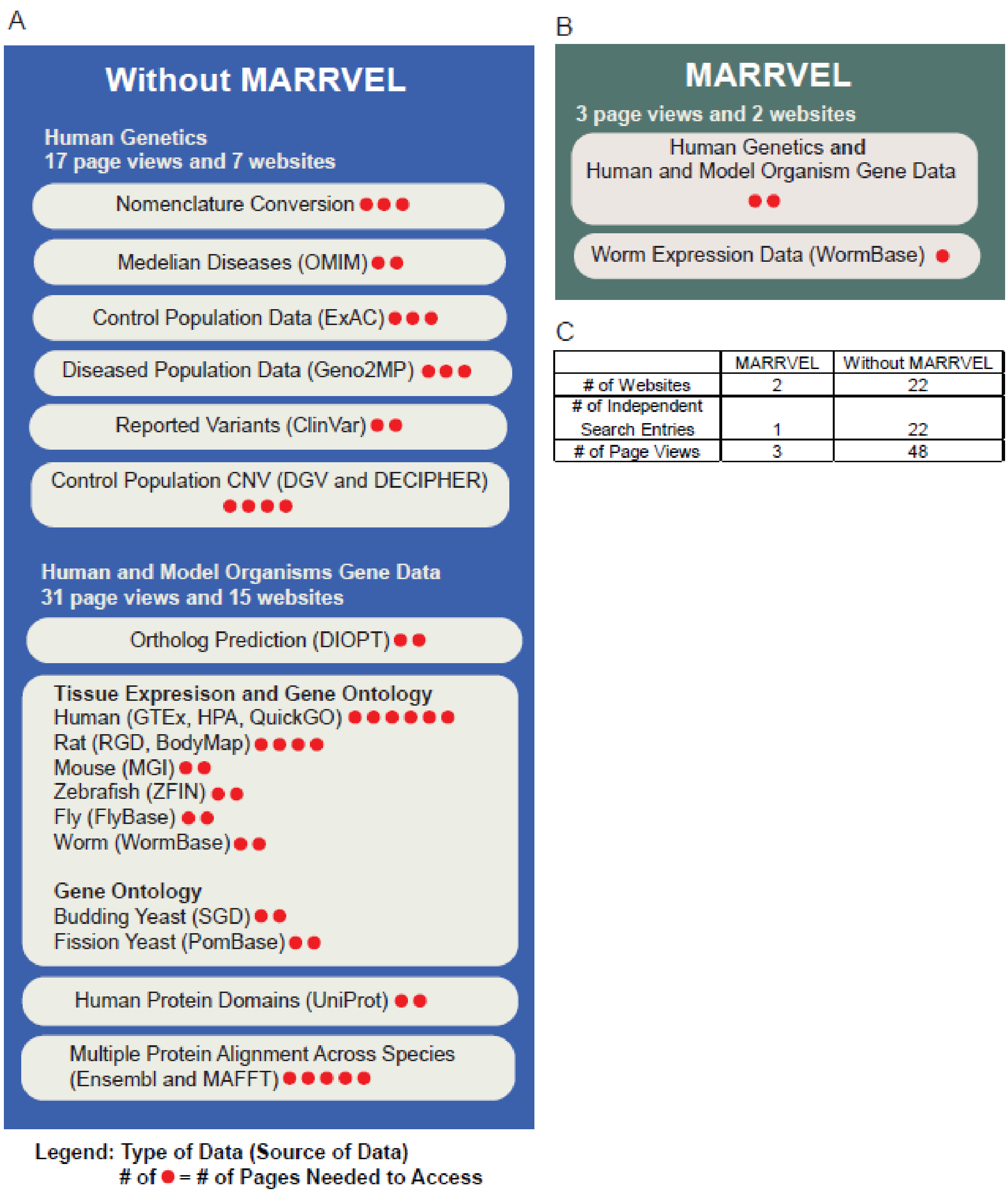
The American Journal of Human Genetics, Volume *100*

Supplemental Data

MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome

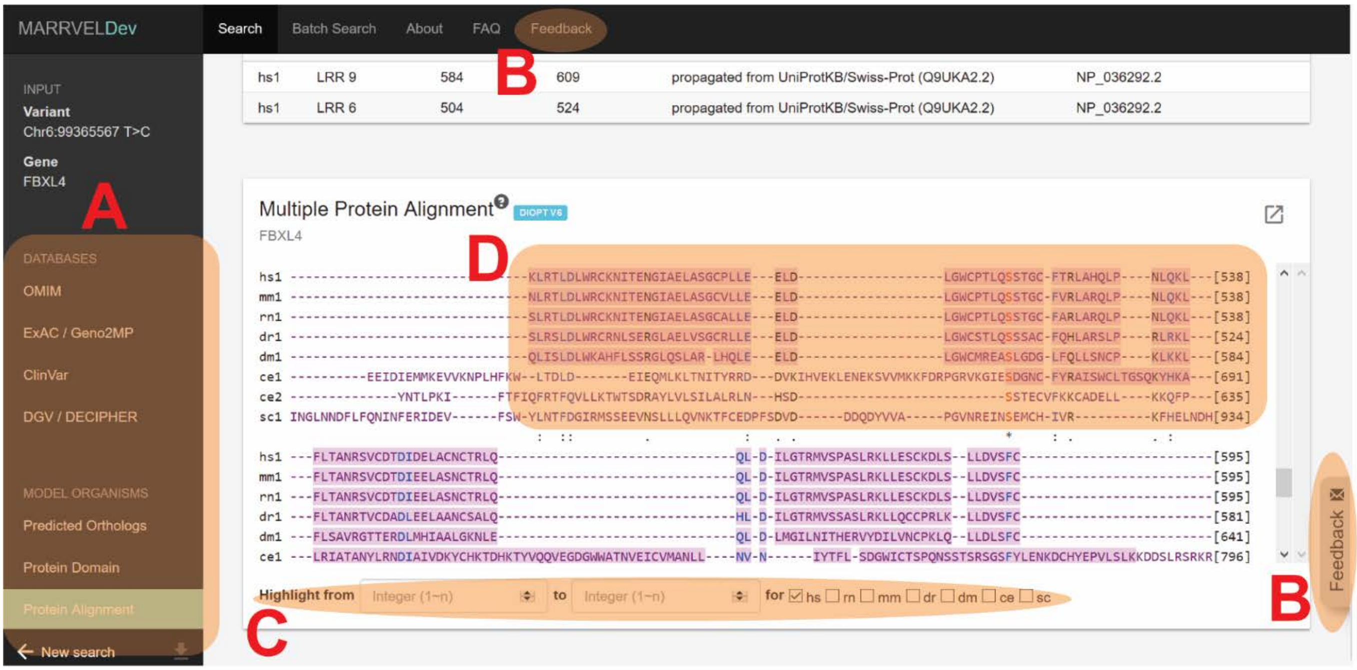
Julia Wang, Rami Al-Ouran, Yanhui Hu, Seon-Young Kim, Ying-Wooi Wan, Michael F. Wangler, Shinya Yamamoto, Hsiao-Tuan Chao, Aram Comjean, Stephanie E. Mohr, UDN, Norbert Perrimon, Zhandong Liu, and Hugo J. Bellen


A
Without MARRVEL
Human Genetics
17 page views and 7 websites
Nomenclature Conversion ●●●
Medelian Diseases (OMIM) ●●
Control Population Data (ExAC) ●●●
Diseased Population Data (Geno2MP) ●●●
Reported Variants (ClinVar) ●●
Control Population CNV (DGV and DECIPHER) ●●●●
Human and Model Organisms Gene Data
31 page views and 15 websites
Ortholog Prediction (DIOPT) ●●
Tissue Expresison and Gene Ontology
Human (GTEx, HPA, QuickGO) ●●●●●●
Rat (RGD, BodyMap) ●●●●
Mouse (MGI) ●●
Zebrafish (ZFIN) ●●
Fly (FlyBase) ●●
Worm (WormBase) ●●
Gene Ontology
Budding Yeast (SGD) ●●
Fission Yeast (PomBase) ●●
Human Protein Domains (UniProt) ●●
Multiple Protein Alignment Across Species (Ensembl and MAFFT) ●●●●●
B
MARRVEL
3 page views and 2 websites
Human Genetics and Human and Model Organism Gene Data ●●
Worm Expression Data (WormBase) ●
C
Legend: Type of Data (Source of Data)
# of ● = # of Pages Needed to Access

	MARRVEL	Without MARRVEL
# of Websites	2	22
# of Independent Search Entries	1	22
# of Page Views	3	48

Figure S1: Comparison of a one-by-one search of databases vs MARRVEL
A) Multiple databases contain useful data for gene and variant analysis. However, obtaining each piece of data requires navigation throughout multiple websites. **B and C)** MARRVEL aggregates useful data from public databases for variant and gene analysis. Aggregation of information across multiple databases greatly facilitates data analysis and provides a platform for integrating the accumulated knowledge in human genetics and model organism research.

OMIM, Online Mendelian Inheritance in Man; ExAC, The Exome Aggregation Consortium; Geno2MP, Genotype to Mendelian Phenotype; DGV, Database of Genomic Variants; DECIPHER, DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources; DIOPT, DRSC Integrative Ortholog Prediction Tool; SGD, Saccharomyces Genome Database; MGI, Mouse Genome Informatics; HPA, Human Protein Atlas



INPUT

Variant
Chr6:99365567 T>C

Gene
FBXL4

hs1	LRR 9	584	609	propagated from UniProtKB/Swiss-Prot (Q9UKA2.2)	NP_036292.2
hs1	LRR 6	504	524	propagated from UniProtKB/Swiss-Prot (Q9UKA2.2)	NP_036292.2

Multiple Protein Alignment

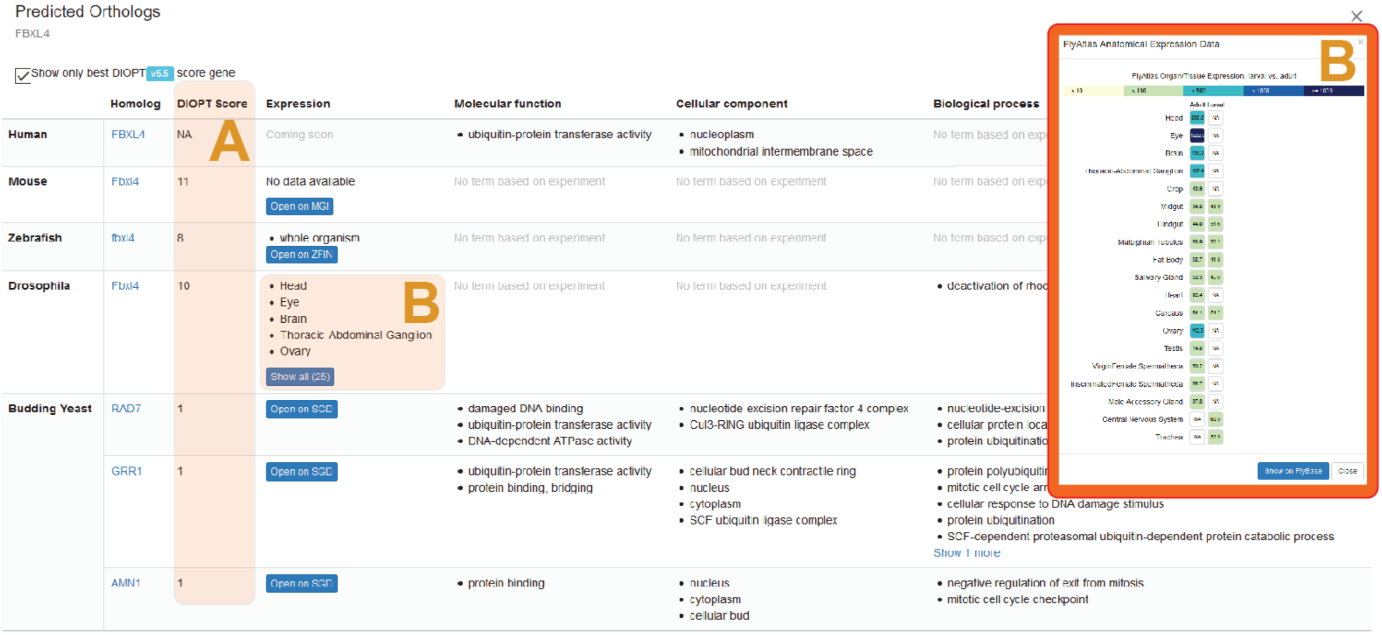
FBXL4

```
hs1 -------------------------------KLRTLDLWRCKNITENGIAELASGCPLLE---ELD-------------------LGWCPTLQSSTGC-FTRLAHQLP----NLQKL---[538]
mm1 -------------------------------NLRTLDLWRCKNITENGIAELASGCVLLE---ELD-------------------LGWCPTLQSSTGC-FVRLARQLP----NLQKL---[538]
rn1 -------------------------------SLRTLDLWRCKNITENGIAELASGCALLE---ELD-------------------LGWCPTLQSSTGC-FARLARQLP----NLQKL---[538]
dr1 -------------------------------SLRSLDLWRCRNLSERGLAELVSGCRLLE---ELD-------------------LGWCSTLQSSSAC-FQHLARSLP----RLRKL---[524]
dm1 -------------------------------QLISLDLWKAHFLSSRGLQSLAR-LHQLE---ELD-------------------LGWCMREASLGDG-LFQLLSNCP----KLKKL---[584]
ce1 ----------EEIDIEMMKEVVKNPLHFKW--LTDLD-------EIEQMLKLTNITYRRD---DVKIHVEKLENEKSVVMKKFDRPGRVKGIESDGNC-FYRAISWCLTGSQKYHKA---[691]
ce2 --------------YNTLPKI------FTFIQFRTFQVLLKTWTSDRAYLVLSILALRLN---HSD---------------------------SSTECVFKKCADELL----KKQFP---[635]
sc1 INGLNNDFLFQNINFERIDEV------FSW-YLNTFDGIRMSSEEVNSLLLQVNKTFCEDPFSDVD------DDQDYVVA-----PGVNREINSEMCH-IVR----------KFHELNDH[934]
                                    :  ::         .            :   . .                           *     : .          . :

hs1 ---FLTANRSVCDTDIDELACNCTRLQ-------------------------------QL-D-ILGTRMVSPASLRKLLESCKDLS--LLDVSFC-------------------------[595]
mm1 ---FLTANRSVCDTDIEELASNCTRLQ-------------------------------QL-D-ILGTRMVSPASLRKLLESCKDLS--LLDVSFC-------------------------[595]
rn1 ---FLTANRSVCDTDIEELASNCTRLQ-------------------------------QL-D-ILGTRMVSPASLRKLLESCKDLS--LLDVSFC-------------------------[595]
dr1 ---FLTANRTVCDADLEELAANCSALQ-------------------------------HL-D-ILGTRMVSSASLRKLLQCCPRLK--LLDVSFC-------------------------[581]
dm1 ---FLSAVRGTTERDLMHIAALGKNLE-------------------------------QL-D-LMGILNITHERVYDILVNCPKLQ--LLDLSFC-------------------------[641]
ce1 ---LRIATANYLRNDIAIVDKYCHKTDHKTYVQQVEGDGWWATNVEICVMANLL----NV-N------IYTFL-SDGWICTSPQNSSTSRSGSFYLENKDCHYEPVLSLKKDDSLRSRKR[796]
```

Highlight from Integer (1~n) to Integer (1~n) for ☑ hs ☐ rn ☐ mm ☐ dr ☐ dm ☐ ce ☐ sc

Figure S2: Overall navigation and multiple protein alignments on MARRVEL
A: The navigation menu on the left allows for a quick jump to a dataset of interest
B: The feedback function is built into each output page to encourage users to submit bug reports and/or suggestions for future updates.
C: The highlight function for the multiple protein alignment allows for quick assessment of the conservation of an amino acid or functional domain of interest
D: Predicted functional domains are highlighted in pink.



Predicted Orthologs
FBXL4

☑ Show only best DIOPT v5.5 score gene

	Homolog	DIOPT Score	Expression	Molecular function	Cellular component	Biological process
Human	FBXL4	NA	Coming soon	• ubiquitin-protein transferase activity	• nucleoplasm • mitochondrial intermembrane space	No term based on exp
Mouse	Fbxl4	11	No data available Open on MGI	No term based on experiment	No term based on experiment	No term based on exp
Zebrafish	fbxl4	8	• whole organism Open on ZFIN	No term based on experiment	No term based on experiment	No term based on exp
Drosophila	Fbxl4	10	• Head • Eye • Brain • Thoracic Abdominal Ganglion • Ovary Show all (25)	No term based on experiment	No term based on experiment	• deactivation of rhod
Budding Yeast	RAD7	1	Open on SGD	• damaged DNA binding • ubiquitin-protein transferase activity • DNA-dependent ATPase activity	• nucleotide excision repair factor 4 complex • Cul3-RING ubiquitin ligase complex	• nucleotide-excision • cellular protein loca • protein ubiquitinatio
	GRR1	1	Open on SGD	• ubiquitin-protein transferase activity • protein binding, bridging	• cellular bud neck contractile ring • nucleus • cytoplasm • SCF ubiquitin ligase complex	• protein polyubiquitin • mitotic cell cycle arr • cellular response to DNA damage stimulus • protein ubiquitination • SCF-dependent proteasomal ubiquitin-dependent protein catabolic process Show 1 more
	AMN1	1	Open on SGD	• protein binding	• nucleus • cytoplasm • cellular bud	• negative regulation of exit from mitosis • mitotic cell cycle checkpoint

FlyAtlas Anatomical Expression Data

FlyAtlas Organ/Tissue Expression, larval vs. adult

< 10 | < 100 | < 500 | < 1000 | >= 1000

	Adult	Larval
Head	335.8	NA
Eye	[illegible]	NA
Brain	133.3	NA
Thoracic-Abdominal Ganglion	165.4	NA
Crop	40.9	NA
Midgut	34.8	43.9
Hindgut	44.8	51.6
Malpighian Tubules	58.9	70.2
Fat Body	52.7	11.8
Salivary Gland	50.3	40.0
Heart	82.4	NA
Carcass	64.1	64.2
Ovary	112.0	NA
Testis	14.8	NA
VirginFemale Spermatheca	59.2	NA
InseminatedFemale Spermatheca	66.7	NA
Male Accessory Gland	87.8	NA
Central Nervous System	NA	60.0
Trachea	NA	53.5

Show on FlyBase | Close

Figure S3: Demonstration of the model organism data section of MARRVEL
A summary of human and model organism gene function information is displayed in a table. The human protein domains are also listed, along with a protein alignment of the human gene with putative orthologs in model organisms. An example of tissue expression data for an ortholog of a human gene in *Drosophila* is shown.
A: DIOPT score indicates the number of individual ortholog prediction tools that report a given ortholog pair. The maximum score depends on the number of ortholog prediction tools that include that species in analysis.
B: A display of gene expression levels will appear by clicking on “show all,” as exemplified here for *Drosophila*.

Table S1:

1. Database Name and URL	2. Version/ Date Accessed	3. Method of Data Access	4. Data Extracted	5. Interpretation	6. Number of Entries
OMIM omim.org	On demand	API http://api.omim.org/api/entry?mimNumber={{omimNumber}}	Gene Description, Gene-Phenotype Relationship (Disease association), reported Allelic Variants	Gene-Phenotype Relationships report any known disease associations reported in the literature. If the individual of interest's phenotype matches the disease/phenotype described here, then the case is likely solved by this gene. The variant can be further analyzed to see if it was previously reported as benign/pathogenic.	15,553 gene descriptions. 8,377 phenotype associations
ExAC exac.broadinstitute.org/	Release 0.3.1	VCF files ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3.1/	Homozygous/Hemizygous count, Allele count, Total Allele number, Allele frequency	Homozygous/Hemizygous count of the variant of interest in a control population indicates how likely this gene can cause recessively inherited disease Allele frequency in ExAC is an estimate of how common this allele is in the control population.	10,195,872 entries
	On demand	Gene table http://exac.broadinstitute.org/gene/{{ensemblId}}	Expected vs. Observed no. of variants and Constraint Metrics (z-scores and pLI)	pLI scores indicate probability of Loss of Function intolerance and indicates the likelihood that this gene can cause dominantly inherited disease.	20,313 genes
ClinVar https://www.ncbi.nlm.nih.gov/clinvar/	Every two weeks	MARRVEL Crawler: Search ClinVar by HGNC identifier for human gene, every two weeks	Variant, Location, Condition(s), Frequency, Clinical Significance, and Review Status	Variants with interpretations reported by researchers and clinicians are valuable for analyzing how likely a variant is pathogenic.	426,009 records with interpretation
Geno2MP geno2mp.gs.washington.edu	October 10, 2016	Geno2MP.variants.vcf(http://geno2mp.gs.washington.edu/Geno2MP/#/terms)	Number of individuals homozygous or heterozygous for the variant of interest.	A summary of number of individuals with allele of interest. If the user searches gene-only, the MARRVEL displays the sum of all alleles (heterozygous or homozygous) found in Geno2MP.	20,313 genes, 392,583 entries
	October 15 15:46 UTC, 2016	Phenotype Information: MARRVEL Crawler	Human phenotype ontology (HPO) profiles of individuals containing variant of interest.	The HPO terms specify an individual's phenotype. The HPO profiles of individuals with the variants of interest are important clues to whether or not a variant is disease causing. If the individual in Geno2MP with the variant of interest displays phenotypes similar to the individual of interest, ie genotype and phenotype are consistent, then the variant of interests is more likely linked to the individual's disease.	5,012,286 entries
DGV (Database of Genomic Variants) dgv.tcag.ca/dgv/app/home	May 15, 2016	http://dgv.tcag.ca/dgv/docs/GRCh37_hg19_variants_2016-05-15.txt	Copy number variations in a control population that contains the gene of interest	The number of individuals with loss of copy number variations that contain the gene of interests may provide insight into how critical the gene is to normal function and whether or not haploinsuffciency may be a mechanism of disease. If there is a high number of individuals with deletions that contain the gene of interest, then it is less likely that haploinsufficiency is the disease mechanism.	392,583
DECIPHER Control Data decipher.sanger.ac.uk	September 7, 2016	Population Copy-Number Variation Frequencies: https://decipher.sanger.ac.uk/about#downloads/data	Copy number variations in a control population that contains the variant of interest	This dataset, similar to DGV, also contains copy number variations in non-disease cohorts.	58,146
Ensembl http://grch37.rest.ensembl.org	September 7, 2016	Ensembl GRCh37 REST API http://grch37.rest.ensembl.org/	Ensembl ID	N/A	34,544
HGNC genenames.org	September 7, 2016	HGNC REST API http://www.genenames.org/help/rest-web-service-help	Official HGNC gene name	N/A	26,307
Mutalyzer	On demand	API https://mutalyzer.nl/json/numberConversion?build=hg19&variant={{variant}}	HGVS conversion to chromosome location and Name checker assists users to input correct nomenclature	N/A	N/A
PubMed https://www.ncbi.nlm.nih.gov/pubmed	On demand	PubMed search by "Links from Gene" https://www.ncbi.nlm.nih.gov/pubmed?LinkName=gene_pubmed&from_uid={{EntrezID}}	URL to the page	N/A	N/A

Table S2:

1. Database Name and URL	2. Version/ Date Accessed	3. Method of Data Access	4. Data Extracted	5. Interpretation	6. Number of Entries

DIOPT (DRSC Integrative Ortholog Prediction Tool)
http://www.flyrnai.org/diopt
	Version 6.0.1 (Jan 2017)	MARRVEL Crawler	DIOPT is an online tool that uses multiple ortholog prediction tools to provide a score of how many prediction tools report a gene as an ortholog of the gene of interest. MARRVEL selects and displays multiple protein alignment of DIOPT's Predicted Best Orthologs and human protein domains. Multiple protein alignment across organisms are generated via DIOPT by using MAFFT FFT-NS-2 (v7.305b) aligner.	Multiple protein alignment of orthologs across model organisms can be used to asses the conservation of the amino acid change of interest and the conservation of protein domains. Highly conserved amino acids and amino acid changes located in protein domains are more likely to cause disrupt protein function and cause disease.	45022 (mouse)
					28118(fly)
					11719(yeast)
					5952(fissionYeast)
					52195(worm)
					36387(zebrafish)

SGD www.yeastgenome.org/	Nov 18 22:11 UTC , 2016	MARRVEL Crawler: http://www.yeastgenome.org/locus/{{SGD ID}}/overview	S. cerevisiae GO terms: EXP/IDA/IEP/IGI/IMP/IPI Tissue Expression: Direct Link	Biological and Molecular function of orthologs of the gene of interest may inform the gene's likelihood to cause the phenotype in an individual of interest. Tissue expression and subcellular localization data can be helpful to draw parallels between human disease and model organism phenotypes.	3818 distinct genes / 24195 human gene - homolog yeast gene pairs
PomBase https://www.pombase.org/	Nov 18 23:04 UTC , 2016	MARRVEL crawler: http://www.pombase.org/spombe/result/{{pombase id	S. pombe GO terms: EXP/IDA/IEP/IGI/IMP/IPI Tissue Expression: Direct Link		2992 / 13959 human gene - homolog yeast gene pairs
WormBase www.wormbase.org	Nov 9 22:17 UTC , 2016	MARRVEL crawler WormBase REST API http://www.wormbase.org/about/userguide/for_developers/API-REST/Go_term#0--10	C. elegans GO terms: EXP/IDA/IEP/IGI/IMP/IPI Expressions from REST API Tissue Expression: Direct Link		3793 genes 15189 human gene - homolog worm gene pairs
FlyBase flybase.org	Oct 26 20:01 UTC, 2016 Nov 7 18:52 , 2016	MARRVEL Crawler: http://flybase.org/reports/{{flybase id}}.html:	D. melanogaster GO Terms: "Terms Based on Experimental Evidence.“ TSV: Only when "Back-to-back Scales" or "Heatmap" is available.		4718 genes 34126 human gene – homolog fly gene pairs
ZFIN zfin.org	Nov 18 21:56 UTC , 2016	MARRVEL Crawler: http://zfin.org/action/marker/marker-go-view/{{ZFIN ID}}) (http://zfin.org/downloads) , "Expression data for wildtype fish"	D. rerio GO terms: EXP/IDA/IEP/IGI/IMP/IPI Tissue Expression: Expression data for wildtype fish		3650 genes 12739 human gene - homolog fish gene pairs
MGI www.informatics.jax.org	Nov 8 22:17 UTC, 2016	MARRVEL Crawler: http://www.informatics.jax.org/marker/gograph/{{MGI ID}}	M. musculus GO terms: EXP/IDA/IEP/IGI/IMP/IPI Tissue Expression: Expression data for wild-type		10280 / 68070 human gene - homolog mouse gene pairs
RGD rgd.mcw.edu	Mar 3, 2017	GOterms: ftp://ftp.rgd.mcw.edu/pub/ontology/annotated_rgd_objects_by_ontology/rattus_genes_go Expression:http://www.ebi.ac.uk/gxa/experiments/E-GEOD-53960?ref=aebrowse	R. norvegicus GO terms: EXP/IDA/IEP/IGI/IMP/IPI Tissue expression: All expression data from rat BodyMap		67650 human gene - homolog rat gene pairs 44914 gene - GO pairs
Protein Atlas www.proteinatlas.org	Dec 14 17:18 UTC, 2016	ProteinAtlas API http://www.proteinatlas.org/{{ensemblID}}.xml	H. sapiens Tissue Expression	Human tissue expression of a gene can help increase or decrease the likelihood that the gene is causative of a set of phenotypes in an individual.	12901 distinct human genes 580545 gene-organ expression level pairs
GTEx	V6	http://www.gtexportal.org/static/datasets/gtex_analysis_v6p/rna_seq_data/GTEx_Analysis_v6p_RNA-seq_RNA-SeQCv1.1.8_gene_median_rpkm.gct.gz	H. sapiens Median RPKM by tissue type		1768504 gene - tissue pairs
EMBL-EBI QuickGO https://www.ebi.ac.uk/QuickGO/	Nov 28, 2016, 9:35:00 AM	ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/HUMAN/goa_human.gaf.gz	H. sapiens Gene Ontology terms EXP/IDA/IEP/IGI/IMP/IPI	GO terms can provide potential disease mechanisms and consistency with model organism GO terms can assist in deciding which model organism can be used for further study.	13350 distinct human genes 86263 gene-GO pairs

Table S1: Description of core human genetics databases

MARRVEL selects information from six human genetics databases (OMIM, ExAC, ClinVar, Geno2MP, DGV, and DECIPHER) and displays data that are important for analyzing human genes and variants. Ensembl and HGNC are resources used to link the databases based on each gene's Ensembl ID and official HGNC gene name. Mutalyzer is used to provide more flexibility for variant input. PubMed links are provided to connect users to all relevant publications. Column 1 describes the name and URL (web addresses) of each database. Columns 2 and 3 describe when and how the data from each database is accessed. Columns 4 and 5 detail what specific data are extracted from each database and displayed on MARRVEL and how these data can be used to analyze variants and genes of interest. Column 6 documents the number of entries that are extracted from each database.

Table S2: Description of model organism and human gene function databases

MARRVEL displays a summary of gene functions of human genes and their model organism homologs. For each gene, when available, expression of protein and mRNA in specific tissues and Gene Ontology (GO) terms are obtained from the databases listed in column 1. Columns 2 and 3 describe when and how the data are obtained. Column 4 describes the type of data obtained from each database. Column 5 discusses how data extracted from each database can be used to analyze candidate genes. Column 6 documents the number of entries that are extracted from each database.

Inferred from Experiment (EXP)/Inferred from Direct Assay (IDA)/Inferred from Expression Pattern (IEP)/Inferred from Genetic Interaction (IGI)/Inferred from Mutant Phenotype (IMP)/Inferred from Physical Interaction (IPI)

Gene Name	Variant	MARRVEL Output Summary
OGDHL	10:50946295_G>A	No OMIM phenotype association. Microtubule and Mitochondrion associated protein. Highly expressed in human cerebellum. Highly conserved amino acid from yeast to human and located in the enzymatic domain
KIAA1632 (EPG5)	18:43496517_G>C	Vici Syndrome – Partial phenotypic match. Involved in autophagy and endosomes. Poorly conserved, Q in mouse and zebrafish.
CCT8	21:30428834_T>G	No OMIM phenotype association. Regulates telomerase, protein binding, cell-cell adhesion. Widely expressed in mouse. Highly expressed in human bronchus, hippocampus, stomach. Poorly conserved. Located outside of protein domain (TCP-1).
TIAM1	21:32624256_C>T	No OMIM phenotype association. 64 DGV loss alleles. Involved in Actin cytoskeleton organization, regulation of GTPase, cell migration, neuron projection development. Poorly conserved. Located outside of protein domains.
WASL	7:123329207_T>A	No OMIM phenotype association. Involved in cytoskeleton, spindle localization, cell migration, actin organization. Highly expressed in most human tissues. Outside of coding region
ARAP1	11:72437677_C>T	No OMIM phenotype association. Involved in cell migration, regulate GTPase. Highly expressed in human cerebellum, nasopharynx, placenta, and thyroid gland. Poorly conserved. Located outside of protein domains and in alignment gaps.
ATP8B1	18:55315737_G>A	Cholestasis – No phenotypic match.Transmembrane transport, sterol metabolic process, inner ear development. Widely expressed in human tissue. Poorly conserved. Located outside of protein domains and in alignment gaps.
ARL13B	3:93769712_C>G	Joubert Syndrome – No phenotypic match. 4 homozygous individuals in ExAC. Reported Benign by ClinVar. Involved in heart looping, left/right symmetry, dorsal/ventral patterning, cilium. Highly expressed in human adrenal gland, colon, endometrium, gallbladder. Intermediately conserved from zebrafish to humans. Located outside of protein domains.
Compound het variants unique in proband in OGDHL family		
AP2A2	11:984758_C>G	No OMIM phenotype association, 477 deletions found in DGV. Involved in dorsal/ventral patterning, protein binding and transportation, endocytosis, neurogenesis. Expressed in mice nervous system. Highly expressed widely in human tissue. Variant 1: amino acid Outside of coding region.
	11:988619_A>G	Variant 2: 3 homozygotes found in ExAC, 42 het / 20 HPO in DGV. Amino acid Highly conserved from yeast to human and located in the (adaptin) domain.
LAMA2	6:129786384_A>G	Muscular dystrophy – No phenotypic match. 774 Deletions found in DGV. Both alleles seen in ClinVar. Involved in axon guidance, cholinergic synaptic transmission, muscle development, localized to basement membrane. Expressed in human cerebral cortex and heart muscle. Variant 1: 1 het / 2 HPO (Integument and head/neck abnormality) in Geno2MP
	6:129601231_C>T	Variant 2: 3 homozygotes in ExAC, 39 het / 19 HPO in Geno2MP.
OBSCN	1:228456440_G>A 1:228461966_C>T	No OMIM phenotype association, 82 DGV deletions. Involved in muscle development, localized to M band of sarcomere, Rho GTPase binding. Widely expressed in human tissue, highly expressed in skeletal muscle.
VWA3A	16:22142902_G>A	No OMIM phenotype association, 5 deletions found in DECIPHER, Expressed in mouse nervous system. Variant 1: Amino acid Outside of coding region.
	16:22157653_T>A	Variant 2: Amino acid Conserved from zebrafish to humans. Located outside of protein domain.
De Novo variants unique in proband in OGDHL family		
PTCHD2 (DISP3)	1:11575517_G>T	No OMIM phenotype association, pLI score of 0, involved in regulation of neuron differentiation, cell migration,lipid metabolism.Conserved from zebrafish to humans but is F in worms and flies.

Table S3: 13 candidate genes and variants from a case study

From a case study in Yoon et al. 2017[13], 13 candidate variants were reported and subsequently filtered to identify a single variant in *OGDHL* prioritized for further study in *Drosophila*. We re-analyzed these variants with output from MARRVEL. Light blue font indicates key pieces of information that were interpreted as decreasing the likelihood that a variant is pathogenic.

Acknowledgements

The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at http://exac.broadinstitute.org/about.

The authors would like to thank the University of Washington Center for Mendelian Genomics and all contributors to Geno2MP for use of data included in Geno2MP.

This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.