Tagging Cheat Sheet for BaseSpace® Correlation Engine

Tagging How-To

Required tags are biosource, biodesign, and tissue

- **Biosource**—Required to describe how a sample was derived. It includes cell lines compiled from resources such as ATCC, HPA, Tumorscape, DSMZ, hESCreg, ISCR, JCRB, CellBank Australia, COSMIC, NIH Human Embryonic Stem Cell Registry, and RIKEN BRC.
- **Biodesign**—Required to describe the nature of the comparison. Tag the biodesigns that most describe the driving differences in the bioset.
- **Tissue**—Required to define the specific organ/tissue/cell type. Tissue ontologies are derived from MeSH.

Add disease, compound, gene, genemode, and biogroup as relevant

- **Disease**—Assigned only if a sample corresponds to a disease state. Disease ontologies are derived from SNOMED CT.
- **Compound**—Assigned only if a sample was affected by a compound. Compound ontologies are derived from MeSH.
- **Gene**—Assigned only if a gene in a sample was modified or served as the key differentiating marker between experimental groups (eg, ER- vs. ER+ breast cancer). Sources include NCBI Entrez Gene, Unigene, and GenBank, EMBL-EBI Ensembl, and others.
- **Genemode**—Required to describe a gene modification. Genemode cannot be assigned without being linked to a specific gene.
- **Biogroup**—Assigned only if applicable. Biogroups should only be used as tags when no other vocabulary above provides relevant terminology. Biogroups are derived from resources such as MSigDB, GO, EMBL-EBI InterPro, PMAP, and TargetScan.

All ontologies are modified, curated versions of accredited biomedical ontologies. Shown here for quick reference are BaseSpace Correlation Engine custom tags. All tags can be browsed in the ontology editor or selected through auto-complete.

Introduction

Tagging is the association of biosets with terms from biomedical ontologies. The tagging step ensures that sample type, sample source, experiment type, and other relevant concepts—such as phenotype, treatments, and gene modifications—are captured for each signature with standardized vocabulary. Tagging enables biological interpretation of correlation and meta-analysis. The framework also takes advantage of hierarchical relationships to return accurate, enriched search results.

Genemode

Cell Marker
- Negative
- Positive

Gene Overexpression
- Conditional
- Constitutive
- Ectopic
- Epigenetic
- Knock-in
- Mimic overexpression

Gene Knockdown
- Epigenetic
- Morpholino
- RNA interference
  - shRNA knockdown
  - siRNA knockdown
- ncRNA knockdown
- miRNA knockdown

Gene Knockout
- Conditional
- Irreversible

Gene Mutation
- Amplification
- Deletion
- Fusion
- Insertion
- Inversion
- Translocation
- Amorphic
- Neomorphic
- Hypermorphic
- Hypomorphic
- Antimorphic
  - Dominant-negative

Immunoprecipitation—co-IP
- ChIP antibody target
- RIP antibody target

Protein Treatment
- Antibody target—inhibitory
- Antibody target—stimulatory

Biodesign

Clinical Research
- Clinical study
  - Clinical outcome

Data Validation
- Below threshold significance
- Insufficient replicates
- Insufficient sequence reads

Demographic Comparison
- Age comparison
- Gender comparison
- Ethnicity comparison

Disease Comparison
- Disease vs. normal
- Disease vs. disease
- Disease resistant vs. sensitive

Genetic Perturbation
- Mutant vs. wildtype
- Mutant vs. mutant

Growth Conditions
- Environmental conditions
- Compound withdrawal
- Treatment deprivation

Pharmacological Response
- Response to a drug
  - Drug nonresponse vs. complete response
  - Drug nonresponse vs. partial response
  - Drug partial vs. complete response
- Drug resistant vs. sensitive

Timecourse
- Circadian time course
- Developmental time course
- Treatment time course

Treatment Comparison
- Dose response
- Treatment vs. control
- Treatment vs. treatment

Other Comparison Types
- Biomarker comparison
- Biosource comparison
- Method comparison
- Normal vs. normal
- Quantitative trait analysis
- Species comparison
- Strain comparison

Biosource

- Blood fraction
- Bone marrow fraction
- Cell line (specific if available)
- Cell lysate
- Primary cells
- Primary cells—cultured
- Primary cells—laser-capture
- Primary tissue—FFPE
- Primary tissue—fresh or fresh frozen
- Whole blood
- Whole body

Abbreviations: Human Protein Atlas (HPA), Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), Human Pluripotent Stem Cell Registry (hESCreg), Institute for Scientific Computing Research (ISCR), Japanese Collection of Research Bioresources (JCRB), Catalogue Of Somatic Mutations In Cancer (COSMIC), National Institutes of Health (NIH), RIKEN BioResource Center (RIKEN BRC), Medical Subject Headings (MeSH), Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), ecdysone receptor (ER), National Center for Biotechnology Information (NCBI), European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute (EBI), Proteolysis Map (PMAP), immunoprecipitation (IP), chromatin immunoprecipitation (ChIP), RNA immunoprecipitation (RIP), formalin-fixed, paraffin-embedded (FFPE)

For Research Use Only. Not for use in diagnostic procedures.

illumina®