MS12-02 | Darkness in the Human Gene and Protein Function Space Despite Big Omics Data and Decline in Molecular Mechanism Discovery after 2000

Eisenhaber, Birgit (Bioinformatics Institute Singapore, Singapore, SGP); Eisenhaber, Frank (Bioinformatics Institute (BII) A*STAR, Singapore, SGP)

It is generally believed that full human genome sequencing was a watershed event in human history that boosted biomedical research, biomolecular mechanism discovery and life science applications. At the same time, researchers in the field of genome annotation see that there is a persisting, substantial body of functionally insufficiently or completely not characterized genes (for example, ~10,000 protein-coding in the human genome) despite the availability of full genome sequences. A survey of the biomedical literature shows that the number of reported new protein functions had been steadily growing until 2000 but the trend reversed to a dramatic decline thereafter [1,2] when, at the same time, the annual amount of new life science publications doubled between 2000 and 2017. Examples of gene function discovery for chronic myeloid leukaemia and from the GPI lipid biosynthesis pathway [3, 4] show the difficulties and the application potential when molecular functions are really understood.

[1] Darkness in the human gene and protein function space. Sinha et al. Proteomics 2018, 10.1002/pmic.201800093
[2] A decade after the first full human genome sequencing: when will we understand our own genome? Eisenhaber F. JBCB 2012 10.1142/S0219720012710011
[3] Transamidase subunit GAA1/GPAA1 is a M28 family metallo-peptide-synthetase that catalyzes the peptide bond formation between the substrate protein's omega-site and the GPI lipid anchor's phosphoethanolamine. Eisenhaber et al., Cell Cycle. 2014, 10.4161/cc.28761
[4] Function of a membrane-embedded domain evolutionarily multiplied in the GPI lipid anchor pathway proteins PIG-B, PIG-M, PIG-U, PIG-W, PIG-V, and PIG-Z. Eisenhaber et al. Cell Cycle. 2018, 10.1080/15384101.2018.1456294