Appendix A12: Ontology Evidence Codes

Manual codes used at RGD are highlighted in turquoise.

Evidence codes for Gene Ontology (GO), Pathway Ontology (PW), Disease Ontology (RDO), Mammalian Phenotype (MP), and Human Phenotype Ontology (HPO/HP)

(updated 2018)

Code	Stands for	Use for	Notes
IDA	inferred from **direct assay**	• Enzyme assays • In vitro reconstitution (e.g. transcription) • Immunodetection (for CC) • Cell fractionation (CC) • Physical interaction/binding assay (CC or MF) • Other experiments providing direct evidence specifically for the function, process or component indicated by the GO term • For RDO annotations use when the molecular mechanism or treatment is involved • use of recombinant or purified gene product as treatment for disease	• Need to be careful in that an experiment considered as direct assay for one ontology may be a different kind of evidence for the other ontologies • For functions such as protein binding, a binding assay is simultaneously IPI and IDA; Use IDA only when no identifier can be placed in the with/from column; when there is an appropriate ID for the with/from column, use IPI if the evidence shows direct binding between two entities
IAGP	Inferred by association of genotype and phenotype	• polymorphism or segregation of genetic markers (SNPs, mutations, RFLPs, microsatellites) • polymorphism or segregation of physical markers (FISH, centromeric, heterochromatic regions, chromosomal banding patterns) • detection of polymorphisms in inbred stock • used with any natural gene variant	• Used for phenotype or disease annotation to gene variants • Used for phenotype or disease annotation to rat and human QTLs • Used for phenotype or disease annotation to strains
IMP	inferred from **mutant phenotype**	• • Any artificial gene mutation/knockout • Overexpression or ectopic expression of wild-type or mutant genes	• • anything that is concluded from looking at artificial variations of a single gene of interest. This includes non-mutational variations such as inhibition with

		• Anti-sense experiments • RNAi experiments • Specific protein inhibitors	antibodies or other inhibitors.
IPI	inferred from **physical interaction**	• 2-hybrid interactions • Co-purification • Co-immunoprecipitation • Ion/protein binding experiments	• Does not include antibody binding. For experiments where antibody binding alters a function use IMP (see above) • Include an identifier for the "other" protein involved in the interaction in the “WITH” column. For multiple “WITH” entries use “,” to mean “and”, “|” (pipe) to mean “or” • For functions such as protein binding, a binding assay is simultaneously IPI and IDA, but IPI should be used if the binding partner is identified.
IGI	inferred from **genetic interaction**	• "Traditional" genetic interactions such as suppressors, synthetic lethals, etc. • Functional complementation • Rescue experiments • Inference about one gene drawn from the phenotype of a mutation in a different gene • Use for RDO when the effect of one gene’s variant on a disease is dependent on another gene’s variant involved in the disease	• Includes any combination of alterations in the sequence (mutation) or expression of more than one gene/gene product • Use when redundant copies of a gene must all be mutated to see an informative phenotype. • Also use for situations where a mutation in one gene (gene A) provides information about the function, process, or component of another. • Include an identifier for the "other" gene involved in the interaction in the “WITH” column—this may include gene(s) from other species in the case of complementation
IEP	inferred from **expression pattern**	• cases where the annotation is inferred from the timing or location of expression of a gene • for microarray experiments, if results were verified by other methods • Transcript levels (e.g. Northerns, microarray data) • Protein levels (e.g. Western blots)	• Most useful biomarker designation in disease annotations • Use of this code is not encouraged and annotations with this evidence code are generally considered “weak”.
ND	**no biological data** available	• Annotations to top level terms: molecular_function, biological_process, or cellular_component	• RGD does not use the “unknown” annotations in GO and therefore does not use this evidence code except for indicating a gene has been curated but no experimental rat data was found: Use reference RGD

			ID 1598407 and put “MM/YYYY: no relevant rat data” in free text box in curation tool.
ISS	inferred from **sequence or structural similarity**	• Sequence similarity (e.g. homolog of/most closely related to…) • Recognized domains • Structural similarity • Southern blotting	• Use for BLAST (or other sequence similarity detection method) results that have been reviewed for accuracy by a curator. • Currently not assigned manually for GO at RGD. • Used for automated annotations to the “other” RGD species when inferring an annotation from a manual annotation made to a human ortholog.
ISO	Inferred from sequence orthology	• automated assignment of annotations to orthologs of genes that have imported annotations • “automated” assignment of disease or pathway annotations to orthologs of primary gene that has been manually annotated	•
IC	Inferred by curator	• Use when annotation is not supported by any direct evidence, but can be reasonably inferred by a curator from other GO annotations, for which evidence is available	• Example: gene annotated to “transcription factor activity” or “DNA binding”, annotate to CC term “nucleus” with evidence code IC • the With/From field should always be filled in with a GO ID when using this evidence code. • Reference is the same as for the original GO term, e.g. DNA binding • Rarely used by RGD
RCA	inferred from **reviewed computational analysis**	• Predictions based on large-scale experiments (e.g. genome-wide two-hybrid, genome-wide synthetic interactions) • Predictions based on integration of large-scale datasets of several types • Text-based computation (e.g. text mining)	• used for annotations based on a non-sequence-based computational method, where the results have been reviewed by an author or a curator • For microarray results alone, IEP is preferred, but RCA is used when microarray results are combined with results of other types of large-scale experiments
IEA	inferred from **electronic annotation**	• Annotations transferred from database records, if not reviewed by curators	• Used for annotations that depend directly on computation or automated transfer of annotations from a database • Used when no curator has checked the annotation to verify its accuracy
			•

TAS	**traceable** author statement	• Review article or intro where the original experiments are traceable • Anything found in a text book or dictionary which could be considered “common knowledge” (i.e. “everybody knows…”)	• **NO LONGER USED AT RGD -** go to the original article to annotate information in question.

NAS	**non-traceable** author statement	• Database entries that don't cite a paper (e.g. UniProt Knowledgebase records) • Statements in papers (abstract, introduction, or discussion) that a curator cannot trace to another publication	• **NO LONGER USED AT RGD**

For examples and more information, see the GO Consortium website at: http://www.geneontology.org/GO.evidence.shtml