How to Use module_corr.r

Man Kam Kwong (May 6, 2017)

1. The only essential file for the program to work is module_corr.r. Place your data files in the same folder.
2. Prepare the data file.
   - Delete extra information, leaving only the labels of the samples in Row 1, the gene names in Column A, and module assignments in Column B as shown below.
   - Make sure that the word "module" in cell B1 is spelled exactly like that.
   - It is recommended to only use modules over a certain size, say 25 genes.

Symbol	module	TCGA-4Z-A	TCGA-4Z-A	TCGA-4Z-A	TCGA-4Z-A	TCGA-4Z-A	TCGA-4Z-A	TCGA-BL-A	TCGA-BL-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A	TCGA-BT-A
A1BG	23	3.631	2.34	0.397	4.862	16.166	0.925	2.211	5.14	0.374	11.812	1.489	0.549	2.752	1.195	1.885	0.316	3.67	2.886
A1CF	24	264.9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A2BP1	7	1	1	1	1	1	1	80.9	1	54.1	1	1	1	370.6	77.6	1	1	1	1
A2LD1	25	1.552	2.197	5.728	0.568	1.926	1.527	0.561	0.859	0.771	1.094	1.477	0.481	0.815	1.47	0.269	2.275	1.299	1.212
A2M	26	1.684	0.804	0.236	0.999	0.505	0.954	3.16	5.732	0.67	1.175	2.69	0.481	0.094	1.028	1.439	0.345	2.937	0.437
A2ML1	17	18.879	5.397	42.885	2.636	0.824	10.746	0.002	0.128	80.058	4.876	0	7.079	23.817	17.201	20.562	2.542	32.58	2.292
A4GALT	27	0.354	0.411	2.518	1.189	0.421	1.361	0.366	1.658	1.566	0.475	0.544	1.032	1.773	0.831	7.261	1.65	2.158	0.54
A4GNT	28	0.023	0.023	0.023	1.153	0.023	1.573	0.023	0.023	1.27	7.167	2.373	0.023	0.023	0.023	0.023	8.981	0.603	3.129
AAA1	29	1	1	1	1	1	1	1	137.5	1	61	1	1	1	1	1	1	1	1
AAAS	30	0.826	0.829	0.796	0.796	0.773	0.61	1.449	1.177	1.085	1.204	1.001	0.976	1.473	1.041	1.332	0.809	1.086	1.186
AACS	4	1.443	2.93	1.203	2.094	0.842	1.402	0.825	0.939	1.755	4.508	0.928	2.248	1.658	1.432	1.139	0.676	1.364	1.506
AACSL	7	14.426	0.02	0.02	0.02	0.02	1.327	1.602	2.723	0.02	10.881	1.002	10.343	0.02	0.02	26.653	90.927	11.717	3.519
AADAC	34	1.014	0.317	0.001	9.297	0.001	0.001	0.098	0.406	22.914	5.655	4.48	6.31	0.099	0.031	0.001	0.154	3.295	9.769
AADACL2	31	264.9	1	1	49.1	1	1	1	1329.2	1	183.1	1	130.6	247.1	1	1	1	77.2	1
AADACL3	32	1	1	1	1	52.1	2745.2	1	1	54.1	1	1	1	1	1	49.8	510.2	1	1
AADACL4	33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	51.4	1
AADAT	2	0.259	0.056	0.237	0.296	0.034	0.767	0.836	0.534	0.297	0.222	0.911	0.16	0.258	1.463	1.248	0.285	0.203	0.226
AAGAB	35	1.065	1.353	0.966	0.728	0.668	1.144	0.755	0.851	1.357	0.861	0.838	1.172	1.373	0.766	1.033	1.077	1.297	0.678
AAK1	36	1.404	0.697	1.715	1.932	1.883	1.43	1.084	1.294	1.026	0.717	1.106	1.211	1.158	0.874	3.161	1.268	0.872	0.932
AAMP	37	1.095	1.189	1.397	0.949	0.97	0.747	1.045	0.943	1.373	1.082	0.853	1.372	0.81	0.659	1.116	0.963	0.846	0.604
AANAT	38	11.968	3.948	2.213	3.946	1.046	1.345	1.624	0.02	0.02	1.225	4.06	1.311	0.02	3.894	6.006	0.02	2.066	2.677
AARS	3	0.715	0.861	0.462	1.642	0.578	1.175	1.605	1.104	1.224	1.255	1.678	1.191	2.858	1.395	1.427	1.56	0.789	0.836
AARS2	39	1.341	1.242	1.064	0.898	0.975	0.654	0.846	0.779	1.023	1	0.837	0.98	1.088	0.259	0.683	1.387	0.831	0.825
AARSD1	40	0.683	1.26	1.03	0.449	0.614	0.693	0.926	1.081	0.61	0.866	0.6	1.178	0.645	0.785	0.766	0.485	0.708	0.715
AASDH	42	1.255	0.837	0.443	0.94	0.83	0.797	1.126	0.947	1.332	0.976	1.248	1.512	1.105	1.456	0.569	1.279	0.799	0.618

   - File -> Save As in a CSV format, for example data.csv.
3. In R, use the command

   source('module_corr.r')

   to load the program. You only need to do this one time for the Project.
4. To analyze the data stored in data.csv, use the command

   mmsc('data')

   The output will be in the file MS20 0.6 output.csv.
5. This will produce a file with the following aligned elements:
   1. At the top, the collapsed value for all modules
   2. At the left, the original input values
   3. At the right, the Spearman correlation values for all genes to all modules