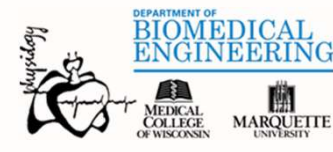




# Ontology Analysis of Coronavirus Associated Human Genes at the COVID-19 Disease Portal

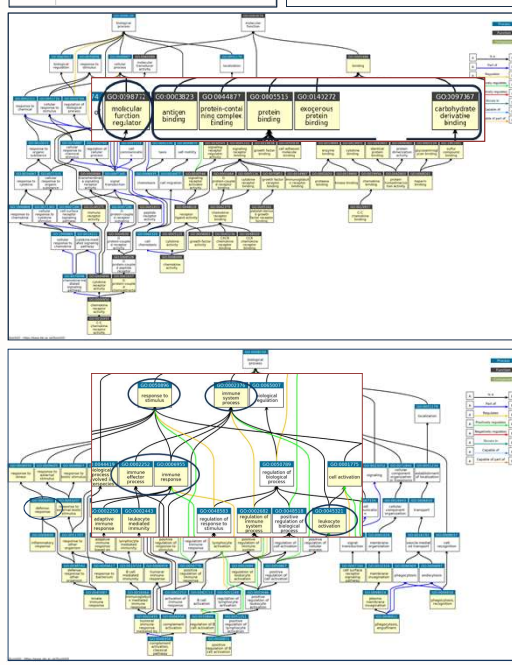
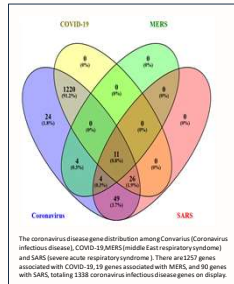
Shur-Jen Wang<sup>1</sup>, Jeffrey L De Pons<sup>1</sup>, Wendy Demos<sup>1</sup>, G Thomas Hayman<sup>1</sup>, Morgan Hill<sup>1</sup>, Mary L Kaldunski<sup>1</sup>, Stan JF Lauderkind<sup>1</sup>, Jennifer R Smith<sup>1</sup>, Monika Tutaj<sup>1</sup>, Mahima VEDI<sup>1</sup>, Melinda R Dwinell\*, Anne E Kwitek<sup>1</sup> and the RGD team<sup>1</sup>  
<sup>1</sup>Rat Genome Database, Department of Biomedical Engineering, Medical College of Wisconsin; \*Department of Physiology, Medical College of Wisconsin Milwaukee, WI, 53226, USA



**Abstract**  
 The Disease Portal at Rat Genome Database (RGD) is a premier disease resource where disease-associated genome objects are integrated with their associated data in one place according to disease areas. In response to SARS-CoV-2 pandemic, RGD developed a "COVID-19 Disease Portal"  
<https://rgd.mcw.edu/rgdweb/portal/home.jsp?p=14>. In the COVID-19 Portal, gene-disease associations are established by manual curation of PubMed literature. All the functional annotations and genome data associated with these disease genes are accessible from the portal. We performed analyses on the COVID-19 disease gene set using tools developed at RGD. The Disease Ontology term enrichment analysis showed that the COVID-19 disease gene set was highly enriched with coronavirus infectious disease and related diseases. Several less related disease areas such as liver disease, and rheumatic disease were also highly enriched. Using the comparison heatmap, we showed that close to 60 percent of the COVID-19 genes were associated with nervous system disease and 40 percent were associated with gastrointestinal disease. Our analysis confirms the role of the immune system in COVID-19 pathogenesis as shown by substantial enrichment of immune system-related Gene Ontology terms. Using the integrated data sets in the RGD COVID-19 Portal will help elucidate mechanisms of COVID-19 and ultimately lead to prevention or treatments.

Table 1. Evidence code analysis of COVID-19 disease annotations and associated genes

A. Evidence code (ECO) distribution			B. Unique ECO among genes		
ECO type	Annotated count	Gene count	Gene /unique ECO	Gene count	Gene count
IAGP	21	19	1	1221	
IDA	4	3	2	28	
IMP	145	63	3	7	
HEP	1195	1175	4	1	
EXP	37	37	3		
ISO	2	2	4		
Total	1407	1257			



The Gene Ontology Ancestor chart showing the relations among the enriched molecular function and biological process terms associated with COVID-19 genes. The enriched GO terms are highlighted in yellow, and the circled terms are enriched high level terms

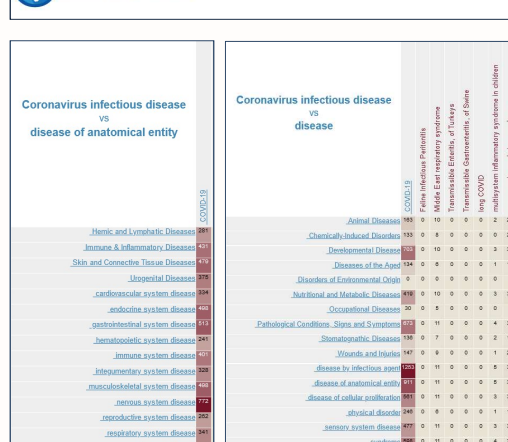


Table 3. COVID-19 disease gene distribution within organ systems

nervous system disease (772)	Gastrointestinal system disease (513)	Endocrine system disease (498)	musculoskeletal system disease (498)	skin and connective tissue disease (479)	Immune & inflammatory disease (431)
central nervous system disease (676)	gastrointestinal disease (401)	liver disease (335)	connective tissue neoplasm (376)	bone disease (270)	primary immunodeficiency disease (348)
sensory system disease (477)	digestive system neoplasm (339)	endocrine gland neoplasm (301)	bone disease (270)	connective tissue neoplasm (161)	autoimmune disease (246)
neurologic Manifestation (356)	liver disease (335)	diabetes Mellitus (134)	neuromuscular disease (174)	breast disease (161)	lymphatic system disease (176)
neurodegenerative disease (278)	intestinal disease (248)	pancreas disease (119)	musculoskeletal anomalies (157)	rheumatic disease (123)	rheumatic disease (133)
peripheral nervous system disease (232)	mouth disease (120)	gonadal disease (109)	autoimmune disease of musculoskeletal	genetic skin disease (119)	allergic disease (121)
nervous system neoplasm (148)	stomach disease (99)	parathyroid gland disease (83)	joint disease (156)	interstitial lung disease (85)	immunoproliferative disorders (127)
nervous system malformation (122)	gastroenteritis (97)	autoimmune disease of endocrine	muscular disease (78)	dermatitis (78)	immune system cancer (57)
nervous system trauma (98)	biliary tract disease (66)	thyroid gland disease (53)	musculoskeletal system cancer (93)	collagen disease (65)	gastroenteritis (97)
autoimmune disease of the nervous system	esophageal disease (56)	dwarfism (29)	jaw disease (37)	eczematous skin disease (57)	dermatitis (78)
congenital nervous system abnormality (75)	autoimmune disease of gastrointestinal tract	adrenal gland disease (24)	musculoskeletal system benign neoplasm	infectious skin disease (62)	pneumonia (79)

Table 2. Top 40 enriched disease ontology terms from MOET and Set Analyzer

A. MOET				B. Set Analyzer			
Rank	Disease term (ID)	p*	count	Ref count	Rank	Mesh term (ID)	Ref count
1	Respiratory tract infection (D0090086)	1.52E-40	155	153	1	Neoplasms (MESH: D0090093)	5,225,126
2	liver disease (D00405)	2.35E-35	335	2638	2	Digestive System Diseases (MESH: D004056)	1,525,117
3	hepatocellular carcinoma (D004884)	9.75E-34	161	853	3	Liver Diseases (MESH: D008207)	1,226,114
4	liver carcinoma (D004886)	1.13E-33	161	854	4	Neoplasms by Site (MESH: D0090094)	1,476,599
5	hepatobiliary disease (D003118)	1.49E-32	337	2746	5	Immune System Diseases (MESH: D007514)	5,038,891
6	liver cancer (D003571)	5.53E-32	164	908	6	Carcinoma (MESH: D002277)	1,288,646
7	hepatobiliary system cancer (D0048835)	7.52E-32	171	975	7	Digestive System Neoplasms (MESH: D004007)	8,856,739
8	adenocarcinoma (D00299)	1.51E-28	228	1617	8	Respiratory Tract Diseases (MESH: D002148)	2,386,171
9	Liver Neoplasms (D0090018)	1.05E-27	176	1100	9	Carcinoma, Hepatocellular (MESH: D006528)	7,716,176
10	rheumatic disease (D001578)	1.88E-27	133	698	10	Adenocarcinoma (MESH: D00239)	1,005,744
11	autoimmune disease of musculoskeletal system (D006023)	17.15E-27	156	522	11	Liver Neoplasms (MESH: D008113)	1,476,172
12	allergic disease (D001205)	3.92E-26	122	624	12	Skin and Connective Tissue Diseases (MESH: D0237)	7,707,172
13	pneumonia (D0052)	1.98E-25	79	292	13	Lung Diseases (MESH: D008371)	1,136,751
14	nephritis (D001092)	7.85E-25	91	388	14	Urogenital Diseases (MESH: D00091642)	1,086,700
15	bacterial infectious disease (D001204)	4.26E-24	130	728	15	Nervous System Diseases (MESH: D000642)	3,176,619
16	respiratory allergy (D004096)	2.25E-23	90	397	16	Female Urogenital Diseases and Pregnancy Complications (MESH: D000361)	1,426,665
17	glomerulonephritis (D002921)	2.89E-23	82	337	17	Respiratory Tract Infections (MESH: D021213)	2,296,653
18	Rheumatoid Arthritis (D003000)	3.76E-23	140	906	18	Infectious (MESH: D000739)	1,629,644
19	Immediate Hypersensitivity (D0090056)	8.89E-23	99	477	19	Skin Disease (MESH: D021871)	2,518,613
20	Bacterial Infections and Mycoses (D0030084)	1.34E-22	149	936	20	Pneumonia (MESH: D00324)	1,533,642
21	Chronic Disease (D000004)	1.36E-22	82	344	21	Female Urogenital Diseases (MESH: D003776)	3,366,642
22	Inflammation (D0030072)	3.57E-22	297	2605	22	Viruses (MESH: D001024)	1,026,577
23	paratuberculous infectious disease (D002789)	3.64E-22	56	170	23	Fibrosis (MESH: D00329)	4,376,517
24	thymic atrophy (D000004)	3.65E-22	94	444	24	Male Urogenital Diseases (MESH: D002801)	2,566,564
25	obstructive lung disease (D002320)	6.72E-22	96	464	25	COVID-19 (MESH: D00086382)	3,996,567
26	carcinoma (D00385)	1.02E-21	346	3235	26	Urologic Diseases (MESH: D004370)	6,126,555
27	arthritis (D00148)	1.82E-21	140	825	27	Autoimmune Diseases (MESH: D000923)	1,105,613
28	bone inflammation disease (D003342)	5.88E-21	140	910	28	Virus Diseases (MESH: D004777)	5,886,448
29	lung disease (D00189)	9.56E-21	248	2066	29	kidney disease (MESH: D000974)	2,346,417
30	paratuberculous disease (D001204)	1.86E-21	77	327	30	Gastrointestinal Diseases (MESH: D000340)	3,768,447
31	Othomyzomycosis Infections (D00900498)	8.07E-20	49	144	31	Hypersensitivity (MESH: D000697)	6,508,455
32	asthma (D00344)	8.38E-20	80	361	32	Nephritis (MESH: D000939)	4,616,451
33	hepatitis (D00237)	1.05E-19	71	293	33	Gastrointestinal Diseases (MESH: D000340)	3,768,447
34	Thoracic Injuries (D00900354)	1.85E-19	68	272	34	Connective Tissue Diseases (MESH: D000820)	3,258,418
35	bronchial disease (D001176)	1.57E-19	145	962	35	Vascular Diseases (MESH: D004632)	5,728,444
36	lower respiratory tract disease (D0030161)	1.52E-19	248	2110	36	Cardiovascular Diseases (MESH: D000218)	4,118,442
37	Influenza (D00480)	2.29E-19	48	145	37	Connective Tissue Diseases (MESH: D000820)	3,258,418
38	autoimmune disease of central nervous system (D006046)	4.09E-19	72	307	38	Liver Cirrhosis, Experimental (MESH: D008306)	6,806,411
39	respiratory system disease (D001578)	4.16E-19	343	3306	39	respiratory system disease (MESH: D001578)	3,996,448
40	lung injury (D00900310)	1.17E-18	65	267	40	Glomerulonephritis (MESH: D000931)	4,908,440

Table 2. \*The Bonferroni corrected P-values are shown; # the number of the COVID-19 related disease gene annotated with the disease term and its child terms; # the number of genes in the reference genome annotated with the disease term and its child terms.