from given pathogenicity scores (such as CADD) and the frequencies of variants observed in the unaffected population (such as ExAC). A second module uses recently available non-human primate (NHP) genomic sequence to interpret potentially pathogenic human variants



At the conclusion of the initial review, BCM and WashU both assign priority values: “Pass” “Return to Site,” and “Pass with Concerns.” In principle, for cases where more than one gene is submitted, the BCM and WashU Initial Review will also prioritize amongst submitted genes for a single case and assign a “Top Candidate” status to one gene. This status will automatically be assigned to genes for cases with only one submitted gene/variant.

BCM and WashU communicate the results described in C.III into the Gateway. The two groups discuss any difference of opinion, and if there are any concerns for either group, the variant is likely assigned a “Pass with Concerns” designation.

Notifications are provided via the Gateway with an “Initial Review” target turnaround time of 2 weeks from submission. Notification recipients depend on the variant designation as follows:

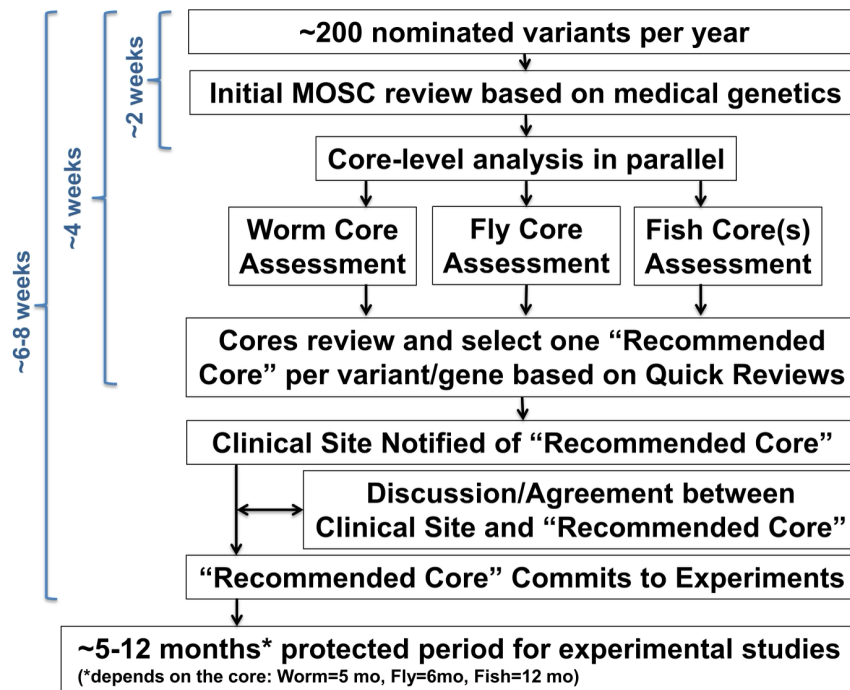
1. “Pass”- notification to Model Organism Core PIs (variants are processed as described in Part D below)
2. “Return to Site”- notification go to the CS. This notification states the reason for the submission return and that the site may contact the MOSC for clarification or additional information (including to make any necessary corrections to genomic coordinates) to resubmit the case. If no action is taken by the CS to resubmit the variant, these variants are not analyzed further.
3. “Pass with Concerns”- notification sent to Model Organism Core PIs (see Part D below).

## D. Variant Assignment to Core

1. The variants that have been designated as “Pass” (with or without concerns) from in the Initial Review (Section C-II) are analyzed at a “Quick Review” by each core (Figure 2). Examples of Quick Review determinations:
  - a. Worm Core might determine whether there is an ortholog and assess for paralogs, and whether the amino acid is conserved in worm so that the variant of interest could be modeled in this system.
  - b. Fly Core might determine whether there is an ortholog and assess for paralogs, determine whether there are published reagents available, whether there is a human cDNA available, and whether there is an alternative strategy to assess gene function other than the primary “humanization” approach.
  - c. Fish Core might determine whether there is an ortholog, assess for paralogs and known phenotypes, and might also assess whether the gene has been studied *in vivo* in mice and/or other vertebrate model organisms. Quick Review is carried out by UO and the results are discussed by the UO and WashU Fish Core teams to achieve a consensus regarding suitability of modeling this gene/variant in zebrafish. For genes/variants deemed suitable for modeling in zebrafish, further discussion leads to a decision about whether the variant would be better modeled at UO or WashU, or in collaboration.
2. Each core answers the question “Can the variant potentially be modeled in the core?” with “Yes”, “No” and “Maybe” based on the “Quick Review” and may add comments.
3. Gateway is configured to provide reminder notifications to ensure this step is completed no later than 4 weeks from submission.
4. Cores selecting “Yes” or “Maybe” will conduct a more thorough review if the variant was also classified as “Top Candidate.”
5. Gateway notifications depend on outcomes, with examples shown below (Table 1).

<b>Worm Core</b>	<b>Fly Core</b>	<b>Fish Core</b>	<b>Gateway Notification</b>
Yes	No	No	Worm Core
No	Yes	No	Fly Core
No	No	Yes	Fish Core
Yes	Yes	No	Worm and Fly Cores
Yes	No	Yes	Worm and Fish Cores
No	Yes	Yes	Fly and Fish Cores
Yes	Yes	Yes	All 3 cores
No	No	No	CS notified that the gene cannot be studied in the MOSC, asked to contact MOSC if they would like further information.

**Table 1. Potential outcomes of the secondary informatics analysis**



**Figure 2. Schematic of the process of variant assignment and commitment**

6. From this informatics analysis, a "Recommended Primary MO Core" is determined. For variants with only one "Yes", the "Recommended Core" is automatic (**Table 1**). For variants with more than one "Yes", the cores communicate within 2 weeks from the completion of the Quick Review, and prior to the next MOWG call, to decide which site is the "Recommended Primary MO Core".
7. The CS is notified that a "Recommended Core" has been nominated. Once the core 'commits' to experimental studies, the variant is officially "assigned" to the specific core. This commitment and the transition from "Recommended Primary MO Core" to "Assigned Core" can either occur 1) immediately after the assignment of the "Recommended Primary MO Core" for some cases when a Core may unconditionally begin to work on the specific gene/variant, or 2) after a discussion between the Core and CS prior to the transition to "Assigned Core" is preferred and should be arranged. This discussion involves exchanging information regarding participant phenotypes, existing data regarding orthologous genes in MOs, the actual experimental strategies proposed by the Recommended Core and additional questions and answers the two parties may have regarding the case. A notification to the CS with instructions for establishing contact sent from the Gateway to prompt the initiation of the discussion. Upon the identification of the "Recommended Primary MO Core", the CSs are responsible for establishing contact with the representatives of the specific MO Core, which is indicated in the Gateway. Ideally, the communication should take place within 2-4 weeks after the selection of the "Recommended Primary MO Core."
  - a. Phone calls or web meetings could be arranged by the CS to discuss the case with the Recommended MO Core.

- b. The discussion could take place as part of the monthly Model Organism Working Group (MOWG) conference call. Other UDN Sites and Cores that are part of the MOWG may be involved in this discussion and may provide some additional input.
- c. The case discussion can also take place through direct personal contact between the CS and Recommended Core representatives. Such communication can take place at an in-person meeting. E-mail communication may be sufficient for some cases (e.g., if Recommended Core is committed to the case unconditionally), but both the CS and core must be in agreement for the Recommended Cores to commit to the case experimentally. At minimum, the two parties should agree on the experimental strategies and the timeline (see section E).
- d. Due to the need for time to perform model organism experiments, each core sets a recommended “protected period” (5 months for worm, 6 months for fly, 12 months for fish) during which the CSs are asked not to publish variants that are under study in the MOSC. After this period, the sites are asked to communicate directly with the MOSC core about the progress.

Furthermore, the MOSC collaborates with the Knockout Mouse Phenotyping Project (KOMP) at BCM to knockout the orthologous gene in the mouse if a null allele has not been published in the literature. These mouse knockout lines will undergo systematic phenotyping based on the standards of the International Mouse Phenotyping Consortium (IMPC; <http://www.mousephenotype.org/>). In addition, for specific genes in which experts in the broader *Drosophila*, *C. elegans* and zebrafish communities have established functional assays and are willing to collaborate with the UDN, the MOSC will facilitate communication between the CS and the model organism investigator on a case-by-case basis. The goal of these efforts is to provide new tools, accelerate future research on the gene for the model organism community, and ultimately to benefit the UDN by creating larger collections of models that were initiated from the UDN.

## E. Model Organism Studies Timeline and Throughput

1. The Worm Core requires 5 months from the date of assignment, during which the sites cannot publish variants under study.
2. The Fly core will require 6 months from the date of assignment, during which the sites cannot publish variants under study.
3. The Fly Core will also generate genetic and molecular resources (e.g. T2A-GAL4 lines, UAS-human cDNA transgenic flies and expression vectors) for genes that are not immediately actionable in the fly core to generate resources for the community to study the gene of interest as described in the original proposal.
4. The Fish core at Oregon will require 12 months from the date of assignment, during which the sites cannot publish variants under study.
5. The Fish core at WashU will require 12 months from the date of assignment, during which the sites cannot publish variants under study.
6. When a variant is recommended to the Fish Core, a discussion between Oregon and WashU fish cores will determine in which of the two cores the variant will be functionally studied. The factors will include balancing the number of variants modeled in each core, and the expertise of each core:
  - a. Oregon: neural, skeletal, muscular and reproductive phenotypes among others.

- b. WashU: developmental, cardiovascular, kidney, and liver phenotypes among others.

## F. *C. elegans* Core Operations

The goal of the *C. elegans* Core is to determine if the candidate patient gene-variant, when examined in the worm, displays a phenotype(s). Observation of a phenotype provides experimental evidence that the gene-variant disrupts gene function in an animal. Additionally, the analysis can provide evidence supporting the participant's mode of inheritance (dominant, recessive or *de novo*) or allele status (homozygous, heterozygous, or compound heterozygous), and can potentially provide information on the nature of dominance (antimorph, hypermorph). In the analysis of phenotype, we employ the phenolog concept – distinct phenotypes that arise from genetic disruption of orthologous genes (McGary et al., 2010). Often patient presentation and *C. elegans* organismal phenotype are not obviously related; for example, a patient with a *TWIST*(E117V) *de novo* variant displays craniosynostosis, while the orthologous gene-variant in *C. elegans*, *hlh-8*(E29V), results in an egg laying defective phenotype. Nevertheless, both cases have the same underlying molecular defect, antimorphic poisoning of the *TWIST* transcription factor, reducing expression of downstream targets such as the FGF receptor and Notch ligand (Kim et al., 2017). Due to evolutionary distance, about 30% of patient gene-variants are conserved and can be modeled in *C. elegans*.

The assessment strategy involves knocking-in the patient gene-variant into the orthologous *C. elegans* gene, which is facile given the efficiency of CRISPR/Cas9 editing and the short generation time in the worm. The goal is to generate three variant edits and two control edits. Phenotypic analysis is guided by information on [www.WormBase.org](http://www.WormBase.org) and in the literature. If no phenotypic information is available, then a multi-step high-throughput phenotyping pipeline is employed. In most cases, variant edits are compared to null reference mutants, obtained from the *Caenorhabditis* Genetics Center or generated by the *C. elegans* CRISPR Knockout Project. Where appropriate, in some cases the *C. elegans* Core collaborates with *C. elegans* laboratories that have expertise in specific genes, pathways or phenotypes. Following completion of the project in conjunction with the relevant CS, genetic and phenotypic information is deposited into WormBase and the gene-variant and control variant strains are sent to the *Caenorhabditis* Genetics Center.

## G. *Drosophila* Core Operations

The Baylor College of Medicine (BCM) *Drosophila* Core performs bioinformatic analysis of human and *Drosophila* databases. Upon further discussion with the CS that submitted the case, the core directors and scientists of the Fly Core design experimental strategies to pursue in *Drosophila melanogaster*.

The *Drosophila* Core tailors experiments to each gene and variant using the optimal genetic strategy for each individual case. One strategy depends on the versatile engineered transposable element MiMIC (*Minos*-mediated integration cassette) (Venken et al., 2011), which allows “humanization” of *Drosophila* genes (Bellen & Yamamoto, 2015). This pipeline allows for: (1) generation of strong loss of function (LOF) alleles in the fly gene of interest; (2) expression of the yeast GAL4 transcription factor driven by endogenous enhancers of the gene of interest; (3) functional replacement of the fly gene with the reference (wild-type) human gene to test for rescue of the fly mutant phenotype with the orthologous reference human cDNA expressed under control of GAL4 expressed in the proper spatial and temporal pattern; and (4) variant

functional analysis through expression of the human disease allele variant in the same fly line. (1) and (2) are accomplished simultaneously by integrating a T2A-GAL4 cassette into a coding exon of a gene of interest (Chao et al., 2017; Wangler et al., 2017; Yoon et al., 2017) via recombinase-mediated cassette exchange of MiMIC, or via CRISPR-mediated homology directed repair (CRIMIC). Other strategies can also be used depending on the availability of pre-existing reagents (e.g., well characterized LOF alleles, experimentally verified RNAi lines), or when rescue with human cDNA has not been successful (Harel et al., 2016). For example, in cases in which the variant affects an amino acid conserved between the fly and human orthologs, a variant can be engineered in the context of a fly genomic construct to perform rescue experiments using previously characterized alleles with scorable phenotypes (Luo et al., 2017). Regardless of the exact methodology and phenotypes to be scored, the overarching goal of the *Drosophila* Core is to demonstrate whether the variant identified in the patient is functionally important, a knowledge that provides a molecular handle to further explore the molecular mechanism of disease.

## H. Zebrafish Core Operations

The goal of the Zebrafish Core is to mimic in zebrafish the lesion found in the participant's candidate gene and to provide a preliminary characterization of the phenotype of the mutation in zebrafish. The Zebrafish Core is composed of two research teams, one at the University of Oregon (UO) and the other at the Washington University in St. Louis (WashU). The two Zebrafish Core teams operate in a highly integrated manner to maximize communication and leverage the complementary expertise of the investigators. For the variants that have been designated as "Pass" (with or without concerns), during the "Quick Review", the UO team initially performs extensive bioinformatic analyses that identify the zebrafish ortholog (or in a minority of cases, co-orthologs) of the human candidate gene. Next, the UO team checks whether existing mutations or knockdowns have been characterized for the candidate gene in zebrafish and/or mouse, and if so, whether the described phenotypes support or fail to support the conclusion that the participant's clinical manifestation arises from a defect in the candidate gene. If no prior models exist, or if prior models have not been adequately investigated, the Zebrafish Core indicates that the variant potentially can be modeled in the Core. For genes/variants deemed suitable for modeling in zebrafish, initial discussion whether the variant could be modeled at UO or WashU also ensues.

When the Fish Core is the "Recommended Primary MO Core", a discussion between UO and WashU fish cores determines in which of the two cores the variant will be studied functionally. Factors considered include balancing the number of variants modeled in each core, and the expertise of each core. The UO team will investigate variants implicated in neural, skeletal, muscular and reproductive phenotypes among others. The WashU team focuses on genes implicated in early developmental defects, developmental delay, cardiovascular, kidney, and liver phenotypes among others. Typically, individual variants are modeled by either Oregon or WashU teams; in some cases, a fish model can be generated at UO but aspect of its phenotype can be studied at WashU.

Either UO or WashU Fish Core develops knockout mutations for the candidate gene using CRISPR/Cas9 technology. The goal is to make at least two alleles for each gene tested, ideally one has an early stop codon to block the formation of any functional protein, and a second one removes the gene's promoter or the entire coding region. The latter strategy aims at generation of a strong loss-of-function (LOF) allele that eliminates a transcript from this locus or a transcript with a nonsense mutation that can be associated with the compensation phenomenon, which

can confound the phenotypic analyses (Rossi et al., 2015). In the third strategy, CRISPR/Cas9/oligonucleotide strategy is employed to introduce a candidate mutant variant into the endogenous zebrafish gene. The two fish cores can also collaborate with WashU fish core engineering missense mutations in the zebrafish endogenous genes and the UO core generating large deletions and/or performing phenotypic analyses when appropriate. Finally, the Zebrafish Core analyzes mutant phenotypes caused by the induced mutations and compares them to the participant's clinical findings. Results are discussed with the CSs upon completion of the phenotyping.

Future goals include experiments to test for variant function by learning whether the human wild-type cDNA or mRNA can rescue the zebrafish mutant phenotype but the human mutant cDNA cannot. In cases in which the participant's phenotype is thought to arise due to dominant gain-of-function mechanisms, experiments to determine whether over-expression of the variant in wild-type fish results in phenotypes that mimic the participant's symptoms can provide valuable mechanistic information. In addition, precise homologous changes in zebrafish proteins are being made.

## I. Communications

The MOSC summarizes bioinformatic analyses in Gateway which sends the information back to the CSs for discussion. The MOSC also communicates the decision regarding which core will be performing the model generation and characterization.

### Communication with CSs after data generation

After generation of mutant models and characterization of mutant phenotypes, the MOSC contacts CSs to discuss observations. The CSs often raise questions and issues that may be addressed in subsequent experiments. Although the MOSC carries out characterization of the gene in mutant models, the CSs often pursue identification of additional patients to provide more robust evidence that the mutation observed in the participant is indeed associated with the observed phenotype. These data are integrated as part of the characterization to provide a final and compelling diagnosis.

### Throughput and allocation

CSs submit approximately 200 variants per year. The *Drosophila* Core studies approximately 30 variants per year, the UO fish core studies approximately 15 per year, and the WashU Fish Core another 15 per year. The *C. elegans* core has a ramp up period of 5 variants in the first 6 months, 15 variants in the second 6 months, and potentially up to 60 variants per year in subsequent years, dependent on the number of accepted variants that can be tested in the worm.

### Distribution of reagents

All *C. elegans* strains are sent to the *Caenorhabditis* Genetics Center. All the fly stocks are deposited in the Bloomington *Drosophila* Stock Center at Indiana University (BDSC, <http://flystocks.bio.indiana.edu/>). The fish resources are deposited at the Zebrafish International Resource Center (ZIRC, <https://zebrafish.org>). These facilities keep records of stocks they distribute, and these records document the usage of the MOSC materials. In addition, the vectors containing the UAS-human cDNAs are deposited in the *Drosophila* Genome Resource

Center (DGRC) in Indiana and are publicly available. All these distribution centers are supported by the NIH. Finally, the reagents (worms, fish, flies and cDNAs) are or will be available for sharing with the CSs as soon as they are generated. Note that the MOSC acquired a human cDNA library containing about 40,000 unique cDNA clones assembled and generated in the past 6 years by the late Dr. Ken Scott at BCM.

## J. References

- Bellen, H. J., & Yamamoto, S. (2015). Morgan's Legacy: Fruit Flies and the Functional Annotation of Conserved Genes. *Cell*, 163(1), 12-14. doi:10.1016/j.cell.2015.09.009
- Chao, H. T., Davids, M., Burke, E., Pappas, J. G., Rosenfeld, J. A., McCarty, A. J., . . . Malicdan, M. C. (2017). A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. *Am J Hum Genet*, 100(1), 128-137.
- Harel, T., Yoon, W. H., Garone, C., Gu, S., Coban-Akdemir, Z., Eldomery, M. K., . . . Lupski, J. R. (2016). Recurrent De Novo and Biallelic Variation of ATAD3A, Encoding a Mitochondrial Membrane Protein, Results in Distinct Neurological Syndromes. *Am J Hum Genet*, 99(4), 831-845. doi:10.1016/j.ajhg.2016.08.007
- Kim, S., Twigg, S. R. F., Scanlon, V. A., Chandra, A., Hansen, T. J., Alsubait, A., . . . Corsi, A. K. (2017). Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in *Caenorhabditis elegans*. *Hum Mol Genet*, 26(11), 2118-2132. doi:10.1093/hmg/ddx107
- Luo, X., Rosenfeld, J. A., Yamamoto, S., Harel, T., Zuo, Z., Hall, M., . . . Wangler, M. F. (2017). Clinically severe CACNA1A alleles affect synaptic function and neurodegeneration differentially. *PLoS Genet*, 13(7), e1006905.
- McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B., & Marcotte, E. M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A*, 107(14), 6544-6549. doi:10.1073/pnas.0910200107
- Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Holper, S., Kruger, M., & Stainier, D. Y. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*, 524(7564), 230-233. doi:10.1038/nature14580
- Schoch, K., Meng, L., Szelinger, S., Bearden, D. R., Stray-Pedersen, A., Busk, O. L., . . . Shashi, V. (2017). A Recurrent De Novo Variant in NACC1 Causes a Syndrome Characterized by Infantile Epilepsy, Cataracts, and Profound Developmental Delay. *Am J Hum Genet*, 100(2), 343-351. doi:10.1016/j.ajhg.2016.12.013
- Venken, K. J., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., . . . Bellen, H. J. (2011). MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nat Methods*, 8(9), 737-743.
- Wangler, M. F., Yamamoto, S., Chao, H. T., Posey, J. E., Westerfield, M., Postlethwait, J., . . . Bellen, H. J. (2017). Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. *Genetics*, 207(1), 9-27.
- Wilfert, A. B., Chao, K. R., Kaushal, M., Jain, S., Zollner, S., Adams, D. R., & Conrad, D. F. (2016). Genome-wide significance testing of variation from single case exomes. *Nat Genet*, 48(12), 1455-1461. doi:10.1038/ng.3697
- Yoon, W. H., Sandoval, H., Nagarkar-Jaiswal, S., Jaiswal, M., Yamamoto, S., Haelterman, N. A., . . . Bellen, H. J. (2017). Loss of Nardilysin, a Mitochondrial Co-chaperone for alpha-Ketoglutarate Dehydrogenase, Promotes mTORC1 Activation and Neurodegeneration. *Neuron*, 93(1), 115-131. doi:10.1016/j.neuron.2016.11.038

## XIV. Appendices

### APPENDIX 1: Research Concept Sheet, Manuscript Drafts, and Accepted Conference Abstract Submission

To complete online, go to: [https://hms.az1.qualtrics.com/jfe/form/SV\\_3CyZOKOuiEvnCx7](https://hms.az1.qualtrics.com/jfe/form/SV_3CyZOKOuiEvnCx7)

**Concept sheets** must be submitted for case/cohort studies, research projects, data analysis projects, grant applications, and UDN-led manuscripts *before substantial work begins*. This includes but is not limited to: single-site projects, case reports, analysis of Gateway data from participants evaluated at other sites, and biorepository specimen requests from external groups.

**Manuscripts** must be reviewed by the PRC prior to publication. Final manuscripts should be submitted to the PRC and journal of interest simultaneously.

**Accepted conference abstracts** and presentations sharing data from a single site do not require PRC review, but submission is needed to track UDN contributions. Abstracts and presentations including UDN-wide data or data from more than one UDN site require PRC and DMCC review to ensure accurate reporting of UDN-wide data.

Concept sheets, manuscripts, and accepted conference abstracts/presentations should all be submitted via this survey.

**\*Please do NOT include any PII/PHI in this submission.\***

Send to [UDN\\_Convener@hms.harvard.edu](mailto:UDN_Convener@hms.harvard.edu) when complete.

Type of submission:

- Research project (involves the collection of new data)
- Data analysis (involves new analyses of existing data)
- Case/cohort study concept sheet
- Manuscript draft (ready for journal submission)
- Grant application
- Accepted conference abstract

Your contact information:

Name:

Institution:

Phone:

Email:

Are you a UDN Member?

- Yes
- No; please list the name of your UDN contact:

Primary investigator contact information:

Name:

Institution:

Phone:

Email:

Additional/Senior investigator(s) (specify institution):

Is this project being led by the UDN or by an external group?

UDN

External group; please list the name of the external group:

Does this project involve collaborations between more than one UDN site?

Yes

No

If yes, which UDN sites are involved?

Does this project utilize UDN-wide data or data from more than one UDN site?

Yes

No

Project title:

UDN ID(s):

Gene(s):

Brief summary (if conference abstract, paste here):

Timeline:

**If the submission is an accepted conference abstract, please provide the following information:**

Meeting/conference name:

Date of meeting:

Type (poster, platform, plenary, presentation, workshop, TBD, other):

Abstract/poster number:

Presenter name:

Presenter email:

**If the submission is a manuscript draft (ready for journal submission), please provide the following information:**

Does this paper describe a new condition or the phenotypic expansion of a known condition?

Yes: new disease-gene association

Yes: new non-genetic condition

Yes: phenotypic expansion

No

Date (mm/dd/yyyy) of manuscript submission (estimate if unsure):

Intended journal:

**If the submission is a research project concept sheet, data analysis concept sheet, or grant application, please provide the following information:**

*Due to the significant costs involved in file egress, downloading raw genomic data files for the entire cohort from the UDN Gateway is not feasible. Researchers seeking access to a large number of files are encouraged to request access to the UDN project on dbGaP. For more information, please visit the UDN website: [UDN Data Availability](#)*

Specific aims:

Methods:

Funding plan:

Subject eligibility criteria:

Access to data:

Burden on subjects:

Biospecimens required:

Which (if any) UDN Working Groups are (or will be) involved? Include primary Working Group first. (See the [UDN Manual of Operations](#) for a list of Working Groups).

Is additional consent required beyond the UDN consent?

Yes; please describe:

No

Data management and statistical support:

Required

Not required

If support is required, please provide contact information for your local statistician:

Name:

Phone:

Email:

Other needs:

Does this research project involve administering a survey to participants?

Yes (please attach a copy of the instrument to this application)

No

**If you will be administering a survey to participants, please provide the following information:**

Is this research project under the central UDN IRB (NIH IRB), meaning it is covered by the UDN consent form, or will there be a separate IRB protocol and consent form through a local IRB?

Will be under UDN Central IRB

Will be under a separate local IRB

A. Survey Details

1. Survey Instrument

2. Construct measured

3. Description of validation of survey instrument

4. Rationale for using this particular survey instrument (as opposed to other instruments that measure the construct)
5. Translation of measure into other languages

B. Survey Population

1. UDN sites included in the study
2. Participant inclusion/exclusion criteria
3. Sample size

C. Study Logistics

1. Amount of time to administer
2. Method of administration (online [Qualtrics, other?], phone, mail, etc.)
3. Individuals administering survey (study staff, self, etc.)
4. Number of times survey is administered and when
5. Timeline for study: When will survey commence, for how long will you continue the survey, when will it end, when will data be analyzed

D. Please provide a plan for follow up if a problem or concern is identified on responses to the measure.

E. Data Sharing – who will have access to the data?

## APPENDIX 2: Diagnosis Coding Tool v1.0

Patient UDN ID  
Diagnosis Name  
Date Diagnosis Entered  
Diagnosis Entered By  
Phenotype MIM Number  
Gene Name  
Gene MIM Number  
ICD Number  
Additional Diagnosis Notes

### Category 1: Certainty

The overall certainty of the diagnosis.

- Certain
- Highly Likely
- Tentative
- Low (still on differential diagnosis, but lacking clear evidence)

More information:

### Category 2: Chief Method Used to Achieve Diagnosis

The single most important piece of evidence supporting the diagnosis.

- Genome-scale sequencing including genome sequencing and exome sequencing (select all that apply)
  - Reanalysis of prior clinical exome data
  - Reanalysis of prior research exome data
  - Reanalysis of prior clinical genome data
  - Reanalysis of prior research genome data
  - Clinical site analysis of Baylor exome data
  - Clinical site analysis of Baylor genome data
  - Clinical site analysis of HudsonAlpha genome data
  - Long read sequencing
  - RNASeq
  - Baylor – pathogenic
  - Baylor – likely pathogenic
  - Baylor – VUS
  - Baylor – research finding
  - HudsonAlpha – pathogenic
  - HudsonAlpha – likely pathogenic
  - HudsonAlpha – VUS
  - HudsonAlpha – research finding
  - Known prior to enrollment in UDN
  - Other

- Non-sequencing, genome-wide diagnostic assay (for example, SNP array, oligo array or karyotype)
  - Microarray
  - Karyotype
  - Known prior to enrollment in UDN
  - Other
- Directed clinical testing based on phenotype (may be molecular, including gene panels; biochemical; radiologic; or other test type)
  - Panel testing
  - Single gene testing
  - Del/dup testing
  - Known prior to enrollment in UDN
  - Other
- Diagnosis made primarily on clinical grounds, including aggregate assessment of non-specific test results
- Other – please explain below

More information:

#### Category 3: Mechanistic Characterization

The degree to which the underlying mechanisms of disease pathology and causation are established.

- Biologically well-characterized disorder, e.g., high-specificity testing available; or, the gene, pathway, and/or cell biology is understood.
- Descriptive (no definitive testing available, specific mechanism unknown).
- Other – please explain below

More information:

#### Category 4: Timing and Source of Diagnosis

The source and timing of the diagnosis relative to evaluation by the UDN.

- Diagnosis made before UDN case submission
- Diagnosis made by non-UDN entity after case submission but before UDN evaluation
- Diagnosis made by UDN after case submission but before evaluation
- Diagnosis made during UDN evaluation
- Diagnosis made after UDN evaluation by UDN
- Diagnosis made after UDN evaluation by non-UDN entity
- Other, including misdiagnosis – please explain below

More information:

#### Category 5: Degree to Which This Diagnosis Explains Major Components of Phenotype

- Explains all major features of illness, including asserted phenotype expansion
- Explains some major features of illness, including asserted phenotype expansion

- Diagnosis does not explain any major features of illness
- Other – please explain below

More information:

Category 6: Consequence of Diagnosis

Was the diagnosis important in the future management of the study participant or family's overall health?

- The diagnosis had a major role in the subsequent management of the participant's and/or family's overall health
- The diagnosis had some role in the subsequent management of the participant's and/or family's overall health
- The diagnosis had no significant role in the subsequent management of the participant's and/or family's overall health
- Other – please explain below

More information:

Is there a readily available treatment?

- Yes
- No

If yes, please provide a brief description of the treatment:

More information:

Was the Mosaic analysis platform (previously known as Calypso) used in making this diagnosis?

- Yes
- No

If yes, please provide a brief explanation of how it was used:

Narrative:

# APPENDIX 3: Onboarding Workflow

