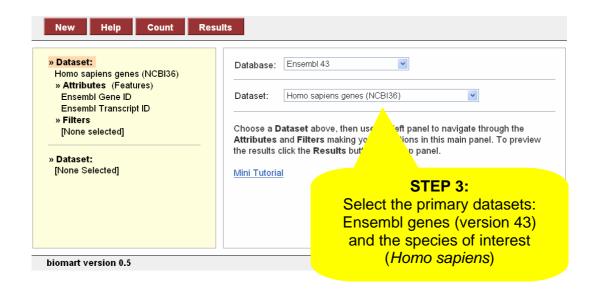
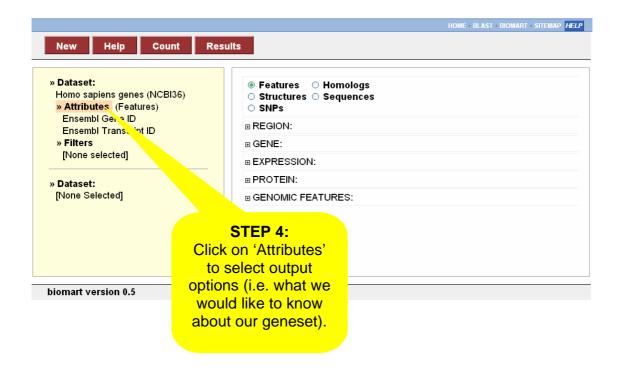
# Data mining in Ensembl with BioMart Worked Example

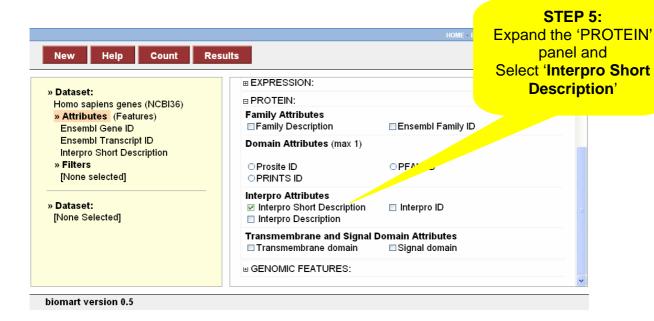
The human gene encoding Glucose-6-phosphate dehydrogenase (G6PD) is located on chromosome X in cytogenetic band q28.

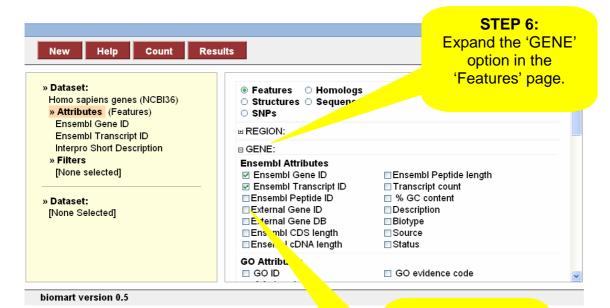
Which other genes related to human diseases locate to the same band? What are their Ensembl Gene IDs and Entrez Gene IDs? Do they have any functions predicted by Interpro?





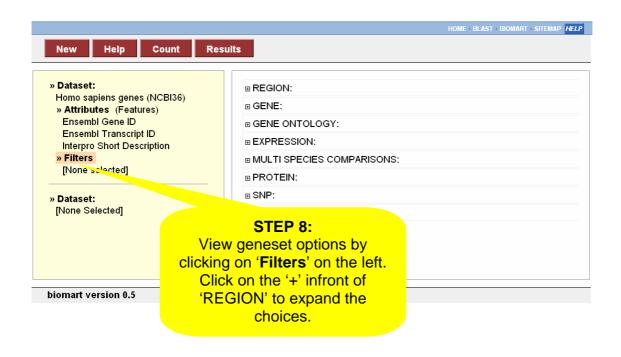


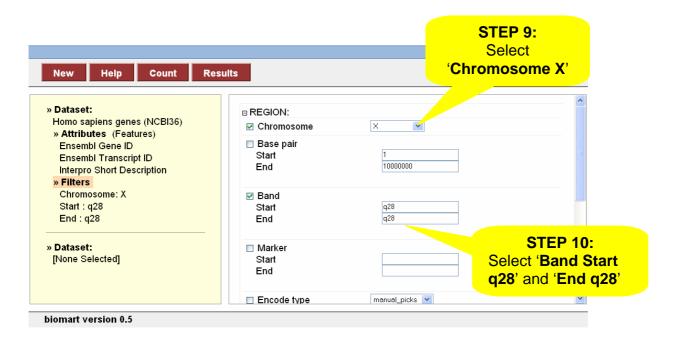


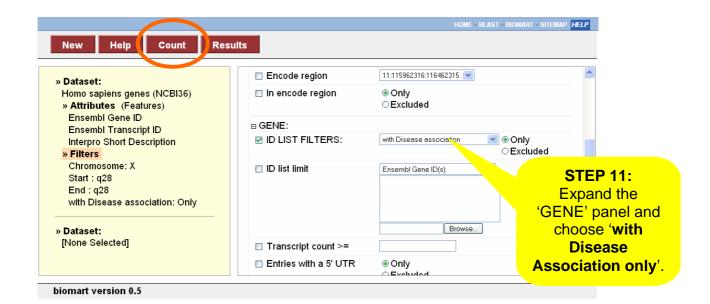


### **STEP 7:**

Select, along with the default options, 'External Gene ID'. Scroll down to select: 'EntrezGene ID' and 'Mim Gene Accession' (this is the OMIM ID).

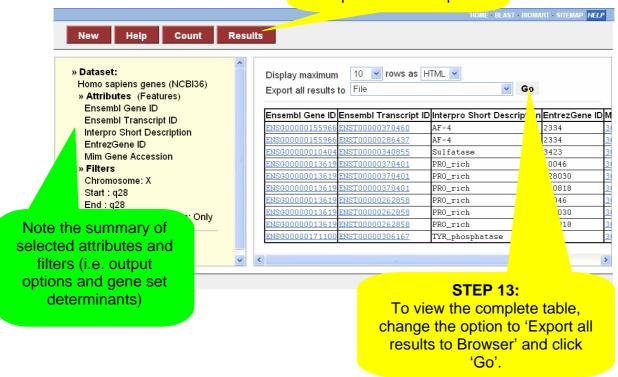






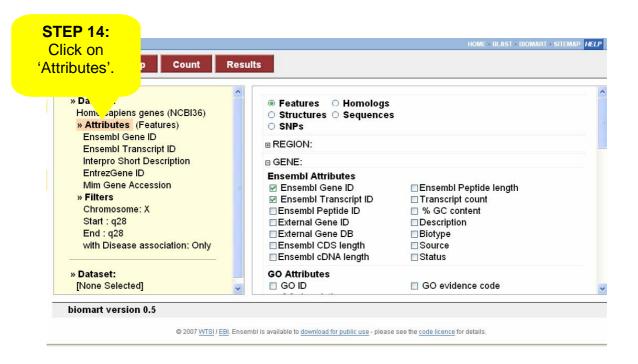
The filters have determined our gene set. Click 'Count' (at the top) to see how many genes have passed these filters.

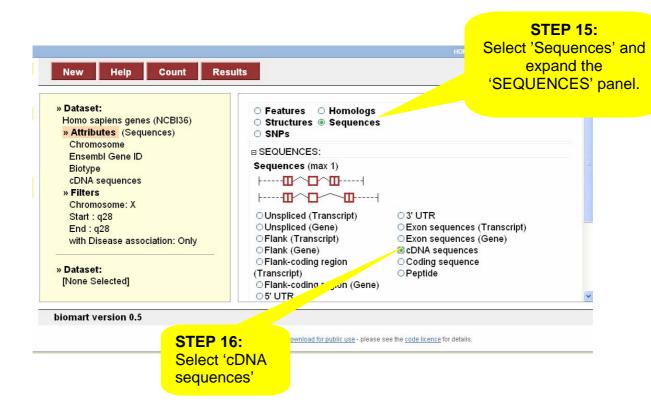
### STEP 12: Click 'RESULTS' at the top to preview the output.

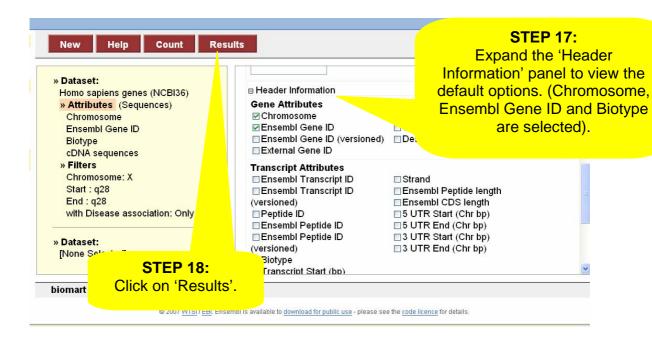


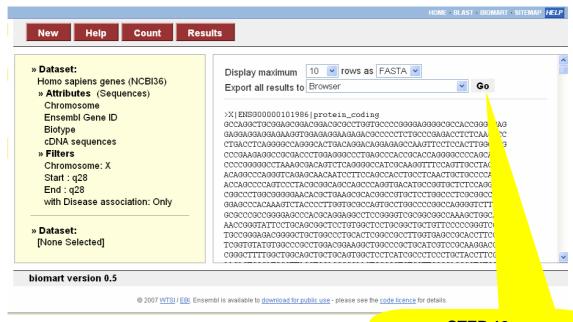
## Result Table 1

Ensembl Gene ID	Ensembl Transcript ID	Interpro Short Description	EntrezGene ID	Mim Gene Accession
ENSG00000155966	ENST00000370460	AF-4	2334	309548
ENSG00000155966	ENST00000286437	AF-4	2334	309548
ENSG00000010404	ENST00000340855	Sulfatase	3423	309900
ENSG00000013619	ENST00000370401	PRO_rich	10046	300120
ENSG00000013619	ENST00000370401	PRO_rich	728030	<u>300120</u>
ENSG00000013619	ENST00000370401	PRO_rich	730818	300120
ENSG00000013619	ENST00000262858	PRO_rich	10046	300120
ENSG00000013619	ENST00000262858	PRO_rich	728030	300120
ENSG00000013619	ENST00000262858	PRO_rich	730818	300120
ENSG00000171100	ENST00000306167	TYR_phosphatase	4534	300415
ENSG00000171100	ENST00000306167	GRAM	4534	300415
ENSG00000171100	ENST00000306167	Myotub-related	4534	300415
ENSG00000147383	ENST00000370274	Epimerase_Dh	50814	<u>300275</u>
ENSG00000147383	ENST00000370274	3Beta_HSD	50814	<u>300275</u>
ENSG00000147383	ENST00000370274	Polysac_CapD	50814	300275
ENSG00000147383	ENST00000370274	Male_sterile_C	50814	<u>300275</u>
ENSG00000130821	ENST00000330048	Na/ntran_symport	6535	<u>300036</u>
ENSG00000130821	ENST00000330048	Crt_transporter	6535	<u>300036</u>
ENSG00000130821	ENST00000253122	Na/ntran_symport	6535	<u>300036</u>
ENSG00000130821	ENST00000253122	Crt_transporter	6535	<u>300036</u>
ENSG00000185825	ENST00000345046	Macro_scav_rcpt	10134	300398
ENSG00000185825	ENST00000345046	Bap31	10134	300398
ENSG00000185825	ENST00000370133	Macro_scav_rcpt	10134	<u>300398</u>
ENSG00000185825	ENST00000370133	Bap31	10134	300398
ENSG00000101986	ENST00000218104	ABC_transp_like	215	<u>300371</u>
ENSG00000101986	ENST00000218104	ABC_Ald_N	215	300371
ENSG00000101986	ENST00000218104	ABC_transp_like	642762	<u>300371</u>
ENSG00000101986	ENST00000218104	ABC_Ald_N	642762	<u>300371</u>
ENSG00000198910	<u>ENST00000370060</u>	FnIII_subd	3897	<u>308840</u>
ENSG00000198910	ENST00000370060	VEGFR_N	3897	<u>308840</u>
ENSG00000198910	<u>ENST00000370060</u>	FN_III	3897	308840
ENSG00000198910	<u>ENST00000370060</u>	Ig_I-set	3897	<u>308840</u>
ENSG00000198910	<u>ENST00000370060</u>	Ig_V-set	3897	<u>308840</u>
	<u>ENST00000370060</u>	Ig	3897	<u>308840</u>
ENSG00000198910	<u>ENST00000370060</u>	Ig-like	3897	<u>308840</u>
ENSG00000198910	<u>ENST00000361699</u>	FnIII_subd	3897	<u>308840</u>
ENSG00000198910	<u>ENST00000361699</u>	VEGFR_N	3897	<u>308840</u>
ENSG00000198910	ENST00000361699	FN_III	3897	308840









### **STEP 19:**

To view all the sequences, export to browser.

### **RESULTS**

#### Header: chromosome, Ensembl Gene ID, Biotype

>X|ENSG00000155966|protein\_coding CCGCTGCCGCCGGCCGCAGCCAGCCAGCGGGGGCCCAGCCGGCTGAGCCCGCA GCGGCTGCCGCGCAGCGTCGGTTCGCTGGGTGCGCGGGCTACCGCGGACCGAGCGGACC CGAGTGGGCGACCAGGCGCTTGCCCGCCCAGTGCCACTGCCGCCGCTTCCTCGCCGGAGC ACAGGACCAGACACCTCCAGCGCCCGCTGCTGCCGATGCGGCCCGGACACTTTTAGC TGGGCGGGAGGCTGGAGAGCCGGGGGCCGCCGAGAACCGCCAGCGAGCTGTGCCGAGAG TGGATCTATTCGACTTTTTCAGAGACTGGGACTTGGAGCAGCAGTGTCACTATGAACAAG ACCGTAGTGCACTTAAAAAAAGGGAATGGGAGCGGAGGAATCAAGAAGTCCAGCAAGAAG ACGATCTCTTTCTTCAGGCTTTGATCTTTTTGGGGAGCCATACAAGGTAGCTGAATATA AAATTCCTTTCCTCTTGGAAGCTTTGGTCATAATATCATGGTTCAATTAAACGGATTCCA CCGGACTTTGTGATGAAAAAGGCTCTGTTAAAATCCAATTGAGTTTCCAAGAGGAAATTG TAGTAGGTCAAGATGCATGAGAGGGGAAGATGGAGGCCACCTCAGCTGGAGAACATGAGCT GAGTTGAGCCCTCAGTGTTGAAGTTGACTTGCTCCAAGCTGCAGTCTAAAACCCTGGGGC CCGTGCCTGGCCTATGCTCCCTCCCAAGTAAGTAGAGGAGCAGAACCATCAGGAACAGCC TGCCTGGCTCCTATGAAGAAAACTTCCTGACGTCCTGTCCCCAAAGGAAGACCCTTTCCC CAAGGGCACCCAGGTGGCCATTAAATTGTGATGATCATTCAGAAAGTGCCCCCTTGGCT TTATGAGAATCCAATTAGTCTTCTGAACCACCTTTTCTTGGGTGCAGATTTCCAACATTC ATGCTCATTGCAGATCCACCAACTGTCACTGTTCTTAACAAGCATGCTCGTCTTGTCAGA TAATGAAATGTTATCTGGTTTTTAAAAGCTGGTTTCATGTGCTTTATGTGTATAAAACTA TATCTGCCTGTGTGGCTTTGCATTTCAAATGTGTGGCGCACAAGCGTTTTGTTGGTGCTT TGTTCTCAGTACAGTAACTCTGTGTACAAACATTTTAATGTGGTTTTGTTGTTTTCCAAC AAGATGTCTCTGTAAAAATGATATTGGCTGAGCTGGTGCGTTGGTTTCTCTCATAGAGGC ATTAACTATACTGCCAATGCATTGAATTATTTAAAAATGCAAAATAAAATTTTTATGAAA

>X|ENSG00000102125|protein coding

GCAGCGCCCCACGGCCTGTGACCCCGGCGACCGCTCCCCAGTGACGAGAGAGCGGGGCC GGGCGCTGCTCCGGCCTGACCTGCGAAGGGACCTCGGTCCAGTCCCCTGTTGCGCCGCGC CCCCTGTCCGTCCGTGCGCGGCCAGTCAGGGGCCAGTGTCTCGAGCGGTCGAGGTCGCA GGGGATGCCTCTGCACGTGAAGTGGCCGTTCCCCGCGGTGCCGCCGCTCACCTGGACCCT GGCCAGCAGCGTCATGGGCTTGGTGGGCACCTACAGCTGCTTCTGGACCAAGTACAT GAACCACCTGACCGTGCACAACAGGGAGGTGCTGTACGAGCTCATCGAGAAGCGAGGCCC GGCCACGCCCTCATCACCGTGTCCAATCACCAGTCCTGCATGGACGACCCTCATCTCTG GGGGATCCTGAAACTCCGCCACATCTGGAACCTGAAGTTGATGCGTTGGACCCCTGCAGC TGCAGACATCTGCTTCACCAAGGAGCTACACTCCCACTTCTTCAGCTTGGGCAAGTGTGT GCCTGTGTGCCGAGGAGCAGAATTTTTCCAAGCAGAGAATGAGGGGAAAGGTGTTCTAGA CACAGGCAGGCACATGCCAGGTGCTGGAAAAAGAAGAAGAGAAAAGGAGATGGCGTCTACCA GAAGGGGATGGACTTCATTTTGGAGAAGCTCAACCATGGGGACTGGGTGCATATCTTCCC AGAAGGGAAAGTGAACATGAGTTCCGAATTCCTGCGTTTCAAGTGGGGAATCGGGCGCCT CGTCCTTCCTAACAGTCCGCCCTACTTCCCCCGCTTTGGACAGAAATCACTGTGCTGAT CGGGAAGCCCTTCAGTGCCCTGCCTGTACTCGAGCGGCTCCGGGCGGAGAACAAGTCGGC TGTGGAGATGCGGAAAGCCCTGACGGACTTCATTCAAGAGGAATTCCAGCATCTGAAGAC CTGGATTCTTGGCCCGCACAGAGCTGGGGCTGAGGGATGGACTGATGCTTTTAGCTCAAA CGTGGCTTTTAGACAGATTTGTTCATAGACCCTCTCAAGTGCCCTCTCCGAGCTGGTAGG CATTCCAGCTCCTCCGTGCTTCCTCAGTTACACAAAGGACCTCAGCTGCTTCTCCCACTT

cDNA<sub>1</sub>

cDNA 2