SJARACNe Tuturial

Version 1.0.0

Alireza Khatamian <akhatami@stjude.org>

August 2018

Gene Regulatory Network Reconstruction

Introduction

SJARACNe is an improved and extended version of Algorithm for the Reconstruction of Accurate Cellular Network. The algorithm uses adaptive partitioning mutual information to calculate the correlation between all pairs of genes to reconstruct the regulatory network. The core application is written in C++ and the pipeline in Python 3.6.X. The following libraries are required for Python:

- numpy (1.14.2)
- argparse (1.1)
- igraph (0.1.11)
- scipy (1.0.1)
- XlsxWriter (1.0.2)
- Pandas (0.22.0)

The entire package is placed on Github (https://github.com/jyyulab/SJARACNE/) for public use alongside pre-compiled Linux and OSX distributions which the program can pick the corresponding distribution according to the platform that is running. In addition, make file is provided to build the platform-specific distribution on user's demand.

SJARACNe Input Format

The main input for SJARACNe is a tab-separated genes/protein by cells/samples expression matrix with the first two columns being ID and symbol. The second required input file is the list of significant genes/proteins IDs to be considered as hubs in the reconstructed network.

isoformId	geneSymbol	MB.0000	MB.0002	MB.0005	MB.0006	MB.0008	MB.0010	MB.0014	MB.0020	MB.0022	MB.0025
ILMN_17258	LOC1005062	5.4731226	5.3422729	5.5996512	5.3761455	5.2879931	5.4493817	4.7832572	5.4350163	5.7684628	5.3870479
ILMN_19101	AK024680	5.9327246	5.6656815	6.044401	5.8880861	5.7339223	6.3817743	5.8724796	5.5289425	6.2460018	6.0491455
ILMN_18041	FCGR2B	6.4790292	6.2619533	6.1958882	6.2059751	6.1841042	6.7108812	7.0014018	6.4945316	6.906629	6.5495275
ILMN_17960	TRIM44	9.0562886	9.6308473	9.2046305	8.9376145	9.2717722	8.2646363	9.008013	9.1674188	8.7406462	8.6706811
ILMN_18119	PGGT1B	5.7914059	6.0927303	6.2750101	6.0355017	6.2272152	6.014328	6.214615	5.8341801	6.1853792	6.3470813
ILMN_16681	AWAT1	5.6141213	5.4203003	5.4115684	5.5776182	5.3892527	5.3766637	5.5858941	5.4314547	5.4528654	5.2441053
ILMN_19122	BP394710	6.6796582	6.1941072	6.3225474	6.3268731	6.9000916	6.1641752	6.8951519	6.2910406	7.3412174	7.741614
ILMN_17937	C15orf39	6.8889265	5.1572421	6.4780979	6.8858174	5.9228874	6.4802885	6.2021206	6.7537573	6.5997965	6.2507022
ILMN_17112	PCDHGC3	5.8403094	5.6788187	6.0904603	5.9290363	5.5082492	5.6322212	5.6639158	5.8423024	5.5599688	5.7304077
ILMN_16827	STAMBPL1	7.9585746	7.469026	6.3285581	7.1279241	7.9419188	7.5609926	8.2737601	7.8143062	8.5001168	8.5684133
ILMN_19062	AW337887	5.7305212	5.703555	5.7415014	5.7924448	5.8125241	5.1782779	5.6644096	5.7208006	5.743174	5.6704024
ILMN_16653	STH	5.9020185	6.3382588	6.0022883	5.8165122	6.0266109	5.8976288	5.898794	5.8343508	5.8027821	6.0501321
ILMN_16572	PALM3	5.3396471	5.3386856	5.7415475	5.4343342	5.6198209	5.3463039	5.1865097	5.5639795	5.3268771	5.2842242
ILMN_16554	SERF1B	10.628982	10.732839	9.899004	10.927851	10.660423	11.143671	10.95203	11.306123	9.5591648	10.775385
ILMN_16791	UGT2B7	7.1117266	6.4425601	6.422447	6.2488961	10.651793	8.7987698	8.6996777	6.5306366	7.0129794	5.9159243
ILMN_17558	UGT2B7	6.8883141	6.2854723	6.3634162	6.0437625	10.602512	8.8435281	8.9014164	6.3628421	7.3295001	5.8390545
ILMN_21213	GORAB	7.8060874	8.087587	7.7908332	7.4476351	7.9535614	7.9716036	8.8447454	7.6431182	7.9015873	8.7856515

ILMN_1720944
ILMN_1670130
ILMN_2275455
ILMN_1814002
ILMN_1815649
ILMN_1736555
ILMN_2399503
ILMN_2290089
ILMN_1697634
ILMN_1681234
ILMN_1710329
ILMN_1752622
ILMN_1658074
ILMN_1761348
ILMN_1656196
ILMN_1657627
ILMN_1751607

SJARACNe Output

The program will produce two info files, three network files, plus one graphml visualization file. The network files come in tabular format, simple and advanced. In the simple network file, the network appears as a tab-separated edge network with source, target and mutual information value as the columns, and the edges as each row. The advanced network comes in the tab-separated .txt and Excel sheet; in which the rows are edges and the columns are as source ID, target ID, source

symbol, target symbol, mutual information, Pearson and Spearman correlation coefficients, regression line slope and p-value.

source	target	source.symbol	target.symbol	MI	pearson	spearman	rho	p-value
ILMN_1651262	ILMN_1653599	HNRNPAB	ATP5D	0.3276	0.1562	0.2124	0.1669	0.1206
ILMN_1651262	ILMN_1654109	HNRNPAB	EGFLAM	0.2879	-0.2467	-0.2937	-0.6090	0.0133
ILMN_1651262	ILMN_1654331	HNRNPAB	HOXB4	0.3358	-0.3098	-0.3332	-0.4498	0.0017
ILMN_1651262	ILMN_1654639	HNRNPAB	HERC6	0.3110	0.2096	0.2463	0.5185	0.0363
ILMN_1651262	ILMN_1655311	HNRNPAB	C15orf61	0.3071	0.0682	-0.0510	0.0743	0.5005
ILMN_1651262	ILMN_1655922	HNRNPAB	SAPS2	0.3132	-0.3071	-0.2566	-0.4560	0.0019
ILMN_1651262	ILMN_1656368	HNRNPAB	ALDH4A1	0.3145	0.1716	0.0524	0.2895	0.0878
ILMN_1651262	ILMN_1656501	HNRNPAB	DUSP5	0.2792	0.1073	0.1311	0.2158	0.2878
ILMN_1651262	ILMN_1656628	HNRNPAB	WDR4	0.3307	0.4944	0.4832	0.6951	0.0000
ILMN_1651262	ILMN_1657317	HNRNPAB	POLR2J	0.3657	0.2541	0.3098	0.2853	0.0107
ILMN_1651262	ILMN_1657729	HNRNPAB	ZNF254	0.3174	-0.1685	-0.1700	-0.1523	0.0937
ILMN_1651262	ILMN_1658058	HNRNPAB	nan	0.3366	0.0819	0.1075	0.0372	0.4180
ILMN_1651262	ILMN_1658121	HNRNPAB	CFP	0.3841	-0.1448	-0.1755	-0.0800	0.1507
ILMN_1651262	ILMN_1658396	HNRNPAB	SPEF2	0.3229	0.1892	0.2002	0.0739	0.0594
ILMN_1651262	ILMN_1658632	HNRNPAB	NLRP7	0.3008	-0.1045	-0.1702	-0.0472	0.3007
ILMN_1651262	ILMN_1659339	HNRNPAB	ELP4	0.3853	0.0273	0.0900	0.0220	0.7874
ILMN_1651262	ILMN_1659649	HNRNPAB	SGCG	0.4407	-0.1954	-0.2502	-0.3302	0.0514