**The ClinGen Individual Level Database (ILDB)**

Working Definition, Aims and Objectives

Draft 3

August 16th, 2015

Contributions by Snehit Prabhu, Heidi Rehm, Sharon Plon and Carlos Bustamante

On behalf of

The ClinGen ILDB workgroup

The ClinGen ILDB project is an umbrella initiative to help patients, laboratories and healthcare entities contribute clinical data (both genomic and phenomic) for the purpose of gene and variant interpretation in the context of disease. The overarching aim of the ILDB workgroup will be to facilitate the transparent and seamless use of (and conversely, explicitly define the conditions for non-use of) genomic, phenomic and other relevant health data – acquired in a clinical context of delivering patient care – to improve the interpretation and actionability of the genomic markers.

Genomic information generated in clinical settings today is often locked away in data silos at their source institutions. Despite its immense scientific value and potential, such data is unable to contribute meaningfully towards patient-health oriented research (often contrary to the intentions of the patients from whose samples the data were derived in the first place). A variety of causes lie at the root of this status quo, including (a) legal hurdles that discourage organizations from contributing sensitive data, (b) regulatory hurdles that dis-incentivize their participation, and (c) technical hurdles pertaining to the ease with which such data might be securely deposited at a single site. The existing state of affairs has given rise to several organic, inefficient and potentially error-prone mechanisms through which clinicians, each with access to a single institutional data silo, share their limited insights pertaining to the molecular mechanisms of disease.

For the task of clinical interpretation, an illustration of the predominant workflow in use today for pathogenicity assessment of genomic variation is provided here:

<https://github.com/snehitp/ILDB/blob/e614badc96b8d8fe44b89de6eddb405199d651e9/Documents/Variant-Interpretation-Today.pdf>

In contrast, we envision a new workflow by which patient data relevant to the task of interpretation can easily and efficiently be shared among the research community. An illustration of this workflow is provided here:

<https://github.com/snehitp/ILDB/blob/e614badc96b8d8fe44b89de6eddb405199d651e9/Documents/Variant-Interpretation-ClinGen.pdf>

The immediate focus of the pilot ILDB project will be to support various ClinGen disease-domain working groups in their respective variant interpretation and curation efforts. In this regard, the ILDB will provide a superior alternative to *ad-hoc* solutions that are in use today, whereby collaborative research around genomic interpretation is mostly done using simple worksheets. In current practice, investigators often manually “scrub” case-level clinical data using rough intuition-based guidelines to protect their patients’ identities. Such data is then shared between peers, often via email, in the interest of discovering relevant patient commonalities or to bolster evidence for/against a VUS’s pathogenicity. The ILDB will provide a more unified, robust and safe solution to replace these practices.

The ILDB is therefore a critical piece in ClinGen’s broader mandate of efficient and accurate genome interpretation. By making large sets of clinical data available to the variant curation and interpretation community in a safe, secure and robust manner, the ILDB will allow each domain-specific workgroup to process a deluge of relevant variation in the context of their respective disease. The quality of interpretations will also directly benefit from the increased evidence made available by the ILDB, resulting in more biological insight and greater clinical actionability of genomic tests.

In order to achieve these goals, the ILDB workgroup will propose and develop new solutions to the following issues (or identify and borrow appropriate mechanisms where other teams have already developed these):

1. Consent standards
2. Data representation models
3. IT infrastructure, including secure storage repositories
4. Data submission, QC and validation protocols
5. Quantitative and qualitative privacy guarantee mechanisms
6. Data usage guidelines, and lastly,
7. Regulatory standards and oversight measures for ClinGen curators

Logistical details and an action-plan for implementation will be described in a separate document.

The ILDB is a resource that has the potential to directly and positively impact patient care, in exchange for minimal to no additional loss in patient privacy and security than those afforded by current practice. Through research and development of the mechanisms enumerated above, the ILDB workgroup will provide a comprehensive solution through which appropriate individual-level clinical data might be incorporated into such research. Besides variant/gene interpretation, extensions to other areas of medical and genomic research may also be charted out, but are of secondary importance.