Ontology for Genetic Susceptibility Factors (OGSF)
-- an Ontological Framework to Link Genomics with Adverse Events

“Asiyah” Yu Lin, M.D., M.S., Ph.D. ORISE Fellow, OTS, CDER, FDA, Sliver Spring, MD, USA.

Norihiro Sakamoto, M.D., Ph.D. Kobe University Hospital, Kobe, Hyogo, Japan

Yongqun “Oliver" He, D.V.M., Ph.D. University of Michigan Medical School, Ann Arbor, MI, USA.

Background

The concept of genetic susceptibility, also called genetic predisposition, describes an increased likelihood or chance of developing a particular disease (*e.g.*, diabetes) or conditions (*e.g.*, adverse event) due to the presence of one or more gene mutations with or without a family history of an increased risk. The fact of drug induced adverse event (AE) manifesting stronger in certain populations than others indicates the role of genetic variants.

The general methodology to identify the genetic susceptibility to complex disease is a combination of linkage and association studies, which leads to heterogeneous genomic-disease data existing in different formats. In addition, many genetic association studies, including GWAS (Genome-Wide Association Study) studies, have been conducted to test drug-gene interaction models. With the easy access of high-throughput Next Generation Sequencing (NGS) technologies, genomic association data is growing into big volume. In order to study the relationship between gene and disease in a holistic manner, there is a need to integrate and organize the heterogeneous genomics data and their related disease systematically.

Ontology, a set of computer- and human- interpretable terms and relations that represent entities in a specific domain and how these entities relate to each other, has been widely used for data integration and knowledge in biomedical informatics.

Method

In this study, ontology based approach is adopted to unify and organize the aforementioned heterogeneous genomic and disease data.

An **Ontology for Genetic Susceptibility Factors (OGSF)** was developed as a framework to represent genetic susceptibility, genetic susceptibility factors, population, studies and phenotypes, such as diseases or AEs.

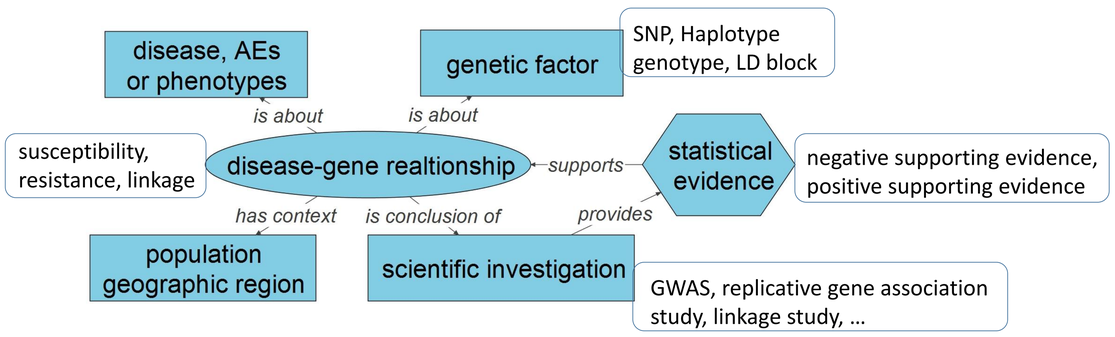
The methodology of developing OGSF follows a hybrid bottom-up and top-down process. **Bottom-up:** Starting with reviewing literature to identify the common element for describing a disease genetic study, we identified primary blocks for genetic susceptibility to common disease by analyzing the titles of the corpus [1]. Based on the basic high level model showed in the figure, more detailed information about each box, such as the subclasses of those concepts in the boxes, were listed out. And the next step is to find the existing ontologies to set a proper and logically sound relationship among those subclasses. **Top-down:** After investigating all the top-ontologies, Basic Formal Ontology (BFO) was chosen to host all the concepts in the boxes, and the subclasses.

OGSF was developed in Web Ontology Language (OWL) using protégé OWL editor.

Google project site, as well as GitHub (https://github.com/linikujp/OGSF), were used for version control and open development environment for the OGSF project.

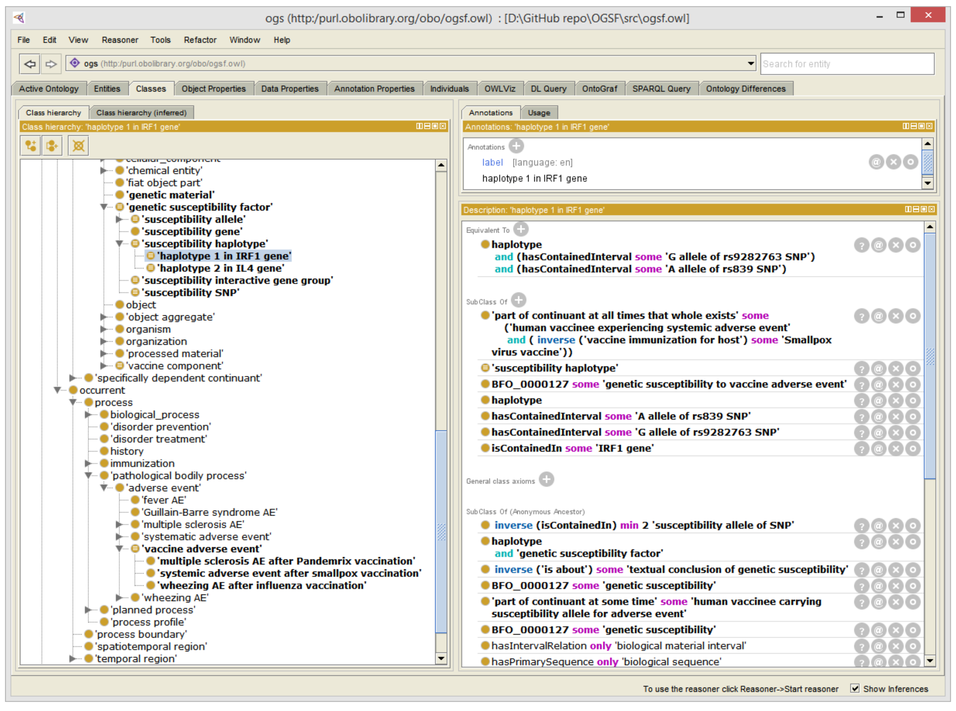
Result

1. Basic high level mode for modeling genetic susceptibility to disease

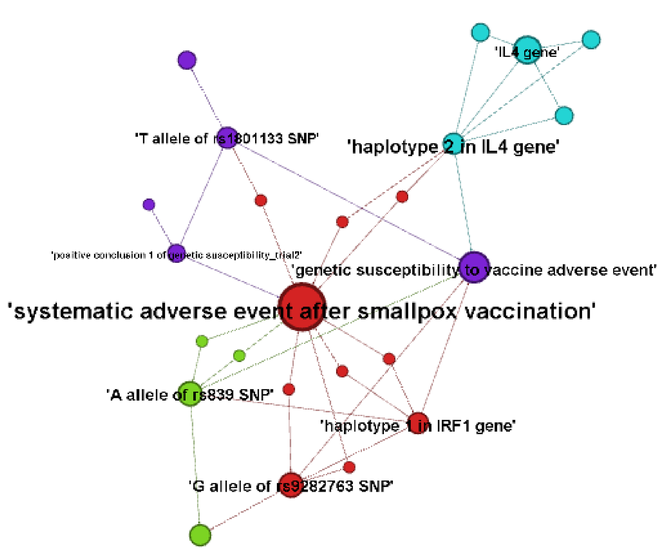


2. Protégé screen shot: define haplotype 1 in IRF1 gene as susceptibility to smallpox systematic adverse event

The haplotype 1 encompassed of two SNPs (rs9282763 and rs839) is contained in a IRF1 gene, which is related with genetic susceptibility to systematic adverse event after smallpox vaccination.

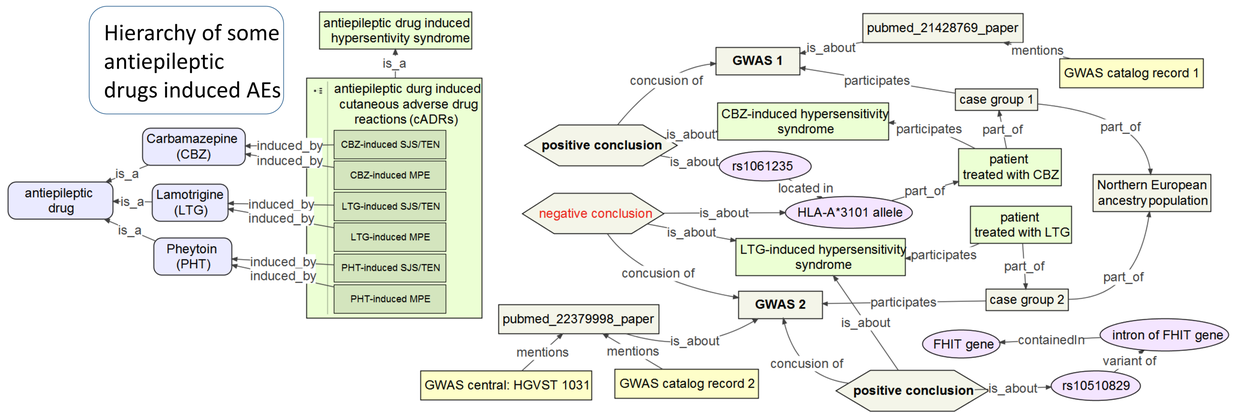


3. Ontology based smallpox vaccination adverse event network analysis



Legend: Genetic factors, such as haplotypes and SNPs associated with systematic adverse events after smallpox vaccination were constructed as a network based on OGSF’s framework. Different colors shows the modularity or community analysis of this network. It is shown that genes and genetic factors are grouped together. Size of the nodes indicates the numbers of connections from the node to other nodes. Since all the nodes are about systematic adverse event after smallpox vaccination, it is the biggest node.

4. Ontological modeling of OGSF shows example to link GWAS data from different studies



Legend: Results from two GWAS studies (PMID:21428769, 22379998) conducted by same group were represented using OGSF’s framework. The GWAS 1 study (PMID:21428769) concluded that a SNP rs1061235, located in HLA-A*3101, is associated with Carbamazepine induced hypersensitivity syndrome, a group of skin and live AEs. However, GWAS 2 study (PMID:22379998) didn’t find any significant HLA-A*3101 SNPs, but 12 other SNPs associated with the Lamotrigine induced hypersensitivity syndrome. One of those 12 SNPs is related with FHIT gene. GWAS 1 study’s statistical result can be found in GWAS catalog database; GWAS 2 study’s result was found both in GWAS catalog database and GWAS central database.

Reference: 1. Lin Y, Sakamoto N: Ontology driven modeling for the knowledge of genetic susceptibility to disease. Kobe J Med Sci 2009, 55(3):E53-66. 2. Lin Y, He Y : The ontology of genetic susceptibility factors (OGSF) and its application in modeling genetic susceptibility to vaccine adverse events Journal of Biomedical Semantics.2014, 5:19