



Association Rule Mining on RNAseq Data
Trends in Bioinformatics

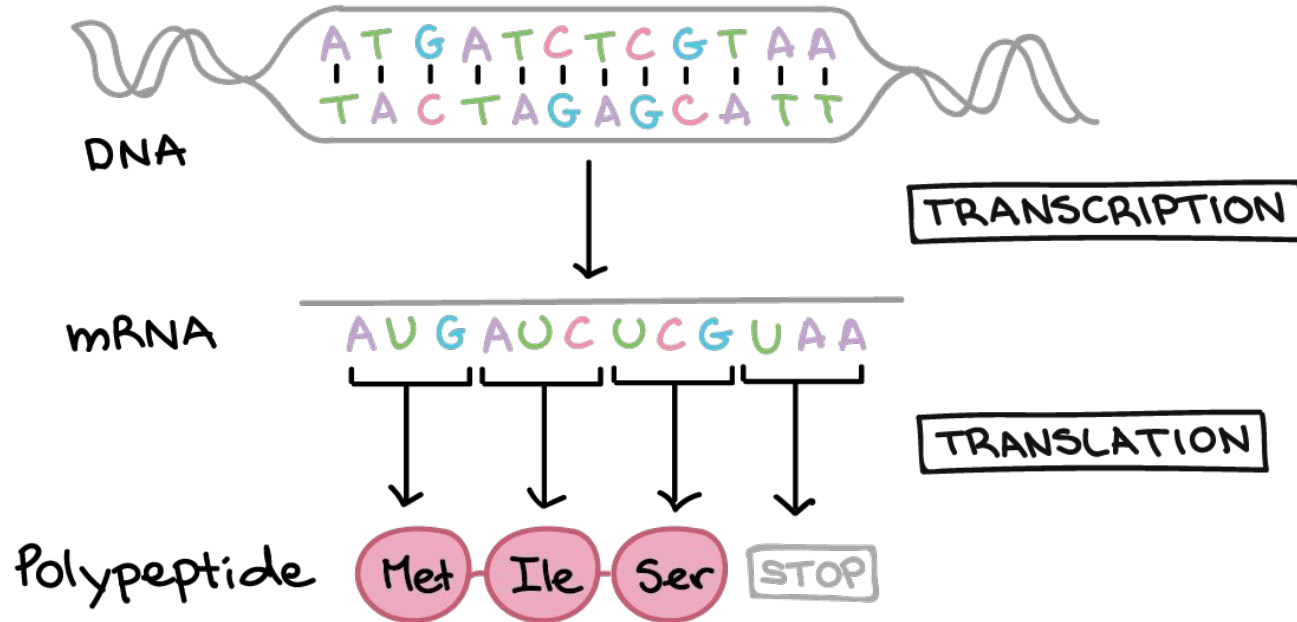
Lukas Ehrig

Nov 27th, 2018

Agenda

1. **Gene Expression**
 - 1.1. Biology Recap - Gene Expression
 - 1.2. Gene Expression (RNAseq) Data
2. **Association Rule Mining**
3. **Challenges**
4. **Next Steps**

1. Gene Expression - Biology Recap



1.2 Gene Expression Data

	ENSG00000000003	ENSG00000000005	ENSG00000000419	ENSG00000000457	ENSG00000000460	ENSG00000000938	ENSG00000000971
TCGA-AB-2841-03B-01T-0760-13	2.76593712005648	1.53324829122169	9.44768726226856	9.88963899009968	9.01877187872937	10.0044407998509	5.69054493055702
TCGA-AB-2818-03A-01T-0734-13	3.58782825427584	1.53324829122169	9.55352739900002	10.0638203943037	9.29156522722249	13.9702018104788	7.95058193099284
TCGA-AB-2976-03A-01T-0734-13	5.56652701424824	2.36734789254045	9.85781380064584	10.1265586746988	9.89295472452555	11.1659358903234	10.2161491140698
TCGA-AB-2867-03A-01T-0734-13	4.4918360998841	1.53324829122169	10.0061015845077	9.45892520299894	9.46391961321915	12.1901941421212	5.59465912838687
TCGA-AB-2839-03A-01T-0734-13	3.40824045628991	1.53324829122169	9.62551673128297	10.3152334019153	9.54101744165605	10.1323934935717	5.07141926665851
TCGA-AB-2881-03A-01T-0735-13	3.42559579476776	1.53324829122169	9.74891883105389	9.91097948514503	9.56450157736863	12.7930569152652	4.618680063385
TCGA-AB-2930-03A-01T-0740-13	2.27803665949948	1.53324829122169	9.6816982615742	9.71862882984999	9.64947557455084	12.0464061583375	6.01813583832155
TCGA-AB-2805-03A-01T-0734-13	1.53324829122169	2.34250535063557	9.93508308009364	9.96642727753544	9.31678991757321	13.875243652276	5.81045856465353
TCGA-AB-2996-03A-01T-0735-13	6.05839766222684	1.53324829122169	9.55610632482587	10.1691079834718	9.98138653010994	12.5302776175638	9.2085392999297
TCGA-AB-2919-03A-01T-0740-13	3.46760139549888	1.53324829122169	9.29621107542366	9.77898483303739	9.96210185508133	8.61696651298998	10.6811598992821
TCGA-AB-2835-03A-01T-0736-13	3.377277044747	1.53324829122169	9.94358747142781	9.09156510537673	9.38790768053172	14.4233187250113	6.60952893809161
TCGA-AB-3012-03A-01T-0736-13	4.79182824926881	1.53324829122169	9.58651623209581	9.47610029001618	8.91034347700386	6.54603396338371	8.19140587613801
TCGA-AB-2842-03A-01T-0734-13	2.3417952587717	1.53324829122169	9.90045946764349	9.99847440855063	9.89905015159871	13.083077404624	6.91285435764463
TCGA-AB-2871-03A-01T-0735-13	4.86264603376552	1.53324829122169	9.11316315950706	10.1237732844438	9.44379726363504	11.0480577028576	7.91698643471225
TCGA-AB-2938-03A-01T-0736-13	7.76952703554482	3.08094347413991	9.11453456835315	9.59464464736721	9.5739925090916	10.6037544757665	12.8757119981577
TCGA-AB-2872-03A-01T-0735-13	4.52927203623542	1.53324829122169	9.20454742690193	9.97858962193787	9.73547163360154	12.3742662794351	7.34163819942233
TCGA-AB-2915-03A-01T-0740-13	3.12189696041535	1.53324829122169	8.80098454848967	10.025719893437	9.27747891349068	13.5059131187465	9.17470944915273
TCGA-AB-2955-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.58945310291011	10.7579572497123	10.5387178077725	9.87641595250127	5.4787977515495
TCGA-AB-2943-03A-01T-0740-13	5.22960414611959	1.53324829122169	9.70534280328187	10.7978009292204	10.2219842837729	11.5693070725036	8.42607343333665
TCGA-AB-2944-03A-01T-0740-13	6.33990221174568	1.53324829122169	9.83248320662875	10.3914671930852	9.6338330041123	10.6956349170102	7.67048168940104
TCGA-AB-3007-03A-01T-0736-13	2.39574027632512	1.53324829122169	9.46332775511307	9.72548645853732	9.29976156052294	7.97681514130557	7.73120968977164
TCGA-AB-2918-03A-01T-0740-13	4.02180530189553	2.33375756783456	9.76440494885102	9.93039644697546	9.70243132635133	11.7858714722287	6.45899654419247
TCGA-AB-2882-03A-01T-0740-13	5.71438877422215	1.53324829122169	9.39106288036326	9.99797579780632	8.97343332402397	12.2208333908276	9.71892498902613
TCGA-AB-2914-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.74888425462813	10.0583377042729	9.71350215738624	11.6649534697277	5.64142817614806
TCGA-AB-2912-03A-01T-0734-13	3.3499271308279	1.53324829122169	10.1816015557517	9.98719613593839	10.2389083180146	12.3412140645443	7.59980012230915

1.2 Gene Expression Data

	ENSG00000000003	ENSG00000000005	ENSG00000000419	ENSG00000000457	ENSG00000000460	ENSG00000000938	ENSG00000000971
TCGA-AB-2841-03B-01T-0760-13	2.76593712005648	1.53324829122169	9.44768726226856	9.88963899009968	9.01877187872937	10.0044407998509	5.69054493055702
TCGA-AB-2818-03A-01T-0734-13	3.58782825427584	1.53324829122169	9.55352739900002	10.0638203943037	9.29156522722249	13.9702018104788	7.95058193099284
TCGA-AB-2976-03A-01T-0734-13	5.56652701424824	2.36734789254045	9.85781380064584	10.1265586746988	9.89295472452555	11.1659358903234	10.2161491140698
TCGA-AB-2867-03A-01T-0734-13	4.4918360998841	1.53324829122169	10.0061015845077	9.45892520299894	9.46391961321915	12.1901941421212	5.59465912838687
TCGA-AB-2839-03A-01T-0734-13	3.40824045628991	1.53324829122169	9.62551673128297	10.3152334019153	9.54101744165605	10.1323934935717	5.07141926665851
TCGA-AB-2881-03A-01T-0735-13	3.42559579476776	1.53324829122169	9.74891883105389	9.91097948514503	9.56450157736863	12.7930569152652	4.618680063385
TCGA-AB-2930-03A-01T-0740-13	2.27803665949948	1.53324829122169	9.6816982615742	9.71862882984999	9.64947557455084	12.0464061583375	6.01813583832155
TCGA-AB-2805-03A-01T-0734-13	1.53324829122169	2.34250535063557	9.93508308009364	9.96642727753544	9.31678991757321	13.875243652276	5.81045856465353
TCGA-AB-2996-03A-01T-0735-13	6.05839766222684	1.53324829122169	9.55610632482587	10.1691079834718	9.98138653010994	12.5302776175638	9.2085392999297
TCGA-AB-2919-03A-01T-0740-13	3.46760139549888	1.53324829122169	9.29621107542366	9.77898483303739	9.96210185508133	8.61696651298998	10.6811598992821
TCGA-AB-2835-03A-01T-0736-13	3.377277044747	1.53324829122169	9.94358747142781	9.09156510537673	9.38790768053172	14.4233187250113	6.60952893809161
TCGA-AB-3012-03A-01T-0736-13	4.79182824926881	1.53324829122169	9.58651623209581	9.47610029001618	8.91034347700386	6.54603396338371	8.19140587613801
TCGA-AB-2842-03A-01T-0734-13	2.3417952587717	1.53324829122169	9.90045946764349	9.99847440855063	9.89905015159871	13.083077404624	6.91285435764463
TCGA-AB-2871-03A-01T-0735-13	4.86264603376552	1.53324829122169	9.11316315950706	10.1237732844438	9.44379726363504	11.0480577028576	7.91698643471225
TCGA-AB-2938-03A-01T-0736-13	7.76952703554482	3.08094347413991	9.11453456835315	9.59464464736721	9.5739925090916	10.6037544757665	12.8757119981577
TCGA-AB-2872-03A-01T-0735-13	4.52927203623542	1.53324829122169	9.20454742690193	9.97858962193787	9.73547163360154	12.3742662794351	7.34163819942233
TCGA-AB-2915-03A-01T-0740-13	3.12189696041535	1.53324829122169	8.80098454848967	10.025719893437	9.27747891349068	13.5059131187465	9.17470944915273
TCGA-AB-2955-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.58945310291011	10.7579572497123	10.5387178077725	9.87641595250127	5.4787977515495
TCGA-AB-2943-03A-01T-0740-13	5.22960414611959	1.53324829122169	9.70534280328187	10.7978009292204	10.2219842837729	11.5693070725036	8.42607343333665
TCGA-AB-2944-03A-01T-0740-13	6.33990221174568	1.53324829122169	9.83248320662875	10.3914671930852	9.6338330041123	10.6956349170102	7.67048168940104
TCGA-AB-3007-03A-01T-0736-13	2.39574027632512	1.53324829122169	9.46332775511307	9.72548645853732	9.29976156052294	7.97681514130557	7.73120968977164
TCGA-AB-2918-03A-01T-0740-13	4.02180530189553	2.33375756783456	9.76440494885102	9.93039644697546	9.70243132635133	11.7858714722287	6.45899654419247
TCGA-AB-2882-03A-01T-0740-13	5.71438877422215	1.53324829122169	9.39106288036326	9.99797579780632	8.97343332402397	12.2208333908276	9.71892498902613
TCGA-AB-2914-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.74888425462813	10.0583377042729	9.71350215738624	11.6649534697277	5.64142817614806
TCGA-AB-2912-03A-01T-0734-13	3.3499271308279	1.53324829122169	10.1816015557517	9.98719613593839	10.2389083180146	12.3412140645443	7.59980012230915

Gene

1.2 Gene Expression Data

	ENSG00000000003	ENSG00000000005	ENSG00000000419	ENSG00000000457	ENSG00000000460	ENSG00000000938	ENSG00000000971
TCGA-AB-2841-03B-01T-0760-13	2.76593712005648	1.53324829122169	9.44768726226856	9.88963899009968	9.01877187872937	10.0044407998509	5.69054493055702
TCGA-AB-2818-03A-01T-0734-13	3.58782825427584	1.53324829122169	9.55352739900002	10.0638203943037	9.29156522722249	13.9702018104788	7.95058193099284
TCGA-AB-2976-03A-01T-0734-13	5.56652701424824	2.36734789254045	9.85781380064584	10.1265586746988	9.89295472452555	11.1659358903234	10.2161491140698
TCGA-AB-2867-03A-01T-0734-13	4.4918360998841	1.53324829122169	10.0061015845077	9.45892520299894	9.46391961321915	12.1901941421212	5.59465912838687
TCGA-AB-2839-03A-01T-0734-13	3.40824045628991	1.53324829122169	9.62551673128297	10.3152334019153	9.54101744165605	10.1323934935717	5.07141926665851
TCGA-AB-2881-03A-01T-0735-13	3.42559579476776	1.53324829122169	9.74891883105389	9.91097948514503	9.56450157736863	12.7930569152652	4.618680063385
TCGA-AB-2930-03A-01T-0740-13	2.27803665949948	1.53324829122169	9.6816982615742	9.71862882984999	9.64947557455084	12.0464061583375	6.01813583832155
TCGA-AB-2805-03A-01T-0734-13	1.53324829122169	2.34250535063557	9.93508308009364	9.96642727753544	9.31678991757321	13.875243652276	5.81045856465353
TCGA-AB-2996-03A-01T-0735-13	6.05839766222684	1.53324829122169	9.55610632482587	10.1691079834718	9.98138653010994	12.5302776175638	9.2085392999297
TCGA-AB-2919-03A-01T-0740-13	3.46760139549888	1.53324829122169	9.29621107542366	9.77898483303739	9.96210185508133	8.61696651298998	10.6811598992821
TCGA-AB-2835-03A-01T-0736-13	3.377277044747	1.53324829122169	9.94358747142781	9.09156510537673	9.38790768053172	14.4233187250113	6.60952893809161
TCGA-AB-3012-03A-01T-0736-13	4.79182824926881	1.53324829122169	9.58651623209581	9.47610029001618	8.91034347700386	6.54603396338371	8.19140587613801
TCGA-AB-2842-03A-01T-0734-13	2.3417952587717	1.53324829122169	9.90045946764349	9.99847440855063	9.89905015159871	13.083077404624	6.91285435764463
TCGA-AB-2871-03A-01T-0735-13	4.86264603376552	1.53324829122169	9.11316315950706	10.1237732844438	9.44379726363504	11.0480577028576	7.91698643471225
TCGA-AB-2938-03A-01T-0736-13	7.76952703554482	3.08094347413991	9.11453456835315	9.59464464736721	9.5739925090916	10.6037544757665	12.8757119981577
TCGA-AB-2872-03A-01T-0735-13	4.52927203623542	1.53324829122169	9.20454742690193	9.97858962193787	9.73547163360154	12.3742662794351	7.34163819942233
TCGA-AB-2915-03A-01T-0740-13	3.12189696041535	1.53324829122169	8.80098454848967	10.025719893437	9.27747891349068	13.5059131187465	9.17470944915273
TCGA-AB-2955-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.58945310291011	10.7579572497123	10.5387178077725	9.87641595250127	5.4787977515495
TCGA-AB-2943-03A-01T-0740-13	5.22960414611959	1.53324829122169	9.70534280328187	10.7978009292204	10.2219842837729	11.5693070725036	8.42607343333665
TCGA-AB-2944-03A-01T-0740-13	6.33990221174568	1.53324829122169	9.83248320662875	10.3914671930852	9.6338330041123	10.6956349170102	7.67048168940104
TCGA-AB-3007-03A-01T-0736-13	2.39574027632512	1.53324829122169	9.46332775511307	9.72548645853732	9.29976156052294	7.97681514130557	7.73120968977164
TCGA-AB-2918-03A-01T-0740-13	4.02180530189553	2.33375756783456	9.76440494885102	9.93039644697546	9.70243132635133	11.7858714722287	6.45899654419247
TCGA-AB-2882-03A-01T-0740-13	5.71438877422215	1.53324829122169	9.39106288036326	9.99797579780632	8.97343332402397	12.2208333908276	9.71892498902613
TCGA-AB-2914-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.74888425462813	10.0583377042729	9.71350215738624	11.6649534697277	5.64142817614806
TCGA-AB-2912-03A-01T-0734-13	3.3499271308279	1.53324829122169	10.1816015557517	9.98719613593839	10.2389083180146	12.3412140645443	7.59980012230915

Sample

1.2 Gene Expression Data

	ENSG00000000003	ENSG00000000005	ENSG00000000419	ENSG00000000457	ENSG00000000460	ENSG00000000938	ENSG00000000971
TCGA-AB-2841-03B-01T-0760-13	2.76593712005648	1.53324829122169	9.44768726226856	9.88963899009968	9.01877187872937	10.0044407998509	5.69054493055702
TCGA-AB-2818-03A-01T-0734-13	3.58782825427584	1.53324829122169	9.55352739900002	10.0638203943037	9.29156522722249	13.9702018104788	7.95058193099284
TCGA-AB-2976-03A-01T-0734-13	5.56652701424824	2.36734789254045	9.85781380064584	10.1265586746988	9.89295472452555	11.1659358903234	10.2161491140698
TCGA-AB-2867-03A-01T-0734-13	4.4918360998841	1.53324829122169	10.0061015845077	9.45892520299894	9.46391961321915	12.1901941421212	5.59465912838687
TCGA-AB-2839-03A-01T-0734-13	3.40824045628991	1.53324829122169	9.62551673128297	10.3152334019153	9.54101744165605	10.1323934935717	5.07141926665851
TCGA-AB-2881-03A-01T-0735-13	3.42559579476776	1.53324829122169	9.74891883105389	9.91097948514503	9.56450157736863	12.7930569152652	4.618680063385
TCGA-AB-2930-03A-01T-0740-13	2.27803665949948	1.53324829122169	9.6816982615742	9.71862882984999	9.64947557455084	12.0464061583375	6.01813583832155
TCGA-AB-2805-03A-01T-0734-13	1.53324829122169	2.34250535063557	9.93508308009364	9.96642727753544	9.31678991757321	13.875243652276	5.81045856465353
TCGA-AB-2996-03A-01T-0735-13	6.05839766222684	1.53324829122169	9.55610632482587	10.1691079834718	9.98138653010994	12.5302776175638	9.2085392999297
TCGA-AB-2919-03A-01T-0740-13	3.4676013954938	1.53324829122169	9.2962110754230	9.77898483303730	9.96210185508133	8.61696651298998	10.6811598992821
TCGA-AB-2835-03A-01T-0736-13	3.37727704474	1.53324829122169	9.9358721178	9.9156705467	9.38790768053172	14.4233187250113	6.60952893809161
TCGA-AB-3012-03A-01T-0736-13	4.79182824926	1.53324829122169	9.5651632058	9.762429004418	8.91034347700386	6.54603396338371	8.19140587613801
TCGA-AB-2842-03A-01T-0734-13	2.3417952587717	1.53324829122169	9.90045946764349	9.99847440855063	9.89905015159871	13.083077404624	6.91285435764463
TCGA-AB-2871-03A-01T-0735-13	4.86264603376552	1.53324829122169	9.11316315950706	10.1237732844438	9.44379726363504	11.0480577028576	7.91698643471225
TCGA-AB-2938-03A-01T-0736-13	7.76952703554482	3.08094347413991	9.11453456835315	9.59464464736721	9.5739925090916	10.6037544757665	12.8757119981577
TCGA-AB-2872-03A-01T-0735-13	4.52927203623542	1.53324829122169	9.20454742690193	9.97858962193787	9.73547163360154	12.3742662794351	7.34163819942233
TCGA-AB-2915-03A-01T-0740-13	3.12189696041535	1.53324829122169	8.80098454848967	10.025719893437	9.27747891349068	13.5059131187465	9.17470944915273
TCGA-AB-2955-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.58945310291011	10.7579572497123	10.5387178077725	9.87641595250127	5.4787977515495
TCGA-AB-2943-03A-01T-0740-13	5.22960414611959	1.53324829122169	9.70534280328187	10.7978009292204	10.2219842837729	11.5693070725036	8.42607343333665
TCGA-AB-2944-03A-01T-0740-13	6.33990221174568	1.53324829122169	9.83248320662875	10.3914671930852	9.6338330041123	10.6956349170102	7.67048168940104
TCGA-AB-3007-03A-01T-0736-13	2.39574027632512	1.53324829122169	9.46332775511307	9.72548645853732	9.29976156052294	7.97681514130557	7.73120968977164
TCGA-AB-2918-03A-01T-0740-13	4.02180530189553	2.33375756783456	9.76440494885102	9.93039644697546	9.70243132635133	11.7858714722287	6.45899654419247
TCGA-AB-2882-03A-01T-0740-13	5.71438877422215	1.53324829122169	9.39106288036326	9.99797579780632	8.97343332402397	12.2208333908276	9.71892498902613
TCGA-AB-2914-03A-01T-0734-13	1.53324829122169	1.53324829122169	9.74888425462813	10.0583377042729	9.71350215738624	11.6649534697277	5.64142817614806
TCGA-AB-2912-03A-01T-0734-13	3.3499271308279	1.53324829122169	10.1816015557517	9.98719613593839	10.2389083180146	12.3412140645443	7.59980012230915

Discretize

1.2 Gene Expression Data

	ENSG00000000003	ENSG00000000005	ENSG00000000049	ENSG000000000457	ENSG000000000460	ENSG000000000938	ENSG000000000971
TCGA-AB-2841-03B-01T-0760-13	-1.0	0.0	0.0	0.0	0.0	0.0	-1.0
TCGA-AB-2818-03A-01T-0734-13	-1.0	0.0	0.0	1.0	0.0	1.0	0.0
TCGA-AB-2976-03A-01T-0734-13	0.0	0.0	0.0	1.0	1.0	0.0	0.0
TCGA-AB-2867-03A-01T-0734-13	0.0	0.0	0.0	0.0	0.0	0.0	-1.0
TCGA-AB-2839-03A-01T-0734-13	-1.0	0.0	0.0	1.0	0.0	0.0	-1.0
TCGA-AB-2881-03A-01T-0735-13	-1.0	0.0	0.0	0.0	0.0	1.0	-1.0
TCGA-AB-2930-03A-01T-0740-13	-1.0	0.0	0.0	0.0	0.0	0.0	-1.0
TCGA-AB-2805-03A-01T-0734-13	-1.0	0.0	0.0	1.0	0.0	1.0	-1.0
TCGA-AB-2996-03A-01T-0735-13	0.0	0.0	0.0	1.0	1.0	1.0	0.0
TCGA-AB-2919-03A-01T-0740-13	-1.0	0.0	-1.0	0.0	1.0	0.0	0.0
TCGA-AB-2835-03A-01T-0736-13	-1.0	0.0	0.0	0.0	0.0	1.0	-1.0
TCGA-AB-3012-03A-01T-0736-13	0.0	0.0	0.0	0.0	0.0	-1.0	0.0
TCGA-AB-2842-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	1.0	0.0
TCGA-AB-2871-03A-01T-0735-13	0.0	0.0	-1.0	1.0	0.0	0.0	0.0
TCGA-AB-2938-03A-01T-0736-13	0.0	0.0	-1.0	0.0	0.0	0.0	1.0
TCGA-AB-2872-03A-01T-0735-13	0.0	0.0	-1.0	1.0	1.0	1.0	0.0
TCGA-AB-2915-03A-01T-0740-13	-1.0	0.0	-1.0	1.0	0.0	1.0	0.0
TCGA-AB-2955-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	0.0	-1.0
TCGA-AB-2943-03A-01T-0740-13	0.0	0.0	0.0	1.0	1.0	0.0	0.0
TCGA-AB-2944-03A-01T-0740-13	0.0	0.0	0.0	1.0	0.0	0.0	0.0
TCGA-AB-3007-03A-01T-0736-13	-1.0	0.0	0.0	0.0	0.0	-1.0	0.0
TCGA-AB-2918-03A-01T-0740-13	0.0	0.0	0.0	1.0	1.0	0.0	-1.0
TCGA-AB-2882-03A-01T-0740-13	0.0	0.0	-1.0	1.0	0.0	0.0	0.0
TCGA-AB-2914-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	0.0	-1.0
TCGA-AB-2912-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	1.0	0.0

2. Association Rule Mining

Table 1. Market basket transactions

<i>Transaction ID</i>	<i>Items Bought</i>
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

{Headset,Laptop}

2. Association Rule Mining

Table 1. Market basket transactions

<i>Transaction ID</i>	<i>Items Bought</i>
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

A: Laptop \Rightarrow Headset

sup(A) = 4

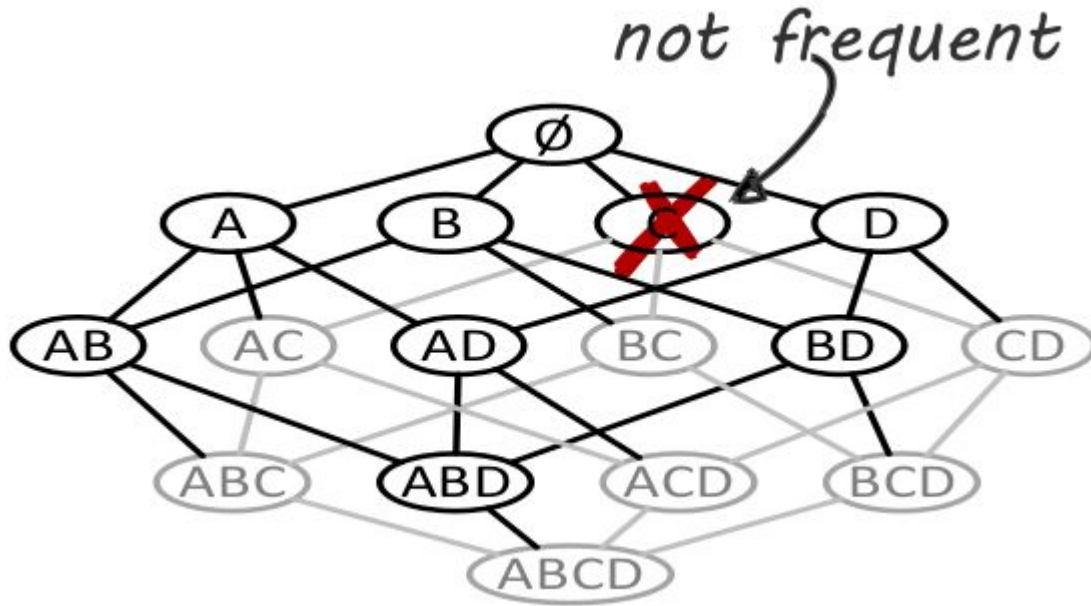
conf(A) = 0.8

B: Headset \Rightarrow Laptop

sup(B) = 4

conf(B) = 0.6

2. Association Rule Mining



2. Association Rule Mining

	ENSG00000000003	ENSG00000000005	ENSG00000000049	ENSG000000000457	ENSG000000000460	ENSG000000000938	ENSG000000000971
TCGA-AB-2841-03B-01T-0760-13	-1.0	0.0	0.0	0.0	0.0	0.0	-1.0
TCGA-AB-2818-03A-01T-0734-13	-1.0	0.0	0.0	1.0	0.0	1.0	0.0
TCGA-AB-2976-03A-01T-0734-13	0.0	0.0	0.0	1.0	1.0	0.0	0.0
TCGA-AB-2867-03A-01T-0734-13	0.0	0.0	0.0	0.0	0.0	0.0	-1.0
TCGA-AB-2839-03A-01T-0734-13	-1.0	0.0	0.0	1.0	0.0	0.0	-1.0
TCGA-AB-2881-03A-01T-0735-13	-1.0	0.0	0.0	0.0	0.0	1.0	-1.0
TCGA-AB-2930-03A-01T-0740-13	-1.0	0.0	0.0	0.0	0.0	0.0	-1.0
TCGA-AB-2805-03A-01T-0734-13	-1.0	0.0	0.0	1.0	0.0	1.0	-1.0
TCGA-AB-2996-03A-01T-0735-13	0.0	0.0	0.0	1.0	1.0	1.0	0.0
TCGA-AB-2919-03A-01T-0740-13	-1.0	0.0	-1.0	0.0	1.0	0.0	0.0
TCGA-AB-2835-03A-01T-0736-13	-1.0	0.0	0.0	0.0	0.0	1.0	-1.0
TCGA-AB-3012-03A-01T-0736-13	0.0	0.0	0.0	0.0	0.0	-1.0	0.0
TCGA-AB-2842-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	1.0	0.0
TCGA-AB-2871-03A-01T-0735-13	0.0	0.0	-1.0	1.0	0.0	0.0	0.0
TCGA-AB-2938-03A-01T-0736-13	0.0	0.0	-1.0	0.0	0.0	0.0	1.0
TCGA-AB-2872-03A-01T-0735-13	0.0	0.0	-1.0	1.0	1.0	1.0	0.0
TCGA-AB-2915-03A-01T-0740-13	-1.0	0.0	-1.0	1.0	0.0	1.0	0.0
TCGA-AB-2955-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	0.0	-1.0
TCGA-AB-2943-03A-01T-0740-13	0.0	0.0	0.0	1.0	1.0	0.0	0.0
TCGA-AB-2944-03A-01T-0740-13	0.0	0.0	0.0	1.0	0.0	0.0	0.0
TCGA-AB-3007-03A-01T-0736-13	-1.0	0.0	0.0	0.0	0.0	-1.0	0.0
TCGA-AB-2918-03A-01T-0740-13	0.0	0.0	0.0	1.0	1.0	0.0	-1.0
TCGA-AB-2882-03A-01T-0740-13	0.0	0.0	-1.0	1.0	0.0	0.0	0.0
TCGA-AB-2914-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	0.0	-1.0
TCGA-AB-2912-03A-01T-0734-13	-1.0	0.0	0.0	1.0	1.0	1.0	0.0

Item: ENSG0..03--1

Transaction ID

Transaction

3. Challenges

Number of features infeasible

Exponential number of possible frequent item sets

→ Runtime complexity

Huge **number of** possible / output **rules**

n frequent items yield $2^n - 2$ association rules

→ which rules are relevant/interesting?

4. Next Steps

Possible research questions to investigate:

Feasibility of ARM

- Dimensionality Reduction / Feature Engineering
- Not keeping irrelevant rules/item sets
- Seeds (fix one side of ARs)

Biological relevance of rules

- Using co-expression DBs
- Evaluating different interestingness measures

4. Next Steps

Possible research questions to investigate:

Feasibility of ARM

Dimensionality Reduction / Feature Engineering

Not keeping irrelevant rules/item sets

Seeds (fix one side of ARs)

adapt algorithm



Biological relevance of rules

Using co-expression DBs

Evaluating different interestingness measures

4. Next Steps


Possible research questions to investigate:

Feasibility of ARM

- Dimensionality Reduction / Feature Engineering
- Not keeping irrelevant rules/item sets
- Seeds (fix one side of ARs)

Biological relevance of rules

- Using co-expression DBs
- Evaluating different interestingness measures



integrating external knowledge
(coexpress-db, GeneFriends)

4. Next Steps

Possible research questions to investigate:

Feasibility of ARM

- Dimensionality Reduction / Feature Engineering
- Not keeping irrelevant rules/item sets
- Seeds (fix one side of ARs)

Biological relevance of rules

- Using co-expression DBs
- Evaluating different interestingness measures



Association Rule Mining on RNAseq Data
Trends in Bioinformatics

Lukas Ehrig

Nov 27th, 2018