ClinGen Data Exchange

Purpose

The ClinGen Data Exchange provides a platform for query and distribution of ClinGen curations in a structured, machine readable form.

Enables

Sharing of ClinGen data within ClinGen

- ClinGen Website
- Evidence Repository
- Gene Curation Tracking System

Sharing of data with the community

- **ClinVar** (via ClinVar submission tool)
- Variant annotation tools
- Other research efforts

Components

Data Standards

ClinVar Submission Generator

Validation Tools

APIs

Streams

Data Standards

Broad/Geisinger

Purpose

Represents a contract between producers and consumers of data. Provides a consistent structure for sharing of data using commonly understood terms (defined in external ontologies).

Formalized data models are necessary for sharing information with the community and are necessary for richly structured data.

Ad-hoc data models are used as a stopgap for sharing data internally while data models are formalized.

Data Standards

Broad/Geisinger

Current status and progress

- Variants
  - Allele (completed)
  - Copy Number Variants (in progress--new)
- Variant Pathogenicity
  - Sequence Variants (completed)
  - Copy Number Variants (need curation standard to start work)
- Gene Dosage (beta release--new)
- Gene Disease Validity (ad-hoc)
- Clinical Actionability (ad-hoc, Baylor has started exploratory work on formal model)

Evidence Repository

Baylor

Purpose

Provides a machine queryable interface to ClinGen data--returns structured data in the standard ClinGen data models.

Used to respond to interfaces that have a specific question to ask, for example, "What does ClinGen know about TSC1:c.3469A>G"

Streams

Broad/Geisinger

Purpose

Provides a stream of all structured curated data produced by ClinGen. Also provides a persistent archive of all structured data, including historical curations.

Used to feed data to consumers that maintain their own backing store, such as the ClinGen Website. Also used to notifiy consumers that the information relating to a topic has changed.

Status and progress

- Variant Pathogenicity (in progress)
- Gene Dosage (beta--new)
- Gene Disease Validity (via ad-hoc model)
- Clinical Actionability (via ad-hoc model)

ClinVar Submission Generator

Purpose

Transform standard ClinGen Variant Pathogenicity data into a ClinVar submission

Status and progress

Beta version available, currently in testing.

Validation tools

Purpose

Validate compliance of data with ClinGen standard data models

Status and progress

- JSON Schema based validation exists for Variant Pathogenicity model.
- Next steps are to develop schemas for other models.
- Need to develop graph-based schemas to validate RDF Transformation.

Next work

Gene Disease Validity -- formal model

- Need to coordinate and prioritize efforts of both Data Exchange and Stanford teams to complete this item.

Validation

- Extend schema-based validation to Gene Dosage model.
- Establish structure and standards for RDF validation of model content.