

[GSEA Home](#)[Downloads](#)[Molecular Signatures Database](#)[Documentation](#)[Contact](#)[MSigDB Home](#)[About Collections](#)[Browse Gene Sets](#)[Search Gene Sets](#)[Investigate Gene Sets](#)[View Gene Families](#)[Help](#)

MSigDB Collections

The 17779 gene sets in the Molecular Signatures Database (MSigDB) are divided into 8 major collections, and several sub-collections. See the table below for a brief description of each, and the [MSigDB Collections: Details and Acknowledgments](#) page for more detailed descriptions. See also the [MSigDB Statistics](#) and the [MSigDB Release Notes](#).

Click on the "browse gene sets" links in the table below to view the gene sets in a collection. Or download the gene sets in a collection by clicking on the links below the "Download GMT Files" headings. For a description of the [GMT file format](#) see the [Data Formats](#) in the [Documentation](#) section. The gene sets can be downloaded as Entrez Gene Identifiers or HUGO Gene Symbols. An [XML file](#) containing all the MSigDB gene sets is available on the [Downloads](#) page.

H: hallmark gene sets (browse 50 gene sets)	Hallmark gene sets summarize and represent specific well-defined biological states or processes and display coherent expression. These gene sets were generated by a computational methodology based on identifying overlaps between gene sets in other MSigDB collections and retaining genes that display coordinate expression. details	Download GMT Files gene symbols entrez genes ids
C1: positional gene sets (browse 326 gene sets)	Gene sets corresponding to each human chromosome and each cytogenetic band that has at least one gene. details	Download GMT Files gene symbols entrez genes ids
C2: curated gene sets (browse 4731 gene sets)	Gene sets curated from various sources such as online pathway databases, the biomedical literature, and knowledge of domain experts. The gene set page for each gene set lists its source. The C2 collection is divided into two sub-collections: CGP and CP. details	Download GMT Files gene symbols entrez genes ids
CGP: chemical and genetic perturbations (browse 3402 gene sets)	Gene sets represent expression signatures of genetic and chemical perturbations. A number of these gene sets come in pairs: xxx_UP (and xxx_DN) gene set representing genes induced (and repressed) by the perturbation.	Download GMT Files gene symbols entrez genes ids
CP: Canonical pathways (browse 1329 gene sets)	Gene sets from pathway databases. Usually, these gene sets are canonical representations of a biological process compiled by domain experts.	Download GMT Files gene symbols entrez genes ids
CP:BIOCARTA: BioCarta gene sets (browse 217 gene sets)	Gene sets derived from the BioCarta pathway database.	Download GMT Files gene symbols entrez genes ids
CP:KEGG: KEGG gene sets (browse 186 gene sets)	Gene sets derived from the KEGG pathway database.	Download GMT Files gene symbols entrez genes ids
CP:REACTOME: Reactome gene sets (browse 674 gene sets)	Gene sets derived from the Reactome pathway database.	Download GMT Files gene symbols entrez genes ids
C3: motif gene sets (browse 836 gene sets)	Gene sets representing potential targets of regulation by transcription factors or microRNAs. The sets consist of genes grouped by short sequence motifs they share in their non-protein coding regions. The motifs represent known or likely cis-regulatory elements in promoters and 3'-UTRs. The C3 collection is divided into two sub-collections: MIR and TFT details	Download GMT Files gene symbols entrez genes ids
MIR: microRNA targets (browse 221 gene sets)	Gene sets that contain genes sharing putative target sites (seed matches) of human mature miRNA in their 3'-UTRs.	Download GMT Files gene symbols entrez genes ids
TFT: transcription factor targets (browse 615 gene sets)	Gene sets that share upstream cis-regulatory motifs which can function as potential transcription factor binding sites. Based on work by Xie et al. 2005	Download GMT Files gene symbols entrez genes ids
C4: computational gene sets (browse 858 gene sets)	Computational gene sets defined by mining large collections of cancer-oriented microarray data. The C4 collection is divided into two sub-collections: CGN and CM. details	Download GMT Files gene symbols entrez genes ids
CGN: cancer gene neighborhoods (browse 427 gene sets)	Gene sets defined by expression neighborhoods centered on 380 cancer-associated genes. This collection is described in Subramanian, Tamayo et al. 2005	Download GMT Files gene symbols entrez genes ids

CM: cancer modules (browse 431 gene sets)	Gene sets defined by Segal et al. 2004 . Briefly, the authors compiled gene sets ('modules') from a variety of resources such as KEGG, GO, and others. By mining a large compendium of cancer-related microarray data, they identified 456 such modules as significantly changed in a variety of cancer conditions.	Download GMT Files gene symbols entrez genes ids
C5: GO gene sets (browse 5917 gene sets)	Gene sets that contain genes annotated by the same GO term. The C5 collection is divided into three sub-collections based on GO ontologies: BP, CC, and MF. details	Download GMT Files gene symbols entrez genes ids
BP: GO biological process (browse 4436 gene sets)	Gene sets derived from the GO Biological Process Ontology.	Download GMT Files gene symbols entrez genes ids
CC: GO cellular component (browse 580 gene sets)	Gene sets derived from the GO Cellular Component Ontology.	Download GMT Files gene symbols entrez genes ids
MF: GO molecular function (browse 901 gene sets)	Gene sets derived from the GO Molecular Function Ontology.	Download GMT Files gene symbols entrez genes ids
C6: oncogenic signatures (browse 189 gene sets)	Gene sets that represent signatures of cellular pathways which are often dis-regulated in cancer. The majority of signatures were generated directly from microarray data from NCBI GEO or from internal unpublished profiling experiments involving perturbation of known cancer genes. details	Download GMT Files gene symbols entrez genes ids
C7: immunologic signatures (browse 4872 gene sets)	Gene sets that represent cell states and perturbations within the immune system. The signatures were generated by manual curation of published studies in human and mouse immunology. details	Download GMT Files gene symbols entrez genes ids