GENE NOMENCLATURE

Sai Sreya Tummala, IUPUI

Genes are the unit of hereditary in humans. They are made up of DNA and account for 20,000-25,000 in number (MedlinePlus, 2021). Gene nomenclature stands for scientific naming of these genes. The need for standardization of gene nomenclature is crucial as they account for improved communication, effective organization, and ease of exchange of biological information (Bruford et al., 2020). Added to this, vocabulary and bibliographic control is established with this process of standardization which can aid in spreading of knowledge. Standardization further expands scope for innovation and also helps in saving a lot of time.

The initial guidelines for naming genes were formed by Human Gene Nomenclature Committee in 1979 (Bruford et al., 2020). Later this committee was placed under Human Genome Organization, and it emerged as HUGO Gene Nomenclature Committee (HGNC) (Bruford et al., 2020). Revisions to the original nomenclature were made in the subsequent years. Till date, HGNC was able to name 40,000 human loci and according to most of the resources, there are 20,000 protein coding genes in human genome (Bruford et al., 2020). HGNC believes that nomenclature of genes is a continuous evolution as it should merge with the emerging technical advancements.

Each gene is assigned a unique symbol, descriptive name, and ID by HGNC. Gene symbols are ideally short, pronounceable, and memorable (Bruford et al., 2020). Symbols are composed of uppercased Latin letters and Arabic numerals (Bruford et al., 2020). Names are in

fact the long descriptions of these symbols and be written in American English. Usually, the names contain the character of function of the genes. The symbol of each gene is unique, and isoforms are not named by HGNC (Bruford et al., 2020). However, in special circumstances, symbols are provided for gene segments located at complex foci. HGNC endorses the usage of italics for genes, alleles and RNAs to distinguish from proteins (Bruford et al., 2020). However, the ongoing research needs for constant changing and adding alternate names and symbols to genes being researched. In order to resolve this ambiguity, a distinctive ID is assigned to each gene and has a format to follow (Example: gene symbol *BRAF*, HGNC ID HGNC:1097) (Bruford et al., 2020). These IDs are based on the gene sequence and do not undergo changes unless the structure is severely altered. This facilitates effective organization of genes in the databanks that ensures reliable information is fetched regardless of nomenclature changes. The database containing these gene symbols and names is maintained by HGNC. This database can be accessed online as search engine containing simple and advanced search options and custom downloads can be performed in the user preferred format (Bruford et al., 2020). The data from this database has enormous importance as it can be used in research to detect trends in diseases, detect new ones and allows for the study of past conditions. This data can also aid in the advancements in the field of pharmacogenomics. It is predicted that HGNC will further improve the website by providing a direct online form to submit new sequences to the database to ensure streamlining of data flow. Advancements in search engines are expected in the future (Eyre et al., 2006).

Standardization of genes helped to study human genome across the globe and led to further improvements. One such result is the promise of personalized medicine with improved

diagnosis, treatment, and prevention of disease (Shoenbill et al., 2013). All this requires incorporation of genetic data into Electronic Health Records. Research has opened gates to include genetic tests in the patient records which help to identify variants responsible for a particular condition in population ((Shoenbill et al., 2013). Accurate and efficient interpretation of data can be done from EHRs using standardized genetic data. This facilitates further scope to develop computerized CDS tools. Health information technology plays a key role in the development of these tools and allow for interpretation and reinterpretation of this genetic data within EHRs promoting personalized medicine (Shoenbill et al., 2013). LOINC standard coupled with genetic nomenclature facilitates interoperability during transitions and variations (Deckard et al., 2015). This facilitates the clinicians to unlock potential information to deliver personalized care (Deckard et al., 2015). Furthermore, this standardization helps to bring improvements in clinical pharmacogenetics by facilitating the ease of reporting and sharing the test results. Applications of genetic nomenclature are paramount in the field of research. Scoping through the database for discoveries is made possible with this standardization. Also, it is made possible to employ machine learning models and artificial intelligence on these codes to detect patterns and discover new diseases. This contributes to improved patient diagnosis and public health. Standardization helps to universally share data across the connect systems fulfilling the goal of Health Information Exchange. Predictive analysis using this standardized nomenclature contributes to the success of CDSS.

Gene nomenclature always holds importance in healthcare as the number of genes being analyzed is an ongoing trend. This in turn draws attention to the databases and the use of

harmonized nomenclature is paramount to extract apt information (Tack et al., 2016). Without a doubt, this standard has a positive impact on the health care as there were improvements in diagnosis and treatment. As mentioned above, personalized medicine was promoted with the aid of this standardization. It has also expanded the scope of research. It also improved the communication and sharing of information among various organizations supporting interoperability.

References:

Bruford, E. A., Braschi, B., Denny, P., Jones, T., Seal, R. L., & Tweedie, S. (2020). Guidelines for human gene nomenclature. *Nature genetics*, *52*(8), 754–758. https://doi.org/10.1038/s41588-020-0669-3

Deckard, J., McDonald, C. J., & Vreeman, D. J. (2015). Supporting interoperability of genetic data with LOINC. *Journal of the American Medical Informatics Association: JAMIA*, *22*(3), 621–627. https://doi.org/10.1093/jamia/ocu012

Eyre, T. A., Ducluzeau, F., Sneddon, T. P., Povey, S., Bruford, E. A., & Lush, M. J. (2006). The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Research*, *34*(suppl_1), D319-D321. https://doi.org/10.1093/nar/gkj147

MedlinePlus. (2021). *What is a gene?* https://medlineplus.gov/genetics/understanding/basics/gene/

Shoenbill, K., Fost, N., Tachinardi, U., & Mendonca, E. A. (2013). Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. *Journal of the American Medical Informatics Association*, *21*(1), 171-180. https://doi.org/10.1136/amiajnl-2013-001694

Tack, V., Deans, Z. C., Wolstenholme, N., Patton, S., & Dequeker, E. M. C. (2016). What's in a Name? A Coordinated Approach toward the Correct Use of a Uniform Nomenclature to Improve Patient Reports and Databases. *Human Mutation*, *37*(6), 570–575. https://doi.org/10.1002/humu.22975