

Genome-wide co-expression distributions as a metric to prioritize genes of functional importance

Additional File 1

Table S1. Expected bin proportions, mean, standard deviation (SD), skewness and kurtosis for the 8 shapes.

	Expected 0.25-Bin Proportions, %								Mean	SD	Skewness	Kurtosis
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8				
Shape 1	5	10	15	20	20	15	10	5	0	0.456	0	-0.660
Shape 2	2	5	15	28	28	15	5	2	0	0.354	0	0.065
Shape 3	15	30	25	15	9	3	2	1	-0.387	0.382	0.857	0.640
Shape 4	1	2	3	9	15	25	30	15	0.387	0.382	-0.857	0.640
Shape 5	28	15	5	2	2	5	15	28	0	0.752	0	-1.771
Shape 6	5	25	15	5	5	15	25	5	0	0.566	0	-1.535
Shape 7	5	30	20	10	5	10	15	5	-0.137	0.532	0.5	-1.111
Shape 8	5	15	10	5	10	20	30	5	0.137	0.532	-0.5	-1.111

Table S2. Datasets summary.

Dataset	Number of tissues	Number of libraries	Number of genes	Gene filtering criteria	Reference
Cattle feed efficiency	5	86	11,662	Average log2FPKM > 1 across all tissues	Alexandre et al. (2015)
Cattle puberty	5	60	16,978	Average FPKM \geq 0.2 in at least 1 tissue	Fortes et al. (2016) Nguyen et al. (2017) Fortes et al. (2018) Nguyen et al. (2018)
Drosophila embryogenesis	1 (14 time-points)	68 averaged by time point	7,771	More than 0 counts in at least 20 samples and more than 100 counts on average	Becker et al. (2018)
Duck preadipocyte	1 (6 time-points)	36 averaged by time point	13,322	FPKM > 0 in all samples	Wang et al. (2019)
Human	1	15	19,206	TPM > 0 in all samples	Lonsdale et al. (2013)

FPKM - fragments per kilobase of gene per million mapped reads

TPM – transcripts per million reads

Table S3. Number of genes assigned to each distribution shape in datasets with contrasting phenotypes. Numbers in parentheses represent the P-values for enrichment (above) and depletion (below). The categories include differentially expressed (DE) and regulator (REG) genes.

	Cattle Feed Efficiency Dataset					Cattle Puberty Dataset				
	REG	DE	DE-REG	Others	Total	REG	DE	DE-REG	Others	Total
Shape 1	(1) 122 (1.62E-07)	(2.93E-05) 91 (1)	(0.31) 6 (0.83)	(0.03) 1,784 (0.98)	2,003	(1) 449 (4.87E-13)	(8.70E-59) 1,223 (1)	(5.14E-06) 82 (1)	(1) 5,003 (4.19E-21)	6,757
Shape 2	(1) 48 (3.61E-04)	(9.20E-09) 56 (1)	(0.12) 4 (0.96)	(0.80) 717 (0.23)	825	(1) 31 (4.39E-03)	(2.11E-20) 153 (1)	(0.49) 5 (0.68)	(1) 372 (1.82E-10)	561
Shape 3	(1) 31 (2.10E-08)	(0.08) 40 (0.95)	(0.85) 1 (0.45)	(3.73E-03) 708 (1)	780	(1) 0 (0.84)	(1) 0 (0.76)	(1) 0 (0.98)	(0.60) 2 (1)	2
Shape 4	(5.94E-17) 818 (1)	(1) 157 (1.94E-19)	(0.86) 16 (0.25)	(0.98) 6,835 (0.02)	7,826	(0.16) 1 (0.99)	(1) 0 (0.76)	(1) 0 (0.98)	(0.95) 1 (0.40)	2
Shape 5	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0
Shape 6	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0
Shape 7	(0.74) 7 (0.40)	(7.63E-111) 6 (1)	(1) 0 (1)	(0.76) 80 (0.35)	93	(1) 182 (4.22E-11)	(3.11E-36) 636 (1)	(1.52E-06) 50 (1)	(1) 2,315 (7.78E-14)	3,183
Shape 8	(0.03) 19 (0.98)	(2.47E-08) 5 (1)	(1) 0 (1)	(0.98) 111 (0.04)	135	(5.86E-38) 781 (1)	(1) 183 (4.99E-251)	(1) 3 (5.12E-25)	(1.43E-75) 5,506 (1)	6,473
Total	1,045	355	27	10,235	11,662	346	236	9	1,952	16,978

Table S4. Number of genes assigned to each distribution shape in the cattle feed efficiency dataset split by phenotype. Numbers in parentheses represent the P-values for enrichment (above) and depletion (below). The categories include differentially expressed (DE) and regulator (REG) genes.

	Low Feed Efficiency					High Feed Efficiency				
	REG	DE	DE-REG	Others	Total	REG	DE	DE-REG	Others	Total
Shape 1	(1)	(7.67E-07)	(0.05)	(0.12)	2,137	(1)	(3.18E-05)	(0.58)	(0.01)	1,729
	134	102	9	1,892		98	81	4	1,546	
	(3.49E-07)	(1)	(0.98)	(0.89)		(2.97E-08)	(1)	(0.63)	(0.99)	
Shape 2	(1)	(1.39E-10)	(0.56)	(0.96)	800	(0.97)	(7.37E-08)	(0.04)	(0.97)	830
	52	59	2	687		60	54	5	711	
	(0.01)	(1)	(0.72)	(0.05)		(0.04)	(1)	(0.99)	(0.03)	
Shape 3	(1)	(0.04)	(0.83)	(1.96E-04)	732	1	0.01	0.85	5.89E-04	786
	28	31	1	672		32	35	1	718	
	(2.14E-08)	(0.98)	(0.49)	(1)		(3.66E-08)	(0.99)	(0.45)	(1)	
Shape 4	(3.75E-18)	(1)	(0.80)	(0.97)	7,192	(3.66E-16)	(1)	(0.87)	(0.99)	8,267
	771	127	15	6,279		850	179	17	7,221	
	(1)	(1.16E-23)	(0.32)	(0.03)		(1)	(1.42E-16)	(0.24)	(0.02)	
Shape 5	(1)	(1)	(1)	(1)	0	(1)	(1)	(1)	(1)	0
	0	0	0	0		0	0	0	0	
	(1)	(1)	(1)	(1)		(1)	(1)	(1)	(1)	
Shape 6	(1)	(1)	(1)	(1)	0	(1)	(1)	(1)	(1)	0
	0	0	0	0		0	0	0	0	
	(1)	(1)	(1)	(1)		(1)	(1)	(1)	(1)	
Shape 7	(1)	(2.08E-04)	(1)	(0.52)	248	(1)	(4.21E-04)	(1)	(0.90)	33
	11	19	0	218		0	6	0	27	
	(4.71E-03)	(1)	(0.56)	(0.56)		(0.04)	(1)	(0.93)	(0.21)	
Shape 8	(0.56)	(0.52)	(1)	(0.44)	553	(0.01)	(1)	(1)	(0.99)	17
	49	17	0	487		5	0	0	12	
	(0.50)	(0.58)	(0.27)	(0.61)		(1)	(0.59)	(0.96)	(0.05)	
Total	1,045	355	27	10,235	11,662	1,045	355	27	10,235	11,662

Table S5. Number of genes assigned to each distribution shape in the cattle puberty dataset split by phenotype. Numbers in parentheses represent the P-values for enrichment (above) and depletion (below). The categories include differentially expressed (DE) and regulator (REG) genes.

	Pre-puberty					Post-puberty				
	REG	DE	DE-REG	Others	Total	REG	DE	DE-REG	Others	Total
Shape 1	(1) 397 (1.61E-15)	(6.66E-66) 1,179 (1)	(2.05E-05) 76 (1)	(1) 4,620 (1.61E-22)	6,272	(1) 478 (1.32E-09)	(4.66E-64) 1,254 (1)	(6.50E-08) 88 (1)	(1) 5,046 (6.50E-28)	6,866
Shape 2	(1) 31 (5.45E-04)	(9.98E-23) 169 (1)	(0.75) 4 (0.42)	(1) 412 (1.86E-10)	616	(1) 23 (2.81E-03)	(4.12E-14) 118 (1)	(0.89) 2 (0.27)	(1) 314 (3.77E-06)	457
Shape 3	(1) 0 (0.64)	(0.13) 2 (0.98)	(1) 0 (0.96)	(0.92) 3 (0.31)	5	(1) 0 (0.54)	(0.01) 4 (1)	(1) 0 (0.94)	(0.99) 3 (0.05)	7
Shape 4	(1) 0 (0.84)	(0.02) 2 (1)	(1) 0 (0.98)	(1) 0 (0.05)	2	(1) 0 (0.92)	(0.13) 1 (1)	(1) 0 (0.99)	(1) 0 (0.22)	1
Shape 5	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0
Shape 6	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0
Shape 7	(1) 196 (1.70E-09)	(3.96E-35) 646 (1)	(3.58E-09) 57 (1)	(1) 2,373 (2.75E-15)	3,272	(1) 167 (2.62E-14)	(3.76E-33) 622 (1)	(7.22E-06) 48 (1)	(1) 2,323 (2.57E-10)	3,160
Shape 8	(1.17E-40) 819 (1)	(1) 198 (5.71E-265)	(1) 3 (6.54E-27)	(7.56E-81) 5,791 (1)	6,811	(7.00E-36) 775 (1)	(1) 197 (7.66E-240)	(1) 2 (1.47E-26)	(2.93E-74) 5,513 (1)	6,487
Total	1,443	2,196	140	13,199	16,978	1,443	2,196	140	13,199	16,978

Table S6. Number of genes assigned to each distribution shape in time-series datasets. Numbers in parentheses represent the P-values for enrichment (above) and depletion (below). The categories include differentially expressed (DE) and regulator (REG) genes – refer to methods for specific categories in Drosophila dataset.

	Drosophila Embryogenesis Dataset							Duck Preadipocyte Dataset				
	Down/down	Down/up	Up/down	Up/up	REG	Others	Total	REG	DE	DE-REG	Others	Total
Shape 1	(4.29E-16)	(0.94)	(0.01)	(8.74E-04)	(1)	(1)	1,433	(0.97)	(1.17E-10)	(0.15)	(1)	1,983
	292	305	76	122	106	638		63	584	29	1,307	
	(1)	(0.07)	(0.99)	(1)	(4.17E-05)	(8.45E-13)		(0.05)	(1)	(0.90)	(1.47E-08)	
Shape 2	(0.16)	(0.96)	(0.02)	(0.63)	(0.98)	(0.69)	98	(0.43)	(1)	(0.83)	(6.66E-10)	262
	17	16	9	6	5	50		11	21	2	228	
	(0.90)	(0.08)	(0.99)	(0.53)	(0.06)	(0.39)		(0.69)	(1.71E-11)	(0.39)	(1)	
Shape 3	(0.14)	(0.41)	(1)	(1)	(1)	(0.87)	10	1	1	1	(0.26)	4
	3	3	0	0	0	4		0	0	0	4	
	(0.96)	(0.83)	(0.65)	(0.51)	(0.34)	(0.31)		(0.85)	(0.34)	(0.95)	1	
Shape 4	(1)	(1)	(1)	(1)	(1)	(1)	0	(1)	(1)	(1)	(1)	0
	0	0	0	0	0	0		0	0	0	0	
	(1)	(1)	(1)	(1)	(1)	(1)		(1)	(1)	(1)	(1)	
Shape 5	(1)	(1)	(1)	(1)	(1)	(1)	0	(1)	(1)	(1)	(1)	0
	0	0	0	0	0	0		0	0	0	0	
	(1)	(1)	(1)	(1)	(1)	(1)		(1)	(1)	(1)	(1)	
Shape 6	(1)	(1.38E-18)	(0.02)	(1)	(0.07)	(0.12)	4,390	(0.58)	(0.83)	(0.06)	(0.32)	1,425
	468	1,159	202	210	467	2,351		54	324	24	1,023	
	(8.76E-17)	(1)	(0.98)	(3.46E-13)	(0.94)	(0.89)		(0.47)	(0.18)	(0.97)	(0.71)	
Shape 7	(1)	(1)	(0.20)	(4.75E-51)	(2.62E-05)	(6.07E-29)	790	(0.90)	(5.42E-11)	(9.71E-04)	(1)	3,332
	1	16	38	171	115	564		117	930	58	2,227	
	(3.27E-51)	(1.6E-69)	(0.84)	(1)	(1)	(1)		(0.12)	(1)	(1)	(1.06E-10)	
Shape 8	(1.43E-29)	(0.01)	(1)	(1)	(0.85)	(1)	1,050	(0.01)	(1)	(1)	(1.05E-14)	6,316
	267	271	1	2	98	509		270	1,302	47	4,697	
	(1)	(0.99)	(5.10E-20)	(1.48E-30)	(0.18)	(9.72E-04)		(0.99)	(5.3E-16)	(2.2E-06)	(1)	
Total	1,048	1,770	326	511	791	4,116	7,771	515	3,161	160	9,486	13,322

Table S7. Number of genes assigned to each distribution shape in the human dataset. Numbers in parentheses represent the P-values for enrichment (above) and depletion (below). The categories include differentially expressed (DE), regulator (REG) and tissue enriched (TE) genes.

	Human Dataset					
	REG	DE	DE-REG	TE	Others	Total
Shape 1	(0.55) 215 (0.48)	(2.48E-04) 182 (1)	(0.02) 24 (0.99)	(0.09) 55 (0.93)	(1) 3,411 (1.53E-03)	3,881
Shape 2	(9.31E-06) 344 (1)	(3.28E-10) 262 (1)	(2.92E-04) 35 (1)	(0.99) 46 (0.01)	(1) 4,382 (6.34E-13)	5,065
Shape 3	(0.92) 1 (0.30)	(4.82E-03) 6 (1)	(0.16) 1 (0.99)	(1) 0 (0.60)	(0.96) 35 (0.09)	43
Shape 4	(0.37) 1 (0.93)	(1) 0 (0.74)	(1) 0 (0.97)	(1) 0 (0.91)	(0.79) 7 (0.60)	8
Shape 5	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0
Shape 6	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	0
Shape 7	(1) 79 (4.00E-03)	(1) 48 (2.94E-03)	(0.88) 5 (0.22)	(7.21E-34) 91 (1)	(9.40E-01) 1,640 (0.07)	1,860
Shape 8	(0.99) 435 (0.01)	(1) 217 (5.79E-14)	(1) 13 (3.59E-07)	(1) 39 (2.65E-18)	(3.58E-24) 7,733 (1)	8,427
Total	1,075	715	78	231	17,208	19,284

Table S8. Average number of connections per gene (Log10) falling in each shape according to the dataset and P-value for the one-way ANOVA.

Dataset	Average number of connections per gene (Log10)								P-value
	Shape 1	Shape 2	Shape 3	Shape 4	Shape 5	Shape 6	Shape 7	Shape 8	
Cattle feed efficiency	3.04	2.81	3.30	3.49	-	-	3.15	3.20	<2E-16
Cattle puberty	3.14	2.52	1.40	1.87	-	-	3.57	3.58	<2E-16
Drosophila embryogenesis	2.58	2.26	2.39	-	-	3.03	2.92	2.87	<2E-16
Duck preadipocyte	2.33	2.06	1.82	-	-	2.60	2.60	2.66	<2E-16
Human	2.60	2.34	2.46	2.40	-	-	3.17	3.44	<2E-16

Table S9. Number of genes assigned to each distribution shape in all datasets based on genes being on the top or bottom 5% when ranked by degree (number of significant correlations to other genes). Numbers in parentheses represent the P-values for enrichment (above) and depletion (below).

	Feed Efficiency		Puberty		Drosophila		Duck		Human	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Shape 1	(1) 0 (8.44E-50)	(4.72E-24) 198 (1)	(1) 0 (3.27E-194)	(1.56E-37) 518 (1)	(1) 0 (3.74E-36)	(7.1E-126) 280 (1)	(1) 0 (1.18E-48)	(9.3E-110) 335 (1)	(1) 0 (9.50E-98)	(1) 61 (3.78E-35)
Shape 2	(1) 0 (8.41E-20)	(4.27E-252) 335 (1)	(1) 0 (1.93E-13)	(8.64E-275) 314 (1)	(1) 0 (6.31E-03)	(8.84E-44) 53 (1)	(1) 0 (1.27E-06)	(2.3E-280) 231 (1)	(1) 0 (1.90E-132)	(<1E-999) 894 (1)
Shape 3	(1) 5 (1.54E-12)	(1) 18 (6.76E-05)	(1) 0 (0.90)	(2.50E-03) 2 (1)	(1) 0 (5.98E-01)	(2.68E-06) 6 (1)	(1) 0 (0.81)	(6.19E-06) 4 (1)	(1) 0 (0.11)	(4.02E-05) 10 (1)
Shape 4	(1.38E-94) 578 (1)	(1) 31 (6.97E-233)	(1) 0 (0.90)	(2.50E-03) 2 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (0.66)	(0.34) 1 (0.94)
Shape 5	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)
Shape 6	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1) 0 (1)	(1.31E-100) 389 (1)	(1) 27 (3.3E-101)	(1) 0 (2.41E-34)	(1) 1 (2.06E-32)	(1) 0 (1)	(1) 0 (1)
Shape 7	(1) 0 (8.32E-03)	(0.99) 1 (0.05)	(1.54E-31) 299 (1)	(1) 5 (5.83E-70)	(1) 0 (2.48E-19)	(1) 13 (1.99E-07)	(0.99) 142 (0.01)	(1) 45 (2.35E-36)	(1) 1 (2.13E-42)	(1) 0 (1.94E-44)
Shape 8	(1) 0 (9.45E-04)	(1) 0 (9.45E-04)	(1.77E-58) 550 (1)	(1) 8 (3.03E-166)	(1) 0 (6.20E-26)	(1) 10 (2.14E-14)	(1.31E-64) 524 (1)	(1) 50 (8.6E-118)	(<1E-999) 965 (1)	(1) 0 (2.81E-250)
Total	583	583	849	849	389	389	666	666	966	966

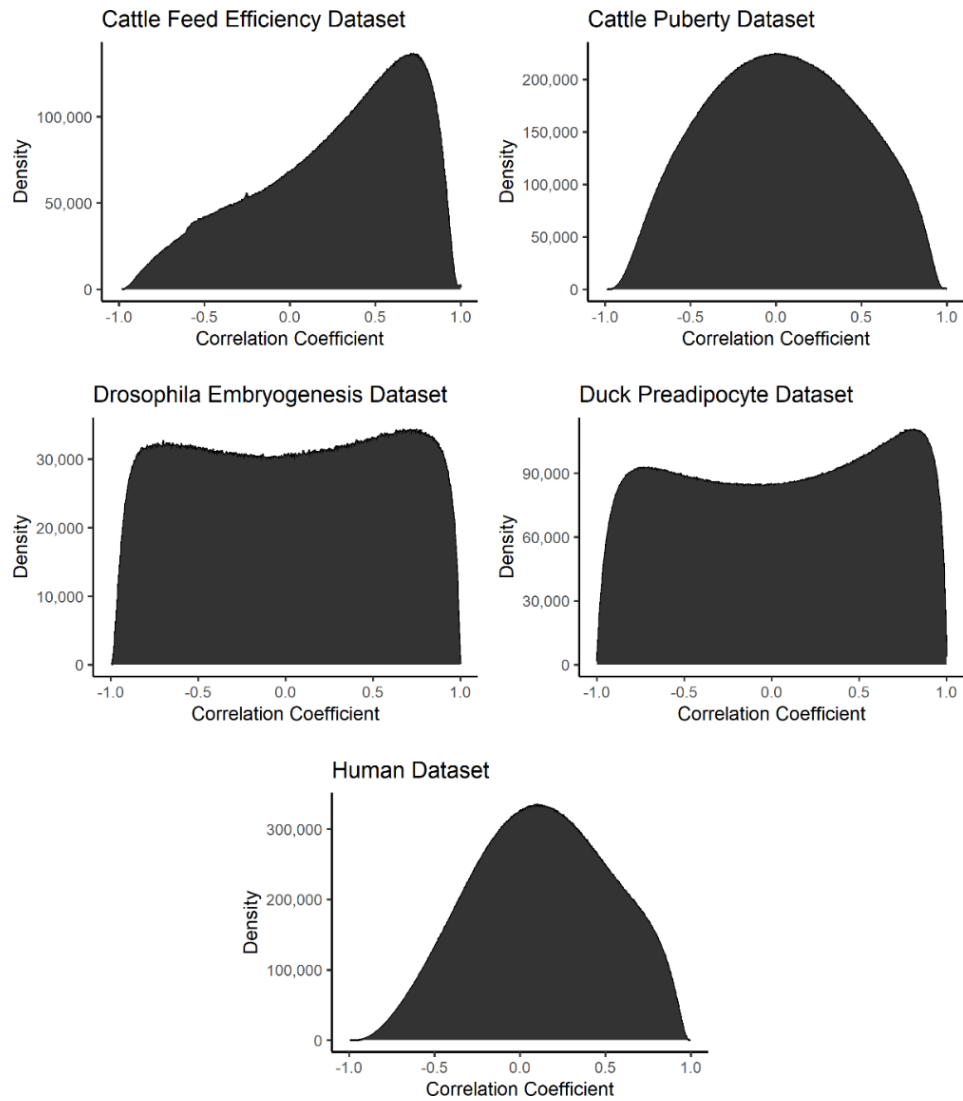


Figure S1. Frequency distributions of all co-expression correlation coefficients in each of the five RNA-Sequence datasets. Datasets include: 1) Cattle Feed Efficiency (11,662 genes and 67,995,291 correlations); 2) Cattle Puberty (16,978 genes and 144,117,753 correlations); 3) Drosophila Embryogenesis (7,771 genes and 30,190,335 correlations); 4) Duck Preadipocyte (13,322 genes and 88,731,181 correlations); and 5) Human (19,311 genes and 186,447,705 correlations).

Cattle Feed Efficiency Dataset		Null		Unimodal		Symmetric	
		Yes	No	Yes	No	Yes	No
DE	Yes	157	225	371	11	157	225
	No	2,671	8,609	11,063	217	2,671	8,609
REG	Yes	180	892	1,046	26	180	892
	No	2,648	7,942	10,388	202	2,648	7,942

Cattle Puberty Dataset		Null		Unimodal		Symmetric	
		Yes	No	Yes	No	Yes	No
DE	Yes	1,463	872	1,463	872	1,463	872
	No	5,855	8,788	5,859	8,784	5,855	8,788
REG	Yes	567	1,017	568	1,016	567	1,017
	No	6,751	8,643	6,754	8,640	6,751	8,643

Drosophyla Embryogenesis Dataset		Null		Unimodal		Symmetric	
		Yes	No	Yes	No	Yes	No
DE	Yes	842	2,805	842	2,805	2,881	766
	No	689	3,435	699	3,425	3,040	1,084
REG	Yes	111	686	117	680	578	219
	No	1,420	5,554	1,424	5,550	5,343	1,631

Duck Preadipocyte Dataset		Null		Unimodal		Symmetric	
		Yes	No	Yes	No	Yes	No
DE	Yes	636	2685	636	2685	984	2686
	No	1,609	8,392	1,613	8,388	2,337	7,315
REG	Yes	105	570	105	570	183	492
	No	2,140	10,507	2,144	10,503	3,487	9,160

Human Non-diseased Dataset		Null		Unimodal		Symmetric	
		Yes	No	Yes	No	Yes	No
DE	Yes	503	290	510	283	503	290
	No	8,443	10,048	8,487	10,004	8,443	10,048
REG	Yes	618	535	621	532	618	535
	No	8,328	9,803	8,376	9,755	8,328	9,803

Figure S2. Number of genes assigned to different distribution shapes. The tables show the number of differentially expressed genes (DE) and regulator (REG) genes assigned to null (Shapes 1 and 2), unimodal (Shapes 1 to 4) and symmetric (Shapes 1, 2, 5 and 6) distributions. Blue cells represent the type of distribution in which DE or REG were found in higher number than expected by chance where variables were found to be dependent ($P < 0.05$).