

Universidade de Lisboa
Faculdade de Ciências
Departamento de Informática



AIDS/SIDA

Home assignment 2: Sequence alignment methods

André Filipe Bernardes Oliveira, 45648

Mestrado em Informática (MI)

Report of **Bioinformática**

to teacher: André Falcão

1 Index

1	Index.....	1
2	Introduction.....	2
3	Methods.....	2
4	Results and Discussion.....	3
5	Conclusion.....	4
6	Bibliography.....	4
7	Appendix.....	6
7.1	BLAST results.....	6
7.1.1	CCR5.....	6
7.1.2	CXCR4	14
7.2	PSI-BLAST results.....	23
7.2.1	CCR5.....	23
7.2.2	CXCR4	31
7.3	Multiple sequence alignment.....	39
7.3.1	CCR5.....	39
7.3.2	CXCR4	41
7.4	HMM Results	43
7.4.1	CCR5.....	43
7.4.2	CXCR4	45

2 Introduction

Following the conclusions from the previous report, the most relevant proteins in Acquired Immunodeficiency Syndrome (AIDS) are: *CCR5* (P51681) and *CXCR4* (P61073), the two coreceptors used as HIV entry point. In this report, sequence alignment methods will be used to retrieve more information about these two proteins, such as finding the homologous genes for each (including distant homologs) and checking for conserved domains.

3 Methods

Protein alignment (BLAST and PSI-BLAST) was performed on the NCBI website, alternatively it could be performed locally using BLAST+ tools and SwissProt database. The defined cut-off value (E-value) was 1×10^{-7} .

The first step of this work is to perform a sequence alignment with BLAST using all default parameters to find homologous sequences. It was also experimented with different parameters to compare the results e.g.: word size – 2, 3 or 6 (default); substitution matrix – BLOSUM 45, BLOSUM 50, BLOSUM 62 (default), BLOSUM 80 or BLOSUM 90. The displayed results used computed from default parameters.

To find distant homologs, PSI-BLAST was used with the default parameters. One version considered all the sequences while in the other some sequences were ignored: sequences with low BLAST score, whose identity was not confirmed or that did not correspond to receptors.

Protein annotation was obtained with PANTHER and tools from Princeton University (Generic GO Term Finder and LAGO) since ProteinON was not working and DAVID was hard to work with. Resulting schemas are exhibited.

A multiple sequence alignment (MSA) was performed with the BLAST result of each protein, using ClustalOmega default values. Which was then exported to Stockholm in order to create HMMs (Hidden Markov Models). The models created were observed as Logos using Skyline. Each model was validated with proteins present in the results of the PSI-BLAST alignment, but that were not used to create HMM model (that is was not present in BLAST alignment results). HMM was created locally using HMMER3.

MSA results were confirmed using Ugene software to allow easier identification of conserved domains. This and Skyline yielded results like what was observed using JalView.

4 Results and Discussion

Even though there was some experimenting with the parameters (described in the previous section) to compare the results obtained, in the end the results used to the next steps and shown here were obtained using the default parameters. This fact can be easily explained by the fact that this alternative provided what seems to be good results. Since there was no reason that deemed using different parameters, the defaults values were used because the results seem to be appropriate.

As it should be expected all results were very similar in booth proteins because the proteins themselves are very similar and have related functions. Therefore, is normal that the multiple sequence alignments obtained and the annotations obtained in the alignments are roughly the same.

It should be noted that since ProteinON was not working until the time of the delivery, it was used another alternative, which as only used with annotations for humans. Since there results referred several organisms besides humans, one should be suspicious of the GO terms displayed in this report because only a relatively small percentage of proteins were human.

Given the extension of results, the most relevant ones are available in the appendix section and indexed for easier search.

BLAST results (Table 7.1, Table 7.5) provided a list of proteins that were related to *CCR5* and *CXCR4*, mostly homologs in other species. Despite the problem with annotations already mentioned, human's annotations (Table 7.2, Table 7.3, Table 7.4, Table 7.6, Table 7.7, Table 7.8) allow to infer that the proteins identified are mostly receptors involved in signaling pathways and all are part of the plasma membrane. Booth results identified two interesting groups: C-C chemokine receptor activity and C-X-C chemokine receptor activity which was expected since the 2 main proteins are related.

An interesting fact in the BLAST results is that the most similar proteins to the queried one are simian proteins (e.g. Chimpanzee proteins) which makes sense since AIDS was transmitted from simians to humans (evolution of SIV virus to HIV).

PSI-BLAST results (Table 7.9, Table 7.13) identified proteins involved in different processes but sill receptors. Those proteins, according to their annotations (Table 7.10, Table 7.11, Table 7.12, Table 7.14, Table 7.15, Table 7.16), were related to hormones and more basic biological processes like eating and system cAMP synthesis (involved in energy production and innumerous regulation processes).

These results exposed the existence of conserved domains which may be important to receptor activity (that is that act of recognizing another molecule). The conservation was in line with what would be expected: variable regions in the ends of the protein and highly conserved

regions inside, as can be seen in the logos presented (Figure 7.3, Figure 7.8) and the overview of amino acid conservation (Figure 7.2, Figure 7.7). It was not possible to reproduce the full logo because the image was too big.

Hidden Markov models seem to be adequate, considering its high results as it would be expected (Table 7.17, Table 7.18). It was tested using proteins that were not used to create the model but were related to the model (as proved by the PSI-BLAST results). The results obtained for those proteins were high with a great insurance that are not random. These results were mostly caused by the score of the best domain which proves the theory that there is some domain conserved among all these proteins. It also proved that these proteins are related, as it would be expected from the PSI-BLAST results. Therefore, the two HMM produced seem to be adequate to evaluate if other proteins belong to this group (are related to these ones) or not.

Most of this work could be automated using programming languages like R. It is possible to make R interact with the system and run the BLAST and PSI-BLAST locally and process the results. To perform the multiple sequence alignment some alternative to ClustalOmega should be used to allow automation. Again, for the creation and testing of HMM models, R system tools could be used. Finally using R is also possible to get GO terms/protein annotations as seen in one of the lab classes. So, creating a R script could automate all these tasks and then it would be a matter of analyzing the results.

These results provided interesting data about related proteins to the two that were selected and identified domains shared among a group of receptors. Despite that it didn't shed light into regions that may be target to treat/avoid AIDS, since the identified domains maybe the too broad and find in many important receptors to the normal behavior of the organism.

5 Conclusion

This report allowed to explore and retrieve more information about 2 most interesting proteins related to AIDS: *CCR5* and *CXCR4*. It identified homologous proteins including distant ones that can to interesting to research about because they may to provide some new information that is relevant in AIDS therapy.

Another contribution of this work is the creating of two Hidden Markov models that can identify the conserved domains among these proteins and thus accessing if some other proteins is related to this one (considering if it has the domain present or not).

6 Bibliography

- "Unipro UGENE: a unified bioinformatics tool" Okonechnikov; golosova; furs; the UGENE team. Bioinformatics 2012 28: 1166-1167"
- PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.

Huaiyu Mi, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D. Thomas
Nucl. Acids Res. (2016) doi: 10.1093/nar/gkw1138

- Large-scale gene function analysis with the PANTHER classification system. Huaiyu Mi, Anushya Muruganujan, John T Casagrande and Paul D Thomas *Nature Protocols* 8, 1551 - 1566 (2013) doi: 10.1038/nprot.2013.092.
- E.I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, G. Sherlock, GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes, *Bioinformatics*. 20 (2004) 3710–3715. doi:10.1093/bioinformatics/bth456. Available at: <http://go.princeton.edu/cgi-bin/GOTermFinder>
- Clements, J. (2017). Skyline. [online] Skyline.org. Available at: <http://skyline.org/logo/7FBA779E-21FF-11E7-ADC9-280777086C57> [Accessed 14 Apr. 2017].
- Ncbi.nlm.nih.gov. (2017). PSSM Viewer. [online] Available at: https://www.ncbi.nlm.nih.gov/Class/Structure/pssm/pssm_viewer.cgi [Accessed 12 Apr. 2017].

7 Appendix

7.1 BLAST results

7.1.1 CCR5

Table 7.1 - Protein BLAST most relevant results for query protein CCR5_HUMAN (P51681). Score equals booth Max and Total Score. Default parameters were used with cut-off value of 1×10^{-7} , limited to 500 hits.

Gene	Accession	Score	E value	Ident
CCR5_HUMAN	P51681.1	723	0.0	100%
CCR5_PANTR	P56440.1	721	0.0	99%
CCR5_PONPY	O97881.1	719	0.0	99%
CCR5_GORGO	P56439.1	717	0.0	99%
CCR5_SYMSY	Q95NC5.1	717	0.0	99%
CCR5_TRAPH	O97879.1	715	0.0	98%
CCR5_HYMLL	Q95NC0.1	714	0.0	99%
CCR5_LOPAT	P61755.1	714	0.0	98%
CCR5_NASLA	Q95NC7.1	714	0.0	98%
CCR5_TRAFR	O97878.1	713	0.0	98%
CCR5_NOMLE	O97883.1	713	0.0	98%
CCR5_COLPO	Q95NC8.1	713	0.0	98%
CCR5_MACMU	P61813.1	713	0.0	98%
CCR5_RHIBE	O97880.1	712	0.0	98%
CCR5_PYGNE	O97882.1	712	0.0	97%
CCR5_RHIAV	O97962.1	712	0.0	97%
CCR5_THEGE	Q95NC1.1	712	0.0	98%
CCR5_MIOTA	Q95NC3.1	711	0.0	98%
CCR5_TRAJO	Q95NC6.1	711	0.0	97%
CCR5_CERGA	Q9TV49.1	711	0.0	98%
CCR5_MACAR	O97975.1	711	0.0	97%
CCR5_CHLPG	Q9TV42.1	709	0.0	97%
CCR5_MANLE	Q95ND2.1	709	0.0	97%
CCR5_CHLTN	Q95NE8.1	709	0.0	97%
CCR5_CERAT	O62743.2	708	0.0	97%
CCR5_MANSF	Q95ND1.1	707	0.0	97%
CCR5_ERYPA	Q95ND0.1	707	0.0	97%
CCR5_CHLAE	P56493.1	705	0.0	97%
CCR5_CHLSB	Q9TV43.1	704	0.0	97%
CCR5_CERCP	Q9TV47.1	704	0.0	97%
CCR5_CERLH	Q9XT76.1	704	0.0	96%
CCR5_CERAS	Q9TV48.1	701	0.0	97%
CCR5_CERSO	Q9BGN6.1	697	0.0	95%
CCR5_CERNF	Q9TV45.1	697	0.0	96%
CCR5_CALMO	Q95NC2.1	655	0.0	92%
CCR5_CALHU	Q6WN98.1	648	0.0	92%
CCR5_ATEGE	Q95NC4.1	644	0.0	91%
CCR5_SAISC	Q8HZT9.1	642	0.0	90%
CCR5_BOVIN	Q2HJ17.1	641	0.0	86%

Gene	Accession	Score	E value	Ident
CCR5_ALOSE	Q95NC9.1	626	0.0	92%
CCR5_CANFA	Q5ECR9.1	612	0.0	83%
CCR5_MOUSE	P51682.3	609	0.0	82%
CCR5_RABIT	Q1ZY22.1	588	0.0	80%
CCR5_RAT	O08556.1	585	0.0	82%
CCR2_MOUSE	P51683.2	536	0.0	74%
CCR2_RAT	O55193.1	533	0.0	74%
CCR2_MACMU	O18793.2	525	0.0	75%
CCR2_HUMAN	P41597.1	475	7E-168	79%
CCR1_HUMAN	P32246.1	409	5E-142	57%
CCR3_CANFA	Q64H34.1	398	1E-137	56%
CCR3_MOUSE	P51678.2	398	1E-137	57%
CCR3_RAT	O54814.1	398	1E-137	58%
CCR1_MOUSE	P51675.2	376	5E-129	56%
CCR4_CANFA	Q8MJW8.1	362	3E-123	50%
CCR1_MACFA	Q2Y2P0.1	355	9E-121	55%
CCR1_MACMU	P56482.1	354	2E-120	54%
CCR4_HUMAN	P51679.1	353	7E-120	49%
CCR4_MOUSE	P51680.2	352	1E-119	51%
CCR3_HUMAN	P51677.1	349	1E-118	53%
CCR3_MACMU	P56483.1	346	3E-117	54%
CCR3_MACFA	Q9BDS8.1	340	7E-115	53%
CCR3_CAVPO	Q9Z2I3.1	339	1E-114	53%
CCR3_CHLAE	P56492.1	339	2E-114	52%
CC1L1_MOUSE	P51676.2	310	3E-103	45%
CCR8_MOUSE	P56484.1	283	2E-92	43%
VGE1_EHV2	Q89609.1	276	1E-89	48%
CCR8_HUMAN	P51685.1	266	9E-86	41%
CCR8_MACMU	O97665.1	263	9E-85	41%
CCRL2_HUMAN	O00421.2	253	8E-81	45%
CCRL2_MACMU	Q9XSD7.1	251	4E-80	41%
CCRL2_BOVIN	Q0II78.1	236	4E-74	39%
CCRL2_PIG	Q75ZH0.1	229	9E-72	38%
CCR9_HUMAN	P51686.2	225	1E-69	39%
CCR9_MOUSE	Q9WUT7.1	221	2E-68	35%
CX3C1_MOUSE	Q9Z0D9.1	220	4E-68	44%
CX3C1_RABIT	Q2KTE1.1	219	1E-67	42%
CCR9_SHEEP	Q1WLP9.1	218	4E-67	38%
CX3C1_HUMAN	P49238.1	213	2E-65	43%
CX3C1_RAT	P35411.1	213	2E-65	45%
CX3C1_BOVIN	A6QNL7.1	211	2E-64	44%
CCR6_HUMAN	P51684.2	207	1E-62	35%
CCR6_MOUSE	O54689.1	206	2E-62	36%

Gene	Accession	Score	E value	Ident
CXCR3_BOVIN	Q5MD61.1	198	3E-59	36%
CXCR3_CAPHI	Q867B2.1	197	4E-59	36%
XCR1_HUMAN	P46094.1	196	9E-59	36%
CCR7_HUMAN	P32248.2	195	4E-58	38%
CCR7_MOUSE	P47774.2	194	7E-58	38%
ACKR2_RAT	O09027.1	194	2E-57	34%
CXCR1_MACMU	Q2YEG0.1	191	1E-56	34%
CXCR4_CERAT	O62747.1	190	2E-56	34%
CCR7_BOVIN	Q5MD62.1	191	2E-56	37%
CXCR4_MACFA	Q28474.2	189	5E-56	34%
CXCR2_BOVIN	Q28003.1	189	6E-56	34%
CXCR4_PANTR	P61072.1	188	9E-56	34%
CXCR1_RABIT	P21109.2	188	1E-55	36%
CXCR4_PAPAN	P56491.1	188	1E-55	33%
CXCR2_MOUSE	P35343.1	188	1E-55	37%
CXCR4_CANFA	Q3LSL6.1	187	2E-55	34%
CXCR4_RAT	O08565.1	187	2E-55	33%
ACKR4_BOVIN	P35350.1	187	3E-55	35%
CXCR2_RAT	P35407.2	187	3E-55	35%
CXCR4_MACMU	P79394.1	187	3E-55	33%
CXCR4_CALJA	Q8HZU1.1	186	4E-55	33%
CXCR6_MOUSE	Q9EQ16.1	186	5E-55	38%
CXCR4_TUPCH	Q7YS92.1	186	6E-55	33%
CXCR4_CHLAE	Q9TSQ8.1	186	6E-55	33%
CXCR4_SAISC	Q8HZU0.1	186	8E-55	33%
CXCR1_RAT	P70612.1	186	9E-55	36%
CXCR4_BOVIN	P25930.1	185	1E-54	34%
CXCR4_FELCA	P56498.1	185	2E-54	34%
CXR4B_XENLA	Q7ZXJ7.1	185	2E-54	33%
CXCR4_PIG	Q764M9.1	183	1E-53	33%
CCRL2_MOUSE	O35457.2	183	1E-53	35%
CXR32_DANRE	E9QJ73.2	183	2E-53	33%
CXCR2_RABIT	P35344.1	180	1E-52	36%
ACKR2_HUMAN	O00590.2	181	2E-52	38%
XCR1_MOUSE	Q9R0M1.1	178	3E-52	36%
CXCR4_MOUSE	P70658.2	179	4E-52	32%
ACKR2_MOUSE	O08707.2	179	4E-52	34%
ACKR4_HUMAN	Q9NPB9.1	177	2E-51	33%
CXCR6_PANTR	Q9TV16.1	176	2E-51	35%
CXCR6_HUMAN	O00574.1	175	8E-51	35%
VQ3L_SHEVK	Q86917.1	176	1E-50	35%
CXCR6_MACNE	O19024.1	175	1E-50	35%
ACKR4_MOUSE	Q92413.1	175	1E-50	33%
CXCR2_CANFA	O97571.1	174	2E-50	33%
CXCR4_XENTR	Q07FZ4.1	174	3E-50	33%
CXCR6_CHLAE	O18983.1	173	4E-50	36%
CXCR5_MOUSE	Q04683.2	174	6E-50	30%
CXCR6_CERAT	Q9N0Z0.1	172	8E-50	36%

Gene	Accession	Score	E value	Ident
CXCR6_MACMU	Q9XT45.1	171	2E-49	36%
CXR4A_XENLA	Q9YGC3.1	171	4E-49	33%
AGTRA_RAT	P25095.1	171	4E-49	31%
AGTRB_MOUSE	P29755.1	171	4E-49	31%
CXCR1_MOUSE	Q810W6.1	171	6E-49	37%
AGTRB_RAT	P29089.1	171	6E-49	31%
CXCR5_RAT	P34997.1	170	1E-48	32%
CXCR6_MACFA	Q9BDS6.1	169	2E-48	35%
AGTR1_SHEEP	O77590.2	169	2E-48	33%
AGTRA_MOUSE	P29754.1	169	2E-48	31%
AGTR1_CANFA	P43240.1	169	3E-48	31%
VK02_SWPVK	Q08520.1	169	4E-48	34%
AGTR1_BOVIN	P25104.1	168	5E-48	33%
AGTR1_MERUN	O35210.1	168	6E-48	31%
CXCR3_HUMAN	P49682.2	168	6E-48	36%
AGTR1_CAVPO	Q9WV26.1	168	7E-48	31%
AGTR1_PIG	P30555.1	167	1E-47	33%
CXCR3_RAT	Q9JII9.1	167	2E-47	34%
AGTR1_PANTR	Q9GLN9.1	166	2E-47	31%
AGTR1_RABIT	P34976.1	166	3E-47	30%
AGTR1_HUMAN	P30556.1	166	3E-47	31%
CXCR3_MOUSE	O88410.2	164	2E-46	34%
CXCR5_HUMAN	P32302.1	162	2E-45	31%
CXCR2_HUMAN	P25025.2	159	1E-44	35%
CXCR1_PANTR	P55920.1	159	2E-44	34%
CXCR2_PANTR	Q28807.2	159	2E-44	35%
CXCR1_HUMAN	P25024.2	159	2E-44	34%
CXCR1_GORGO	P55919.1	158	3E-44	34%
CXCR1_HOOHO	Q2YEG2.1	158	3E-44	34%
US28_HCMVM	F5HF62.1	158	4E-44	27%
APJ_XENTR	Q4VA82.1	158	5E-44	30%
CXCR1_PONPY	Q2YEF9.1	157	9E-44	34%
CXCR3_CANFA	Q5KSK8.1	157	1E-43	34%
CXCR2_MACMU	Q28519.1	156	2E-43	34%
CXCR2_GORGO	Q28422.1	155	3E-43	35%
APJA_XENLA	P79960.2	155	5E-43	30%
AGTRB_XENLA	P35373.1	154	2E-42	29%
AGTRA_XENLA	P32303.2	153	3E-42	29%
AGTR1_CHICK	P79785.1	152	5E-42	31%
AGTR1_MELGA	P33396.1	152	5E-42	31%
US28_HCMVA	P69332.1	152	5E-42	29%
OPRK_RAT	P34975.1	153	6E-42	30%
APJB_XENLA	Q2TAD5.1	152	1E-41	30%
OPRK_MOUSE	P33534.1	152	1E-41	30%
OPRK_CAVPO	P41144.1	151	2E-41	31%
LPAR6_CHICK	P32250.1	148	9E-41	29%
SSR4_RAT	P30937.1	149	1E-40	31%
ACKR3_RAT	O89039.2	148	3E-40	29%

Gene	Accession	Score	E value	Ident
ACKR3_MOUSE	P56485.2	147	5E-40	29%
ACKR3_HUMAN	P25106.3	145	5E-39	28%
LPAR6_MOUSE	Q8BMC0.1	144	7E-39	28%
SSR4_MOUSE	P49660.2	145	7E-39	31%
LPAR6_RAT	Q4G072.1	144	7E-39	28%
VC03_SWPVK	P32229.1	140	3E-38	32%
OPRK_HUMAN	P41145.2	142	4E-38	31%
APJB_DANRE	A0T2N3.1	141	1E-37	27%
OPRK_BOVIN	Q2KIP6.1	141	1E-37	30%
LPAR6_HUMAN	P43657.3	139	3E-37	28%
SSR4_HUMAN	P31391.2	140	3E-37	32%
GP183_HUMAN	P32249.3	136	6E-36	27%
OPRM_MACFA	Q95M54.1	137	6E-36	27%
OPRX_CAVPO	P47748.1	136	1E-35	29%
APJA_DANRE	Q7SZP9.2	135	2E-35	27%
GP183_RAT	D4A7K7.1	133	8E-35	27%
OPRM_MACMU	Q9MYW9.2	134	1E-34	26%
CXCR4_SHEEP	Q28553.1	127	2E-34	37%
OPRX_HUMAN	P41146.1	131	5E-34	29%
ACKR3_CANFA	P11613.1	131	5E-34	28%
BKRB1_HUMAN	P46663.3	131	5E-34	29%
OPRX_MOUSE	P35377.1	131	6E-34	29%
OPRX_RAT	P35370.1	131	8E-34	29%
APJ_MOUSE	Q9WV08.1	131	9E-34	29%
OPRX_PIG	P79292.1	130	2E-33	29%
SSR5_HUMAN	P35346.3	129	2E-33	29%
OPRM_MOUSE	P42866.1	130	3E-33	27%
BKRB1_MACMU	Q8HZP2.1	129	3E-33	29%
SSR1_HUMAN	P30872.1	130	3E-33	28%
SSR1_CANFA	Q49LX5.1	129	4E-33	28%
SSR1_RAT	P28646.1	129	4E-33	28%
OPRM_SAIBB	Q5IS84.1	129	4E-33	26%
SSR1_MOUSE	P30873.1	129	4E-33	29%
BKRB1_CHLPG	Q8HZP3.1	128	7E-33	29%
OPRM_HUMAN	P35372.2	129	7E-33	27%
OPRD_RAT	P33533.1	128	7E-33	28%
OPRM_BOVIN	P79350.2	129	8E-33	27%
BKRB1_TUPMI	Q8HZP1.1	128	8E-33	27%
BKRB1_CHLAE	Q95L01.1	127	9E-33	29%
OPRM_PANTR	Q5IS39.1	129	9E-33	27%
OPRM_PIG	Q95247.1	129	1E-32	27%
GP183_BOVIN	Q1RMI1.1	127	1E-32	27%
APJ_MACMU	O97666.1	128	1E-32	28%
BKRB1_MACFA	Q3BCU0.1	127	1E-32	29%
OPRD_MOUSE	P32300.1	127	2E-32	28%
APJ_RAT	Q9JHG3.1	127	2E-32	27%
OPRM_RAT	P33535.1	127	2E-32	27%
BKRB2_HUMAN	P30411.2	127	3E-32	27%

Gene	Accession	Score	E value	Ident
BKRB1_CANFA	Q9BDQ5.1	126	4E-32	29%
GP183_MOUSE	Q3U6B2.1	125	7E-32	27%
APJ_HUMAN	P35414.1	126	7E-32	28%
C3AR_DANRE	P0C7U4.1	124	2E-31	26%
BKRB2_RABIT	Q28642.1	124	2E-31	26%
BKRB1_PIG	Q8HZN9.1	123	4E-31	28%
BKRB2_RAT	P25023.4	124	5E-31	27%
OPRD_HUMAN	P41143.4	123	5E-31	27%
AGTR2_MERUN	Q9Z0Z6.1	123	6E-31	29%
CLTR2_RAT	Q924T9.1	122	7E-31	27%
AGTR2_HUMAN	P50052.1	122	8E-31	29%
FPR1_MOUSE	P33766.1	122	1E-30	26%
CLTR1_RAT	Q924T8.1	122	1E-30	27%
CLTR1_MOUSE	Q99JA4.1	121	2E-30	27%
CML1_BOVIN	B9VR26.1	121	3E-30	26%
BKRB2_MOUSE	P32299.4	122	3E-30	27%
AGTR2_MOUSE	P35374.1	121	3E-30	29%
AGTR2_RAT	P35351.1	121	3E-30	29%
GP181_DANRE	B3G515.1	120	6E-30	25%
C5AR1_RAT	P97520.1	120	6E-30	26%
GPR17_RAT	Q09QM4.1	120	6E-30	27%
GP182_MOUSE	P43142.1	120	8E-30	27%
GPR17_MOUSE	Q6NS65.1	118	2E-29	26%
BKRB1_RABIT	P48748.1	119	2E-29	27%
SSR5_BOVIN	F1MV99.2	119	2E-29	31%
LPAR4_HUMAN	Q99677.1	118	4E-29	25%
C5AR1_MOUSE	P30993.2	117	4E-29	25%
GPR17_HUMAN	Q13304.2	117	8E-29	26%
SSR2_PIG	P34994.1	117	9E-29	29%
GPR15_MOUSE	Q0VDU3.2	117	9E-29	28%
LPAR4_MOUSE	Q8BLG2.2	117	1E-28	25%
CCR10_HUMAN	P46092.3	117	1E-28	34%
SSR2_BOVIN	P34993.1	117	1E-28	29%
SSR2_MOUSE	P30875.1	117	1E-28	29%
PTAFR_CAVPO	P21556.2	116	1E-28	28%
FPR2_PANTR	P79242.1	116	2E-28	29%
CLTR2_MOUSE	Q920A1.2	115	2E-28	26%
GP181_MOUSE	Q8BMP4.2	116	2E-28	27%
SSR2_HUMAN	P30874.1	116	2E-28	28%
NPBW1_MOUSE	P49681.2	115	2E-28	27%
SSR2_RAT	P30680.1	116	2E-28	29%
SSR2_CANFA	Q49LX6.1	116	2E-28	27%
CLTR1_CAVPO	Q2NNR5.1	115	3E-28	26%
PAR1_XENLA	P47749.1	117	3E-28	25%
CLTR2_HUMAN	Q9NS75.1	115	4E-28	27%
GALR2_RAT	O08726.1	115	5E-28	27%
NPBW1_HUMAN	P48145.2	114	5E-28	28%
NPBW2_BOVIN	Q8MJV2.1	114	6E-28	27%

Gene	Accession	Score	E value	Ident
CCR10_MOUSE	Q9JL21.2	115	6E-28	33%
NPBW1_RAT	Q56UD9.1	114	8E-28	26%
FPRS6_MOUSE	Q3SXG2.1	114	9E-28	28%
NPBW1_BOVIN	Q8MJV3.1	114	1E-27	27%
FPR2_HUMAN	P25090.2	114	1E-27	26%
GP182_MOUSE	F7EQ49.1	114	1E-27	26%
GP182_HUMAN	B0F9W3.1	114	2E-27	24%
FPR2_MOUSE	O88536.1	113	2E-27	30%
GP83A_DANRE	A5PLE7.1	114	2E-27	27%
P2RY4_RAT	O35811.1	114	2E-27	24%
BKRB1_MOUSE	Q61125.1	113	2E-27	27%
P2RY1_MOUSE	P49650.1	113	2E-27	26%
GP182_HUMAN	Q99527.1	113	3E-27	25%
FPR2_GORGO	P79177.1	112	4E-27	29%
GP182_RAT	O08878.1	112	5E-27	27%
CLTR1_PIG	Q95N02.1	112	5E-27	24%
AR_BOMMO	Q8WPA2.1	112	6E-27	24%
FPR1_MACMU	P79189.1	112	7E-27	28%
BKRB2_PIG	Q9GLX8.1	112	7E-27	25%
P2RY1_MELGA	P49652.1	112	8E-27	26%
FPR1_PONPY	P79235.1	111	8E-27	27%
P2RY1_CHICK	P34996.1	112	9E-27	26%
GALR2_MOUSE	O88854.2	112	1E-26	27%
BKRB2_CAVPO	O70526.1	112	1E-26	26%
PTAFR_HUMAN	P25105.1	111	1E-26	27%
P2RY4_MOUSE	Q9JJS7.1	111	1E-26	23%
OGR1_MOUSE	Q8BFQ3.1	111	2E-26	25%
GP182_RAT	P31392.3	111	2E-26	26%
P2RY1_RAT	P49651.1	110	2E-26	26%
GPR4_RAT	Q4KLH9.1	110	3E-26	26%
GPR4_MOUSE	Q8BUD0.1	110	3E-26	26%
GPR1_RAT	P46090.1	110	3E-26	27%
GPR1_MOUSE	Q8K087.1	109	4E-26	27%
OXGR1_RAT	Q6Y1R5.1	109	4E-26	25%
P2RY1_HUMAN	P47900.1	110	5E-26	26%
FPRS7_MOUSE	Q71MR7.2	109	5E-26	26%
FPR1_GORGO	P79176.1	109	5E-26	28%
RL3R1_MOUSE	Q8BGE9.1	111	5E-26	26%
CML1_MOUSE	P97468.1	109	6E-26	27%
FPR1_HUMAN	P21462.3	109	6E-26	28%
P2RY1_CAVPO	P59902.1	109	7E-26	26%
C5AR1_PANTR	P79240.2	108	7E-26	24%
P2RY1_BOVIN	P48042.1	109	8E-26	26%
FPRS3_MOUSE	O88537.2	108	1E-25	26%
BKRB1_RAT	P97583.2	108	1E-25	26%
MCHR1_RAT	P97639.3	108	1E-25	27%
GPR4_BOVIN	Q1JQB3.1	108	2E-25	26%
CML1_RAT	O35786.1	108	2E-25	26%

Gene	Accession	Score	E value	Ident
FPR1_PANTR	P79241.1	108	2E-25	28%
PTAFR_BOVIN	Q9TTY5.2	107	2E-25	27%
OGR1_HUMAN	Q15743.1	108	2E-25	24%
GP182_HUMAN	O15218.1	108	3E-25	28%
GPR4_HUMAN	P46093.2	107	3E-25	26%
GPR4_PIG	P50132.1	107	3E-25	25%
SSR5_MOUSE	O08858.3	107	4E-25	30%
C5AR1_HUMAN	P21730.2	107	4E-25	24%
PTAFR_PIG	Q9XSD4.2	107	4E-25	27%
CLTR2_PIG	Q95N03.1	106	5E-25	26%
PTAFR_MOUSE	Q62035.1	106	5E-25	27%
GP183_SALSA	B5X337.1	107	5E-25	24%
FPR2_MACMU	P79190.1	106	6E-25	28%
FPR2_PONPY	P79236.1	106	6E-25	26%
GALR2_HUMAN	O43603.1	107	6E-25	26%
MCHR1_MACMU	Q8MJ89.2	106	9E-25	26%
OXGR1_HUMAN	Q96P68.1	105	9E-25	25%
SSR5_RAT	P30938.2	106	1E-24	30%
MCHR1_MOUSE	Q8JZL2.2	107	1E-24	27%
FPR3_PONPY	P79237.1	105	1E-24	26%
OXGR1_MOUSE	Q6IYF8.1	105	1E-24	26%
C5AR1_GORGO	P79175.2	105	1E-24	24%
P2RY4_HUMAN	P51582.1	105	1E-24	22%
CML1_HUMAN	Q99788.2	105	2E-24	27%
PAR1_PAPHA	P56488.1	106	2E-24	24%
C3AR_ONCMY	Q2WED0.1	105	2E-24	25%
C5AR1_PONPY	P79234.1	104	3E-24	23%
PAR1_HUMAN	P25116.2	105	3E-24	23%
PTAFR_CAPHI	Q9GK76.1	104	3E-24	27%
CLTR1_HUMAN	Q9Y271.1	104	3E-24	27%
MCHR1_HUMAN	Q99705.2	105	3E-24	27%
PTAFR_RAT	P46002.1	103	5E-24	26%
PAR1_RAT	P26824.1	104	7E-24	25%
PAR3_HUMAN	O00254.1	103	7E-24	25%
C5AR1_RABIT	Q9TUE1.1	103	9E-24	24%
PAR3_RAT	Q920E1.1	103	9E-24	26%
NPBW2_HUMAN	P48146.2	102	1E-23	28%
PAR3_BOVIN	Q58D85.1	103	1E-23	27%
RL3R1_HUMAN	Q9NSD7.1	104	1E-23	25%
FPR1_RABIT	Q05394.1	103	1E-23	27%
PAR1_MOUSE	P30558.2	103	1E-23	25%
CML1_PIG	B1PHQ8.1	102	1E-23	26%
FPR3_GORGO	P79178.1	102	2E-23	25%
OGR1_BOVIN	O46685.1	102	3E-23	25%
RYAR_DROME	P25931.2	103	3E-23	25%
PAR1_CRILO	Q00991.2	102	4E-23	24%
C5AR1_MACMU	P79188.1	100	1E-22	23%
FPRS4_MOUSE	A4FUQ5.2	99.0	2E-22	26%

Gene	Accession	Score	E value	Ident
FPR3_PANTR	P79243.1	99.4	3E-22	23%
NPFF2_MOUSE	Q924H0.2	100	3E-22	26%
GPR54_ORENI	Q6BD04.1	99.4	3E-22	26%
GP83B_DANRE	B0UXR0.1	98.6	6E-22	22%
P2RY8_CHICK	Q5ZI82.1	98.6	6E-22	26%
US27_HCMVA	P09703.1	98.6	6E-22	26%
US27_HCMVM	F5HDK1.1	98.2	7E-22	26%
RYAR_TRICA	G4WMX4.1	97.8	1E-21	25%
FPR3_MACMU	P79191.1	97.4	1E-21	24%
GPR35_MOUSE	Q9ES90.1	96.7	1E-21	27%
PAR1_BOVIN	A7YY44.1	98.2	1E-21	23%
SUCR1_RAT	Q6IYF9.1	96.3	2E-21	26%
SSR3_HUMAN	P32745.1	97.1	3E-21	29%
HCAR2_HUMAN	Q8TDS4.1	96.3	3E-21	24%
HCAR3_HUMAN	P49019.3	96.7	3E-21	25%
OPRD_PIG	P79291.1	93.6	4E-21	27%
PAR2_RAT	Q63645.1	96.3	4E-21	26%
PAR2_MOUSE	P55086.1	96.3	5E-21	25%
GPR15_MACFA	Q9BDS7.1	95.5	6E-21	25%
C5AR1_CANFA	P30992.1	95.5	6E-21	23%
C5AR1_DANRE	P0C7U5.1	95.1	7E-21	25%
NMBR_RAT	P24053.1	95.9	7E-21	24%
PAR4_RAT	Q920E0.1	95.1	1E-20	26%
SSR3_RAT	P30936.1	95.1	2E-20	28%
GPR15_MACMU	O97663.1	94.0	2E-20	25%
SSR3_MOUSE	P30935.2	94.4	3E-20	28%
GPR35_HUMAN	Q9HC97.4	92.8	3E-20	26%
GALR1_RAT	Q62805.1	93.6	3E-20	26%
HCAR2_MOUSE	Q9EP66.1	93.6	3E-20	24%
GPR15_MACNE	P56412.1	93.2	4E-20	25%
P2Y12_RAT	Q9EPX4.1	93.2	4E-20	28%
GALR1_HUMAN	P47211.3	93.2	4E-20	27%
NPR9_CAEEL	Q23497.1	93.6	6E-20	24%
SUCR1_MOUSE	Q99MT6.2	92.0	6E-20	23%
NPFF2_HUMAN	Q9Y5X5.2	94.0	7E-20	24%
AGTR2_SHEEP	Q28929.1	90.1	1E-19	28%
GPR15_CHLAE	O18982.1	91.7	1E-19	25%
C5AR1_CAVPO	O70129.1	91.3	2E-19	24%
P2Y12_HUMAN	Q9H244.1	91.3	2E-19	27%
P2RY3_CHICK	Q98907.1	90.9	2E-19	23%
GALR1_MOUSE	P56479.1	91.3	2E-19	25%
GPR87_HUMAN	Q9BY21.1	91.3	2E-19	27%
GPR87_MOUSE	Q99MT7.2	91.3	2E-19	28%
KISSR_RAT	Q924U1.2	91.7	2E-19	24%
RL3R2_MOUSE	Q7TQP4.1	91.3	3E-19	26%
GPR34_HUMAN	Q9UPC5.2	90.9	3E-19	24%
GPR25_MOUSE	P0C5I1.1	90.5	4E-19	27%
P2RY3_MELGA	O93361.1	90.1	4E-19	23%

Gene	Accession	Score	E value	Ident
GPR15_HUMAN	P49685.1	90.1	5E-19	24%
PAR4_MOUSE	O88634.2	90.5	5E-19	26%
P2Y14_HUMAN	Q15391.1	89.4	9E-19	26%
GPR34_MOUSE	Q9R1K6.1	89.4	9E-19	22%
GNRHR_OCTVU	Q2V2K5.1	89.7	1E-18	27%
P2Y12_MOUSE	Q9CPV9.1	89.0	1E-18	28%
GPR32_HUMAN	O75388.1	89.0	1E-18	24%
KISSR_MOUSE	Q91V45.1	89.0	2E-18	25%
TLR2_DROME	P30975.2	89.4	2E-18	23%
NMBR_MOUSE	O54799.1	88.6	2E-18	23%
TLR1_DROME	P30974.2	89.0	3E-18	25%
HCAR2_RAT	Q80Z39.1	87.8	3E-18	25%
P2RY6_HUMAN	Q15077.1	87.0	5E-18	23%
KISSR_HUMAN	Q969F8.2	86.7	1E-17	25%
GALR3_HUMAN	O60755.1	86.3	1E-17	25%
MCHR2_MACFA	Q8SQ54.1	85.9	1E-17	25%
GPR31_MOUSE	F8VQN3.1	85.5	1E-17	25%
PKR2_RAT	Q8R415.2	86.3	1E-17	26%
MCHR2_MACMU	Q8MJ88.1	85.9	1E-17	25%
PAR2_HUMAN	P55085.1	86.3	2E-17	24%
SSRL_TAKRU	O42179.1	84.3	2E-17	31%
GALR3_MOUSE	O88853.2	85.5	2E-17	25%
NMBR_HUMAN	P28336.2	85.5	3E-17	23%
GALR3_RAT	O88626.1	84.7	4E-17	24%
NPR11_CAEEL	Q18179.2	85.1	4E-17	24%
GPR18_HUMAN	Q14330.2	83.6	7E-17	24%
LT4R1_RAT	Q9R0Q2.1	83.6	8E-17	29%
AVT_CATCO	Q90352.1	84.0	1E-16	24%
GPR18_MACFA	Q4R613.1	82.4	2E-16	26%
LT4R1_MOUSE	O88855.1	82.8	2E-16	28%
NK3R_HUMAN	P29371.1	83.6	2E-16	26%
PKR1_BOVIN	Q8SPN2.1	82.8	2E-16	26%
NK2R_BOVIN	P05363.1	82.8	2E-16	23%
GP174_MOUSE	Q3U507.1	82.0	3E-16	21%
P2RY6_MOUSE	Q9ERK9.1	81.6	3E-16	21%
MCHR2_HUMAN	Q969V1.1	82.0	3E-16	25%
GP174_HUMAN	Q9BXC1.1	81.3	4E-16	22%
NK2R_CANFA	Q5DUB2.1	81.6	4E-16	23%
PKR2_BOVIN	Q8SPN1.1	81.6	5E-16	25%
OPSD_RAT	P51489.1	81.3	5E-16	25%
PKR1_HUMAN	Q8TCW9.1	81.6	5E-16	26%
GPR18_RAT	A1A5S3.1	80.9	6E-16	23%
NK2R_CAVPO	Q64077.1	81.6	6E-16	23%
GP171_BOVIN	Q3ZBK9.1	80.5	7E-16	22%
P2RY6_RAT	Q63371.1	80.1	1E-15	22%
PD2R2_HUMAN	Q9Y5Y4.3	80.9	1E-15	26%
OPN4_FELCA	Q5YKK9.1	80.1	2E-15	27%
SUCR1_HUMAN	Q9BXA5.2	79.3	2E-15	24%

Gene	Accession	Score	E value	Ident	Gene	Accession	Score	E value	Ident
PAR2_BOVIN	Q2HJA4.1	79.7	3E-15	23%	V027_FOWPN	Q9J5H4.1	76.6	2E-14	22%
GP171_HUMAN	O14626.1	78.6	3E-15	22%	NK3R_RABIT	O97512.2	77.4	2E-14	24%
HCAR1_MOUSE	Q8C131.1	79.0	3E-15	24%	GP171_MOUSE	Q8BG55.1	76.6	2E-14	23%
PD2R2_RAT	Q6XKD3.1	79.3	3E-15	26%	V206_FOWPN	Q9J529.1	76.3	2E-14	24%
PKR2_HUMAN	Q8NFI6.1	79.0	4E-15	25%	GPR18_MOUSE	Q8K126.1	76.3	2E-14	22%
SIFAR_DROME	Q8IN35.2	79.7	5E-15	24%	NK3R_MOUSE	P47937.2	77.0	2E-14	24%
NK2R_RABIT	P79218.1	79.0	5E-15	23%	TRISR_DROME	Q9VML9.4	77.4	2E-14	28%
NK2R_MOUSE	P30549.1	78.6	6E-15	24%	PKR2_MOUSE	Q8K458.1	76.6	2E-14	25%
RL3R2_HUMAN	Q8TDU9.1	78.2	7E-15	26%	OPSD_MACFA	Q28886.1	76.3	2E-14	23%
HRH2_CANFA	P17124.1	77.8	8E-15	27%	OPSD_RABIT	P49912.1	76.3	3E-14	23%
PAR4_HUMAN	Q96RI0.3	77.8	9E-15	27%	MTR1B_MOUSE	Q8CIQ6.1	76.3	3E-14	23%
NK2R_HUMAN	P21452.3	78.2	9E-15	24%	NK3R_RAT	P16177.1	76.6	3E-14	24%
OPSD_ALLMI	P52202.1	77.8	1E-14	25%	GPR33_PANTR	Q49SQ3.1	75.9	3E-14	26%
CCH1R_DROME	A1ZAX0.2	78.2	1E-14	22%	NK2R_RAT	P16610.1	76.3	3E-14	23%
OPSD_MOUSE	P15409.2	77.4	1E-14	24%	MTR1A_CHICK	P49285.1	75.9	4E-14	24%
C5AR1_ONCMY	Q6UNA4.1	77.0	2E-14	25%	TAA8A_RAT	Q923X9.2	75.9	4E-14	24%
PD2R2_MOUSE	Q9Z2J6.1	77.0	2E-14	26%	MTR1A_SHEEP	P48040.1	75.9	4E-14	24%

7.1.1.1 GO terms

7.1.1.1.1 Molecular Function

Table 7.2 - GO terms of the Molecular Function Ontology for the CCR5 Protein BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	412	82.4%	11207	23.7%	5.16e-169
GO:0008528	G-protein coupled peptide receptor activity	61	12.2%	185	0.4%	1.23e-71
GO:0001653	peptide receptor activity	61	12.2%	187	0.4%	2.67e-71
GO:0004930	G-protein coupled receptor activity	87	17.4%	1166	2.5%	4.43e-45
GO:0001637	G-protein coupled chemoattractant receptor activity	23	4.6%	41	0.1%	3.71e-33
GO:0004950	chemokine receptor activity	23	4.6%	41	0.1%	3.71e-33
GO:0004888	transmembrane signaling receptor activity	88	17.6%	1924	4.1%	2.99e-29
GO:0099600	transmembrane receptor activity	88	17.6%	2023	4.3%	1.12e-27
GO:0038023	signaling receptor activity	88	17.6%	2105	4.5%	1.90e-26
GO:0042923	neuropeptide binding	15	3.0%	22	0.0%	2.99e-23
GO:0008188	neuropeptide receptor activity	19	3.8%	53	0.1%	1.76e-22
GO:0004872	receptor activity	88	17.6%	2540	5.4%	7.65e-21
GO:0060089	molecular transducer activity	88	17.6%	2540	5.4%	7.65e-21
GO:0004871	signal transducer activity	88	17.6%	2641	5.6%	9.99e-20
GO:0004896	cytokine receptor activity	23	4.6%	147	0.3%	2.90e-18
GO:0016493	C-C chemokine receptor activity	11	2.2%	16	0.0%	6.89e-17
GO:0019956	chemokine binding	12	2.4%	31	0.1%	2.01e-14
GO:0001608	G-protein coupled nucleotide receptor activity	9	1.8%	18	0.0%	6.85e-12
GO:0045028	G-protein coupled purinergic nucleotide receptor activity	9	1.8%	18	0.0%	6.85e-12
GO:0004994	somatostatin receptor activity	5	1.0%	5	0.0%	1.28e-08

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:0016494	C-X-C chemokine receptor activity	6	1.2%	11	0.0%	5.96e-08
GO:0001614	purinergic nucleotide receptor activity	9	1.8%	46	0.1%	1.19e-07
GO:0016502	nucleotide receptor activity	9	1.8%	46	0.1%	1.19e-07
GO:0019955	cytokine binding	12	2.4%	111	0.2%	2.62e-07
GO:0004985	opioid receptor activity	6	1.2%	14	0.0%	3.77e-07
GO:0035586	purinergic receptor activity	9	1.8%	54	0.1%	5.34e-07
GO:0019957	C-C chemokine binding	5	1.0%	11	0.0%	5.64e-06
GO:0019958	C-X-C chemokine binding	5	1.0%	11	0.0%	5.64e-06
GO:0042277	peptide binding	19	3.8%	418	0.9%	1.47e-05
GO:0015057	thrombin-activated receptor activity	4	0.8%	7	0.0%	4.18e-05
GO:0033218	amide binding	19	3.8%	462	1.0%	6.71e-05
GO:0001631	cysteinyl leukotriene receptor activity	3	0.6%	3	0.0%	0.00011
GO:0071791	chemokine (C-C motif) ligand 5 binding	3	0.6%	3	0.0%	0.00011
GO:0004966	galanin receptor activity	3	0.6%	4	0.0%	0.00046
GO:0015065	uridine nucleotide receptor activity	3	0.6%	4	0.0%	0.00046
GO:0071553	G-protein coupled pyrimidinergic nucleotide receptor activity	3	0.6%	4	0.0%	0.00046
GO:0004945	angiotensin type II receptor activity	3	0.6%	7	0.0%	0.00395
GO:0001595	angiotensin receptor activity	3	0.6%	8	0.0%	0.00627
GO:0004974	leukotriene receptor activity	3	0.6%	9	0.0%	0.00934
GO:0031681	G-protein beta-subunit binding	3	0.6%	9	0.0%	0.00934

7.1.1.1.2 Cellular Component

Table 7.3 - GO terms of the Cellular Component Ontology for the CCR5 Protein BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	412	82.4%	11988	25.4%	1.32e-157
GO:0005887	integral component of plasma membrane	76	15.2%	1917	4.1%	3.54e-21
GO:0031226	intrinsic component of plasma membrane	76	15.2%	1982	4.2%	2.57e-20
GO:0044459	plasma membrane part	77	15.4%	3409	7.2%	2.46e-08
GO:0032590	dendrite membrane	4	0.8%	21	0.0%	0.00597

7.1.1.1.3 Biological Process

Table 7.4 - GO terms of the Biological Process Ontology for the CCR5 Protein BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	412	82.4%	12072	25.6%	2.76e-155
GO:0007186	G-protein coupled receptor signaling pathway	88	17.6%	1702	3.6%	4.46e-32
GO:0070098	chemokine-mediated signaling pathway	23	4.6%	97	0.2%	1.54e-21
GO:0007218	neuropeptide signaling pathway	22	4.4%	116	0.2%	3.10e-18

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0007200	phospholipase C-activating G-protein coupled receptor signaling pathway	20	4.0%	89	0.2%	5.48e-18
GO:0007204	positive regulation of cytosolic calcium ion concentration	29	5.8%	262	0.6%	9.27e-18
GO:0051480	regulation of cytosolic calcium ion concentration	29	5.8%	283	0.6%	8.01e-17
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	24	4.8%	185	0.4%	5.21e-16
GO:0006874	cellular calcium ion homeostasis	29	5.8%	382	0.8%	2.72e-13
GO:0055074	calcium ion homeostasis	29	5.8%	396	0.8%	7.03e-13
GO:0072503	cellular divalent inorganic cation homeostasis	29	5.8%	409	0.9%	1.63e-12
GO:0072507	divalent inorganic cation homeostasis	29	5.8%	429	0.9%	5.64e-12
GO:0006875	cellular metal ion homeostasis	30	6.0%	511	1.1%	7.72e-11
GO:0035589	G-protein coupled purinergic nucleotide receptor signaling pathway	9	1.8%	18	0.0%	8.83e-11
GO:0055065	metal ion homeostasis	30	6.0%	583	1.2%	2.23e-09
GO:0030003	cellular cation homeostasis	30	6.0%	588	1.2%	2.76e-09
GO:0006873	cellular ion homeostasis	30	6.0%	602	1.3%	4.97e-09
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	16	3.2%	158	0.3%	2.06e-08
GO:0035588	G-protein coupled purinergic receptor signaling pathway	9	1.8%	31	0.1%	3.24e-08
GO:0006935	chemotaxis	30	6.0%	659	1.4%	4.59e-08
GO:0042330	taxis	30	6.0%	660	1.4%	4.77e-08
GO:0055080	cation homeostasis	30	6.0%	677	1.4%	8.82e-08
GO:0098771	inorganic ion homeostasis	30	6.0%	692	1.5%	1.49e-07
GO:0038169	somatostatin receptor signaling pathway	5	1.0%	5	0.0%	1.66e-07
GO:0038170	somatostatin signaling pathway	5	1.0%	5	0.0%	1.66e-07
GO:0055082	cellular chemical homeostasis	31	6.2%	758	1.6%	3.03e-07
GO:0050801	ion homeostasis	30	6.0%	756	1.6%	1.20e-06
GO:0035590	purinergic nucleotide receptor signaling pathway	9	1.8%	47	0.1%	1.88e-06
GO:0002407	dendritic cell chemotaxis	7	1.4%	22	0.0%	2.70e-06
GO:0007193	adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway	10	2.0%	74	0.2%	8.06e-06
GO:0030815	negative regulation of cAMP metabolic process	8	1.6%	41	0.1%	1.33e-05
GO:0036336	dendritic cell migration	7	1.4%	27	0.1%	1.34e-05
GO:0035587	purinergic receptor signaling pathway	9	1.8%	60	0.1%	1.81e-05
GO:0038003	opioid receptor signaling pathway	6	1.2%	17	0.0%	1.95e-05
GO:0030800	negative regulation of cyclic nucleotide metabolic process	8	1.6%	46	0.1%	3.48e-05
GO:0019725	cellular homeostasis	31	6.2%	935	2.0%	4.08e-05
GO:0044060	regulation of endocrine process	8	1.6%	47	0.1%	4.15e-05
GO:0006954	inflammatory response	26	5.2%	731	1.5%	0.00013
GO:0030818	negative regulation of cAMP biosynthetic process	7	1.4%	37	0.1%	0.00014
GO:0007194	negative regulation of adenylate cyclase activity	6	1.2%	24	0.1%	0.00019
GO:0030814	regulation of cAMP metabolic process	11	2.2%	130	0.3%	0.00019
GO:0006968	cellular defense response	8	1.6%	59	0.1%	0.00026
GO:0019221	cytokine-mediated signaling pathway	25	5.0%	706	1.5%	0.00026

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0030803	negative regulation of cyclic nucleotide biosynthetic process	7	1.4%	41	0.1%	0.00029
GO:0007631	feeding behavior	10	2.0%	108	0.2%	0.00031
GO:0030809	negative regulation of nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00042
GO:1900372	negative regulation of purine nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00042
GO:0031280	negative regulation of cyclase activity	6	1.2%	28	0.1%	0.00053
GO:1900543	negative regulation of purine nucleotide metabolic process	8	1.6%	67	0.1%	0.00071
GO:0048878	chemical homeostasis	32	6.4%	1123	2.4%	0.00071
GO:0045980	negative regulation of nucleotide metabolic process	8	1.6%	68	0.1%	0.00080
GO:0051350	negative regulation of lyase activity	6	1.2%	30	0.1%	0.00083
GO:0045761	regulation of adenylate cyclase activity	8	1.6%	70	0.1%	0.00100
GO:0046058	cAMP metabolic process	11	2.2%	154	0.3%	0.00110
GO:0030595	leukocyte chemotaxis	12	2.4%	191	0.4%	0.00138
GO:0002685	regulation of leukocyte migration	11	2.2%	161	0.3%	0.00170
GO:0030799	regulation of cyclic nucleotide metabolic process	11	2.2%	162	0.3%	0.00181
GO:0051930	regulation of sensory perception of pain	6	1.2%	36	0.1%	0.00258
GO:0051931	regulation of sensory perception	6	1.2%	36	0.1%	0.00258
GO:0071345	cellular response to cytokine stimulus	26	5.2%	856	1.8%	0.00267
GO:0002687	positive regulation of leukocyte migration	9	1.8%	110	0.2%	0.00356
GO:0030817	regulation of cAMP biosynthetic process	9	1.8%	114	0.2%	0.00479
GO:0050886	endocrine process	8	1.6%	86	0.2%	0.00489
GO:0050900	leukocyte migration	16	3.2%	385	0.8%	0.00571
GO:0070374	positive regulation of ERK1 and ERK2 cascade	11	2.2%	184	0.4%	0.00624
GO:0031279	regulation of cyclase activity	8	1.6%	90	0.2%	0.00688
GO:0070493	thrombin-activated receptor signaling pathway	4	0.8%	12	0.0%	0.00731
GO:0051339	regulation of lyase activity	8	1.6%	92	0.2%	0.00811

7.1.2 CXCR4

Table 7.5 - Protein BLAST most relevant results for query protein CXCR4_HUMAN (P61073). Score equals Total Score. Default parameters were used with cut-off value of 1×10^{-7} , limited to 500 hits.

Gene	Accession	Score	E value	Ident
CXCR4_PANTR	P61072.1	723	0.0	100%
CXCR4_PAPAN	P56491.1	715	0.0	99%
CXCR4_SAISC	Q8HZU0.1	714	0.0	99%
CXCR4_MACFA	Q28474.2	714	0.0	99%
CXCR4_CALJA	Q8HZU1.1	712	0.0	98%
CXCR4_MACMU	P79394.1	711	0.0	98%
CXCR4_CHLAE	Q9TSQ8.1	711	0.0	98%
CXCR4_TUPCH	Q7YS92.1	710	0.0	98%
CXCR4_CERAT	O62747.1	708	0.0	98%

Gene	Accession	Score	E value	Ident
CXCR4_CANFA	Q3LSL6.1	660	0.0	95%
CXCR4_PIG	Q764M9.1	655	0.0	94%
CXCR4_FELCA	P56498.1	655	0.0	94%
CXCR4_BOVIN	P25930.1	648	0.0	92%
CXCR4_RAT	O08565.1	629	0.0	90%
CXCR4_MOUSE	P70658.2	629	0.0	90%
CXR4B_XENLA	Q7ZXJ7.1	518	0.0	74%
CXCR4_XENTR	Q07FZ4.1	498	2E-177	74%
CXR4A_XENLA	Q9YGC3.1	496	3E-176	74%
CXCR4_SHEEP	Q28553.1	370	3E-129	91%
CCR4_HUMAN	P51679.1	228	4E-71	40%
CXCR3_CAPHI	Q867B2.1	223	8E-69	35%
CXCR3_BOVIN	Q5MD61.1	223	8E-69	38%

Gene	Accession	Score	E value	Ident
CCR4_CANFA	Q8MJW8.1	222	8E-69	38%
CCR4_MOUSE	P51680.2	221	3E-68	38%
CXCR1_RABIT	P21109.2	217	1E-66	35%
CXCR2_RAT	P35407.2	216	2E-66	37%
CCR9_HUMAN	P51686.2	214	2E-65	34%
CCR6_MOUSE	O54689.1	213	3E-65	37%
CXCR1_MACMU	Q2YEG0.1	213	3E-65	37%
CCR9_MOUSE	Q9WUT7.1	211	3E-64	34%
CCR9_SHEEP	Q1WLP9.1	203	2E-61	34%
CCR7_MOUSE	P47774.2	203	3E-61	34%
CXCR5_MOUSE	Q04683.2	202	5E-61	34%
CXR32_DANRE	E9QJ73.2	202	6E-61	34%
CCR7_HUMAN	P32248.2	200	5E-60	33%
CXCR2_CANFA	O97571.1	199	1E-59	36%
CCR2_HUMAN	P41597.1	199	1E-59	32%
CXCR5_RAT	P34997.1	198	2E-59	34%
CXCR2_MOUSE	P35343.1	197	3E-59	37%
CXCR1_HOOHO	Q2YEG2.1	197	4E-59	39%
CXCR2_RABIT	P35344.1	196	7E-59	35%
CCR7_BOVIN	Q5MD62.1	197	1E-58	34%
CCR2_MACMU	O18793.2	196	1E-58	35%
CXCR2_BOVIN	Q28003.1	196	1E-58	34%
CXCR1_RAT	P70612.1	196	1E-58	38%
CCR2_RAT	O55193.1	195	5E-58	35%
CCR5_BOVIN	Q2HJ17.1	193	1E-57	35%
CCR6_HUMAN	P51684.2	194	1E-57	33%
CXCR6_PANTR	Q9TV16.1	193	1E-57	33%
CCR2_MOUSE	P51683.2	194	1E-57	35%
CCR5_CERSO	Q9BGN6.1	192	2E-57	34%
CXCR5_HUMAN	P32302.1	193	2E-57	34%
CXCR6_HUMAN	O00574.1	192	2E-57	33%
CCR5_CERLH	Q9XT76.1	192	2E-57	34%
CXCR1_PANTR	P55920.1	192	2E-57	39%
CXCR6_CHLAE	O18983.1	192	3E-57	33%
CXCR6_CERAT	Q9N0Z0.1	192	3E-57	33%
CCR8_MOUSE	P56484.1	192	3E-57	33%
CXCR6_MACNE	O19024.1	191	4E-57	33%
CXCR1_GORGO	P55919.1	191	5E-57	39%
CXCR3_HUMAN	P49682.2	191	1E-56	37%
CXCR1_HUMAN	P25024.2	189	3E-56	38%
CXCR3_RAT	Q9JII9.1	189	6E-56	37%
CXCR6_MACMU	Q9XT45.1	188	6E-56	32%
CXCR1_PONPY	Q2YEF9.1	188	1E-55	39%
CXCR3_MOUSE	O88410.2	188	2E-55	36%
ACKR4_HUMAN	Q9NPB9.1	187	2E-55	35%
CXCR6_MACFA	Q9BDS6.1	187	2E-55	32%
CCR5_NOMLE	O97883.1	187	2E-55	35%
CCR5_CANFA	Q5ECR9.1	187	3E-55	34%

Gene	Accession	Score	E value	Ident
ACKR4_BOVIN	P35350.1	186	5E-55	35%
CXCR6_MOUSE	Q9EQ16.1	186	5E-55	34%
AGTR1_BOVIN	P25104.1	186	1E-54	34%
CCR5_CHLAE	P56493.1	185	2E-54	34%
AGTR1_PIG	P30555.1	185	2E-54	35%
CCR5_CERCP	Q9TV47.1	185	2E-54	34%
CCR5_GORGO	P56439.1	184	2E-54	34%
CCR5_ERYPA	Q95ND0.1	184	3E-54	34%
CCR5_MACMU	P61813.1	184	3E-54	34%
CCR5_MACAR	O97975.1	184	3E-54	34%
CCR5_MIOTA	Q95NC3.1	184	3E-54	34%
CCR8_MACMU	O97665.1	184	4E-54	32%
CXCR3_CANFA	Q5KSK8.1	184	4E-54	38%
CCR8_HUMAN	P51685.1	184	4E-54	33%
CCR5_CHLTN	Q95NE8.1	184	4E-54	34%
CCR5_MANLE	Q95ND2.1	184	4E-54	34%
CCR5_CHLSB	Q9TV43.1	184	4E-54	34%
CCR5_CERGA	Q9TV49.1	184	4E-54	34%
ACKR4_MOUSE	Q924I3.1	184	5E-54	35%
CCR5_LOPAT	P61755.1	184	5E-54	34%
CCR5_MANSP	Q95ND1.1	184	5E-54	34%
CCR5_SYMSY	Q95NC5.1	184	6E-54	34%
CCR5_CHLPG	Q9TV42.1	184	6E-54	34%
CCR5_CERAS	Q9TV48.1	183	6E-54	34%
CCR5_PONPY	O97881.1	183	6E-54	34%
CCR5_THEGE	Q95NC1.1	183	6E-54	34%
CCR5_CERAT	O62743.2	183	7E-54	34%
CCR5_RHIAV	O97962.1	183	9E-54	34%
CCR5_SAISC	Q8HZT9.1	183	1E-53	34%
CCR5_PANTR	P56440.1	182	1E-53	34%
CCR5_TRAFR	O97878.1	182	1E-53	34%
CCR5_TRAJO	Q95NC6.1	182	2E-53	34%
CCR5_COLPO	Q95NC8.1	182	2E-53	34%
CCR5_TRAPH	O97879.1	182	2E-53	34%
CCR1_HUMAN	P32246.1	182	2E-53	32%
CCR5_NASLA	Q95NC7.1	182	2E-53	34%
CCR5_RHIBE	O97880.1	182	2E-53	34%
CCR5_CALMO	Q95NC2.1	182	2E-53	33%
CCR5_HUMAN	P51681.1	182	2E-53	34%
CCR5_MOUSE	P51682.3	182	3E-53	33%
CCR5_PYGNE	O97882.1	182	3E-53	33%
AGTR1_SHEEP	O77590.2	182	3E-53	34%
AGTR1_PANTR	Q9GLN9.1	181	5E-53	34%
AGTR1_HUMAN	P30556.1	181	5E-53	34%
CCR3_MOUSE	P51678.2	181	6E-53	31%
CCR5_CALHU	Q6WN98.1	181	8E-53	33%
CCR5_HYLML	Q95NC0.1	181	9E-53	34%
CCR3_CAVPO	Q9Z2I3.1	181	1E-52	34%

Gene	Accession	Score	E value	Ident
CCR3_CANFA	Q64H34.1	181	1E-52	31%
CCR5_ATEGE	Q95NC4.1	180	1E-52	33%
ACKR2_RAT	O09027.1	181	1E-52	32%
CCR5_CERNS	Q9TV45.1	179	2E-52	34%
AGTR1_CANFA	P43240.1	179	3E-52	34%
CX3C1_HUMAN	P49238.1	179	3E-52	34%
AGTR1_MERUN	O35210.1	179	6E-52	34%
AGTRB_RAT	P29089.1	178	8E-52	33%
AGTR1_CAVPO	Q9WV26.1	178	1E-51	34%
AGTRA_MOUSE	P29754.1	178	1E-51	34%
AGTR1_RABIT	P34976.1	177	1E-51	33%
CXCR2_MACMU	Q28519.1	177	1E-51	39%
AGTRB_XENLA	P35373.1	177	3E-51	32%
AGTRA_RAT	P25095.1	176	3E-51	33%
CXCR2_HUMAN	P25025.2	176	5E-51	38%
CXCR2_PANTR	Q28807.2	176	5E-51	38%
AGTRA_XENLA	P32303.2	176	6E-51	33%
AGTRB_MOUSE	P29755.1	176	7E-51	33%
CXCR2_GORGO	Q28422.1	176	8E-51	38%
CX3C1_RABIT	Q2KTE1.1	176	8E-51	35%
CCR5_RAT	O08556.1	174	3E-50	34%
CCR5_RABIT	Q1ZY22.1	173	8E-50	33%
CCR3_RAT	O54814.1	171	3E-49	30%
CXCR1_MOUSE	Q810W6.1	171	4E-49	35%
CC1L1_MOUSE	P51676.2	171	6E-49	32%
ACKR3_HUMAN	P25106.3	170	8E-49	27%
CCR3_CHLAE	P56492.1	170	9E-49	32%
ACKR2_MOUSE	O08707.2	171	1E-48	31%
CCR3_MACMU	P56483.1	170	1E-48	31%
CCR5_ALOSE	Q95NC9.1	169	1E-48	33%
ACKR3_MOUSE	P56485.2	170	1E-48	28%
ACKR3_RAT	O89039.2	169	2E-48	29%
CCR3_MACFA	Q9BDS8.1	169	3E-48	31%
CCR3_HUMAN	P51677.1	168	5E-48	31%
XCR1_HUMAN	P46094.1	167	1E-47	32%
CX3C1_MOUSE	Q9Z0D9.1	166	2E-47	34%
US28_HCMVM	F5HF62.1	165	6E-47	30%
ACKR2_HUMAN	O00590.2	165	2E-46	33%
CCR1_MOUSE	P51675.2	162	1E-45	34%
VGE1_EHV2	Q89609.1	162	2E-45	32%
SSR4_RAT	P30937.1	161	4E-45	33%
APJB_XENLA	Q2TAD5.1	160	5E-45	29%
CX3C1_RAT	P35411.1	160	5E-45	33%
US28_HCMVA	P69332.1	160	6E-45	31%
APJA_XENLA	P79960.2	159	1E-44	29%
SSR1_MOUSE	P30873.1	160	2E-44	31%
SSR1_RAT	P28646.1	160	2E-44	31%
APJA_DANRE	Q7SZP9.2	157	7E-44	31%

Gene	Accession	Score	E value	Ident
SSR4_MOUSE	P49660.2	157	2E-43	32%
APJ_XENTR	Q4VA82.1	156	2E-43	29%
FPR3_PONPY	P79237.1	154	1E-42	33%
ACKR3_CANFA	P11613.1	154	1E-42	29%
FPR3_GORGO	P79178.1	154	1E-42	32%
SSR1_HUMAN	P30872.1	154	2E-42	31%
SSR1_CANFA	Q49LX5.1	154	3E-42	31%
FPR3_MACMU	P79191.1	152	7E-42	31%
FPR2_PANTR	P79242.1	152	7E-42	33%
FPR3_PANTR	P79243.1	152	8E-42	32%
SSR4_HUMAN	P31391.2	152	1E-41	32%
XCR1_MOUSE	Q9ROM1.1	150	1E-41	31%
FPR2_MACMU	P79190.1	150	2E-41	34%
APJB_DANRE	A0T2N3.1	150	4E-41	28%
FPR3_HUMAN	P25089.2	150	5E-41	33%
APJ_RAT	Q9JHG3.1	150	6E-41	29%
FPRS7_MOUSE	Q71MR7.2	149	9E-41	32%
APJ_MOUSE	Q9WV08.1	149	9E-41	30%
FPRS6_MOUSE	Q3SXG2.1	148	2E-40	32%
FPR2_HUMAN	P25090.2	148	2E-40	32%
CX3C1_BOVIN	A6QNL7.1	148	3E-40	32%
FPR2_GORGO	P79177.1	147	3E-40	33%
FPR2_MOUSE	O88536.1	146	9E-40	31%
FPR1_MACMU	P79189.1	146	1E-39	34%
GPR15_MOUSE	Q0VDU3.2	146	1E-39	29%
APJ_MACMU	O97666.1	144	7E-39	29%
FPR1_HUMAN	P21462.3	144	9E-39	34%
VK02_SWPVK	Q08520.1	144	1E-38	28%
SSR5_HUMAN	P35346.3	144	1E-38	32%
FPR2_PONPY	P79236.1	143	2E-38	33%
FPR1_PONPY	P79235.1	143	2E-38	33%
FPR1_PANTR	P79241.1	142	2E-38	33%
APJ_HUMAN	P35414.1	143	3E-38	28%
FPR1_MOUSE	P33766.1	142	6E-38	31%
OXGR1_RAT	Q6Y1R5.1	139	3E-37	29%
AGTR1_MELGA	P33396.1	140	3E-37	32%
AGTR1_CHICK	P79785.1	140	3E-37	32%
C3AR_DANRE	P0C7U4.1	140	3E-37	29%
FPRS4_MOUSE	A4FUQ5.2	139	4E-37	32%
FPR1_GORGO	P79176.1	139	4E-37	32%
GPR1_MOUSE	Q8K087.1	139	8E-37	32%
OXGR1_MOUSE	Q6IYF8.1	138	1E-36	30%
C5AR1_RAT	P97520.1	138	2E-36	31%
VQ3L_SHEVK	Q86917.1	138	2E-36	30%
SSR2_BOVIN	P34993.1	138	2E-36	30%
SSR2_MOUSE	P30875.1	138	2E-36	29%
FPR1_RABIT	Q05394.1	137	2E-36	31%
SSR5_BOVIN	F1MV99.2	137	5E-36	32%

Gene	Accession	Score	E value	Ident
SSR2_RAT	P30680.1	136	7E-36	30%
SSR2_PIG	P34994.1	135	1E-35	31%
CCR1_MACFA	Q2Y2P0.1	135	1E-35	32%
BKRB2_RAT	P25023.4	136	1E-35	28%
BKRB2_CAVPO	O70526.1	135	2E-35	29%
SSR5_MOUSE	O08858.3	135	2E-35	32%
CCR1_MACMU	P56482.1	135	2E-35	32%
SSR2_HUMAN	P30874.1	135	2E-35	31%
GPR1_RAT	P46090.1	134	4E-35	33%
SSR2_CANFA	Q49LX6.1	134	4E-35	30%
CML1_BOVIN	B9VR26.1	133	1E-34	30%
C5AR1_MOUSE	P30993.2	132	1E-34	28%
SSR5_RAT	P30938.2	132	2E-34	31%
BKRB2_HUMAN	P30411.2	132	3E-34	29%
FPRS3_MOUSE	O88537.2	131	3E-34	31%
OPRX_CAVPO	P47748.1	131	5E-34	31%
LPAR6_CHICK	P32250.1	129	9E-34	30%
OPRK_MOUSE	P33534.1	131	1E-33	28%
GPR15_MACMU	O97663.1	130	1E-33	29%
OXGR1_HUMAN	Q96P68.1	129	1E-33	29%
FPRS1_MOUSE	O08790.2	129	2E-33	31%
OPRK_RAT	P34975.1	129	3E-33	28%
GPR15_MACFA	Q9BDS7.1	129	4E-33	28%
BKRB1_CANFA	Q9BDQ5.1	129	4E-33	28%
OPRK_CAVPO	P41144.1	129	5E-33	27%
P2RY4_RAT	O35811.1	128	1E-32	28%
GPR15_MACNE	P56412.1	127	1E-32	28%
C5AR1_PANTR	P79240.2	127	1E-32	27%
OPRX_PIG	P79292.1	127	1E-32	28%
BKRB2_MOUSE	P32299.4	128	2E-32	29%
BKRB2_PIG	Q9GLX8.1	127	2E-32	30%
OPRX_RAT	P35370.1	127	3E-32	30%
C3AR_ONCMY	Q2WED0.1	126	3E-32	27%
P2RY4_MOUSE	Q9JJS7.1	126	4E-32	27%
OPRX_MOUSE	P35377.1	126	4E-32	30%
OPRX_HUMAN	P41146.1	126	4E-32	30%
GPR15_HUMAN	P49685.1	125	6E-32	29%
P2RY1_MELGA	P49652.1	125	7E-32	28%
P2RY1_CHICK	P34996.1	125	8E-32	28%
GPR15_CHLAE	O18982.1	125	9E-32	28%
CML1_MOUSE	P97468.1	125	1E-31	28%
LPAR6_RAT	Q4G072.1	124	1E-31	30%
GPR1_MACFA	Q95LH1.1	124	2E-31	28%
GP182_MOUSE	P43142.1	125	2E-31	26%
LPAR6_MOUSE	Q8BMC0.1	123	3E-31	30%
CCRL2_MACMU	Q9XSD7.1	123	4E-31	28%
OPRK_BOVIN	Q2KIP6.1	124	4E-31	28%
CML1_RAT	O35786.1	123	5E-31	29%

Gene	Accession	Score	E value	Ident
RL3R1_MOUSE	Q8BGE9.1	125	5E-31	25%
GPR1_MACMU	O97664.1	123	6E-31	28%
C5AR1_GORGO	P79175.2	123	6E-31	27%
LPAR4_HUMAN	Q99677.1	123	7E-31	30%
GP182_RAT	P31392.3	123	9E-31	26%
C5AR1_HUMAN	P21730.2	122	9E-31	26%
BKRB2_RABIT	Q28642.1	122	1E-30	28%
CLTR1_MOUSE	Q99JA4.1	122	1E-30	29%
CCRL2_BOVIN	Q0II78.1	122	1E-30	26%
CCRL2_HUMAN	O00421.2	122	1E-30	28%
OPRK_HUMAN	P41145.2	122	1E-30	28%
CML1_PIG	B1PHQ8.1	122	1E-30	29%
CLTR1_RAT	Q924T8.1	121	2E-30	30%
VC03_SWPVK	P32229.1	119	2E-30	27%
GP183_HUMAN	P32249.3	122	2E-30	28%
V027_FOWPN	Q9J5H4.1	120	3E-30	27%
LPAR6_HUMAN	P43657.3	120	3E-30	30%
C5AR1_RABIT	Q9TUE1.1	120	4E-30	28%
BKRB1_TUPMI	Q8HZP1.1	120	4E-30	31%
LPAR4_MOUSE	Q8BLG2.2	120	5E-30	30%
CLTR2_MOUSE	Q920A1.2	119	9E-30	30%
P2RY4_HUMAN	P51582.1	120	9E-30	26%
CML1_HUMAN	Q99788.2	120	1E-29	29%
GP183_RAT	D4A7K7.1	119	1E-29	24%
C5AR1_PONPY	P79234.1	119	2E-29	25%
BKRB1_RABIT	P48748.1	119	2E-29	29%
P2RY3_MELGA	O93361.1	118	2E-29	28%
OPRM_MACFA	Q95M54.1	119	2E-29	28%
P2RY3_CHICK	Q98907.1	117	3E-29	28%
NPBW1_MOUSE	P49681.2	117	3E-29	28%
P2RY1_MOUSE	P49650.1	118	3E-29	27%
BKRB1_MOUSE	Q61125.1	117	3E-29	30%
BKRB1_HUMAN	P46663.3	118	4E-29	28%
RL3R1_HUMAN	Q9NSD7.1	119	4E-29	25%
P2RY4_XENLA	P79928.1	119	9E-29	29%
NPBW1_RAT	Q56UD9.1	116	9E-29	28%
OGR1_HUMAN	Q15743.1	117	9E-29	29%
OPRM_MACMU	Q9MYW9.2	117	1E-28	28%
BKRB1_MACMU	Q8HZP2.1	117	1E-28	29%
P2RY1_CAVPO	P59902.1	117	1E-28	26%
GP182_HUMAN	O15218.1	117	1E-28	25%
BKRB1_CHLAE	Q95L01.1	116	1E-28	28%
GP183_SALSA	B5X337.1	117	1E-28	28%
MCHR1_MOUSE	Q8JZL2.2	117	1E-28	27%
CCRL2_PIG	Q75ZH0.1	116	1E-28	27%
NPBW2_BOVIN	Q8MJV2.1	116	2E-28	28%
P2RY1_RAT	P49651.1	116	2E-28	25%
C5AR1_MACMU	P79188.1	115	2E-28	26%

Gene	Accession	Score	E value	Ident
BKRB1_MACFA	Q3BCU0.1	115	2E-28	29%
CLTR1_CAVPO	Q2NNR5.1	115	2E-28	29%
GALR2_RAT	O08726.1	116	3E-28	27%
OGR1_MOUSE	Q8BFQ3.1	115	3E-28	28%
GPR1_HUMAN	P46091.2	115	5E-28	28%
GPR54_ORENI	Q6BD04.1	115	5E-28	26%
GP183_MOUSE	Q3U6B2.1	115	6E-28	27%
MCHR1_MACMU	Q8MJ89.2	115	8E-28	27%
GP183_BOVIN	Q1RMI1.1	114	8E-28	27%
OPRD_RAT	P33533.1	114	9E-28	25%
OPRM_MOUSE	P42866.1	115	9E-28	29%
MCHR1_RAT	P97639.3	114	1E-27	28%
BKRB1_CHLPG	Q8HZP3.1	114	1E-27	27%
P2RY1_BOVIN	P48042.1	114	1E-27	26%
C5AR1_DANRE	P0C7U5.1	113	2E-27	30%
BKRB1_RAT	P97583.2	113	2E-27	29%
CLTR2_PIG	Q95N03.1	112	3E-27	31%
GP1R1_DANRE	B3G515.1	112	4E-27	29%
OPRD_MOUSE	P32300.1	112	4E-27	27%
OPRM_PIG	Q95247.1	113	4E-27	29%
OPRD_HUMAN	P41143.4	112	5E-27	27%
OPRM_RAT	P33535.1	113	5E-27	29%
P2RY1_HUMAN	P47900.1	112	6E-27	25%
CCR10_HUMAN	P46092.3	112	6E-27	31%
MCHR1_HUMAN	Q99705.2	113	6E-27	27%
OPRM_HUMAN	P35372.2	112	9E-27	28%
OPRM_SAIBB	Q5IS84.1	112	1E-26	28%
OPRM_BOVIN	P79350.2	112	1E-26	28%
OPRM_PANTR	Q5IS39.1	112	1E-26	28%
CLTR1_HUMAN	Q9Y271.1	110	2E-26	32%
C5AR1_CANFA	P30992.1	109	4E-26	28%
CLTR1_PIG	Q95N02.1	109	5E-26	30%
SSR3_HUMAN	P32745.1	110	5E-26	29%
OGR1_BOVIN	O46685.1	109	5E-26	27%
GPR87_HUMAN	Q9BY21.1	109	5E-26	28%
NPBW1_HUMAN	P48145.2	108	8E-26	27%
GALR2_MOUSE	O88854.2	109	8E-26	27%
AR_BOMMO	Q8WPA2.1	108	1E-25	28%
CLTR2_RAT	Q924T9.1	107	2E-25	29%
GPR17_HUMAN	Q13304.2	108	2E-25	27%
GP1R1_MICUN	B0F9W3.1	108	2E-25	29%
NPBW2_HUMAN	P48146.2	107	2E-25	28%
GALR2_HUMAN	O43603.1	108	2E-25	28%
CLTR2_HUMAN	Q9NS75.1	107	2E-25	31%
GPR87_MOUSE	Q99MT7.2	107	3E-25	27%
GP1R1_MACMU	F7EQ49.1	107	4E-25	27%
GP1R1_HUMAN	Q99527.1	107	6E-25	27%
GPR25_MOUSE	P0C5I1.1	106	9E-25	27%

Gene	Accession	Score	E value	Ident
NPBW1_BOVIN	Q8MJV3.1	105	1E-24	29%
P2Y12_RAT	Q9EPX4.1	105	2E-24	25%
SSR3_RAT	P30936.1	106	2E-24	28%
SSR3_MOUSE	P30935.2	106	2E-24	28%
GPR4_RAT	Q4KLH9.1	104	3E-24	26%
GP1R1_MOUSE	Q8BMP4.2	105	3E-24	28%
GALR1_RAT	Q62805.1	104	4E-24	28%
GP83A_DANRE	A5PLE7.1	104	5E-24	29%
CCR10_MOUSE	Q9JL21.2	104	5E-24	31%
P2RY6_HUMAN	Q15077.1	103	6E-24	27%
GPR17_RAT	Q09QM4.1	103	7E-24	26%
GPR25_HUMAN	O00155.2	103	8E-24	29%
GALR1_MOUSE	P56479.1	103	9E-24	29%
GPR17_MOUSE	Q6NS65.1	103	1E-23	26%
GPR4_BOVIN	Q1JQB3.1	103	1E-23	26%
C5AR2_RAT	Q695P6.1	102	1E-23	26%
P2Y12_HUMAN	Q9H244.1	102	2E-23	25%
GPR4_PIG	P50132.1	102	2E-23	27%
GP83B_DANRE	B0UXR0.1	102	2E-23	26%
GPR4_MOUSE	Q8BUD0.1	102	3E-23	26%
GP1R1_RAT	O08878.1	102	3E-23	27%
GPR4_HUMAN	P46093.2	101	4E-23	26%
AGTR2_HUMAN	P50052.1	101	5E-23	29%
AGTR2_RAT	P35351.1	100	8E-23	30%
AGTR2_MOUSE	P35374.1	100	1E-22	30%
AGTR2_MERUN	Q9Z0Z6.1	99.8	2E-22	29%
C5AR1_CAVPO	O70129.1	99.0	3E-22	28%
PTAFR_CAVPO	P21556.2	98.6	4E-22	27%
GALR1_HUMAN	P47211.3	98.6	5E-22	28%
LT4R1_RAT	Q9R0Q2.1	98.6	5E-22	29%
C5AR1_ONCMY	Q6UNA4.1	97.8	8E-22	32%
PTAFR_HUMAN	P25105.1	97.4	1E-21	27%
PAR2_RAT	Q63645.1	98.2	1E-21	25%
OX2R_HUMAN	O43614.2	98.2	1E-21	25%
GPR34_MOUSE	Q9R1K6.1	97.1	2E-21	29%
PAR3_HUMAN	O00254.1	97.1	2E-21	29%
P2Y12_MACFA	Q95KC3.1	96.7	2E-21	24%
OX2R_CANFA	Q9TUP7.1	97.8	2E-21	25%
PD2R2_HUMAN	Q9Y5Y4.3	97.4	2E-21	26%
SIFAR_DROME	Q8IN35.2	98.2	3E-21	28%
RYAR_DROME	P25931.2	97.4	3E-21	24%
P2RY6_MOUSE	Q9ERK9.1	95.9	3E-21	25%
PAR2_HUMAN	P55085.1	96.7	3E-21	25%
SSRL_TAKRU	O42179.1	95.1	4E-21	31%
LPAR5_MOUSE	Q149R9.2	96.3	4E-21	23%
OX2R_PIG	O62809.2	97.1	4E-21	25%
PTAFR_MOUSE	Q62035.1	95.5	5E-21	27%
OX2R_MOUSE	P58308.2	96.7	5E-21	26%

Gene	Accession	Score	E value	Ident
GPR34_HUMAN	Q9UPC5.2	95.9	6E-21	27%
NK2R_RABIT	P79218.1	95.9	7E-21	27%
LPAR5_HUMAN	Q9H1C0.1	95.5	7E-21	22%
OX2R_RAT	P56719.1	96.3	9E-21	26%
SUCR1_RAT	Q6IYF9.1	94.4	9E-21	27%
LPAR5_BOVIN	Q3ZC80.1	94.4	2E-20	24%
GPR18_MACFA	Q4R613.1	94.0	2E-20	26%
PAR3_BOVIN	Q58D85.1	94.4	2E-20	31%
PAR2_MOUSE	P55086.1	94.4	2E-20	24%
TLR1_DROME	P30974.2	94.7	3E-20	24%
RYAR_TRICA	G4WMX4.1	93.2	4E-20	26%
GPR33_MOUSE	O88416.1	92.8	5E-20	26%
GALR3_RAT	O88626.1	93.2	5E-20	26%
P2RY8_HUMAN	Q86VZ1.1	92.8	5E-20	26%
C5AR2_MOUSE	Q8BW93.1	92.4	7E-20	25%
LT4R1_MOUSE	O88855.1	92.4	8E-20	27%
PTAFR_RAT	P46002.1	92.0	9E-20	27%
PTAFR_PIG	Q9XSD4.2	92.0	1E-19	28%
SUCR1_HUMAN	Q9BXA5.2	91.3	1E-19	26%
OX1R_BOVIN	Q0GBZ5.1	92.4	2E-19	25%
CCH1R_DROME	A1ZAX0.2	92.8	2E-19	26%
RL3R2_HUMAN	Q8TDU9.1	91.3	2E-19	25%
GPR18_HUMAN	Q14330.2	90.5	3E-19	25%
PAR1_XENLA	P47749.1	90.9	5E-19	24%
P2RY6_RAT	Q63371.1	89.7	5E-19	25%
GPR33_HUMAN	Q49SQ1.1	89.7	6E-19	28%
GPR33_PANPA	Q49SQ2.1	89.7	6E-19	28%
GPR18_RAT	A1A5S3.1	89.7	6E-19	27%
OX1R_MOUSE	P58307.3	90.1	7E-19	25%
NK2R_HUMAN	P21452.3	89.7	9E-19	28%
NK2R_CANFA	Q5DUB2.1	89.7	9E-19	26%
NK2R_BOVIN	P05363.1	89.7	1E-18	27%
GPR33_PANTR	Q49SQ3.1	89.0	1E-18	28%
AVT_CATCO	Q90352.1	90.1	1E-18	24%
V021_FOWPN	Q9J5I0.1	88.6	1E-18	30%
NK2R_RAT	P16610.1	89.4	1E-18	24%
NPR9_CAEEL	Q23497.1	89.4	2E-18	26%
OX1R_PIG	O97661.2	89.0	2E-18	25%
NK2R_CAVPO	Q64077.1	89.0	2E-18	27%
GALR3_HUMAN	O60755.1	88.6	2E-18	25%
PTAFR_CAPHI	Q9GK76.1	88.2	2E-18	25%
NK2R_MESAU	P51144.1	88.6	2E-18	25%
GALR3_MOUSE	O88853.2	88.2	2E-18	25%
PTAFR_BOVIN	Q9TTY5.2	87.8	3E-18	25%
SUCR1_MOUSE	Q99MT6.2	87.0	3E-18	26%
HCAR2_MOUSE	Q9EP66.1	87.4	4E-18	24%
PD2R2_MOUSE	Q9Z2J6.1	87.8	4E-18	26%

Gene	Accession	Score	E value	Ident
LT4R1_BOVIN	Q3T181.1	86.7	7E-18	29%
GPR18_MOUSE	Q8K1Z6.1	86.3	9E-18	25%
NK2R_MOUSE	P30549.1	86.7	1E-17	23%
GPR18_BOVIN	Q3T0E9.1	85.9	1E-17	26%
PD2R2_RAT	Q6XKD3.1	86.7	1E-17	26%
GPR33_RATRT	Q49SP8.1	85.9	1E-17	25%
OX1R_RAT	P56718.1	85.9	2E-17	25%
NK3R_HUMAN	P29371.1	86.3	2E-17	24%
GP171_HUMAN	O14626.1	84.7	3E-17	25%
C3AR_PONAB	Q5REI5.1	155	3E-17	34%
PAR1_MOUSE	P30558.2	85.5	3E-17	27%
C3AR_MACFA	Q6TAC8.1	151	3E-17	33%
HCAR2_RAT	Q80Z39.1	84.7	4E-17	24%
C3AR_HUMAN	Q16581.2	152	5E-17	34%
C3AR_RAT	O55197.1	154	5E-17	31%
P2RY8_CHICK	Q5ZI82.1	83.6	8E-17	26%
LT4R2_RAT	Q924U0.1	83.6	8E-17	26%
PAR2_BOVIN	Q2HJA4.1	84.0	9E-17	26%
LT4R1_HUMAN	Q15722.2	83.6	9E-17	27%
OX1R_HUMAN	O43613.2	84.0	1E-16	25%
GASR_BOVIN	P79266.1	84.0	1E-16	29%
GASR_MOUSE	P56481.1	83.6	1E-16	29%
BRS3_SHEEP	O97967.1	83.2	1E-16	28%
GRPR_MOUSE	P21729.2	82.4	3E-16	24%
MCHR2_MACFA	Q8SQ54.1	82.0	3E-16	25%
GASR_MASNA	P30796.1	82.8	3E-16	29%
PAR1_RAT	P26824.1	82.4	3E-16	27%
P2Y13_RAT	Q6GUG4.1	81.6	4E-16	26%
LT4R2_MOUSE	Q9JIL9.1	82.0	4E-16	26%
C5AR2_HUMAN	Q9P296.1	81.3	6E-16	28%
GASR_HUMAN	P32239.1	81.6	7E-16	29%
PAR1_PAPHA	P56488.1	81.6	7E-16	25%
P2Y13_MOUSE	Q9D8I2.2	80.9	7E-16	25%
HCAR1_HUMAN	Q9BXC0.1	80.9	7E-16	23%
BRS3_MOUSE	O54798.1	81.3	8E-16	28%
GP171_BOVIN	Q3ZBK9.1	80.5	8E-16	24%
PKR1_BOVIN	Q8SPN2.1	80.9	9E-16	28%
PSYR_HUMAN	Q8IYL9.1	80.5	1E-15	27%
BRS3_RAT	Q8K418.1	80.9	1E-15	28%
GASR_RAT	P30553.1	80.9	1E-15	28%
MTR1L_MOUSE	O88495.3	81.3	1E-15	24%
MCHR2_MACMU	Q8MJ88.1	80.1	1E-15	25%
C3AR_CAVPO	O88680.1	145	1E-15	30%
TAA7B_MOUSE	Q5QD11.1	80.1	1E-15	27%
PAR1_HUMAN	P25116.2	80.5	2E-15	25%
PSYR_MOUSE	Q61038.2	79.7	2E-15	25%
OPRD_PIG	P79291.1	78.2	2E-15	27%

7.1.2.1 GO terms

7.1.2.1.1 Molecular Function

Table 7.6 - GO terms of the Molecular Function Ontology for the CXCR4 Protein BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	411	82.2%	11206	23.7%	7.85e-168
GO:0008528	G-protein coupled peptide receptor activity	59	11.8%	185	0.4%	3.28e-68
GO:0001653	peptide receptor activity	59	11.8%	187	0.4%	6.89e-68
GO:0004930	G-protein coupled receptor activity	87	17.4%	1166	2.5%	4.52e-45
GO:0001637	G-protein coupled chemoattractant receptor activity	23	4.6%	41	0.1%	3.79e-33
GO:0004950	chemokine receptor activity	23	4.6%	41	0.1%	3.79e-33
GO:0004888	transmembrane signaling receptor activity	89	17.8%	1924	4.1%	5.75e-30
GO:0099600	transmembrane receptor activity	89	17.8%	2023	4.3%	2.27e-28
GO:0038023	signaling receptor activity	89	17.8%	2105	4.5%	4.04e-27
GO:0042923	neuropeptide binding	15	3.0%	22	0.0%	3.05e-23
GO:0004872	receptor activity	89	17.8%	2540	5.4%	1.99e-21
GO:0060089	molecular transducer activity	89	17.8%	2540	5.4%	1.99e-21
GO:0004871	signal transducer activity	89	17.8%	2641	5.6%	2.71e-20
GO:0004896	cytokine receptor activity	23	4.6%	147	0.3%	2.96e-18
GO:0008188	neuropeptide receptor activity	16	3.2%	53	0.1%	2.04e-17
GO:0016493	C-C chemokine receptor activity	11	2.2%	16	0.0%	7.02e-17
GO:0019956	chemokine binding	12	2.4%	31	0.1%	2.05e-14
GO:0001608	G-protein coupled nucleotide receptor activity	9	1.8%	18	0.0%	6.99e-12
GO:0045028	G-protein coupled purinergic nucleotide receptor activity	9	1.8%	18	0.0%	6.99e-12
GO:0004994	somatostatin receptor activity	5	1.0%	5	0.0%	1.31e-08
GO:0016494	C-X-C chemokine receptor activity	6	1.2%	11	0.0%	6.08e-08
GO:0001614	purinergic nucleotide receptor activity	9	1.8%	46	0.1%	1.21e-07
GO:0016502	nucleotide receptor activity	9	1.8%	46	0.1%	1.21e-07
GO:0019955	cytokine binding	12	2.4%	111	0.2%	2.67e-07
GO:0004985	opioid receptor activity	6	1.2%	14	0.0%	3.84e-07
GO:0035586	purinergic receptor activity	9	1.8%	54	0.1%	5.45e-07
GO:0042277	peptide binding	20	4.0%	418	0.9%	3.07e-06
GO:0019957	C-C chemokine binding	5	1.0%	11	0.0%	5.76e-06
GO:0019958	C-X-C chemokine binding	5	1.0%	11	0.0%	5.76e-06
GO:0033218	amide binding	20	4.0%	462	1.0%	1.55e-05
GO:0071791	chemokine (C-C motif) ligand 5 binding	3	0.6%	3	0.0%	0.00011
GO:0017046	peptide hormone binding	6	1.2%	38	0.1%	0.00028
GO:0004966	galanin receptor activity	3	0.6%	4	0.0%	0.00047
GO:0042562	hormone binding	7	1.4%	71	0.2%	0.00106
GO:0004875	complement receptor activity	3	0.6%	6	0.0%	0.00232
GO:0004945	angiotensin type II receptor activity	3	0.6%	7	0.0%	0.00403
GO:0015057	thrombin-activated receptor activity	3	0.6%	7	0.0%	0.00403
GO:0004435	phosphatidylinositol phospholipase C activity	6	1.2%	62	0.1%	0.00512

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:0001595	angiotensin receptor activity	3	0.6%	8	0.0%	0.00640
GO:0004982	N-formyl peptide receptor activity	3	0.6%	8	0.0%	0.00640
GO:0004974	leukotriene receptor activity	3	0.6%	9	0.0%	0.00953
GO:0031681	G-protein beta-subunit binding	3	0.6%	9	0.0%	0.00953

7.1.2.1.2 Cellular Component

Table 7.7 - GO terms of the Cellular Component Ontology for the CXCR4 Protein BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	411	82.2%	11987	25.4%	1.85e-156
GO:0005887	integral component of plasma membrane	73	14.6%	1917	4.1%	2.89e-19
GO:0031226	intrinsic component of plasma membrane	73	14.6%	1982	4.2%	1.89e-18
GO:0044459	plasma membrane part	75	15.0%	3409	7.2%	1.41e-07
GO:0032590	dendrite membrane	4	0.8%	21	0.0%	0.00610

7.1.2.1.3 Biological Process

Table 7.8 - GO terms of the Biological Process Ontology for the CXCR4 Protein BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	411	82.2%	12071	25.5%	3.86e-154
GO:0007186	G-protein coupled receptor signaling pathway	89	17.8%	1702	3.6%	7.63e-33
GO:0070098	chemokine-mediated signaling pathway	23	4.6%	97	0.2%	1.60e-21
GO:0007204	positive regulation of cytosolic calcium ion concentration	30	6.0%	262	0.6%	7.52e-19
GO:0051480	regulation of cytosolic calcium ion concentration	30	6.0%	283	0.6%	7.08e-18
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	25	5.0%	185	0.4%	3.54e-17
GO:0007218	neuropeptide signaling pathway	21	4.2%	116	0.2%	7.26e-17
GO:0007200	phospholipase C-activating G-protein coupled receptor signaling pathway	19	3.8%	89	0.2%	1.57e-16
GO:0006874	cellular calcium ion homeostasis	30	6.0%	382	0.8%	3.35e-14
GO:0055074	calcium ion homeostasis	30	6.0%	396	0.8%	8.99e-14
GO:0072503	cellular divalent inorganic cation homeostasis	30	6.0%	409	0.9%	2.16e-13
GO:0072507	divalent inorganic cation homeostasis	30	6.0%	429	0.9%	7.87e-13
GO:0006875	cellular metal ion homeostasis	31	6.2%	511	1.1%	1.25e-11
GO:0035589	G-protein coupled purinergic nucleotide receptor signaling pathway	9	1.8%	18	0.0%	9.14e-11
GO:0006935	chemotaxis	33	6.6%	659	1.4%	3.69e-10
GO:0042330	taxis	33	6.6%	660	1.4%	3.85e-10
GO:0055065	metal ion homeostasis	31	6.2%	583	1.2%	4.16e-10
GO:0030003	cellular cation homeostasis	31	6.2%	588	1.2%	5.20e-10

Gene Ontology term		Cluster frequency		Genome frequency		P-Value
GOID	TERM	#/500	%	#/47246	%	
GO:0006873	cellular ion homeostasis	31	6.2%	602	1.3%	9.59e-10
GO:0055080	cation homeostasis	31	6.2%	677	1.4%	1.92e-08
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	16	3.2%	158	0.3%	2.13e-08
GO:0098771	inorganic ion homeostasis	31	6.2%	692	1.5%	3.33e-08
GO:0035588	G-protein coupled purinergic receptor signaling pathway	9	1.8%	31	0.1%	3.35e-08
GO:0055082	cellular chemical homeostasis	32	6.4%	758	1.6%	7.19e-08
GO:0038169	somatostatin receptor signaling pathway	5	1.0%	5	0.0%	1.71e-07
GO:0038170	somatostatin signaling pathway	5	1.0%	5	0.0%	1.71e-07
GO:0007631	feeding behavior	13	2.6%	108	0.2%	1.95e-07
GO:0050801	ion homeostasis	31	6.2%	756	1.6%	2.94e-07
GO:0035590	purinergic nucleotide receptor signaling pathway	9	1.8%	47	0.1%	1.95e-06
GO:0006954	inflammatory response	29	5.8%	731	1.5%	2.39e-06
GO:0002407	dendritic cell chemotaxis	7	1.4%	22	0.0%	2.79e-06
GO:0007193	adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway	10	2.0%	74	0.2%	8.34e-06
GO:0019725	cellular homeostasis	32	6.4%	935	2.0%	1.20e-05
GO:0030815	negative regulation of cAMP metabolic process	8	1.6%	41	0.1%	1.38e-05
GO:0036336	dendritic cell migration	7	1.4%	27	0.1%	1.39e-05
GO:0035587	purinergic receptor signaling pathway	9	1.8%	60	0.1%	1.87e-05
GO:0038003	opioid receptor signaling pathway	6	1.2%	17	0.0%	2.01e-05
GO:0030595	leukocyte chemotaxis	14	2.8%	191	0.4%	2.68e-05
GO:0002685	regulation of leukocyte migration	13	2.6%	161	0.3%	2.72e-05
GO:0030800	negative regulation of cyclic nucleotide metabolic process	8	1.6%	46	0.1%	3.60e-05
GO:0030818	negative regulation of cAMP biosynthetic process	7	1.4%	37	0.1%	0.00014
GO:0007194	negative regulation of adenylate cyclase activity	6	1.2%	24	0.1%	0.00020
GO:0030814	regulation of cAMP metabolic process	11	2.2%	130	0.3%	0.00020
GO:0048878	chemical homeostasis	33	6.6%	1123	2.4%	0.00024
GO:0006968	cellular defense response	8	1.6%	59	0.1%	0.00027
GO:0019221	cytokine-mediated signaling pathway	25	5.0%	706	1.5%	0.00027
GO:0050900	leukocyte migration	18	3.6%	385	0.8%	0.00027
GO:0030803	negative regulation of cyclic nucleotide biosynthetic process	7	1.4%	41	0.1%	0.00031
GO:0002687	positive regulation of leukocyte migration	10	2.0%	110	0.2%	0.00038
GO:0030809	negative regulation of nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00043
GO:1900372	negative regulation of purine nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00043
GO:0060326	cell chemotaxis	15	3.0%	280	0.6%	0.00051
GO:0031280	negative regulation of cyclase activity	6	1.2%	28	0.1%	0.00055
GO:1900543	negative regulation of purine nucleotide metabolic process	8	1.6%	67	0.1%	0.00074
GO:0040011	locomotion	43	8.6%	1779	3.8%	0.00075
GO:0044060	regulation of endocrine process	7	1.4%	47	0.1%	0.00082
GO:0045980	negative regulation of nucleotide metabolic process	8	1.6%	68	0.1%	0.00083

Gene Ontology term		Cluster frequency		Genome frequency		P-Value
GOID	TERM	#/500	%	#/47246	%	
GO:0051350	negative regulation of lyase activity	6	1.2%	30	0.1%	0.00086
GO:0070374	positive regulation of ERK1 and ERK2 cascade	12	2.4%	184	0.4%	0.00096
GO:0046058	cAMP metabolic process	11	2.2%	154	0.3%	0.00113
GO:0030799	regulation of cyclic nucleotide metabolic process	11	2.2%	162	0.3%	0.00187
GO:0002688	regulation of leukocyte chemotaxis	9	1.8%	102	0.2%	0.00196
GO:0051930	regulation of sensory perception of pain	6	1.2%	36	0.1%	0.00267
GO:0051931	regulation of sensory perception	6	1.2%	36	0.1%	0.00267
GO:0071345	cellular response to cytokine stimulus	26	5.2%	856	1.8%	0.00276
GO:0002690	positive regulation of leukocyte chemotaxis	8	1.6%	82	0.2%	0.00352
GO:0030817	regulation of cAMP biosynthetic process	9	1.8%	114	0.2%	0.00496
GO:0050921	positive regulation of chemotaxis	9	1.8%	122	0.3%	0.00867

7.2 PSI-BLAST results

7.2.1 CCR5

Table 7.9 - Protein PSI-BLAST results for query protein CCR5_HUMAN (P51681). Results of iteration 8. Default parameters were used with cut-off value of 1×10^{-7} , limited to 500 hits. Results retrieved from the saved file, with no gene name entry available.

Accession	Score	E value
Q2HJ17.1	384	4E-132
Q9XT76.1	376	3E-129
Q5ECR9.1	376	6E-129
P51683.2	375	2E-128
P51682.3	373	6E-128
O55193.1	374	6E-128
Q9BGN6.1	372	8E-128
Q1ZY22.1	365	8E-125
P32246.1	363	6E-124
O18793.2	362	9E-124
Q8MJW8.1	362	1E-123
P61813.1	361	4E-123
Q95NC4.1	361	5E-123
O97883.1	360	5E-123
O62743.2	360	6E-123
Q95ND1.1	360	8E-123
P61755.1	360	8E-123
O97975.1	360	9E-123
P51678.2	360	1E-122
Q95NC1.1	360	1E-122
Q95NE8.1	359	1E-122
Q9TV42.1	359	1E-122

Accession	Score	E value
Q9TV47.1	359	1E-122
Q9TV49.1	359	2E-122
Q95ND2.1	359	2E-122
O97882.1	359	2E-122
P56493.1	359	2E-122
Q95NC2.1	359	2E-122
P56439.1	359	2E-122
Q9TV48.1	359	2E-122
Q9TV43.1	359	3E-122
O97962.1	358	3E-122
O97878.1	358	3E-122
Q6WN98.1	358	4E-122
O97879.1	358	4E-122
Q95ND0.1	358	4E-122
Q95NC8.1	358	4E-122
Q95NC5.1	358	4E-122
O97881.1	358	5E-122
Q95NC3.1	358	5E-122
P56440.1	357	8E-122
Q95NC0.1	357	1E-121
Q95NC6.1	357	1E-121
O97880.1	356	2E-121
Q95NC7.1	356	2E-121
P51681.1	356	2E-121
Q8HZT9.1	356	2E-121
O54814.1	355	7E-121
P51680.2	354	2E-120
O08556.1	352	7E-120
P51679.1	352	1E-119

Accession	Score	E value
Q64H34.1	352	1E-119
P41597.1	352	3E-119
Q95NC9.1	345	3E-117
Q9TV45.1	345	4E-117
P25104.1	335	3E-113
O77590.2	335	5E-113
P30555.1	332	6E-112
O35210.1	332	9E-112
P25095.1	332	1E-111
P29754.1	331	2E-111
P43240.1	330	7E-111
P51675.2	328	2E-110
P34976.1	328	3E-110
Q9GLN9.1	328	4E-110
P32303.2	328	4E-110
P30556.1	327	5E-110
Q9WV26.1	327	6E-110
Q2YEG0.1	326	1E-109
P21109.2	326	1E-109
P29755.1	326	1E-109
P29089.1	325	5E-109
P35373.1	324	2E-108
P51686.2	323	4E-108
Q28003.1	320	3E-107
Q1WLP9.1	319	1E-106
P56483.1	318	3E-106
P35407.2	318	3E-106
Q9Z213.1	317	6E-106
Q3LSL6.1	317	8E-106

Accession	Score	E value
Q9WUT7.1	317	1E-105
Q9BDS8.1	316	2E-105
Q2Y2P0.1	315	3E-105
P56498.1	315	4E-105
P51677.1	315	4E-105
P56482.1	315	5E-105
Q764M9.1	314	7E-105
Q7ZXJ7.1	314	2E-104
Q867B2.1	313	3E-104
P61072.1	312	3E-104
O08565.1	312	3E-104
P25930.1	312	7E-104
P56492.1	311	1E-103
Q7YS92.1	311	1E-103
Q8HZU0.1	311	1E-103
Q5MD61.1	312	1E-103
P56491.1	311	1E-103
Q28474.2	310	2E-103
Q8HZU1.1	310	3E-103
Q9TSQ8.1	309	8E-103
P79394.1	309	1E-102
Q7SZP9.2	308	3E-102
O62747.1	307	4E-102
P51684.2	308	5E-102
P35344.1	307	6E-102
O54689.1	306	1E-101
P70658.2	306	2E-101
Q95M54.1	307	2E-101
P41144.1	305	5E-101
P70612.1	304	8E-101
P32249.3	303	2E-100
Q9MYW9.2	305	2E-100
P35343.1	303	3E-100
E9QJ73.2	304	3E-100
P56484.1	302	4E-100
A0T2N3.1	302	4E-100
P34975.1	303	6E-100
Q9JHG3.1	302	1E-99
P33534.1	302	2E-99
P51676.2	300	3E-99
Q2TAD5.1	300	3E-99
Q4VA82.1	300	5E-99
Q99JA4.1	299	8E-99
P30937.1	300	8E-99
P79928.1	305	1E-98
Q9WV08.1	299	1E-98
O97666.1	299	2E-98
P25023.4	299	2E-98

Accession	Score	E value
O97571.1	298	2E-98
P79960.2	298	3E-98
P49660.2	298	5E-98
D4A7K7.1	297	7E-98
Q07FZ4.1	296	1E-97
P35414.1	297	1E-97
Q2NNR5.1	295	2E-97
P32299.4	296	4E-97
Q9YGC3.1	294	5E-97
Q924T8.1	293	6E-97
Q9NPB9.1	293	1E-96
P35350.1	293	1E-96
P79785.1	293	2E-96
P33396.1	293	3E-96
O35811.1	292	6E-96
Q04683.2	292	1E-95
P41231.4	291	2E-95
P33535.1	292	2E-95
Q99677.1	290	3E-95
Q5IS84.1	291	4E-95
P30411.2	291	4E-95
P35346.3	290	5E-95
P30680.1	290	6E-95
P34996.1	289	7E-95
P49652.1	289	7E-95
P42866.1	290	8E-95
P51685.1	288	8E-95
Q2YEG2.1	288	9E-95
P32250.1	287	1E-94
P34997.1	289	1E-94
Q8BLG2.2	288	3E-94
Q1RMI1.1	287	3E-94
Q4G072.1	287	4E-94
Q5IS39.1	288	4E-94
P30875.1	287	5E-94
Q8BMC0.1	286	5E-94
P35372.2	288	6E-94
P25090.2	286	9E-94
P55920.1	286	1E-93
P49650.1	287	1E-93
P51582.1	286	1E-93
F1MV99.2	286	1E-93
P25024.2	285	2E-93
P55919.1	285	2E-93
Q95247.1	287	2E-93
P48042.1	285	2E-93
P49651.1	285	2E-93
P34994.1	285	3E-93

Accession	Score	E value
O97665.1	285	3E-93
Q2YEF9.1	285	3E-93
P28646.1	285	4E-93
P79350.2	286	4E-93
Q9JJS7.1	284	5E-93
P34993.1	285	5E-93
P30874.1	284	6E-93
O70526.1	284	7E-93
Q13304.2	284	1E-92
P41143.4	284	1E-92
P33766.1	283	1E-92
Q28642.1	283	1E-92
P47900.1	283	1E-92
P30873.1	284	2E-92
O09027.1	283	2E-92
Q49LX6.1	283	3E-92
Q3U6B2.1	282	3E-92
P41145.2	283	4E-92
Q2KIP6.1	282	6E-92
P21462.3	281	7E-92
Q9Z0D9.1	281	9E-92
Q09QM4.1	280	9E-92
P25025.2	281	1E-91
P59902.1	282	1E-91
Q6NS65.1	280	1E-91
P49238.1	280	1E-91
P79177.1	280	1E-91
Q9EQ16.1	280	1E-91
Q9GLX8.1	280	2E-91
Q2KTE1.1	280	2E-91
P79218.1	281	2E-91
Q28807.2	280	3E-91
Q924I3.1	279	3E-91
P79242.1	279	3E-91
Q5YA25.1	280	4E-91
P31391.2	280	5E-91
P43657.3	278	6E-91
P79190.1	278	8E-91
A5PLE7.1	279	9E-91
Q49LX5.1	279	1E-90
P30872.1	279	1E-90
Q9Y271.1	277	1E-90
O18983.1	277	2E-90
P33533.1	278	3E-90
Q5DUB2.1	278	3E-90
P32300.1	278	3E-90
Q95N02.1	277	3E-90
B5X337.1	277	5E-90

Accession	Score	E value
B9VR26.1	276	7E-90
O43603.1	277	8E-90
O08726.1	276	9E-90
Q9XT45.1	275	1E-89
P16610.1	277	1E-89
P79235.1	275	1E-89
Q28519.1	275	2E-89
P32302.1	276	2E-89
P21452.3	277	2E-89
O19024.1	275	2E-89
Q05394.1	275	2E-89
Q9N0Z0.1	275	2E-89
P79236.1	275	2E-89
Q810W6.1	274	3E-89
P32248.2	275	4E-89
P79189.1	274	4E-89
Q8HZP1.1	274	4E-89
P79237.1	273	6E-89
O88854.2	274	6E-89
P0C7U4.1	274	7E-89
P30549.1	274	9E-89
Q9BDS6.1	273	1E-88
P46090.1	273	1E-88
P79241.1	272	1E-88
Q8K087.1	273	1E-88
Q28422.1	272	2E-88
Q64077.1	274	2E-88
P05363.1	273	2E-88
Q9TV16.1	272	2E-88
O00574.1	272	3E-88
O88536.1	272	4E-88
P47774.2	272	4E-88
P79176.1	270	8E-88
P79243.1	270	1E-87
Q5MD62.1	271	1E-87
P25089.2	270	1E-87
O97664.1	270	1E-87
Q0II78.1	269	2E-87
O89039.2	270	3E-87
P97468.1	270	3E-87
Q95LH1.1	269	3E-87
Q8HZN9.1	269	4E-87
Q9BDQ5.1	269	4E-87
O00590.2	270	4E-87
P56485.2	269	5E-87
P49682.2	269	5E-87
P51144.1	269	7E-87
P25106.3	268	8E-87

Accession	Score	E value
Q63645.1	270	8E-87
P35411.1	268	9E-87
Q8BUD0.1	268	1E-86
P55086.1	269	1E-86
Q75ZH0.1	267	1E-86
P46663.3	267	1E-86
P21556.2	267	2E-86
Q4KLH9.1	267	2E-86
P79178.1	267	2E-86
Q8BGE9.1	271	2E-86
P79292.1	267	4E-86
P48748.1	267	4E-86
B1PHQ8.1	267	4E-86
P41232.2	267	4E-86
P47748.1	267	5E-86
P46094.1	265	5E-86
P79191.1	266	7E-86
Q9XSD7.1	265	7E-86
Q8HZP2.1	266	7E-86
A7YY44.1	268	8E-86
P46093.2	266	9E-86
O35786.1	266	1E-85
Q89609.1	266	2E-85
P41146.1	265	2E-85
P97520.1	265	2E-85
P29371.1	268	2E-85
Q6IYF8.1	263	3E-85
Q8HZP3.1	264	3E-85
B0UXR0.1	264	4E-85
O00421.2	263	4E-85
P35370.1	264	4E-85
Q95L01.1	264	4E-85
Q1JQB3.1	264	4E-85
P35377.1	264	6E-85
Q6Y1R5.1	263	6E-85
P26824.1	266	7E-85
Q3BCU0.1	263	7E-85
P47749.1	265	8E-85
Q15743.1	263	9E-85
P30974.2	267	1E-84
P50132.1	262	2E-84
Q15077.1	261	2E-84
Q9ERK9.1	261	3E-84
Q99788.2	262	4E-84
O88410.2	262	4E-84
P25105.1	261	5E-84
P46091.2	260	8E-84
P30938.2	261	9E-84

Accession	Score	E value
Q9NSD7.1	264	1E-83
Q924T9.1	258	1E-83
Q8BFQ3.1	260	2E-83
O08707.2	260	2E-83
F5HF62.1	260	2E-83
O08858.3	260	2E-83
P30993.2	259	2E-83
Q98907.1	258	3E-83
P79175.2	259	3E-83
P49681.2	258	3E-83
P55085.1	260	3E-83
Q9JII9.1	259	4E-83
Q96P68.1	258	5E-83
Q95N03.1	258	6E-83
Q920A1.2	257	6E-83
Q56UD9.1	257	8E-83
Q0VDU3.2	258	8E-83
P35383.1	258	1E-82
P56488.1	260	2E-82
P46002.1	257	2E-82
Q9TUE1.1	256	3E-82
O93361.1	255	3E-82
P79234.1	256	4E-82
P21730.2	256	5E-82
Q61125.1	255	7E-82
P79240.2	255	8E-82
Q62035.1	255	9E-82
Q00991.2	258	1E-81
Q5KSK8.1	255	1E-81
Q2WED0.1	255	1E-81
P25116.2	257	2E-81
P11613.1	255	2E-81
Q9TTY5.2	253	3E-81
O46685.1	254	3E-81
A6QNL7.1	254	4E-81
P47937.2	256	6E-81
Q9NS75.1	253	6E-81
P16177.1	256	7E-81
Q9GK76.1	252	1E-80
Q8MJV2.1	252	1E-80
P97583.2	252	1E-80
Q3SXG2.1	252	1E-80
P0C7U5.1	252	1E-80
P43142.1	253	2E-80
Q9R1K6.1	252	3E-80
P48145.2	250	3E-80
O08790.2	251	4E-80
Q9BY21.1	251	5E-80

Accession	Score	E value
P50052.1	251	5E-80
Q6BD04.1	251	6E-80
Q924H0.2	252	8E-80
O18982.1	250	9E-80
P30975.2	255	9E-80
P30558.2	253	9E-80
O15218.1	251	1E-79
O88537.2	250	1E-79
P49019.3	251	1E-79
Q71MR7.2	249	1E-79
P79188.1	250	1E-79
Q9UPC5.2	251	1E-79
Q9XSD4.2	249	1E-79
Q8K418.1	251	2E-79
Q149R9.2	250	2E-79
Q58D85.1	250	2E-79
Q9H244.1	248	3E-79
Q99MT7.2	249	3E-79
Q9Y5X5.2	254	3E-79
P56412.1	249	3E-79
O97663.1	249	3E-79
Q9BXC1.1	248	4E-79
P32745.1	251	4E-79
P49685.1	248	5E-79
Q95KC3.1	248	5E-79
P35351.1	248	5E-79
Q8TDS4.1	248	6E-79
P35374.1	248	6E-79
Q98982.1	250	8E-79
Q86VZ1.1	248	1E-78
F7EQ49.1	247	2E-78
P69332.1	246	2E-78
O60755.1	247	3E-78
Q9H1C0.1	247	3E-78
Q6H2Y3.1	248	3E-78
Q9Z0Z6.1	246	3E-78
Q9EQD2.1	248	4E-78
A1ZAX0.2	250	4E-78
Q9BDS7.1	246	5E-78
Q91V45.1	247	5E-78
O00254.1	246	6E-78
Q63371.1	245	6E-78
O97512.2	249	8E-78
Q3U507.1	245	8E-78

Accession	Score	E value
O97967.1	246	8E-78
Q99527.1	245	1E-77
Q8WPA2.1	245	1E-77
Q3ZC80.1	244	3E-77
Q99705.2	246	3E-77
Q8BMP4.2	244	3E-77
Q99MT6.2	242	4E-77
O88626.1	244	4E-77
O54798.1	245	4E-77
P32247.1	245	5E-77
Q969F8.2	244	7E-77
Q924U1.2	244	8E-77
Q8MJV3.1	241	1E-76
Q2HJA4.1	243	2E-76
P30936.1	244	2E-76
Q9R0M1.1	240	4E-76
P31392.3	242	4E-76
P56479.1	240	5E-76
Q6IYF9.1	239	5E-76
P30935.2	242	8E-76
Q9EP66.1	240	8E-76
P24053.1	241	8E-76
Q62805.1	240	1E-75
Q9EPX4.1	239	1E-75
O88853.2	240	1E-75
A4FUQ5.2	238	2E-75
Q8MJ89.2	240	2E-75
O08878.1	239	4E-75
Q5Z182.1	238	5E-75
Q920E1.1	238	6E-75
Q80Z39.1	238	7E-75
P47751.2	238	9E-75
Q5DUB1.1	239	1E-74
P48146.2	236	2E-74
P30548.2	238	2E-74
P35371.1	238	2E-74
Q5DUB3.1	238	2E-74
O54799.1	237	2E-74
Q9BXC0.1	236	2E-74
Q86917.1	237	3E-74
Q8JZL2.2	238	3E-74
Q8IN35.2	246	3E-74
P14600.1	238	3E-74
Q9BXA5.2	235	3E-74

Accession	Score	E value
Q9EP86.1	238	4E-74
P30547.1	236	1E-73
Q8C131.1	234	1E-73
P56719.1	238	1E-73
Q9CPV9.1	234	1E-73
O35457.2	235	1E-73
P25931.2	238	1E-73
P25103.1	236	1E-73
P58308.2	237	1E-73
O00155.2	235	1E-73
Q9ESG6.1	233	2E-73
P28336.2	235	2E-73
O00398.1	233	2E-73
B3G515.1	233	4E-73
P47211.3	233	6E-73
B0F9W3.1	233	6E-73
Q3SX17.1	231	9E-73
Q8BFU7.1	231	2E-72
P30992.1	231	2E-72
Q9TUP7.1	234	3E-72
O62809.2	233	4E-72
O08675.2	231	4E-72
P97639.3	228	3E-71
O43614.2	231	4E-71
O88634.2	229	4E-71
P30098.1	231	4E-71
P0C5I1.1	228	7E-71
Q9GZQ6.1	228	3E-70
Q15391.1	225	4E-70
Q6GUG4.1	224	6E-70
Q08520.1	225	9E-70
Q8IYL9.1	224	9E-70
Q8SQ54.1	223	1E-69
O75388.1	223	2E-69
Q8MJ88.1	222	5E-69
Q99678.2	222	7E-69
Q61038.2	221	1E-68
Q920E0.1	223	1E-68
Q9D8I2.2	219	9E-68
O70129.1	219	1E-67
Q23497.1	220	8E-67
P58307.3	219	8E-67
P56718.1	215	1E-65
Q7TQP4.1	215	2E-65

7.2.1.1 GO terms

7.2.1.1.1 Molecular Function

Table 7.10 - GO terms of the Molecular Function Ontology for the CCR5 Protein PSI-BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	408	81.6%	11203	23.7%	2.37e-164
GO:0008528	G-protein coupled peptide receptor activity	60	12.0%	185	0.4%	6.45e-70
GO:0001653	peptide receptor activity	60	12.0%	187	0.4%	1.37e-69
GO:0004930	G-protein coupled receptor activity	91	18.2%	1166	2.5%	6.15e-49
GO:0004888	transmembrane signaling receptor activity	92	18.4%	1924	4.1%	3.51e-32
GO:0001637	G-protein coupled chemoattractant receptor activity	22	4.4%	41	0.1%	4.44e-31
GO:0004950	chemokine receptor activity	22	4.4%	41	0.1%	4.44e-31
GO:0099600	transmembrane receptor activity	92	18.4%	2023	4.3%	1.64e-30
GO:0038023	signaling receptor activity	92	18.4%	2105	4.5%	3.31e-29
GO:0008188	neuropeptide receptor activity	21	4.2%	53	0.1%	5.02e-26
GO:0004872	receptor activity	92	18.4%	2540	5.4%	3.01e-23
GO:0060089	molecular transducer activity	92	18.4%	2540	5.4%	3.01e-23
GO:0042923	neuropeptide binding	15	3.0%	22	0.0%	3.02e-23
GO:0004871	signal transducer activity	92	18.4%	2641	5.6%	4.65e-22
GO:0004896	cytokine receptor activity	22	4.4%	147	0.3%	5.29e-17
GO:0001608	G-protein coupled nucleotide receptor activity	11	2.2%	18	0.0%	4.97e-16
GO:0045028	G-protein coupled purinergic nucleotide receptor activity	11	2.2%	18	0.0%	4.97e-16
GO:0016493	C-C chemokine receptor activity	10	2.0%	16	0.0%	1.21e-14
GO:0019956	chemokine binding	12	2.4%	31	0.1%	2.03e-14
GO:0001614	purinergic nucleotide receptor activity	11	2.2%	46	0.1%	1.59e-10
GO:0016502	nucleotide receptor activity	11	2.2%	46	0.1%	1.59e-10
GO:0035586	purinergic receptor activity	11	2.2%	54	0.1%	1.06e-09
GO:0004994	somatostatin receptor activity	5	1.0%	5	0.0%	1.30e-08
GO:0016494	C-X-C chemokine receptor activity	6	1.2%	11	0.0%	6.02e-08
GO:0019955	cytokine binding	12	2.4%	111	0.2%	2.64e-07
GO:0004985	opioid receptor activity	6	1.2%	14	0.0%	3.81e-07
GO:0042277	peptide binding	21	4.2%	418	0.9%	5.91e-07
GO:0015065	uridine nucleotide receptor activity	4	0.8%	4	0.0%	1.23e-06
GO:0071553	G-protein coupled pyrimidinergic nucleotide receptor activity	4	0.8%	4	0.0%	1.23e-06
GO:0033218	amide binding	21	4.2%	462	1.0%	3.31e-06
GO:0019957	C-C chemokine binding	5	1.0%	11	0.0%	5.70e-06
GO:0019958	C-X-C chemokine binding	5	1.0%	11	0.0%	5.70e-06
GO:0071791	chemokine (C-C motif) ligand 5 binding	3	0.6%	3	0.0%	0.00011
GO:0004966	galanin receptor activity	3	0.6%	4	0.0%	0.00046
GO:0004995	tachykinin receptor activity	3	0.6%	5	0.0%	0.00115
GO:0015057	thrombin-activated receptor activity	3	0.6%	7	0.0%	0.00399
GO:0017046	peptide hormone binding	5	1.0%	38	0.1%	0.00489

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:0004982	N-formyl peptide receptor activity	3	0.6%	8	0.0%	0.00634
GO:0031681	G-protein beta-subunit binding	3	0.6%	9	0.0%	0.00943

7.2.1.1.2 Cellular Component

Table 7.11 - GO terms of the Cellular Component Ontology for the CCR5 Protein PSI-BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	408	81.6%	11984	25.4%	4.41e-153
GO:0005887	integral component of plasma membrane	75	15.0%	1917	4.1%	1.61e-20
GO:0031226	intrinsic component of plasma membrane	75	15.0%	1982	4.2%	1.13e-19
GO:0044459	plasma membrane part	76	15.2%	3409	7.2%	6.15e-08
GO:0032590	dendrite membrane	4	0.8%	21	0.0%	0.00623

7.2.1.1.3 Biological Process

Table 7.12 - GO terms of the Biological Process Ontology for the CCR5 Protein PSI-BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	408	81.6%	12068	25.5%	8.94e-151
GO:0007186	G-protein coupled receptor signaling pathway	92	18.4%	1702	3.6%	3.21e-35
GO:0070098	chemokine-mediated signaling pathway	22	4.4%	97	0.2%	4.89e-20
GO:0007200	phospholipase C-activating G-protein coupled receptor signaling pathway	20	4.0%	89	0.2%	5.79e-18
GO:0007204	positive regulation of cytosolic calcium ion concentration	29	5.8%	262	0.6%	9.79e-18
GO:0051480	regulation of cytosolic calcium ion concentration	29	5.8%	283	0.6%	8.45e-17
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	24	4.8%	185	0.4%	5.50e-16
GO:0007218	neuropeptide signaling pathway	20	4.0%	116	0.2%	1.57e-15
GO:0035589	G-protein coupled purinergic nucleotide receptor signaling pathway	11	2.2%	18	0.0%	6.70e-15
GO:0006874	cellular calcium ion homeostasis	29	5.8%	382	0.8%	2.87e-13
GO:0055074	calcium ion homeostasis	29	5.8%	396	0.8%	7.42e-13
GO:0072503	cellular divalent inorganic cation homeostasis	29	5.8%	409	0.9%	1.72e-12
GO:0072507	divalent inorganic cation homeostasis	29	5.8%	429	0.9%	5.95e-12
GO:0035588	G-protein coupled purinergic receptor signaling pathway	11	2.2%	31	0.1%	1.57e-11
GO:0006875	cellular metal ion homeostasis	30	6.0%	511	1.1%	8.15e-11
GO:0055065	metal ion homeostasis	30	6.0%	583	1.2%	2.35e-09
GO:0035590	purinergic nucleotide receptor signaling pathway	11	2.2%	47	0.1%	2.78e-09
GO:0030003	cellular cation homeostasis	30	6.0%	588	1.2%	2.91e-09

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0006873	cellular ion homeostasis	30	6.0%	602	1.3%	5.25e-09
GO:0035587	purinergic receptor signaling pathway	11	2.2%	60	0.1%	4.83e-08
GO:0055080	cation homeostasis	30	6.0%	677	1.4%	9.31e-08
GO:0098771	inorganic ion homeostasis	30	6.0%	692	1.5%	1.57e-07
GO:0038169	somatostatin receptor signaling pathway	5	1.0%	5	0.0%	1.75e-07
GO:0038170	somatostatin signaling pathway	5	1.0%	5	0.0%	1.75e-07
GO:0006935	chemotaxis	29	5.8%	659	1.4%	2.28e-07
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	15	3.0%	158	0.3%	2.35e-07
GO:0042330	taxis	29	5.8%	660	1.4%	2.36e-07
GO:0055082	cellular chemical homeostasis	31	6.2%	758	1.6%	3.20e-07
GO:0050801	ion homeostasis	30	6.0%	756	1.6%	1.26e-06
GO:0007631	feeding behavior	12	2.4%	108	0.2%	2.59e-06
GO:0002407	dendritic cell chemotaxis	7	1.4%	22	0.0%	2.85e-06
GO:0030815	negative regulation of cAMP metabolic process	8	1.6%	41	0.1%	1.40e-05
GO:0036336	dendritic cell migration	7	1.4%	27	0.1%	1.42e-05
GO:0038003	opioid receptor signaling pathway	6	1.2%	17	0.0%	2.05e-05
GO:0030814	regulation of cAMP metabolic process	12	2.4%	130	0.3%	2.17e-05
GO:0030800	negative regulation of cyclic nucleotide metabolic process	8	1.6%	46	0.1%	3.67e-05
GO:0006954	inflammatory response	27	5.4%	731	1.5%	3.89e-05
GO:0019725	cellular homeostasis	31	6.2%	935	2.0%	4.30e-05
GO:0044060	regulation of endocrine process	8	1.6%	47	0.1%	4.38e-05
GO:0007193	adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway	9	1.8%	74	0.2%	0.00012
GO:0046058	cAMP metabolic process	12	2.4%	154	0.3%	0.00014
GO:0030818	negative regulation of cAMP biosynthetic process	7	1.4%	37	0.1%	0.00015
GO:0019233	sensory perception of pain	10	2.0%	101	0.2%	0.00017
GO:0007194	negative regulation of adenylate cyclase activity	6	1.2%	24	0.1%	0.00021
GO:0002685	regulation of leukocyte migration	12	2.4%	161	0.3%	0.00023
GO:0030799	regulation of cyclic nucleotide metabolic process	12	2.4%	162	0.3%	0.00024
GO:0006968	cellular defense response	8	1.6%	59	0.1%	0.00027
GO:0030803	negative regulation of cyclic nucleotide biosynthetic process	7	1.4%	41	0.1%	0.00031
GO:0002687	positive regulation of leukocyte migration	10	2.0%	110	0.2%	0.00039
GO:0030809	negative regulation of nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00044
GO:1900372	negative regulation of purine nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00044
GO:0050886	endocrine process	9	1.8%	86	0.2%	0.00046
GO:0030817	regulation of cAMP biosynthetic process	10	2.0%	114	0.2%	0.00055
GO:0031280	negative regulation of cyclase activity	6	1.2%	28	0.1%	0.00056
GO:1900543	negative regulation of purine nucleotide metabolic process	8	1.6%	67	0.1%	0.00075
GO:0048878	chemical homeostasis	32	6.4%	1123	2.4%	0.00075

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0045980	negative regulation of nucleotide metabolic process	8	1.6%	68	0.1%	0.00084
GO:0051350	negative regulation of lyase activity	6	1.2%	30	0.1%	0.00087
GO:0019221	cytokine-mediated signaling pathway	24	4.8%	706	1.5%	0.00100
GO:0045761	regulation of adenylate cyclase activity	8	1.6%	70	0.1%	0.00106
GO:0042755	eating behavior	6	1.2%	31	0.1%	0.00108
GO:0070474	positive regulation of uterine smooth muscle contraction	4	0.8%	8	0.0%	0.00112
GO:0006171	cAMP biosynthetic process	10	2.0%	125	0.3%	0.00129
GO:0050900	leukocyte migration	17	3.4%	385	0.8%	0.00134
GO:0030595	leukocyte chemotaxis	12	2.4%	191	0.4%	0.00145
GO:0045987	positive regulation of smooth muscle contraction	6	1.2%	33	0.1%	0.00159
GO:0051930	regulation of sensory perception of pain	6	1.2%	36	0.1%	0.00273
GO:0051931	regulation of sensory perception	6	1.2%	36	0.1%	0.00273
GO:0071345	cellular response to cytokine stimulus	26	5.2%	856	1.8%	0.00282
GO:0070472	regulation of uterine smooth muscle contraction	4	0.8%	10	0.0%	0.00333
GO:0030802	regulation of cyclic nucleotide biosynthetic process	10	2.0%	140	0.3%	0.00365
GO:0051047	positive regulation of secretion	16	3.2%	374	0.8%	0.00417
GO:0040011	locomotion	41	8.2%	1779	3.8%	0.00459
GO:1900542	regulation of purine nucleotide metabolic process	12	2.4%	214	0.5%	0.00477
GO:0008217	regulation of blood pressure	11	2.2%	179	0.4%	0.00505
GO:0070471	uterine smooth muscle contraction	4	0.8%	11	0.0%	0.00519
GO:1900371	regulation of purine nucleotide biosynthetic process	10	2.0%	146	0.3%	0.00532
GO:0030808	regulation of nucleotide biosynthetic process	10	2.0%	147	0.3%	0.00565
GO:0003018	vascular process in circulatory system	11	2.2%	182	0.4%	0.00593
GO:0070374	positive regulation of ERK1 and ERK2 cascade	11	2.2%	184	0.4%	0.00658
GO:0006140	regulation of nucleotide metabolic process	12	2.4%	221	0.5%	0.00663
GO:0035296	regulation of tube diameter	10	2.0%	150	0.3%	0.00677
GO:0097746	regulation of blood vessel diameter	10	2.0%	150	0.3%	0.00677
GO:0031279	regulation of cyclase activity	8	1.6%	90	0.2%	0.00727
GO:0051339	regulation of lyase activity	8	1.6%	92	0.2%	0.00857
GO:0009187	cyclic nucleotide metabolic process	12	2.4%	229	0.5%	0.00952
GO:0003081	regulation of systemic arterial blood pressure by renin-angiotensin	5	1.0%	26	0.1%	0.00959

7.2.2 CXCR4

Table 7.13 - - Protein PSI-BLAST results for query protein CXCR4_HUMAN (P61073). Results of iteration 7. Default parameters were used with cut-off value of 1×10^{-7} , limited to 500 hits. Results retrieved from the saved file, with no gene name entry available

Accession	Score	E value
Q8MJW8.1	357	2E-121
P32246.1	354	2E-120
P51683.2	354	4E-120
P51678.2	353	4E-120
O55193.1	353	8E-120
P51680.2	352	1E-119
Q64H34.1	351	5E-119
Q2HJ17.1	350	5E-119
P61072.1	350	8E-119
O54814.1	349	2E-118
Q8HZU0.1	348	4E-118
P56491.1	348	5E-118
P41597.1	348	6E-118
Q28474.2	347	7E-118
Q3LSL6.1	347	7E-118
P51679.1	347	1E-117
Q8HZU1.1	347	1E-117
Q7YS92.1	347	1E-117
Q9TSQ8.1	346	2E-117
O18793.2	346	3E-117
P79394.1	346	4E-117
O08565.1	344	1E-116
O62747.1	344	2E-116
P56498.1	343	4E-116
Q9XT76.1	342	5E-116
P51682.3	342	9E-116
Q9BGN6.1	341	2E-115
Q764M9.1	341	3E-115
P25930.1	340	4E-115
Q5ECR9.1	339	1E-114
P70658.2	337	8E-114
P21109.2	334	2E-112
Q2YEG0.1	330	4E-111
Q1ZY22.1	330	7E-111
Q7ZXJ7.1	328	5E-110
Q95NC4.1	326	2E-109
P61813.1	325	5E-109
P56493.1	325	5E-109
Q867B2.1	325	6E-109

Accession	Score	E value
O97883.1	325	6E-109
O97975.1	325	6E-109
O62743.2	324	8E-109
Q95NC2.1	324	8E-109
Q95ND2.1	324	8E-109
Q95ND1.1	324	9E-109
P61755.1	324	1E-108
P25104.1	324	1E-108
O77590.2	324	1E-108
Q95NC1.1	324	1E-108
Q95NE8.1	324	1E-108
Q28003.1	324	1E-108
Q6WN98.1	324	2E-108
Q9TV47.1	323	2E-108
Q9TV49.1	323	2E-108
Q95NC3.1	323	2E-108
P56439.1	323	2E-108
Q9TV42.1	323	3E-108
Q5MD61.1	323	4E-108
O08556.1	323	4E-108
O97962.1	322	4E-108
P30555.1	323	4E-108
Q95ND0.1	322	4E-108
Q95NC5.1	322	4E-108
Q9TV43.1	322	5E-108
O97878.1	322	5E-108
Q9TV48.1	322	5E-108
O97881.1	322	5E-108
O97882.1	322	6E-108
O97879.1	322	7E-108
Q95NC8.1	322	7E-108
Q95NC0.1	322	1E-107
P56440.1	321	1E-107
Q95NC6.1	321	1E-107
Q9Z2I3.1	321	2E-107
P51681.1	321	2E-107
O35210.1	321	2E-107
Q8HZT9.1	320	2E-107
P51686.2	321	3E-107
Q95NC7.1	320	3E-107
O97880.1	320	3E-107
P25095.1	320	4E-107
P32303.2	320	5E-107
P35407.2	320	6E-107
P29754.1	319	1E-106
P43240.1	319	2E-106
P51675.2	318	3E-106

Accession	Score	E value
Q9GLN9.1	317	5E-106
P34976.1	317	7E-106
P30556.1	317	7E-106
P70612.1	316	1E-105
Q1WLP9.1	317	1E-105
P35373.1	316	2E-105
P29755.1	316	2E-105
Q9WV26.1	315	3E-105
O54689.1	315	6E-105
E9QJ73.2	315	8E-105
Q9WUT7.1	314	2E-104
Q9TV45.1	313	2E-104
P29089.1	313	3E-104
P56483.1	312	8E-104
Q9NPB9.1	311	1E-103
P35344.1	311	1E-103
Q95NC9.1	310	2E-103
P56492.1	310	3E-103
Q7SZP9.2	310	3E-103
Q07FZ4.1	310	5E-103
Q9BDS8.1	309	7E-103
P35350.1	309	8E-103
P51684.2	310	9E-103
P51677.1	309	1E-102
P79960.2	308	3E-102
Q9YGC3.1	307	4E-102
P30937.1	308	7E-102
P41144.1	307	1E-101
P35343.1	306	1E-101
Q2TAD5.1	305	3E-101
Q2Y2P0.1	305	4E-101
P56482.1	304	5E-101
O97571.1	304	5E-101
Q4VA82.1	305	6E-101
P49660.2	305	7E-101
A0T2N3.1	304	8E-101
P56484.1	304	9E-101
Q95M54.1	305	1E-100
P34975.1	304	1E-100
P33534.1	303	4E-100
Q9MYW9.2	302	2E-99
P79928.1	305	6E-99
Q9JHG3.1	300	7E-99
Q04683.2	299	9E-99
P34997.1	298	3E-98
P32249.3	297	3E-98
P41231.4	297	1E-97

Accession	Score	E value
Q9WV08.1	296	2E-97
Q924I3.1	295	2E-97
P51676.2	295	2E-97
P35346.3	295	2E-97
O97666.1	295	3E-97
P51685.1	294	4E-97
P30875.1	295	5E-97
P35414.1	294	1E-96
P30874.1	294	1E-96
P34994.1	293	2E-96
P30680.1	293	2E-96
F1MV99.2	293	2E-96
Q2YEG2.1	292	3E-96
P25023.4	294	3E-96
Q9EQ16.1	292	3E-96
Q99JA4.1	292	4E-96
D4A7K7.1	291	1E-95
Q49LX6.1	291	2E-95
P55919.1	289	4E-95
P34993.1	289	6E-95
Q924T8.1	288	7E-95
P25024.2	288	1E-94
O97665.1	288	1E-94
P25090.2	288	1E-94
Q2NNR5.1	287	1E-94
P32299.4	289	2E-94
P55920.1	287	2E-94
P33535.1	289	3E-94
Q2YEF9.1	287	4E-94
Q810W6.1	286	7E-94
Q5IS84.1	288	8E-94
O09027.1	287	1E-93
P30411.2	287	1E-93
P51582.1	286	1E-93
P31391.2	287	1E-93
P41143.4	286	1E-93
P28646.1	287	2E-93
P42866.1	287	2E-93
P25025.2	285	2E-93
O35811.1	285	2E-93
O18983.1	285	2E-93
P79177.1	285	2E-93
P79242.1	285	2E-93
P49682.2	285	3E-93
P30873.1	286	3E-93
Q13304.2	285	4E-93
Q9XT45.1	284	4E-93
Q28807.2	284	5E-93

Accession	Score	E value
Q5IS39.1	285	7E-93
P34996.1	284	7E-93
P79785.1	284	7E-93
P49652.1	284	8E-93
Q9N0Z0.1	283	8E-93
P33396.1	283	9E-93
Q2KIP6.1	284	1E-92
P35372.2	285	1E-92
O19024.1	282	2E-92
Q9BDS6.1	282	2E-92
Q5YA25.1	283	3E-92
P41145.2	283	3E-92
Q6NS65.1	281	5E-92
P79190.1	281	5E-92
P79218.1	282	5E-92
Q95247.1	282	9E-92
Q99677.1	282	9E-92
Q28519.1	281	9E-92
P47774.2	282	1E-91
Q09QM4.1	280	1E-91
P32248.2	281	1E-91
Q9TV16.1	280	1E-91
P48042.1	281	1E-91
P33766.1	280	2E-91
P97468.1	281	2E-91
O00574.1	280	2E-91
P79350.2	282	2E-91
Q9JJS7.1	280	2E-91
Q49LX5.1	281	2E-91
P32300.1	280	2E-91
P30872.1	281	2E-91
P21462.3	280	2E-91
P0C7U4.1	280	3E-91
P33533.1	280	3E-91
P32302.1	280	4E-91
P49238.1	279	5E-91
Q2KTE1.1	279	6E-91
P79236.1	278	6E-91
P97520.1	278	7E-91
B9VR26.1	279	7E-91
Q5MD62.1	279	8E-91
P21452.3	280	9E-91
Q8BLG2.2	279	1E-90
Q28642.1	278	1E-90
P49650.1	278	1E-90
P25106.3	278	1E-90
O89039.2	278	2E-90
P49651.1	278	2E-90

Accession	Score	E value
P56485.2	278	2E-90
P47900.1	278	2E-90
O08726.1	278	3E-90
Q28422.1	277	3E-90
P32250.1	275	3E-90
Q1RMI1.1	277	4E-90
Q99788.2	277	5E-90
O70526.1	277	6E-90
O88854.2	277	6E-90
P79189.1	276	6E-90
O43603.1	277	7E-90
Q5DUB2.1	277	7E-90
P79235.1	275	9E-90
Q8BMC0.1	275	1E-89
Q3U6B2.1	275	2E-89
F5HF62.1	275	2E-89
P05363.1	276	2E-89
P46093.2	275	2E-89
O35786.1	275	3E-89
Q9GLX8.1	275	3E-89
P16610.1	276	3E-89
P59902.1	275	3E-89
Q8BUD0.1	275	3E-89
Q4G072.1	274	3E-89
Q4KLH9.1	275	4E-89
Q64077.1	276	4E-89
O88536.1	274	5E-89
Q05394.1	273	6E-89
P79237.1	273	6E-89
Q9Z0D9.1	273	7E-89
P50132.1	273	7E-89
P30549.1	274	9E-89
P79241.1	273	1E-88
Q8K087.1	272	2E-88
Q75ZH0.1	272	2E-88
P79175.2	272	2E-88
P79243.1	272	2E-88
B1PHQ8.1	272	3E-88
O97664.1	272	3E-88
P25089.2	272	3E-88
Q95LH1.1	272	4E-88
P46090.1	272	4E-88
Q1JQB3.1	272	4E-88
P79176.1	271	5E-88
Q9Y271.1	270	7E-88
Q5KSK8.1	271	9E-88
Q9XSD7.1	270	1E-87
Q8HZP1.1	270	2E-87

Accession	Score	E value
O00421.2	270	2E-87
P29371.1	273	2E-87
Q95N02.1	269	2E-87
P79292.1	270	2E-87
P47748.1	270	2E-87
O88410.2	270	3E-87
Q8HZN9.1	269	3E-87
P30938.2	270	3E-87
P21730.2	269	4E-87
P46663.3	269	4E-87
P41232.2	269	5E-87
P30993.2	268	5E-87
P21556.2	268	5E-87
P79178.1	268	5E-87
Q0II78.1	268	5E-87
P51144.1	270	5E-87
Q9BDQ5.1	268	7E-87
P79240.2	268	8E-87
P35370.1	267	2E-86
P41146.1	267	2E-86
Q8BGE9.1	271	3E-86
O00590.2	268	3E-86
Q9TUE1.1	267	3E-86
Q8HZP3.1	267	3E-86
P79191.1	267	3E-86
Q8HZP2.1	267	3E-86
Q95L01.1	267	3E-86
P35377.1	267	4E-86
P69332.1	266	5E-86
Q9JII9.1	267	5E-86
P46094.1	265	8E-86
P43657.3	265	8E-86
Q3BCU0.1	265	1E-85
Q89609.1	266	2E-85
P79234.1	264	2E-85
P48748.1	264	3E-85
O08707.2	265	3E-85
P35383.1	265	4E-85
B5X337.1	264	4E-85
Q15743.1	264	5E-85
P46091.2	263	8E-85
P11613.1	263	1E-84
O08858.3	263	1E-84
P35411.1	263	1E-84
A5PLE7.1	263	1E-84
Q63645.1	264	1E-84
Q9NSD7.1	266	2E-84
Q6IYF8.1	261	2E-84

Accession	Score	E value
P26824.1	264	2E-84
Q95N03.1	261	5E-84
P55086.1	262	6E-84
Q8BFQ3.1	261	7E-84
P49681.2	260	7E-84
P25105.1	260	9E-84
Q924T9.1	258	1E-83
A7YY44.1	262	2E-83
Q3SXG2.1	259	2E-83
Q9Y5X5.2	265	2E-83
Q6Y1R5.1	259	2E-83
Q56UD9.1	258	3E-83
F7EQ49.1	260	3E-83
Q920A1.2	257	4E-83
Q96P68.1	258	5E-83
P79188.1	258	5E-83
P47749.1	260	6E-83
Q61125.1	258	6E-83
Q62035.1	258	6E-83
Q2WED0.1	258	7E-83
Q71MR7.2	257	8E-83
Q99527.1	258	8E-83
Q0VDU3.2	258	1E-82
Q8BMP4.2	258	2E-82
P30974.2	262	2E-82
P0C7U5.1	257	2E-82
P47937.2	260	2E-82
B0UXR0.1	257	3E-82
P16177.1	260	4E-82
O46685.1	257	4E-82
P46002.1	255	6E-82
Q8MJV2.1	255	7E-82
P32745.1	258	7E-82
Q9TUP7.1	258	9E-82
Q00991.2	257	1E-81
O62809.2	257	2E-81
Q91V45.1	256	2E-81
P56719.1	257	3E-81
Q924H0.2	256	3E-81
Q9NS75.1	253	3E-81
P58308.2	257	4E-81
Q9TTY5.2	253	5E-81
O88537.2	253	5E-81
Q9GK76.1	253	7E-81
O08878.1	253	7E-81
P56488.1	255	9E-81
Q9ERK9.1	252	9E-81
P55085.1	254	1E-80

Accession	Score	E value
Q6BD04.1	253	1E-80
Q924U1.2	253	1E-80
O08790.2	252	2E-80
Q9XSD4.2	252	2E-80
O43614.2	255	2E-80
P49019.3	253	2E-80
P43142.1	253	2E-80
P97583.2	251	3E-80
Q15077.1	250	6E-80
P25116.2	252	9E-80
Q98907.1	249	9E-80
Q9R0M1.1	249	1E-79
A6QNL7.1	250	2E-79
P48145.2	248	2E-79
O97512.2	252	3E-79
Q8TDS4.1	249	3E-79
O15218.1	250	4E-79
P30936.1	250	5E-79
Q99705.2	250	6E-79
P30558.2	250	8E-79
Q8K418.1	249	9E-79
Q86VZ1.1	247	1E-78
P30935.2	249	1E-78
Q9EP86.1	249	2E-78
P30992.1	247	2E-78
O93361.1	246	2E-78
Q9BXC1.1	245	3E-78
P56479.1	246	3E-78
Q149R9.2	247	4E-78
O00155.2	246	4E-78
Q9EQD2.1	248	5E-78
Q9UPC5.2	247	5E-78
P50052.1	246	6E-78
Q8WPA2.1	245	7E-78
Q9EP66.1	245	1E-77
Q98982.1	247	1E-77
P30975.2	249	2E-77
Q3U507.1	243	2E-77
Q9H1C0.1	245	2E-77
O18982.1	244	3E-77
P58307.3	245	5E-77
Q62805.1	243	5E-77
O60755.1	243	5E-77
P0C5I1.1	243	7E-77
P35351.1	243	7E-77
O00254.1	243	7E-77
Q8MJ89.2	244	7E-77
Q969F8.2	244	7E-77

Accession	Score	E value
O97967.1	244	8E-77
Q58D85.1	243	8E-77
A4FUQ5.2	242	8E-77
P25931.2	246	9E-77
P49685.1	243	9E-77
Q8IN35.2	253	1E-76
Q3ZC80.1	243	1E-76
O97663.1	242	1E-76
P56412.1	242	1E-76
Q6H2Y3.1	243	1E-76
P35374.1	242	2E-76
P31392.3	243	2E-76
A1ZAX0.2	246	2E-76
O54798.1	243	2E-76
Q9R1K6.1	242	2E-76
B3G515.1	241	2E-76
B0F9W3.1	241	3E-76
Q9BDS7.1	241	3E-76
Q99MT6.2	240	3E-76
Q80Z39.1	241	4E-76
Q0GBZ5.1	243	4E-76
P32247.1	242	5E-76
P56718.1	242	7E-76
Q9Z0Z6.1	240	9E-76
Q8MJV3.1	239	1E-75
O97661.2	242	1E-75
Q8JZL2.2	241	2E-75

Accession	Score	E value
Q6IYF9.1	238	2E-75
O88626.1	240	2E-75
Q9BY21.1	239	3E-75
Q5ZI82.1	238	6E-75
P24053.1	238	8E-75
Q99MT7.2	237	1E-74
P48146.2	236	1E-74
Q9BXA5.2	236	1E-74
Q9H244.1	236	1E-74
P47211.3	236	2E-74
Q63371.1	236	2E-74
Q920E1.1	237	2E-74
O88853.2	236	3E-74
Q95KC3.1	235	3E-74
P30098.1	238	4E-74
Q2HJA4.1	236	5E-74
O54799.1	236	7E-74
Q9GZQ6.1	235	4E-73
P35371.1	234	5E-73
Q86917.1	233	1E-72
P97639.3	232	1E-72
Q9BXC0.1	231	1E-72
Q08520.1	231	5E-72
Q9EPX4.1	229	1E-71
P28336.2	230	1E-71
Q8C131.1	228	2E-71
O75388.1	228	3E-71

Accession	Score	E value
Q5DUB1.1	230	3E-71
O08675.2	229	3E-71
Q5DUB3.1	229	7E-71
P30548.2	229	7E-71
P47751.2	228	1E-70
P14600.1	228	1E-70
O00398.1	226	1E-70
Q8IYL9.1	226	2E-70
Q8BFU7.1	225	4E-70
P30547.1	227	4E-70
P25103.1	227	4E-70
Q8SQ54.1	225	4E-70
Q8BW93.1	224	1E-69
Q77QP4.1	226	1E-69
Q61038.2	223	2E-69
Q8MJ88.1	223	2E-69
O43613.2	225	6E-69
Q96P65.2	224	9E-69
Q9CPV9.1	221	1E-68
Q28553.1	215	2E-68
O70129.1	220	5E-68
Q695P6.1	215	2E-66
Q23497.1	218	3E-66
Q969V1.1	213	1E-65
Q922J6.1	209	2E-63

7.2.2.1 GO terms

7.2.2.1.1 Molecular Function

Table 7.14 - GO terms of the Molecular Function Ontology for the CXCR4 Protein PSI-BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	407	81.4%	11202	23.7%	3.35e-163
GO:0008528	G-protein coupled peptide receptor activity	63	12.6%	185	0.4%	4.30e-75
GO:0001653	peptide receptor activity	63	12.6%	187	0.4%	9.63e-75
GO:0004930	G-protein coupled receptor activity	92	18.4%	1166	2.5%	6.43e-50
GO:0004888	transmembrane signaling receptor activity	93	18.6%	1924	4.1%	6.27e-33
GO:0099600	transmembrane receptor activity	93	18.6%	2023	4.3%	3.09e-31
GO:0001637	G-protein coupled chemoattractant receptor activity	22	4.4%	41	0.1%	4.44e-31
GO:0004950	chemokine receptor activity	22	4.4%	41	0.1%	4.44e-31
GO:0038023	signaling receptor activity	93	18.6%	2105	4.5%	6.51e-30
GO:0008188	neuropeptide receptor activity	23	4.6%	53	0.1%	1.03e-29
GO:0004872	receptor activity	93	18.6%	2540	5.4%	7.26e-24

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0060089	molecular transducer activity	93	18.6%	2540	5.4%	7.26e-24
GO:0042923	neuropeptide binding	15	3.0%	22	0.0%	3.02e-23
GO:0004871	signal transducer activity	93	18.6%	2641	5.6%	1.17e-22
GO:0004896	cytokine receptor activity	22	4.4%	147	0.3%	5.29e-17
GO:0016493	C-C chemokine receptor activity	10	2.0%	16	0.0%	1.21e-14
GO:0019956	chemokine binding	12	2.4%	31	0.1%	2.03e-14
GO:0001608	G-protein coupled nucleotide receptor activity	10	2.0%	18	0.0%	6.53e-14
GO:0045028	G-protein coupled purinergic nucleotide receptor activity	10	2.0%	18	0.0%	6.53e-14
GO:0001614	purinergic nucleotide receptor activity	10	2.0%	46	0.1%	4.67e-09
GO:0016502	nucleotide receptor activity	10	2.0%	46	0.1%	4.67e-09
GO:0004994	somatostatin receptor activity	5	1.0%	5	0.0%	1.30e-08
GO:0042277	peptide binding	23	4.6%	418	0.9%	1.91e-08
GO:0035586	purinergic receptor activity	10	2.0%	54	0.1%	2.54e-08
GO:0016494	C-X-C chemokine receptor activity	6	1.2%	11	0.0%	6.02e-08
GO:0033218	amide binding	23	4.6%	462	1.0%	1.32e-07
GO:0019955	cytokine binding	12	2.4%	111	0.2%	2.64e-07
GO:0004985	opioid receptor activity	6	1.2%	14	0.0%	3.81e-07
GO:0019957	C-C chemokine binding	5	1.0%	11	0.0%	5.70e-06
GO:0019958	C-X-C chemokine binding	5	1.0%	11	0.0%	5.70e-06
GO:0001631	cysteinyi leukotriene receptor activity	3	0.6%	3	0.0%	0.00011
GO:0071791	chemokine (C-C motif) ligand 5 binding	3	0.6%	3	0.0%	0.00011
GO:0017046	peptide hormone binding	6	1.2%	38	0.1%	0.00028
GO:0004966	galanin receptor activity	3	0.6%	4	0.0%	0.00046
GO:0015065	uridine nucleotide receptor activity	3	0.6%	4	0.0%	0.00046
GO:0071553	G-protein coupled pyrimidinergic nucleotide receptor activity	3	0.6%	4	0.0%	0.00046
GO:0042562	hormone binding	7	1.4%	71	0.2%	0.00105
GO:0004995	tachykinin receptor activity	3	0.6%	5	0.0%	0.00115
GO:0015057	thrombin-activated receptor activity	3	0.6%	7	0.0%	0.00399
GO:0004982	N-formyl peptide receptor activity	3	0.6%	8	0.0%	0.00634
GO:0004974	leukotriene receptor activity	3	0.6%	9	0.0%	0.00943
GO:0031681	G-protein beta-subunit binding	3	0.6%	9	0.0%	0.00943

7.2.2.1.2 Cellular Component

Table 7.15 - GO terms of the Cellular Component Ontology for the CXCR4 Protein PSI-BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	407	81.4%	11983	25.4%	5.70e-152
GO:0005887	integral component of plasma membrane	76	15.2%	1917	4.1%	3.69e-21
GO:0031226	intrinsic component of plasma membrane	76	15.2%	1982	4.2%	2.68e-20
GO:0044459	plasma membrane part	77	15.4%	3409	7.2%	2.57e-08

GO:0032590	dendrite membrane	4	0.8%	21	0.0%	0.00623
------------	-------------------	---	------	----	------	---------

7.2.2.1.3 Biological Process

Table 7.16 - GO terms of the Biological Process Ontology for the CXCR4 Protein PSI-BLAST results. Performed on GENERIC GENE ONTOLOGY (GO) TERM FINDER using goa_human ontology

GOID	Gene Ontology term TERM	Cluster frequency		Genome frequency		P-Value
		#/500	%	#/47246	%	
GO:XXXXXXX	unannotated	407	81.4%	12067	25.5%	1.15e-149
GO:0007186	G-protein coupled receptor signaling pathway	93	18.6%	1702	3.6%	5.04e-36
GO:0070098	chemokine-mediated signaling pathway	22	4.4%	97	0.2%	4.92e-20
GO:0007218	neuropeptide signaling pathway	23	4.6%	116	0.2%	1.37e-19
GO:0007200	phospholipase C-activating G-protein coupled receptor signaling pathway	21	4.2%	89	0.2%	1.96e-19
GO:0007204	positive regulation of cytosolic calcium ion concentration	29	5.8%	262	0.6%	9.83e-18
GO:0051480	regulation of cytosolic calcium ion concentration	29	5.8%	283	0.6%	8.49e-17
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	24	4.8%	185	0.4%	5.53e-16
GO:0006874	cellular calcium ion homeostasis	29	5.8%	382	0.8%	2.89e-13
GO:0055074	calcium ion homeostasis	29	5.8%	396	0.8%	7.45e-13
GO:0035589	G-protein coupled purinergic nucleotide receptor signaling pathway	10	2.0%	18	0.0%	8.84e-13
GO:0072503	cellular divalent inorganic cation homeostasis	29	5.8%	409	0.9%	1.73e-12
GO:0072507	divalent inorganic cation homeostasis	29	5.8%	429	0.9%	5.98e-12
GO:0006875	cellular metal ion homeostasis	30	6.0%	511	1.1%	8.19e-11
GO:0035588	G-protein coupled purinergic receptor signaling pathway	10	2.0%	31	0.1%	7.92e-10
GO:0055065	metal ion homeostasis	30	6.0%	583	1.2%	2.36e-09
GO:0030003	cellular cation homeostasis	30	6.0%	588	1.2%	2.93e-09
GO:0006873	cellular ion homeostasis	30	6.0%	602	1.3%	5.27e-09
GO:0035590	purinergic nucleotide receptor signaling pathway	10	2.0%	47	0.1%	7.95e-08
GO:0055080	cation homeostasis	30	6.0%	677	1.4%	9.35e-08
GO:0098771	inorganic ion homeostasis	30	6.0%	692	1.5%	1.58e-07
GO:0038169	somatostatin receptor signaling pathway	5	1.0%	5	0.0%	1.76e-07
GO:0038170	somatostatin signaling pathway	5	1.0%	5	0.0%	1.76e-07
GO:0007631	feeding behavior	13	2.6%	108	0.2%	1.99e-07
GO:0006935	chemotaxis	29	5.8%	659	1.4%	2.29e-07
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	15	3.0%	158	0.3%	2.37e-07
GO:0042330	taxis	29	5.8%	660	1.4%	2.37e-07
GO:0055082	cellular chemical homeostasis	31	6.2%	758	1.6%	3.21e-07
GO:0035587	purinergic receptor signaling pathway	10	2.0%	60	0.1%	1.02e-06
GO:0050801	ion homeostasis	30	6.0%	756	1.6%	1.27e-06
GO:0002407	dendritic cell chemotaxis	7	1.4%	22	0.0%	2.86e-06
GO:0030815	negative regulation of cAMP metabolic process	8	1.6%	41	0.1%	1.41e-05
GO:0036336	dendritic cell migration	7	1.4%	27	0.1%	1.42e-05

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0038003	opioid receptor signaling pathway	6	1.2%	17	0.0%	2.06e-05
GO:0030814	regulation of cAMP metabolic process	12	2.4%	130	0.3%	2.18e-05
GO:0030800	negative regulation of cyclic nucleotide metabolic process	8	1.6%	46	0.1%	3.69e-05
GO:0006954	inflammatory response	27	5.4%	731	1.5%	3.91e-05
GO:0019725	cellular homeostasis	31	6.2%	935	2.0%	4.32e-05
GO:0044060	regulation of endocrine process	8	1.6%	47	0.1%	4.40e-05
GO:0007193	adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway	9	1.8%	74	0.2%	0.00012
GO:0046058	cAMP metabolic process	12	2.4%	154	0.3%	0.00014
GO:0030818	negative regulation of cAMP biosynthetic process	7	1.4%	37	0.1%	0.00015
GO:0019233	sensory perception of pain	10	2.0%	101	0.2%	0.00017
GO:0007194	negative regulation of adenylate cyclase activity	6	1.2%	24	0.1%	0.00021
GO:0002685	regulation of leukocyte migration	12	2.4%	161	0.3%	0.00023
GO:0030799	regulation of cyclic nucleotide metabolic process	12	2.4%	162	0.3%	0.00025
GO:0006968	cellular defense response	8	1.6%	59	0.1%	0.00027
GO:0030803	negative regulation of cyclic nucleotide biosynthetic process	7	1.4%	41	0.1%	0.00031
GO:0002687	positive regulation of leukocyte migration	10	2.0%	110	0.2%	0.00039
GO:0030809	negative regulation of nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00044
GO:1900372	negative regulation of purine nucleotide biosynthetic process	7	1.4%	43	0.1%	0.00044
GO:0050886	endocrine process	9	1.8%	86	0.2%	0.00046
GO:0030817	regulation of cAMP biosynthetic process	10	2.0%	114	0.2%	0.00055
GO:0031280	negative regulation of cyclase activity	6	1.2%	28	0.1%	0.00057
GO:1900543	negative regulation of purine nucleotide metabolic process	8	1.6%	67	0.1%	0.00075
GO:0048878	chemical homeostasis	32	6.4%	1123	2.4%	0.00076
GO:0045980	negative regulation of nucleotide metabolic process	8	1.6%	68	0.1%	0.00085
GO:0051350	negative regulation of lyase activity	6	1.2%	30	0.1%	0.00088
GO:0019221	cytokine-mediated signaling pathway	24	4.8%	706	1.5%	0.00101
GO:0045761	regulation of adenylate cyclase activity	8	1.6%	70	0.1%	0.00106
GO:0042755	eating behavior	6	1.2%	31	0.1%	0.00108
GO:0070474	positive regulation of uterine smooth muscle contraction	4	0.8%	8	0.0%	0.00113
GO:0006171	cAMP biosynthetic process	10	2.0%	125	0.3%	0.00130
GO:0050900	leukocyte migration	17	3.4%	385	0.8%	0.00135
GO:0030595	leukocyte chemotaxis	12	2.4%	191	0.4%	0.00146
GO:0045987	positive regulation of smooth muscle contraction	6	1.2%	33	0.1%	0.00160
GO:0051930	regulation of sensory perception of pain	6	1.2%	36	0.1%	0.00274
GO:0051931	regulation of sensory perception	6	1.2%	36	0.1%	0.00274
GO:0071345	cellular response to cytokine stimulus	26	5.2%	856	1.8%	0.00283
GO:0070472	regulation of uterine smooth muscle contraction	4	0.8%	10	0.0%	0.00334

GOID	Gene Ontology term	Cluster frequency		Genome frequency		P-Value
	TERM	#/500	%	#/47246	%	
GO:0030802	regulation of cyclic nucleotide biosynthetic process	10	2.0%	140	0.3%	0.00366
GO:0051047	positive regulation of secretion	16	3.2%	374	0.8%	0.00419
GO:0040011	locomotion	41	8.2%	1779	3.8%	0.00461
GO:1900542	regulation of purine nucleotide metabolic process	12	2.4%	214	0.5%	0.00479
GO:0008217	regulation of blood pressure	11	2.2%	179	0.4%	0.00507
GO:0070471	uterine smooth muscle contraction	4	0.8%	11	0.0%	0.00521
GO:1900371	regulation of purine nucleotide biosynthetic process	10	2.0%	146	0.3%	0.00534
GO:0030808	regulation of nucleotide biosynthetic process	10	2.0%	147	0.3%	0.00568
GO:0003018	vascular process in circulatory system	11	2.2%	182	0.4%	0.00595
GO:0007268	chemical synaptic transmission	23	4.6%	731	1.5%	0.00619
GO:0098916	anterograde trans-synaptic signaling	23	4.6%	731	1.5%	0.00619
GO:0099536	synaptic signaling	23	4.6%	731	1.5%	0.00619
GO:0099537	trans-synaptic signaling	23	4.6%	731	1.5%	0.00619
GO:0070374	positive regulation of ERK1 and ERK2 cascade	11	2.2%	184	0.4%	0.00661
GO:0006140	regulation of nucleotide metabolic process	12	2.4%	221	0.5%	0.00666
GO:0035296	regulation of tube diameter	10	2.0%	150	0.3%	0.00680
GO:0097746	regulation of blood vessel diameter	10	2.0%	150	0.3%	0.00680
GO:0031279	regulation of cyclase activity	8	1.6%	90	0.2%	0.00730
GO:0051339	regulation of lyase activity	8	1.6%	92	0.2%	0.00860
GO:0009187	cyclic nucleotide metabolic process	12	2.4%	229	0.5%	0.00957
GO:0003081	regulation of systemic arterial blood pressure by renin-angiotensin	5	1.0%	26	0.1%	0.00963

7.3 Multiple sequence alignment

7.3.1 CCR5



Figure 7.1 - Conserved regions overview for the CCR5 Multiple Alignment using Ugene software.

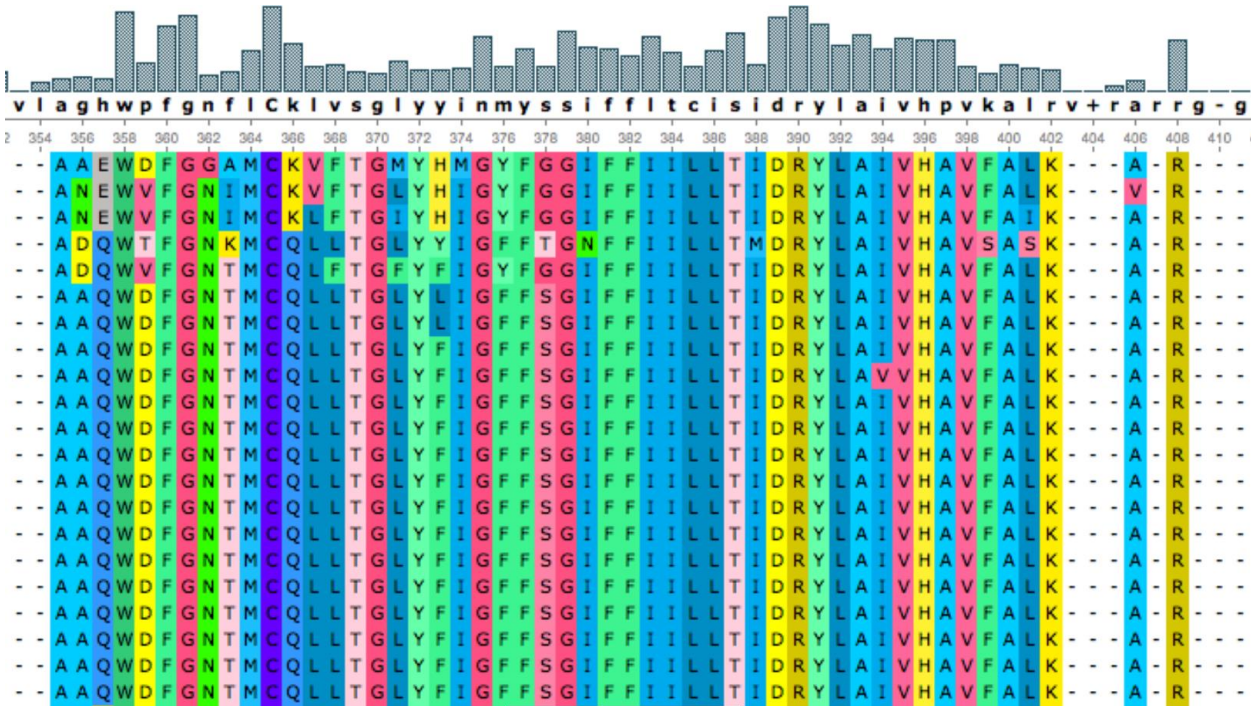


Figure 7.2 – Example of conserved region in the CCR5 Multiple Alignment, using CCR5 as consensus sequence in Ugene software.

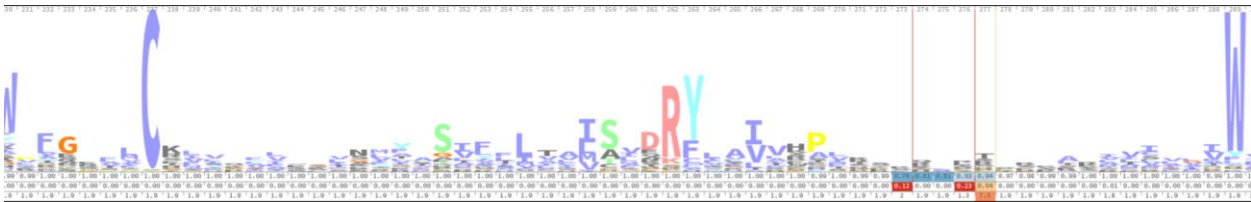


Figure 7.3 – Portion of Logo created on SkyView from the HMM model for the CCR5 multiple sequence alignment where the conservation is visible.

101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
C	K	F	V	H	F	F	F	Y	I	N	M	Y	C	S	I	F	F	L	T
A	R	I	I	S	Y	I	Y	H	M	S	F	F	A	G	V	L	L	M	A
S	Q	L	C	N	A	L	I	F	V	G	L	H	S	C	M	Y	C	I	C
T	T	V	L	G	G	M	L	Q	L	T	V	C	T	E	C	C	T	V	M
V	N	M	M	T	L	V	Q	S	T	A	I	Q	G	A	L	M	I	F	V
I	H	A	S	P	M	A	D	T	C	H	Q	W	V	T	T	W	M	N	G
L	M	Y	T	R	T	C	V	A	F	D	T	A	I	N	A	I	S	A	I
M	S	T	A	A	V	S	H	D	Y	E	Y	G	N	D	S	V	Y	C	S
F	Y	G	F	F	C	T	K	E	A	K	R	L	D	K	F	A	V	T	L
G	E	S	N	I	S	Y	M	G	H	M	A	S	E	Q	N	N	W	W	N
H	I	C	Q	L	I	G	P	I	G	Q	C	N	K	H	Y	S	A	Y	K
N	A	W	H	M	W	K	A	K	S	R	W	V	L	P	D	G	H	E	Q
P	D	E	Y	V	H	Q	C	L	W	C	E	I	M	R	E	H	D	H	D
Q	G	H	K	E	D	W	R	N	E	P	H	M	P	M	G	Q	E	K	E
W	L	K	P	K	E	E	T	M	N	V	K	E	Q	V	H	T	G	P	F
Y	P	N	D	Q	K	H	W	V	K	Y	S	K	R	Y	K	K	K	Q	H
D	F	P	E	D	N	N	E	W	P	F	D	R	F	F	P	R	N	R	P
E	V	Q	G	Y	P	P	N	R	Q	I	G	T	H	I	Q	D	P	S	R
K	W	R	R	C	Q	R	S	C	R	L	N	D	W	L	R	E	Q	D	Y
R	C	D	W	W	R	D	G	P	D	W	P	P	Y	W	W	P	R	G	W

Figure 7.4 – View of the PSSM outputted from PSI-BLAST of the CCR5 multiple alignment, using PSSM viewer, where high conservation is visible (note 101 position).

























1-letter code	3-letter code	Chemistry	Potential H-bonds	Molecular Weight	Isoelectric Point	Hydrophobicity
<u>C</u>	Cys	 -C-C-	0	103	5.0	0.721
<u>A</u>	Ala	 CH ₂ -C-C-	0	71	6.0	0.806
<u>S</u>	Ser	 -C-C-	3	87	5.7	0.601
<u>T</u>	Thr	 -C-C-	3	101	5.6	0.634
<u>V</u>	Val	 CH ₂ -C-C-	0	99	6.0	0.923
<u>I</u>	Ile	 CH ₂ -C-C-	0	113	6.0	1.000
<u>L</u>	Leu	 CH ₂ -C-C-	0	113	6.0	0.918
<u>M</u>	Met	 -C-C-	0	131	5.7	0.811
<u>F</u>	Phe	 CH ₂ 	0	147	5.5	0.951
<u>G</u>	Gly	 CH ₂ -C-C-	0	57	6.0	0.770
<u>H</u>	His	 CH ₂ 	3	137	7.6	0.548
<u>N</u>	Asn	 O=CN -C-C-	5	114	5.4	0.448
<u>P</u>	Pro	 CH ₂ 	0	97	6.3	0.678
<u>Q</u>	Gln	 O=CN -C-C-	5	128	5.7	0.430
<u>W</u>	Trp	 CH ₂ 	1	186	5.9	0.854
<u>Y</u>	Tyr	  	3	163	5.7	0.714
<u>D</u>	Asp	 CH ₂ -C-C-	4	115	3.0	0.417
<u>E</u>	Glu	 CH ₂ -C-C-	4	129	3.2	0.458

Figure 7.5 – Detailed view of position 101, with possible amino acids using PSSM viewer.

7.3.2 CXCR4



Figure 7.6 - Conserved regions overview for the CXCR4 Multiple Alignment using Ugene software.

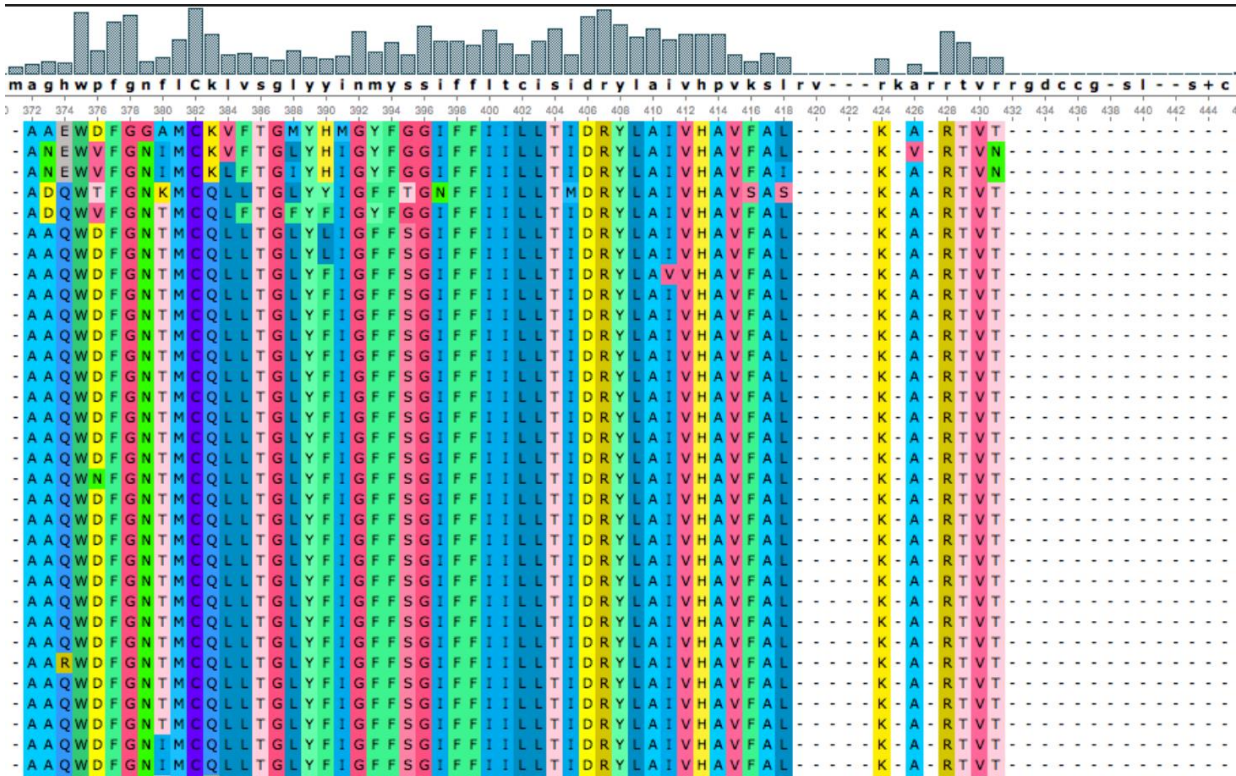


Figure 7.7 - Example of conserved region in the CXCR4 Multiple Alignment, using CXCR4 as consensus sequence in Ugene software.

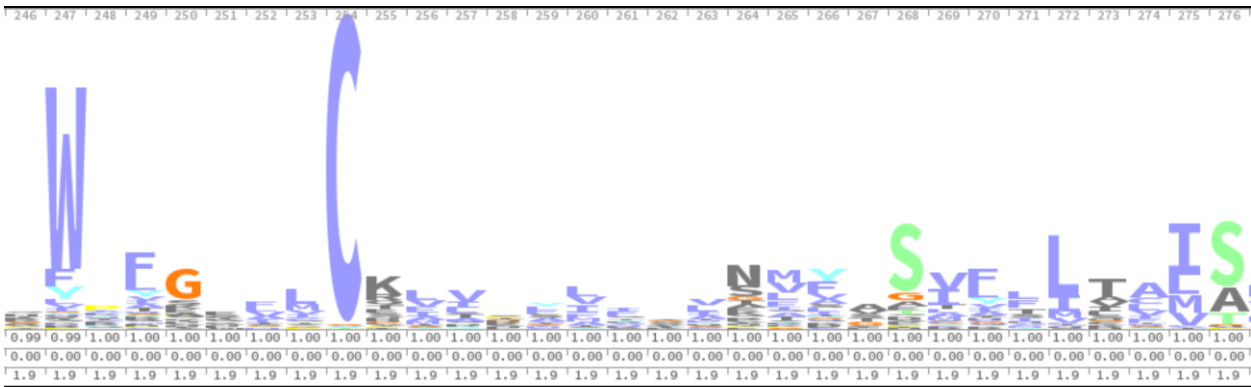


Figure 7.8 - Portion of Logo created on SkyView from the HMM model for the CXCR4 multiple sequence alignment where the conservation is visible.

101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
H	W	P	F	G	D	F	M	C	K	F	V	H	F	F	F	Y	I	N	M
N	Y	I	Y	H	E	A	L	A	R	I	I	S	Y	I	Y	F	M	S	F
D	G	V	L	P	N	V	F	I	Q	L	M	N	A	L	I	H	V	G	L
Q	F	Y	H	S	W	W	I	S	W	V	C	G	G	M	L	Q	L	T	V
E	N	L	M	A	K	I	Y	T	N	M	L	P	L	V	Q	I	T	A	I
G	E	F	W	D	S	G	A	V	T	A	T	T	M	A	D	S	C	D	Q
K	Q	M	I	Q	T	L	P	L	H	T	A	I	C	C	H	T	F	E	T
S	A	T	D	R	A	P	T	M	M	Y	F	R	S	T	K	A	Y	H	Y
T	C	H	T	Y	H	T	V	F	S	G	S	A	T	S	M	E	A	K	R
Y	H	R	V	E	P	C	G	G	Y	S	H	F	V	Y	P	G	W	M	A
A	K	S	G	K	R	M	S	H	E	C	N	L	I	G	T	K	G	Q	C
P	L	A	P	N	Q	Y	W	K	I	W	Q	M	W	K	V	L	H	R	W
R	M	D	A	I	G	H	C	N	A	E	Y	E	H	Q	A	N	S	C	E
F	T	E	C	T	V	N	Q	P	D	H	K	K	D	W	C	D	E	P	H
L	I	G	E	C	Y	S	E	Q	G	K	D	Q	E	E	R	M	N	V	K
M	R	K	K	M	F	E	H	W	L	N	E	V	K	H	W	V	K	Y	N
C	S	Q	N	W	I	K	K	Y	P	P	G	D	N	N	E	W	P	F	S
I	V	C	Q	F	L	Q	N	D	V	Q	P	Y	P	P	N	R	Q	I	D
V	D	N	R	L	M	D	R	E	C	R	R	C	Q	R	S	C	R	L	G
W	P	W	S	V	C	R	D	R	F	D	W	W	R	D	G	P	D	W	P

Figure 7.9 - View of the PSSM outputted from PSI-BLAST of the CXCR4 multiple sequence alignment, using PSSM viewer, where high conservation is visible (note positions 102, 109 and 119).

1-letter code	3-letter code	Chemistry	Potential H-bonds	Molecular Weight	Isoelectric Point	Hydrophobicity
<u>W</u>	Trp	0 CH ₂	1	186	5.9	0.854
<u>Y</u>	Tyr	OH	3	163	5.7	0.714
<u>G</u>	Gly	0 CH ₂ -C-C-	0	57	6.0	0.770
<u>F</u>	Phe	0 CH ₂	0	147	5.5	0.951
<u>N</u>	Asn	O=CN -C-C-	5	114	5.4	0.448
<u>E</u>	Glu	- CH ₂ -C-C-	4	129	3.2	0.458
<u>Q</u>	Gln	O=CN -C-C-	5	128	5.7	0.430
<u>A</u>	Ala	0 CH ₂ -C-C-	0	71	6.0	0.806
<u>C</u>	Cys	S -C-C-	0	103	5.0	0.721
<u>H</u>	His	CH ₂	3	137	7.6	0.548
<u>K</u>	Lys	+ CH ₂ -C-C-	3	128	9.7	0.263
<u>L</u>	Leu	0 CH ₂ -C-C-	0	113	6.0	0.918
<u>M</u>	Met	0 S -C-C-	0	131	5.7	0.811
<u>T</u>	Thr	OH -C-C-	3	101	5.6	0.634
<u>I</u>	Ile	0 CH ₂ -C-C-	0	113	6.0	1.000
<u>R</u>	Arg	+ CH ₂ -C-C-	7	156	10.8	0.000
<u>S</u>	Ser	OH -C-C-	3	87	5.7	0.601
<u>V</u>	Val	0 CH ₂ -C-C-	0	99	6.0	0.923

Figure 7.10 - Detailed view of position 102, with possible amino acids using PSSM viewer.

7.4 HMM Results

7.4.1 CCR5

Table 7.17 – Results of fitting proteins found in PSI-BLAST but not in BLAST for CCR5 to the HMM model created from the CCR5 multi alignment

Accession	full sequence			best domain		
	E-value	score	bias	E-value	score	bias
P30098.1	6.4e-109	356.9	14.0	8,00E-109	356.6	9.7
Q9EQD2.1	4,00E-106	347.7	14.2	5,00E-106	347.4	9.9
P30548.2	7.4e-104	340.2	21.4	9.3e-104	339.9	14.8
P14600.1	1.1e-103	339.7	21.8	1.3e-103	339.3	15.1
Q5DUB1.1	1.8e-103	338.9	18.4	2.3e-103	338.6	12.8
Q9EP86.1	1.4e-102	335.9	7.5	1.7e-102	335.7	5.2
P25103.1	1.9e-102	335.5	20.4	2.3e-102	335.3	14.1
Q5DUB3.1	2.6e-102	335.1	21.0	3.2e-102	334.8	14.6
P30547.1	4.2e-102	334.4	20.6	5.2e-102	334.1	14.3

Acession	full sequence			best domain		
	E-value	score	bias	E-value	score	bias
P51144.1	4.4e-102	334.3	22.9	5.5e-102	334.0	15.9
Q9GZQ6.1	4.8e-102	334.2	6.6	5.8e-102	333.9	4.6
P79928.1	1,00E-97	319.9	7.2	1.4e-97	319.5	5.0
Q98982.1	1.8e-97	319.1	23.1	2.1e-97	318.9	16.0
P25089.2	4,00E-93	304.7	21.8	4.6e-93	304.5	15.1
P56719.1	4.2e-93	304.7	14.7	4.8e-93	304.5	10.2
P58308.2	7.6e-93	303.8	16.5	8.9e-93	303.6	11.4
O62809.2	5.8e-92	300.9	16.2	7.4e-92	300.6	11.2
Q9TUP7.1	9.3e-92	300.3	16.8	1.2e-91	299.9	11.6
O43614.2	9.4e-92	300.2	17.2	1.2e-91	299.9	11.9
P46091.2	3.5e-91	298.4	25.5	4.3e-91	298.0	17.7
P41231.4	4.3e-90	294.8	7.3	5.1e-90	294.5	5.0
O97664.1	4.4e-90	294.7	26.3	5.5e-90	294.4	18.3
Q0GBZ5.1	1.1e-89	293.5	11.3	1.4e-89	293.1	7.8
Q95LH1.1	1.4e-89	293.0	26.1	1.8e-89	292.7	18.1
Q5YA25.1	2.8e-89	292.1	5.2	3.4e-89	291.8	3.6
O08790.2	7.5e-89	290.7	20.1	9.4e-89	290.3	13.9
P35383.1	8,00E-89	290.6	13.2	9.4e-89	290.3	9.2
P58307.3	1.2e-88	290.0	14.8	1.5e-88	289.7	10.3
P41232.2	1.3e-88	289.9	13.5	1.5e-88	289.7	9.4
O97661.2	5,00E-88	287.9	11.3	6.3e-88	287.6	7.9
P56718.1	2.6e-87	285.6	14.7	3.2e-87	285.3	10.2
O00155.2	7.3e-86	280.8	0.0	8.8e-86	280.5	0.0
Q3ZC80.1	1.4e-85	279.9	3.3	1.7e-85	279.6	2.3
Q86VZ1.1	1.6e-84	276.3	18.3	1.9e-84	276.1	12.7
P35371.1	8.5e-83	270.7	10.7	1,00E-82	270.4	7.4
P47751.2	1.9e-82	269.5	17.2	2.3e-82	269.3	11.9
Q8K418.1	3.3e-82	268.7	9.3	4,00E-82	268.5	6.5
Q6H2Y3.1	3.4e-81	265.4	12.6	4,00E-81	265.1	8.7
Q9H1C0.1	9.1e-81	264.0	6.8	1.1e-80	263.7	4.7
Q149R9.2	9.2e-81	264.0	8.4	1.2e-80	263.6	5.8
P32247.1	9.9e-81	263.9	14.8	1.2e-80	263.6	10.2
Q8IYL9.1	1,00E-80	263.8	20.4	1.3e-80	263.5	14.2
O54798.1	1.2e-80	263.6	9.9	1.3e-80	263.4	6.8
Q61038.2	7,00E-80	261.1	21.0	8.5e-80	260.8	14.5
O97967.1	9.8e-80	260.6	12.7	1.1e-79	260.4	8.8
O08675.2	2.5e-76	249.3	29.3	2.9e-76	249.1	20.3
Q95KC3.1	1.7e-73	240.0	18.3	2.2e-73	239.6	12.7
O00398.1	2.5e-73	239.4	30.9	2.9e-73	239.2	21.4
Q8BFU7.1	2.6e-72	236.0	30.1	2.9e-72	235.9	20.8
Q9ESG6.1	1.1e-71	234.0	20.1	1.3e-71	233.8	13.9
Q9BXC0.1	5.6e-71	231.7	15.0	7.1e-71	231.3	10.4
Q3SX17.1	1.8e-66	216.8	23.1	2.2e-66	216.5	16.0

Accession	full sequence			best domain		
	E-value	score	bias	E-value	score	bias
Q6GUG4.1	2.2e-65	213.2	24.0	2.9e-65	212.8	16.7
Q9D8I2.2	9.3e-63	204.5	22.7	1.1e-62	204.3	15.7
Q99678.2	4.8e-61	198.9	18.0	6.9e-60	195.1	12.4

7.4.2 CXCR4

Table 7.18 - Results of fitting proteins found in PSI-BLAST but not in BLAST for CXCR4 to the HMM model created from the CXCR4 multi alignment

Accession	full sequence			best domain		
	E-value	score	bias	E-value	score	bias
O97512.2	5.1e-100	327.8	14.2	6.5e-100	327.4	9.9
P41231.4	2.2e-98	322.4	10.8	2.6e-98	322.1	7.5
P30098.1	3.9e-98	321.5	16.5	5,00E-98	321.2	11.4
P16177.1	8.9e-98	320.4	15.1	1,00E-97	320.2	10.5
P35383.1	1.9e-97	319.3	15.3	2.2e-97	319.1	10.6
P30548.2	2.3e-97	319.0	22.7	3,00E-97	318.6	15.7
Q5DUB3.1	2.4e-97	318.9	21.4	3,00E-97	318.6	14.9
P47937.2	2.8e-97	318.7	15.5	3.2e-97	318.6	10.8
Q5DUB1.1	3,00E-97	318.6	19.5	3.8e-97	318.3	13.5
P41232.2	3.2e-97	318.6	16.0	3.9e-97	318.3	11.1
P14600.1	3.4e-97	318.4	22.9	4.4e-97	318.1	15.9
P30547.1	7.1e-97	317.4	21.3	8.9e-97	317.1	14.8
P25103.1	9,00E-97	317.1	22.1	1,00E-96	316.9	15.3
Q5YA25.1	1.9e-96	316.0	8.0	2.3e-96	315.7	5.5
Q00991.2	9.9e-96	313.6	20.6	1.1e-95	313.5	14.3
Q91V45.1	3.4e-95	311.8	13.0	3.8e-95	311.7	9.0
Q98982.1	8.7e-95	310.5	22.7	1,00E-94	310.3	15.7
P30975.2	1.7e-94	309.5	15.3	2,00E-94	309.3	10.6
Q924U1.2	2.3e-94	309.1	8.2	2.9e-94	308.8	5.7
Q969F8.2	7.2e-93	304.2	7.7	8.1e-93	304.0	5.3
A7YY44.1	9.1e-93	303.8	19.1	1.2e-92	303.4	13.3
Q8HZN9.1	3,00E-92	302.1	17.3	3.5e-92	301.9	12.0
Q9GZQ6.1	4.7e-92	301.5	9.6	5.4e-92	301.3	6.7
Q9EP86.1	2.5e-91	299.1	9.4	2.8e-91	298.9	6.5
Q9Y5X5.2	8.1e-90	294.1	12.0	1,00E-89	293.8	8.3
Q924H0.2	3,00E-88	288.9	16.0	3.2e-88	288.8	11.1
Q9EQD2.1	1.5e-87	286.6	13.8	1.8e-87	286.4	9.6
P35371.1	4.1e-87	285.2	9.4	4.6e-87	285.0	6.5
Q96P65.2	9.8e-86	280.6	13.2	1.2e-85	280.3	9.2
P24053.1	2,00E-85	279.6	10.9	2.5e-85	279.2	7.5
P47751.2	2.6e-85	279.2	16.4	3.1e-85	278.9	11.4

Acession	full sequence			best domain		
	E-value	score	bias	E-value	score	bias
Q6H2Y3.1	1,00E-84	277.2	14.0	1.1e-84	277.1	9.7
O54799.1	1.2e-84	277.0	12.1	1.5e-84	276.6	8.4
P32247.1	2.7e-84	275.8	16.1	3.1e-84	275.6	11.2
P28336.2	5.6e-84	274.8	12.0	6.9e-84	274.5	8.3
Q7TQP4.1	2,00E-83	273.0	18.2	2.2e-83	272.8	12.6
Q969V1.1	7.5e-83	271.1	22.1	8.7e-83	270.8	15.3
Q920E1.1	1.1e-80	263.9	29.2	1.3e-80	263.6	20.3
O08675.2	2.3e-79	259.6	29.8	2.6e-79	259.4	20.7
P49019.3	2.5e-78	256.1	14.5	3.2e-78	255.7	10.0
O75388.1	5.1e-77	251.8	17.6	6.5e-77	251.4	12.2
Q8TDS4.1	2.1e-76	249.8	16.6	2.6e-76	249.4	11.5
Q9CPV9.1	7.9e-75	244.5	19.9	9.5e-75	244.3	13.8
Q9BXC1.1	3.5e-74	242.4	23.3	4,00E-74	242.2	16.2
Q3U507.1	4.4e-74	242.1	23.9	5,00E-74	241.9	16.6
Q8C131.1	1.9e-73	240.0	19.2	2.2e-73	239.8	13.3
Q8BFU7.1	7.1e-73	238.1	30.9	7.6e-73	238.0	21.4
O00398.1	8.5e-73	237.8	31.4	9.5e-73	237.7	21.8
O35457.2	3.6e-54	176.3	4.2	4.3e-54	176.0	2.9