Session 1-3. 중개연구를 위한

Public database, Data repository의 종류와 활용법

2021.06.28

유태경 / 심형태 / 최무림

Databases are essential in genomics studies

Variant level: Allele frequency, Conservation (Pathogenicity predictions)

Sequencing > variant > pathogenic?

Gene level: Function, Expression, Domain & Structure, Disease association

General

- UCSC Genome Browser (https://genome.ucsc.edu/)
- Uniprot (https://www.uniprot.org/)
- Proteinpainter (https://pecan.stjude.cloud/proteinpaint)

Variant level: Frequency

- IGSR: The International Genome Sample Resource (https://www.internationalgenome.org/home)
- gnomAD (Genome Aggregation Database) (https://gnomad.broadinstitute.org/)
- UK10K (https://www.uk10k.org/)

Variant level: Sequence conservation

- MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/)
- HomoloGene (https://www.ncbi.nlm.nih.gov/homologene)

Gene level: Disease association

- OMIM (https://www.omim.org/)
- ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/)

ClinVar

ACMG STANDARDS AND GUIDELINES | Genetics in Medicine

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Gene level: Expression

- GTEx (Genotype Tissue Expression project) (https://gtexportal.org/home/)
- The Human Protein Atlas (https://www.proteinatlas.org/)
- MGI (Mouse Genome Informatics) (http://www.informatics.jax.org/)

Gene ontology

- DAVID (https://david.ncifcrf.gov/)
- ToppGene (https://toppgene.cchmc.org/)
- EnrichR (https://maayanlab.cloud/Enrichr/)

Databases are essential in genomics studies

Variant level: Conservation(Damaging variants prediction), Frequency

Gene level: Function, Domain & Structure, Expression, Disease association