Supplementary Table 1: Resources on genotype-phenotype relationships. To compute statistics, all resources were accessed on June, 2016

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| name | URL | Scope | Organism | Bulk download | Current Statistics | Original reference | Current reference |
| DisGeNET | http://disgenet.org | Gene-disease, and variant-disease associations | Human | Y | 429,036 associations, between 17,381 genes and 15,093 diseases,  46589 SNPs | 2010 (1) | (2) |
| Comparative Toxicogenomics Database  (CTD) | http://ctdbase.org/ | Chemicals, genes, and disease associations | Human and animal models | Y | 33,532 associations between  8,098 genes and 4,983 diseases | 2003 (3) | (4) |
| Online Mendelian Inheritance in Man  (OMIM) | http://www.omim.org | Mendelian diseases and their genes | Human | Y (with license agreement) | 7,096 associations between 6,676 diseases and 4,864 genes | 1998 (5) | (6) |
| Genetic Association Database\*  (GAD) | http://geneticassociationdb.nih.gov/ | Genes, variants, and complex diseases and traits. | Human | Y | 61,861 associations between 2,753 diseases and 9,280 genes | 2004 (7) | (7) |
| UniProt Knowledgebase | http://www.uniprot.org/ | Proteins | Human (disease annotations) | Y | 3,430 associations between 3,189 diseases and 2,420 genes | 2004 (8) | (9) |
| Orphanet | http://www.orpha.net | Rare diseases and their genes | Human | Y | 6,505 associations between 3,455 diseases and 3,516 genes | 1997 (10) | (11) |
| The Monarch Initiative | http://monarchinitiative.org/ | genes, diseases, and animal models | Human and several animal models | N | NA | 2016 (12) | (12) |
| Rat Genome Database  (RDG) | http://rgd.mcw.edu/ | Rat genes and phenotypes | Rattus norvergicus | Y | 35,035 associations between 5,943 diseases and 5,953  genes | 2002 (13) | (14) |
| Mouse Genome Database  (MGD) | http://www.informatics.jax.org/ | Mouse genes and phenotypes | Mus musculus | Y | 7,752 associations between 4,685 diseases and 5,066 (human or mouse ) genes | 1995 (15) | (16) |
| ClinVar | http://www.ncbi.nlm.nih.gov/clinvar/ | Variants and phenotypes | Human | Y | 18,329 associations between 9479 diseases and 5334 genes  (118,215 variants) | 2013 (17) | (18) |
| The NHGRI-EBI Catalog of published GWAS (GWAS Catalog) | http://www.ebi.ac.uk/gwas/ | GWAS studies | Human | Y | 27,114 associations between 1422  diseases and 12,448 genes (18,666 variants) | 2009 (19) | (20) |
| Database of Genotypes and Phenotypes  (dbGaP) | http://www.ncbi.nlm.nih.gov/gap | GWAS studies | Human | Y | 30,572 variants | 2007 (21) | (22) |
| GeneCards human gene database | http://www.genecards.org/ | Genes | Human | Y (with license agreement) | 10,167 disease genes | 1997 (23) | (24) |
| MalaCards Human Disease Database | http://www.malacards.org/ | Diseases | Human | Y (with license agreement) | NA | 2013 (25) | (26) |
| GWAS Central | http://www.gwascentral.org | GWAS studies | Human | Y\*\* | 69,986,326 associations between 2,974,967 SNPs and 829 diseases and traits | 2000 (27) | (28) |

\* Retired on 2014

\*\* Not all the data, restricted to 1,000 per query

References

1. Bauer-Mehren,A., Rautschka,M., Sanz,F. and Furlong,L.I. (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics*, **26**, 2924–6.

2. Pinero,J., Queralt-Rosinach,N., Bravo,A., Deu-Pons,J., Bauer-Mehren,A., Baron,M., Sanz,F. and Furlong,L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, **2015**, bav028-bav028.

3. Mattingly,C.J., Colby,G.T., Forrest,J.N. and Boyer,J.L. (2003) The Comparative Toxicogenomics Database (CTD). *Environ. Health Perspect.*, **111**, 793–5.

4. Davis,A.P., Grondin,C.J., Lennon-Hopkins,K., Saraceni-Richards,C., Sciaky,D., King,B.L., Wiegers,T.C. and Mattingly,C.J. (2015) The Comparative Toxicogenomics Database’s 10th year anniversary: update 2015. *Nucleic Acids Res.*, **43**, D914-20.

5. McKusick,V.A. (1998) Mendelian inheritance in man: a catalog of human genes and genetic disorders JHU Press.

6. Amberger,J., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2009) McKusick’s Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res*, **37**, D793-6.

7. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–2.

8. Leinonen,R., Diez,F.G., Binns,D., Fleischmann,W., Lopez,R. and Apweiler,R. (2004) UniProt archive. *Bioinformatics*, **20**, 3236–7.

9. The UniProt Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204-212.

10. INSERM (1997) Orphanet: an online database of rare diseases and orphan drugs.

11. Rath,A., Olry,A., Dhombres,F., Brandt,M.M., Urbero,B. and Ayme,S. (2012) Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum. Mutat.*, **33**, 803–8.

12. McMurry,J.A., Köhler,S., Washington,N.L., Balhoff,J.P., Borromeo,C., Brush,M., Carbon,S., Conlin,T., Dunn,N., Engelstad,M., *et al.* (2016) Navigating the Phenotype Frontier: The Monarch Initiative. *Genetics*, **203**.

13. Twigger,S., Lu,J., Shimoyama,M., Chen,D., Pasko,D., Long,H., Ginster,J., Chen,C.-F., Nigam,R., Kwitek,A., *et al.* (2002) Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res.*, **30**, 125–128.

14. Shimoyama,M., De Pons,J., Hayman,G.T., Laulederkind,S.J.F., Liu,W., Nigam,R., Petri,V., Smith,J.R., Tutaj,M., Wang,S.-J., *et al.* (2015) The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res.*, **43**, D743-50.

15. Richardson,J.E., Eppig,J.T. and Nadeau,J.H. (1995) Building an integrated mouse genome database: a view from the front line. *IEEE Eng. Med. Biol. Mag.*, **14**, 718–724.

16. Eppig,J.T., Blake,J.A., Bult,C.J., Kadin,J.A. and Richardson,J.E. (2015) The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.*, **43**, D726-36.

17. Landrum,M.J., Lee,J.M., Riley,G.R., Jang,W., Rubinstein,W.S., Church,D.M. and Maglott,D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980-5.

18. Landrum,M.J., Lee,J.M., Benson,M., Brown,G., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Hoover,J., *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, **44**, D862–D868.

19. Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S. and Manolio,T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 9362–7.

20. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L., *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001-6.

21. Mailman,M.D., Feolo,M., Jin,Y., Kimura,M., Tryka,K., Bagoutdinov,R., Hao,L., Kiang,A., Paschall,J., Phan,L., *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–6.

22. Tryka,K.A., Hao,L., Sturcke,A., Jin,Y., Wang,Z.Y., Ziyabari,L., Lee,M., Popova,N., Sharopova,N., Kimura,M., *et al.* (2014) NCBI’s Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.*, **42**, D975-9.

23. Rebhan,M., Chalifa-Caspi,V., Prilusky,J. and Lancet,D. (1997) GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.*, **13**, 163.

24. Fishilevich,S., Zimmerman,S., Kohn,A., Iny Stein,T., Olender,T., Kolker,E., Safran,M. and Lancet,D. (2016) Genic insights from integrated human proteomics in GeneCards. *Database*, **2016**, baw030.

25. Rappaport,N., Nativ,N., Stelzer,G., Twik,M., Guan-Golan,Y., Stein,T.I., Bahir,I., Belinky,F., Morrey,C.P., Safran,M., *et al.* (2013) MalaCards: an integrated compendium for diseases and their annotation. *Database (Oxford).*, **2013**, bat018.

26. Rappaport,N., Twik,M., Nativ,N., Stelzer,G., Bahir,I., Stein,T.I., Safran,M. and Lancet,D. (2014) MalaCards: A Comprehensive Automatically-Mined Database of Human Diseases. *Curr. Protoc. Bioinformatics*, **47**, 1.24.1-1.24.19.

27. Brookes,A.J., Lehväslaiho,H., Siegfried,M., Boehm,J.G., Yuan,Y.P., Sarkar,C.M., Bork,P. and Ortigao,F. (2000) HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res.*, **28**, 356–60.

28. Beck,T., Hastings,R.K., Gollapudi,S., Free,R.C. and Brookes,A.J. (2014) GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet.*, **22**, 949–52.