

Supplementary Data

Table of Contents

Table S1: Nodes in the OpenBiolink Graph.....	2
Table S2: Ontology Edges in the OpenBioLink Graph.....	2
Table S3: Edges in the OpenBioLink Graph.....	3
Figure S1: Nodes in the OBL Graph.....	4
Figure S2: Number of final in the high quality, directed OpenBioLink Graph.....	5
Table S4: Hyperparameter optimisation results	10
References	11

Node Type	Prefix	ID Source
Anatomy	ANATOMY_	Uberon [1]
GO-Term	GO_	GO [2]
Gene / Protein	GENE_	STRING [3]
Disease	DIS_	DO [4]
Drug	DRUG_	PubChem [5]
Pathway	PATHWAY_	KEGG [6] REACTOME [7]
Phenotype/Side Effect	PHENOTYPE_	HPO [8]

Table S1: Nodes in the OpenBioLink Graph *Provided are the (1) node type, (2) prefix of the id within the OpenBioLink graph, (3) the source of the entity id.*

Node Type	Ontology Edge Types	Source DB
Anatomy	IS_A PART_OF	Uberon [1]
GO-Term	IS_A PART_OF	GO [2]
Disease	IS_A	DO [4]
Phenotype/Side Effect	IS_A	HPO [8]

Table S2: Ontology Edges in the OpenBioLink Graph *Provided are the (1) entity type of the ontology, (2) which kind of relationships are present as well as (3) the source data base*

Node Type	Relations	Relation ID	Directional	Source DB	TN	Score
Gene – Go	gene_interaction_go	GENE_GO	yes	GO [2]	no	yes
Gene – Disease	gene_interaction_disease	GENE_DIS	yes	DO [4]	no	yes
Gene – Phenotype	gene_interaction_phenotype	GENE_PHENOTYPE	yes	HPO [8]	no	no
Gene – Pathway	gene_interaction_pathway	GENE_PATHWAY	yes	CTD [9]	no	no
Protein – Protein	protein_interaction_protein protein_activation_protein protein_binding_protein protein_catalysis_protein protein_expression_protein protein_inhibition_protein protein_bindingAndActivating_protein protein_bindingAndInhibiting_protein protein_reaction_protein protein_postTranslationalModifications_protein	GENE_GENE GENE_ACTIVATION_GENE GENE_BINDING_GENE GENE_CATALYSIS_GENE GENE_EXPRESSION_GENE GENE_INHIBITION_GENE GENE_BINDACT_GENE GENE_BINDINH_GENE GENE_REACTION_GENE GENE_PTMOD_GENE	no yes no yes yes yes yes yes yes yes no	STRING [3]	no	yes
Drug – Protein	drug_interaction_protein drug_activation_protein drug_binding_protein drug_catalysis_protein drug_expression_protein drug_inhibition_protein drug_bindingAndActivating_protein drug_bindingAndInhibiting_protein drug_reaction_protein drug_predictedBinding_protein	GENE_DRUG DRUG_ACTIVATION_GENE DRUG_BINDING_GENE DRUG_CATALYSIS_GENE DRUG_EXPRESSION_GENE DRUG_INHIBITION_GENE DRUG_BINDACT_GENE DRUG_BINDINH_GENE DRUG_REACTION_GENE DRUG_PREDBIND_GENE	yes	STITCH [10]	no	yes
Gene – Anatomy	gene_expression_anatomy gene_underexpression_anatomy gene_overexpression_anatomy	GENE_EXPRESSED_ANATOMY GENE_UNDEREXPRESSED_ANATOMY GENE_OVEREXPRESSED_ANATOMY	yes	Bgee [11]	yes	yes
Disease – Drug	disease_indication_drug	DIS_DRUG	yes	DrugCentral [12]	yes	no
Disease – Phenotype	disease_associated_phenotype	DIS_PHENOTYPE	yes	HPO [8]	yes	yes
Drug – SideEffect	drug_associatd_sideEffect	DRUG_PHENOTYPE	yes	SIDER [13]	no	no

Table S3: Edges in the OpenBioLink Graph *Provided are (1) the types of the nodes of an edge type, (2) the relation types, (3) the id of the relation type within the OpenBioLink graph, (4) whether the relation type is directional, (5) the source database of the relation type, (6) whether the source provides corresponding true negative edges, and (7) whether the source provides a confidence score.*

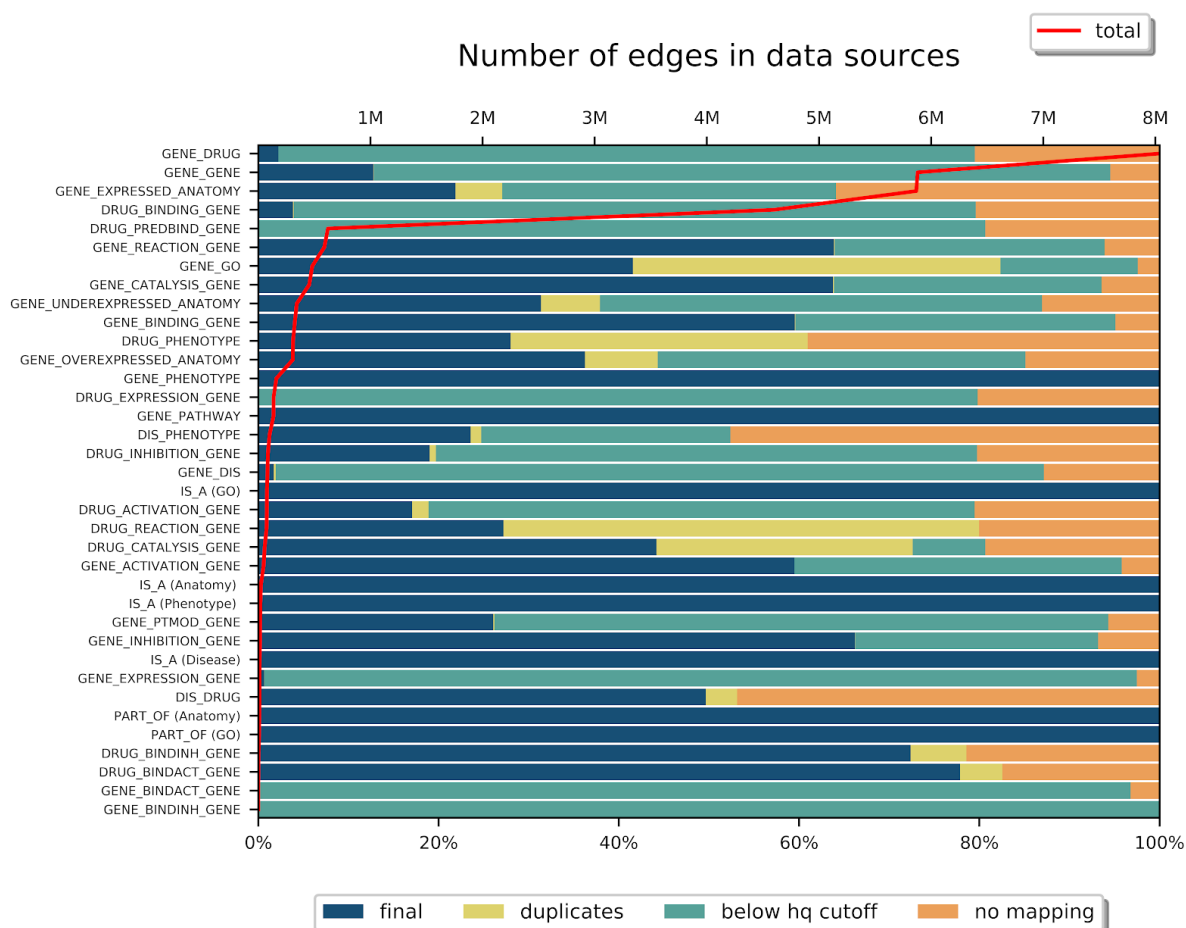


Figure S1: Pre-processing of Edges in the OBL Graph The stacked horizontal bar chart shows the percentage (bottom x-axis) of edges being discarded during the pre-processing steps of the high quality, directed OpenBioLink Graph due to lack of available mappings, quality constraints or duplication. The edge types are sorted by total number before pre-processing (red line and upper x-axis, M =million).

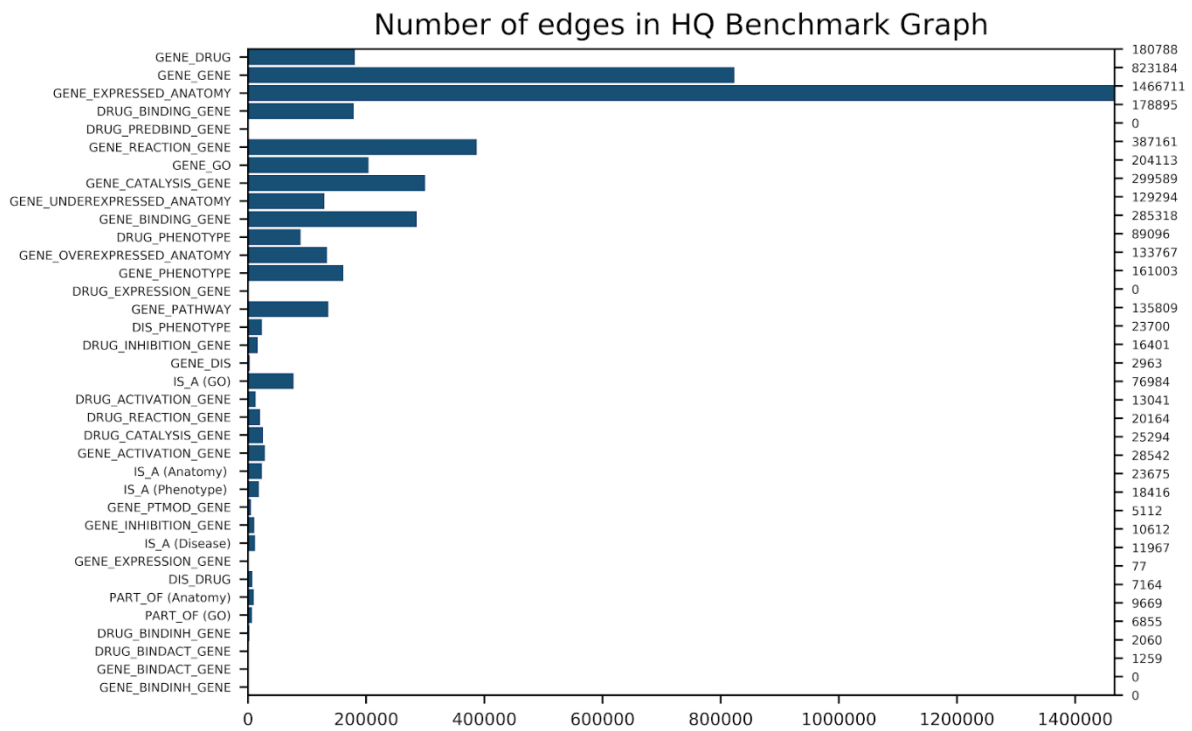


Figure S2: Number of Edges in the OpenBioLink Graph *Provided is the final number of edges per edge type in the directed, high quality OpenBioLink graph, after mapping, quality cutoff and removal of duplicates.*

Model configuration	hits@1	hits@3	hits@5	hits@10	Mean rank	Emb. dim	Margin loss	Learning rate	Epochs	Scoring function	Relation embed. dimension	Norm of entities
transe_0_0	0.0072	0.0171	0.0394	0.0519	8,111.94	50	0.5	0.1	1000	2	-	1
transe_0_1	0.0157	0.0633	0.0925	0.1353	3,721.17	100	2	0.1	1000	1	-	2
transe_0_10	0.0111	0.0365	0.0506	0.0811	5,530.92	50	0.5	0.01	1000	1	-	1
transe_0_11	0.0127	0.0311	0.0436	0.0713	12,705.00	20	2	0.01	1000	2	-	1
transe_0_12	0.0155	0.0360	0.0526	0.0651	6,714.60	100	5	0.1	1000	1	-	1
transe_0_13	0.0069	0.0255	0.0429	0.0657	17,794.95	30	0.5	0.01	1000	1	-	2
transe_0_14	0.0152	0.0459	0.0641	0.0902	10,309.39	50	1	0.01	1000	1	-	2
transe_0_15	0.0273	0.0589	0.0793	0.1112	6,717.38	30	0.5	0.1	1000	2	-	2
transe_0_16	0.0066	0.0235	0.0405	0.0728	8,343.33	50	0.5	0.01	1000	2	-	1
transe_0_17	0.0310	0.0484	0.0581	0.0712	7,302.83	100	15	0.1	1000	2	-	2
transe_0_18	0.0272	0.0461	0.0563	0.0655	5,798.28	30	10	0.1	1000	1	-	1
transe_0_19	0.0066	0.0235	0.0405	0.0728	8,343.33	50	5	0.01	1000	2	-	1
transe_0_2	0.0205	0.0417	0.0610	0.0809	4,252.08	20	1	0.1	1000	1	-	1
transe_0_3	0.0045	0.0189	0.0393	0.0549	7,634.75	100	0.5	0.1	1000	2	-	1
transe_0_4	0.0072	0.0171	0.0394	0.0519	8,111.94	50	5	0.1	1000	2	-	1
transe_0_5	0.0246	0.0409	0.0511	0.0649	7,836.92	50	2	0.1	1000	2	-	2
transe_0_6	0.0210	0.0494	0.0684	0.0879	28,267.83	50	0.5	0.01	1000	2	-	2
transe_0_7	0.0193	0.0403	0.0513	0.0639	7,257.09	30	2	0.01	1000	1	-	1
transe_0_8	0.0188	0.0322	0.0429	0.0566	8,776.40	20	2	0.1	1000	2	-	2
transe_0_9	0.0150	0.0319	0.0508	0.0643	5,972.00	50	10	0.1	1000	1	-	1
transe_1_0	0.0159	0.0343	0.0436	0.0661	6,452.61	100	15	0.01	1000	1	-	1
transe_1_1	0.0197	0.0355	0.0459	0.0548	8,021.54	20	15	0.1	1000	1	-	2
transe_1_10	0.0192	0.0397	0.0503	0.0629	7,257.53	30	10	0.01	1000	1	-	1

transe_1_11	0.0072	0.0171	0.0394	0.0519	8,111.94	50	10	0.1	1000	2	-	1
transe_1_12	0.0106	0.0486	0.0594	0.0777	7,019.38	100	15	0.01	1000	2	-	1
transe_1_13	0.0147	0.0478	0.0709	0.0918	25,958.34	100	10	0.01	1000	2	-	2
transe_1_14	0.0050	0.0198	0.0335	0.0589	16,729.57	20	1	0.01	1000	1	-	2
transe_1_15	0.0194	0.0374	0.0494	0.0608	7,521.01	30	15	0.1	1000	1	-	2
transe_1_16	0.0147	0.0478	0.0709	0.0918	25,958.34	100	15	0.01	1000	2	-	2
transe_1_17	0.0140	0.0336	0.0437	0.0669	6,133.23	50	10	0.01	1000	1	-	1
transe_1_18	0.0272	0.0461	0.0563	0.0655	5,798.28	30	15	0.1	1000	1	-	1
transe_1_19	0.0145	0.0403	0.0645	0.0845	36,530.95	20	1	0.01	1000	2	-	2
transe_1_2	0.0203	0.0379	0.0464	0.0594	8,356.50	30	10	0.1	1000	2	-	2
transe_1_3	0.0045	0.0189	0.0393	0.0549	7,634.75	100	1	0.1	1000	2	-	1
transe_1_4	0.0272	0.0461	0.0563	0.0655	5,798.28	30	5	0.1	1000	1	-	1
transe_1_5	0.0177	0.0401	0.0538	0.0855	3,994.82	30	1	0.1	1000	1	-	1
transe_1_6	0.0148	0.0395	0.0603	0.0992	6,709.35	30	1	0.1	1000	2	-	2
transe_1_7	0.0106	0.0486	0.0594	0.0777	7,019.38	100	1	0.01	1000	2	-	1
transe_1_8	0.0246	0.0409	0.0511	0.0649	7,836.92	50	15	0.1	1000	2	-	2
transe_1_9	0.0045	0.0189	0.0393	0.0549	7,634.75	100	5	0.1	1000	2	-	1
transe_2_0	0.0161	0.0429	0.0644	0.1058	6,268.69	50	1	0.1	1000	2	-	2
transe_2_1	0.0066	0.0235	0.0405	0.0728	8,343.33	50	2	0.01	1000	2	-	1
transe_2_10	0.0150	0.0319	0.0508	0.0643	5,972.00	50	15	0.1	1000	1	-	1
transe_2_11	0.0144	0.0397	0.0562	0.0789	12,917.90	20	10	0.01	1000	1	-	2
transe_2_12	0.0159	0.0343	0.0436	0.0661	6,452.61	100	5	0.01	1000	1	-	1
transe_2_13	0.0032	0.0096	0.0346	0.0466	8,844.22	30	10	0.1	1000	2	-	1
transe_2_14	0.0089	0.0341	0.0697	0.1052	4,984.02	30	1	0.1	1000	1	-	2

transe_2_15	0.0232	0.0413	0.0494	0.0614	8,588.43	20	2	0.01	1000	1	-	1
transe_2_16	0.0161	0.0501	0.0757	0.0987	6,043.79	100	5	0.01	1000	1	-	2
transe_2_17	0.0031	0.0096	0.0347	0.0465	8,844.16	30	0.5	0.1	1000	2	-	1
transe_2_18	0.0188	0.0322	0.0429	0.0566	8,776.40	20	10	0.1	1000	2	-	2
transe_2_19	0.0300	0.0412	0.0517	0.0616	6,074.49	20	10	0.1	1000	1	-	1
transe_2_2	0.0066	0.0235	0.0405	0.0728	8,343.33	50	15	0.01	1000	2	-	1
transe_2_20	0.0157	0.0353	0.0433	0.0658	5,874.67	100	1	0.01	1000	1	-	1
transe_2_3	0.0124	0.0553	0.0848	0.1297	3,549.62	100	5	0.1	1000	1	-	2
transe_2_4	0.0192	0.0397	0.0503	0.0629	7,257.53	30	15	0.01	1000	1	-	1
transe_2_5	0.0122	0.0348	0.0607	0.0812	37,016.86	20	2	0.01	1000	2	-	2
transe_2_6	0.0163	0.0378	0.0442	0.0670	5,877.85	50	1	0.01	1000	1	-	1
transe_2_7	0.0119	0.0274	0.0373	0.0610	8,289.58	20	0.5	0.01	1000	1	-	1
transe_2_8	0.0034	0.0112	0.0179	0.0463	9,120.77	20	2	0.1	1000	2	-	1
transe_2_9	0.0154	0.0533	0.0782	0.1064	5,096.75	50	0.5	0.1	1000	1	-	2
transe_3_0	0.0024	0.0110	0.0194	0.0368	22,135.14	20	0.5	0.01	1000	1	-	2
transe_3_1	0.0233	0.0414	0.0496	0.0612	8,587.03	20	15	0.01	1000	1	-	1
transe_3_10	0.0034	0.0112	0.0179	0.0463	9,120.77	20	5	0.1	1000	2	-	1
transe_3_11	0.0045	0.0189	0.0393	0.0549	7,634.75	100	15	0.1	1000	2	-	1
transe_3_12	0.0140	0.0433	0.0686	0.0886	29,042.81	50	15	0.01	1000	2	-	2
transe_3_13	0.0213	0.0643	0.0818	0.1129	4,425.61	100	0.5	0.1	1000	1	-	2
transe_3_14	0.0271	0.0428	0.0537	0.0669	6,605.98	100	15	0.1	1000	1	-	2
transe_3_15	0.0109	0.0416	0.0532	0.0841	7,950.58	50	5	0.01	1000	1	-	2
transe_3_16	0.0233	0.0414	0.0496	0.0612	8,587.03	20	10	0.01	1000	1	-	1
transe_3_17	0.0106	0.0486	0.0594	0.0777	7,019.38	100	2	0.01	1000	2	-	1

transe_3_18	0.0192	0.0516	0.0669	0.0890	13,517.98	50	0.5	0.01	1000	1	-	2
transe_3_19	0.0110	0.0412	0.0782	0.1170	4,504.40	50	1	0.1	1000	1	-	2
transe_3_2	0.0089	0.0330	0.0622	0.1033	5,038.81	20	2	0.1	1000	1	-	2
transe_3_20	0.0122	0.0348	0.0607	0.0812	37,016.86	20	5	0.01	1000	2	-	2
transe_3_3	0.0300	0.0412	0.0517	0.0616	6,074.49	20	15	0.1	1000	1	-	1
transe_3_4	0.0122	0.0348	0.0607	0.0812	37,016.86	20	10	0.01	1000	2	-	2
transe_3_5	0.0114	0.0474	0.0702	0.1053	4,056.15	50	5	0.1	1000	1	-	2
transe_3_6	0.0048	0.0251	0.0436	0.0782	10,351.99	30	10	0.01	1000	2	-	1
transe_3_7	0.0062	0.0277	0.0651	0.0973	5,463.06	20	1	0.1	1000	1	-	2
transe_3_8	0.0203	0.0379	0.0464	0.0594	8,356.50	30	15	0.1	1000	2	-	2
transe_3_9	0.0131	0.0535	0.0751	0.1217	4,144.90	50	2	0.1	1000	1	-	2
transr_0_0	0.0021	0.0215	0.0383	0.0762	3,139.93	100	10	0.1	1000	1	50	-
transr_0_1	0.0007	0.0102	0.0194	0.0411	5,383.00	20	10	0.01	1000	1	50	-
transr_0_2	0.0150	0.0481	0.0711	0.1130	3,042.02	50	10	0.1	1000	1	50	-
transr_0_3	0.0210	0.0580	0.0825	0.1237	2,848.24	50	5	0.1	1000	1	100	-
transr_0_4	0.0192	0.0580	0.0817	0.1245	2,638.40	50	15	0.1	1000	1	100	-
transr_0_5	0.0133	0.0404	0.0617	0.0960	4,450.07	30	0.5	0.1	1000	1	100	-
transr_0_6	0.0044	0.0249	0.0454	0.0844	4,415.55	20	1	0.1	1000	1	50	-
transr_1_0	0.0119	0.0426	0.0680	0.1047	3,864.74	50	1	0.1	1000	1	50	-
transr_1_1	0.0003	0.0097	0.0191	0.0424	6,937.76	30	0.5	0.01	1000	1	50	-
transr_1_2	0.0062	0.0271	0.0468	0.0863	5,601.22	20	0.5	0.1	1000	1	50	-
transr_1_3	0.0242	0.0607	0.0842	0.1262	2,696.74	50	10	0.1	1000	1	100	-
transr_1_4	0.0004	0.0100	0.0195	0.0429	6,648.55	30	1	0.01	1000	1	50	-
transr_1_5	0.0003	0.0168	0.0326	0.0689	4,840.72	100	5	0.01	1000	1	50	-

transr_1_6	0.0095	0.0378	0.0612	0.1006	4,235.83	50	0.5	0.1	1000	1	50	-
transr_2_0	0.0087	0.0351	0.0570	0.0989	3,242.81	20	5	0.1	1000	1	50	-
transr_2_1	0.0237	0.0575	0.0799	0.1206	2,526.47	30	15	0.1	1000	1	100	-
transr_2_2	0.0003	0.0230	0.0428	0.0839	3,624.07	100	1	0.01	1000	1	100	-
transr_2_3	0.0012	0.0152	0.0293	0.0632	4,120.07	30	15	0.01	1000	1	100	-
transr_2_4	0.0010	0.0142	0.0277	0.0601	5,997.54	30	1	0.01	1000	1	100	-
transr_2_5	0.0010	0.0212	0.0407	0.0836	3,765.19	50	10	0.01	1000	1	100	-
transr_3_0	0.0099	0.0366	0.0595	0.1034	2,864.18	20	5	0.1	1000	1	100	-
transr_3_1	0.0006	0.0132	0.0263	0.0584	5,196.38	50	1	0.01	1000	1	50	-
transr_3_2	0.0139	0.0489	0.0670	0.0980	2,814.09	100	1	0.1	1000	1	100	-
transr_3_3	0.0171	0.0462	0.0679	0.1030	3,930.74	50	0.5	0.1	1000	1	100	-
transr_3_4	0.0151	0.0444	0.0660	0.1008	3,840.52	30	1	0.1	1000	1	100	-

Table S4: Hyperparameter Optimisation Results *Random hyperparameter search with a minor adapted version of pyKEEN [14] for outputting all configuration results was performed for 10 days on 8 GTX1080Ti GPUs (4 GPUs per model type) in unfiltered setting. A ratio of 90:10 was chosen for the train-test-split. Scoring functions and normalisation of entities 1 and 2 correspond to L1 and L2, respectively. Note that pyKEEN performs random splits for train-test splits without removal of trivial edges, therefore, results are likely to be higher than those of OpenBioLink.*

References

- [1] C. J. Mungall, C. Torniai, G. V Gkoutos, S. E. Lewis, and M. A. Haendel, “Uberon , an integrative multi-species anatomy ontology,” *Genome Biol.*, pp. 1–20, 2012.
- [2] The Gene Ontology Consortium, “The Gene Ontology Resource : 20 years and still GOing strong,” *Nucleic Acids Res.*, vol. 47, no. November 2018, pp. 330–338, 2019.
- [3] C. Von Mering *et al.*, “STRING : known and predicted protein-protein associations , integrated and transferred across organisms,” *Nucleic Acids Res.*, vol. 33, pp. 433–437, 2005.
- [4] L. M. Schriml *et al.*, “Human Disease Ontology 2018 update : classification , content and workflow expansion,” *Nucleic Acids Res.*, vol. 47, pp. 955–962, 2019.
- [5] S. Kim *et al.*, “PubChem 2019 update : improved access to chemical data,” vol. 47, no. October 2018, pp. 1102–1109, 2019.
- [6] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, and M. Tanabe, “New approach for understanding genome variations in KEGG,” vol. 47, no. October 2018, pp. 590–595, 2019.
- [7] A. Fabregat *et al.*, “The Reactome Pathway Knowledgebase,” vol. 46, no. November 2017, pp. 649–655, 2018.
- [8] K. Sebastian *et al.*, “Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources,” *Nucleic Acids Res.*, vol. 47, no. November 2018, pp. 1018–1027, 2019.
- [9] A. P. Davis *et al.*, “The Comparative Toxicogenomics Database : update 2019,” *Nucleic Acids Res.*, vol. 47, pp. 948–954, 2019.
- [10] D. Szklarczyk, A. Santos, C. Von Mering, L. J. Jensen, P. Bork, and M. Kuhn, “STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D380–D384, 2016.
- [11] F. Bastian *et al.*, “Bgee : Integrating and Comparing Heterogeneous Transcriptome Data Among Species,” *Springer-Verlag Berlin Heidelb.*, pp. 124–131, 2008.
- [12] O. Ursu *et al.*, “DrugCentral 2018 : an update,” *Nucleic Acids Res.*, vol. 47, pp. 963–970, 2019.
- [13] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, “The SIDER database of drugs and side effects,” *Nucleic Acids Res.*, vol. 44, no. October 2015, pp. 1075–1079, 2016.
- [14] M. Ali, C. T. Hoyt, D. Domingo-Fernández, J. Lehmann, and H. Jabeen, “BioKEEN : A library for learning and evaluating biological knowledge graph embeddings,” *Bioinformatics*, vol. 35, no. 18, pp. 3538–3540, 2019.