Metabolic Phenotypes in the ClinGen Repositories

Karen Eilbeck[1], Donna Maglott[2], Melissa Landrum[2], Sara Velden[3], Marzia Pasquali[3], Elaine Lyon[3], Nicole Ruiz[1], William Craigen[4], Robert Steiner[5], Michael Murray[6], David T. Miller[7], Rong Mao[3]

1) Department of Biomedical Informatics, University of Utah School of Medicine; 2) National Center for Biotechnology Information, National Institutes of Health; 3)ARUP Laboratories, Salt Lake City;4) Molecular and Human Genetics, Baylor College of Medicine, Houston; 5) Marshfield Clinic Research Foundation; 6) Genomic Medicine Institute, Geisinger Health System; 7) Boston Children's Hospital, Harvard Medical School.

Background

Data sharing and reuse is a goal of clinical variant databases such as the NCBI's ClinVar database. To achieve this, the data-model and data-dictionary rely heavily on the use of controlled vocabularies, established nomenclatures and terminologies to define the data entered. Defining metabolites is emerging as an area of development for the database in order to adequately capture the clinical findings necessary to support the interpretation of inborn errors of metabolism. The ClinGen clinical domain working group for metabolic disorders have focused on PKU and MCAD to explore this domain, and were assisted by the ClinGen Phenotype Working Group and the ClinGen IT/Data Standards working group.

Approach

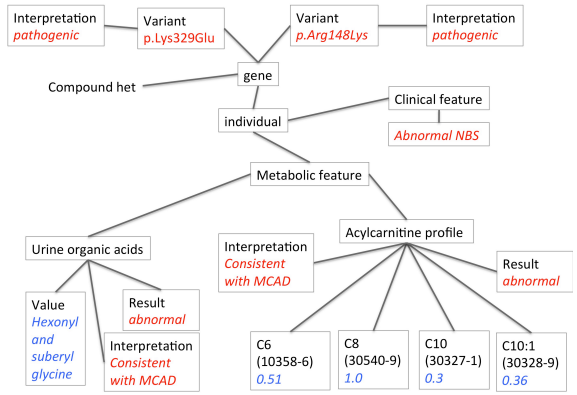
Clinical and laboratory information was collected from patient history forms and from clinical laboratory websites for both genetic and metabolic testing, and the kinds of phenotypes were catalogued. De-identified data for MCAD patients were shared by ARUP. This included sequence variants, metabolite levels (with LOINC codes) reported from the biochemistry lab, phenotypic features, and test interpretations. This data was used to build a model for the submission of this data to ClinVar.

Results

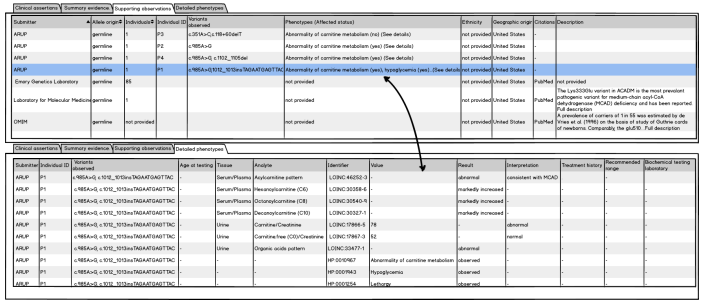
A data model for supporting different levels of metabolic evidence was developed. The ClinVar team developed a new tab delimited form for the submission of metabolite data, and provided mock up screens of how this will be displayed in their system. The model and submission form were iteratively refined. One hundred and eighty cases with metabolic data are in preparation for submission.

Acknowledgments. This research was supported by NIH Grants U41 HG006834-01A1, and U01HG007437 .

Modeling the combined test results



Mockup of a ClinVar display of supporting evidence and detailed phenotypes with test results submitted for NM_000016.4:c.985A>G



Proposed data entry tab for Phenotypic details

#Individual ID	Tissue	Analyte	LOINC code	Result	Interpretation	Biochemical testing laboratory	Clinical feature	Clinical Feature qualifier
Required. Identifier to connect this row in this spreadsheet to the data in the CaseData worksheet	Optional. Do not provide if the LOINC code for the measurement is provided	Optional. Do not provide if the LOINC code for the measurement is provided.	Optional.	Optional: if Tissue/Analyte or LOINC code are provided, either Result or Interpretation must also be supplied. Value obtained for what was measured.	Optional. Your evaluation of the what the results mean. Either result or interpretation must be supplied.	Optional. Name of the laboratory in which the metabolic testing was performed, if different from the submitter.	Optional: HPO id or term	Required if Clinical feature provided. If not provided, will be reported as present.Options are present, absent, not tested, unknown
ARU0001			46252-3	abnormal	consistent with MCAD			
ARU0001			30388-6	0.32				
ARU0001			30540-9	1.82				
ARU0001			30327-1	0.42				
ARU0001			53718-3	Acyl HG 28.8 (0.0-0.8);	abnormal			
ARU0001			33477-1	hexanoyl glycine suber	abnormal			
ARU0001			17866-5	13	abnormal			
ARU0001			17867-3	6	abnormal			

Multiple variants within a gene. The interpretation can be made to a compound heterozygote pair of variants, where the interpretations is made in the variant tab of the submission file. Other variants may be annotated as co-occurring in the same gene, but with no associated assertion. This second mechanism can be used to detail benign variants that were detected in the test.

Annotation of Clinical features. Other clinical features can be described using Human Phenotype Ontology (HPO) terms in the new Phenotypic details tab, clinical feature column. A qualifier enables the submitter to specify if the feature was observed, or looked for but not present. The table shows the collection of HPO terms collected by the lab.

HPO term	HPO ID
Coma	0001259
Encephalopathy	HP:0001298
Hepatomegaly	HP:0002240
Hypoglycemia	HP:0001943
Lethargy	HP:0001254
hepatic failure	HP:0001399
Reye syndrome-like episodes	HP:0006582
Seizures	HP:0001250
Vomiting	HP:0002013

In preparation for ClinGenDB

Annotation of tests and results with the LOINC codes will enable the data to be computed over. These codes link the test values to the correct units and analytes. This will enable the development of calculators such as 'ratio calculators' to be embedded in ClinGenDB to enable normalization across laboratories.

Anatomy of a LOINC code

LOINC number	**Component** - analyte that is being measured	**Property** – how is it being measured? (mass, volume, etc.)	**Time aspect** – the timing of the measurement	**System** – the specimen type	**Scale** – is the measuremnt quantitative, qualitative, nominal etc.
30358-6	Hexanoylcarnitine (C6)	SCnc (substance concentration e.g. nmol/ml)	Pt (point in time)	Ser/Plas (Serum or plasma)	Qn (quantitative)
33477-1	Organic acids pattern	Imp (Impression/Interpretation of Study evaluation over a collection of data)	Pt (point in time)	Urine	nominal

Conclusions

Metabolic phenotypes differ from the clinical features available in existing terminologies for medical genetics in that they contain analyte values. As these metabolites come from laboratory tests, they are most often described using LOINC. This is a universal standard system for reporting clinical observations. It uses a six-part name to uniquely discriminate between tests. Incorporating this system into the database ensures that the values collected remain interpretable. Laboratories may measure single metabolites or they may aggregate into panels. Importantly, different labs have different references ranges for interpretation of the levels of each metabolite. Commonly, to normalize this, they also examine ratios between select metabolites when making an interpretation. At this stage, raw data will enable calculation of ratios and in the future it is expected that ClinGenDB will adopt this model, provide ratio calculators, and mechanisms to normalize data across labs.