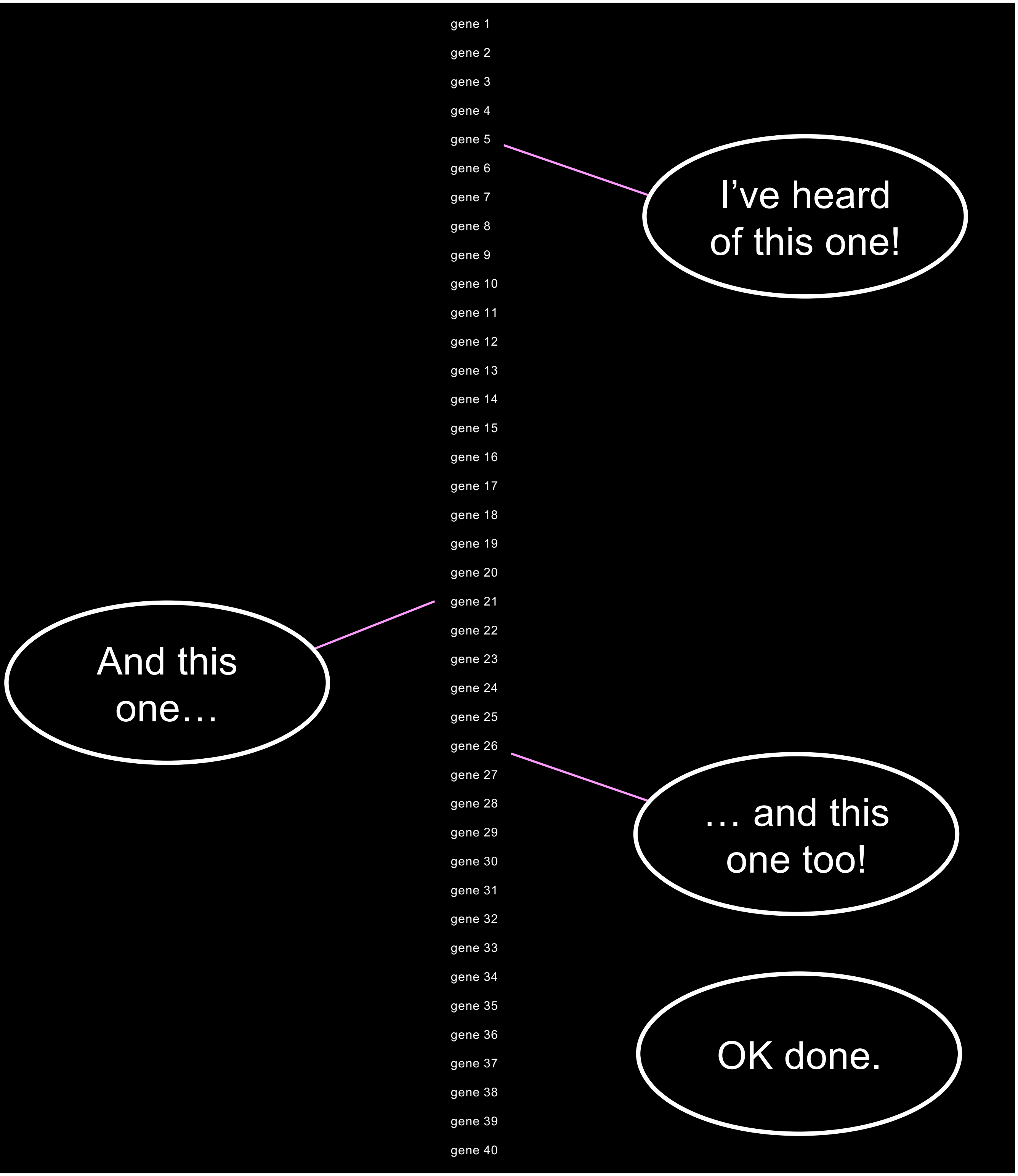
Interpreting your gene list

Steve W. Cole, Ph.D.

UCLA School of Medicine

Division of Hematology-Oncology



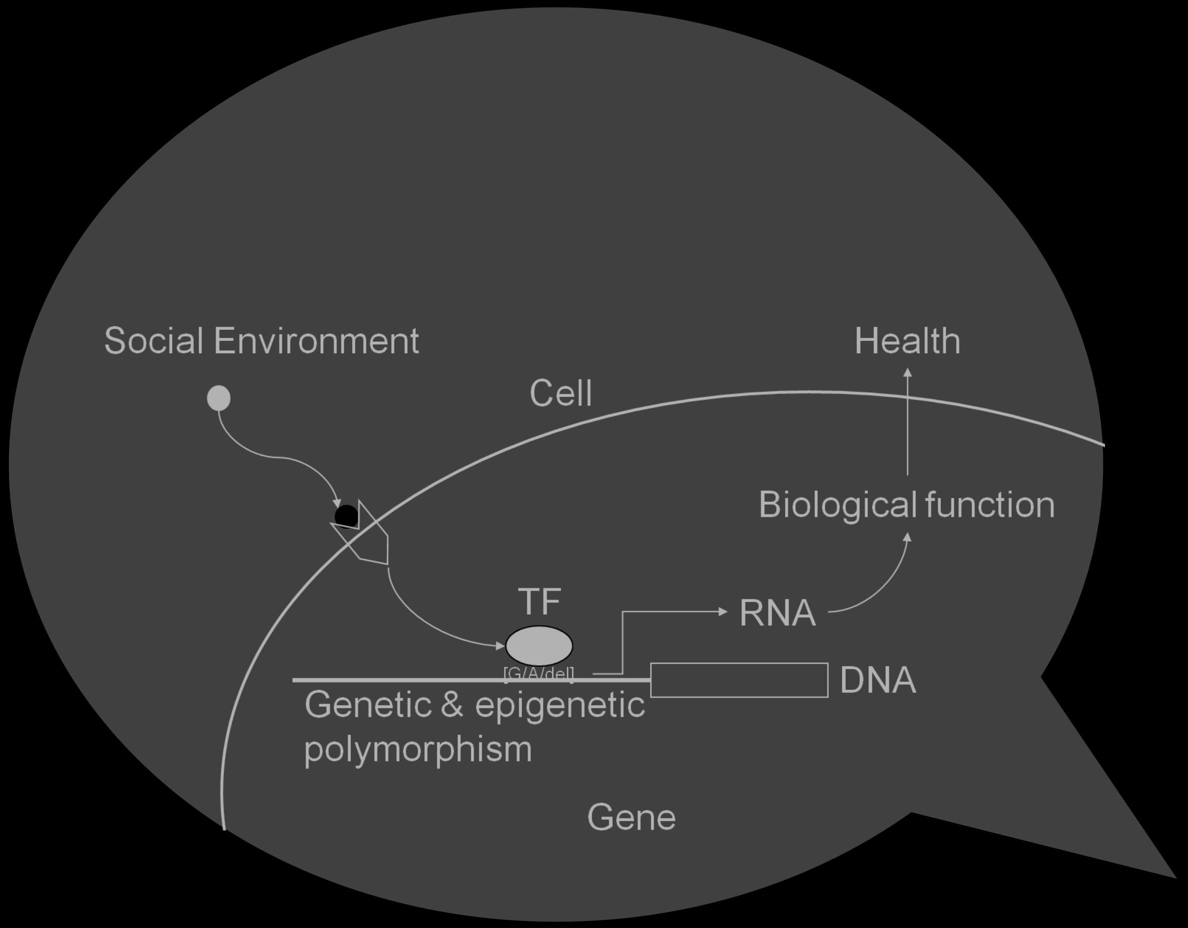
gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40

I’ve heard of this one!

And this one…

… and this one too!

OK done.



Social Environment

Cell

Health

Biological function

TF

RNA

[G/A/del]

DNA

Genetic & epigenetic polymorphism

Gene

gene 1

gene 2

gene 3

gene 4

gene 5

gene 6

gene 7

gene 8

gene 9

gene 10

gene 11

gene 12

gene 13

gene 14

gene 15

gene 16

gene 17

gene 18

gene 19

gene 20

gene 21

gene 22

gene 23

gene 24

gene 25

gene 26

gene 27

gene 28

gene 29

gene 30

gene 31

gene 32

gene 33

gene 34

gene 35

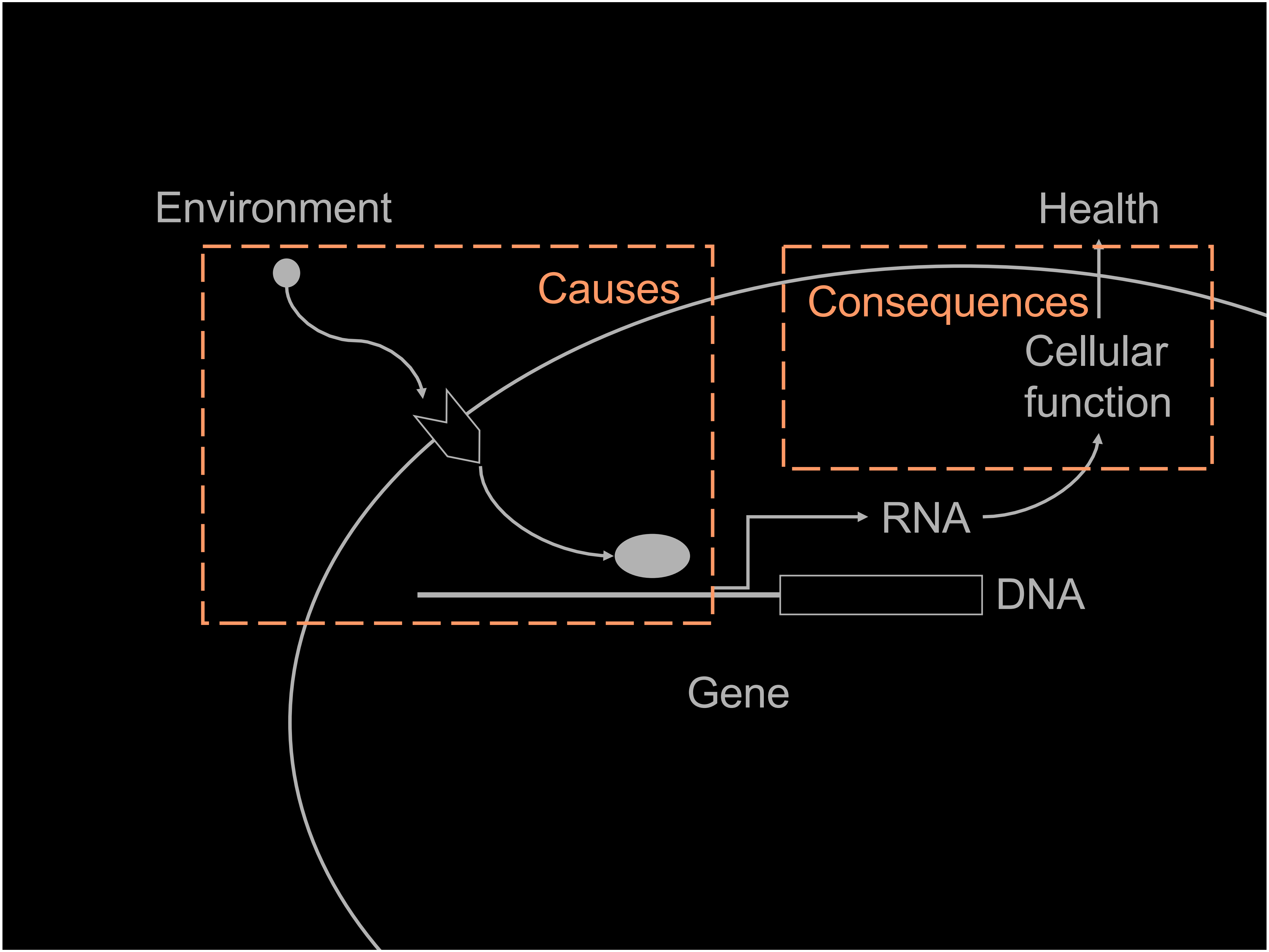
gene 36

gene 37

gene 38

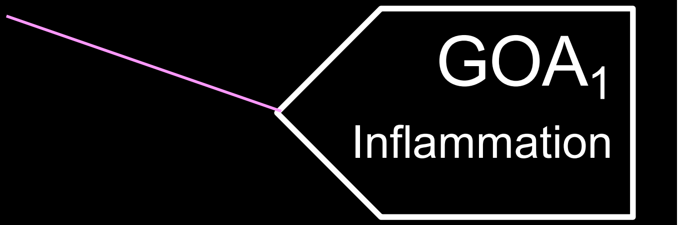
gene 39

gene 40


Environment
Health
Causes
Consequences
Cellular function
RNA
DNA
Gene

Interpreting results within an unbiased
“expert annotation” reference space

gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40



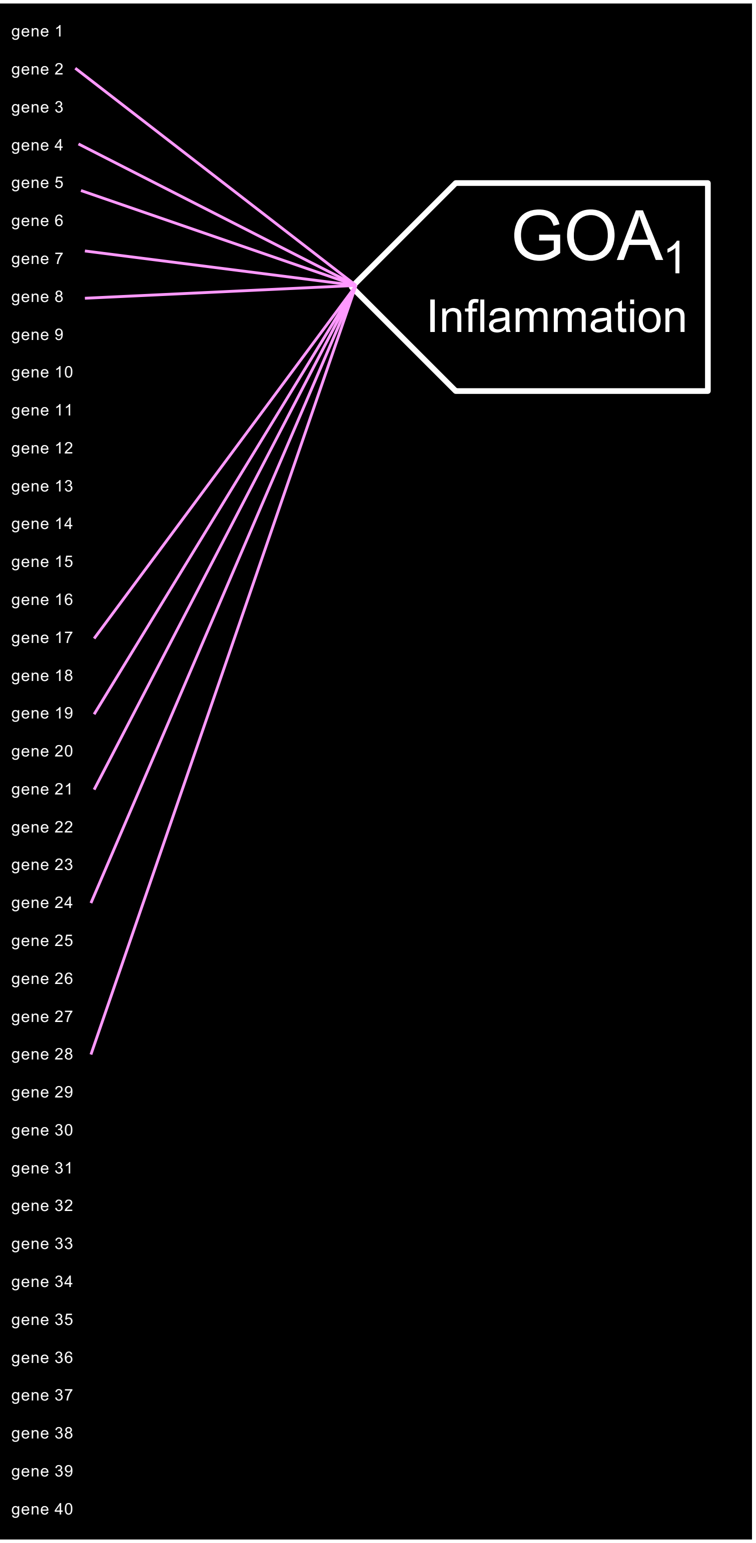
GOA$_1$

Inflammation

Interpreting results within an unbiased “expert annotation” reference space

A statistical caveat: what’s the right reference point for a null hypothesis?

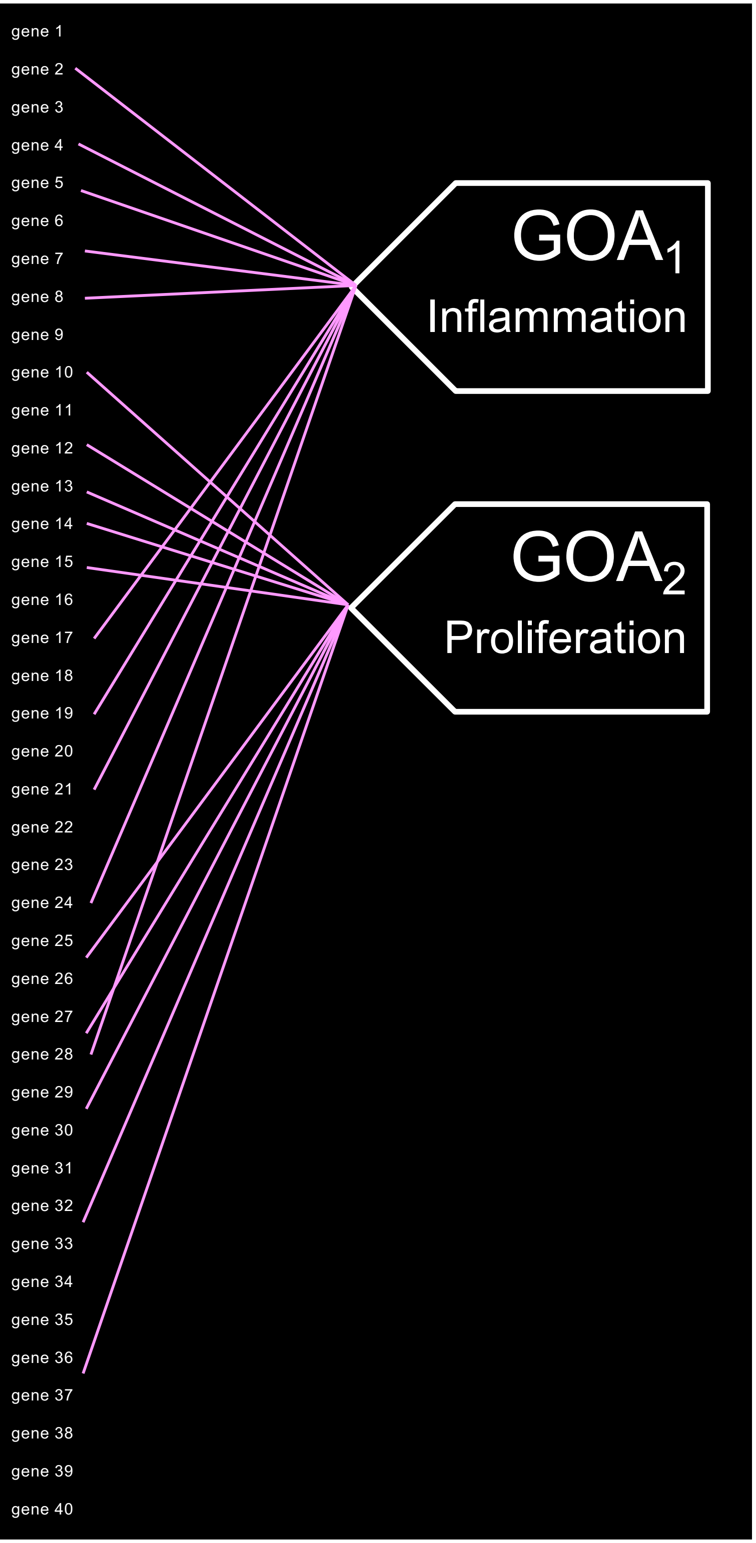
- Genome-wide baseline?
- Expressed transcriptome?
- Up- vs. down-regulated gene sets?



Interpreting results within an unbiased "expert annotation" reference space

A statistical caveat: what's the right reference point for a null hypothesis?

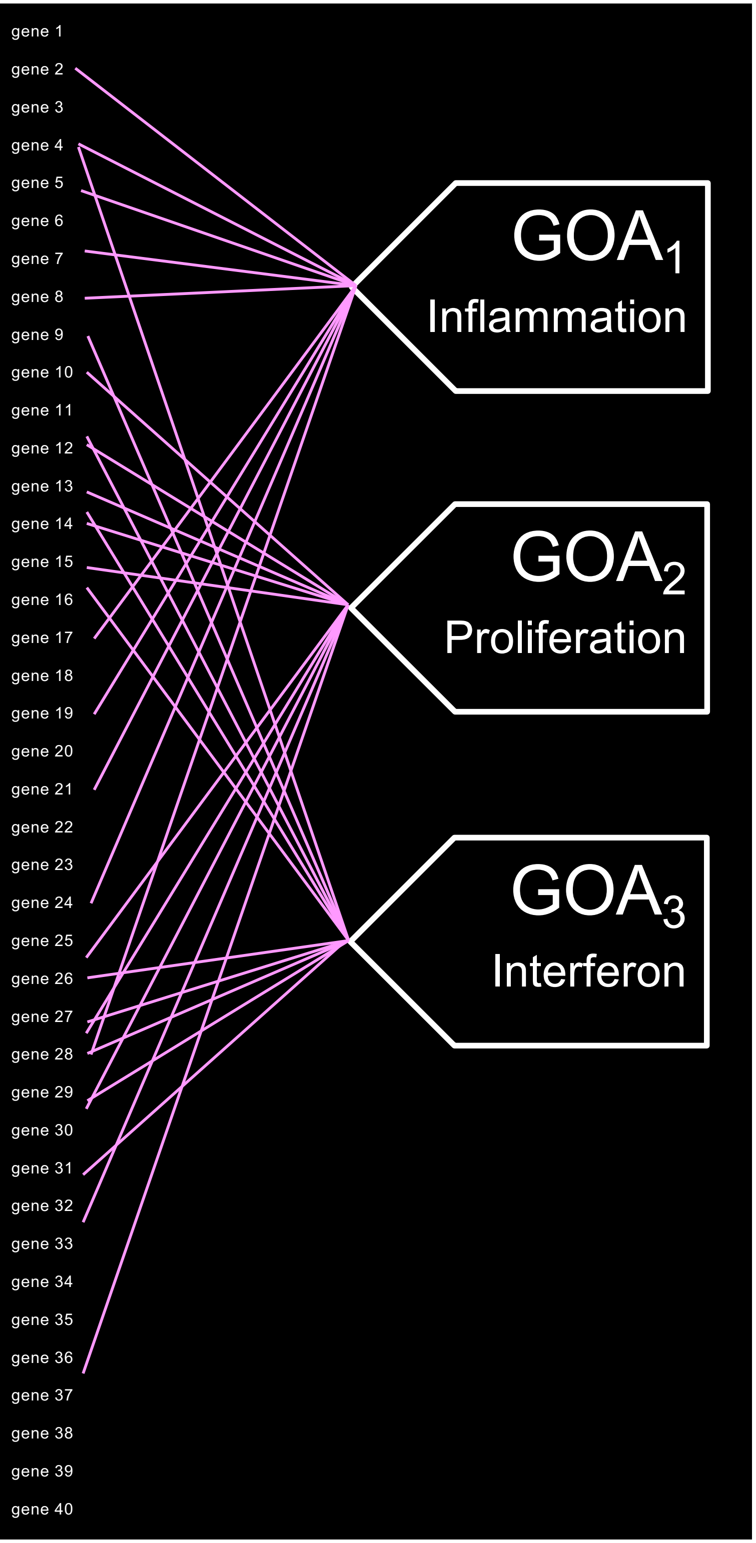
- Genome-wide baseline?
- Expressed transcriptome?
- Up- vs. down-regulated gene sets?



Interpreting results within an unbiased "expert annotation" reference space

A statistical caveat: what's the right reference point for a null hypothesis?

- Genome-wide baseline?
- Expressed transcriptome?
- Up- vs. down-regulated gene sets?



gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40

GOA_1
Inflammation

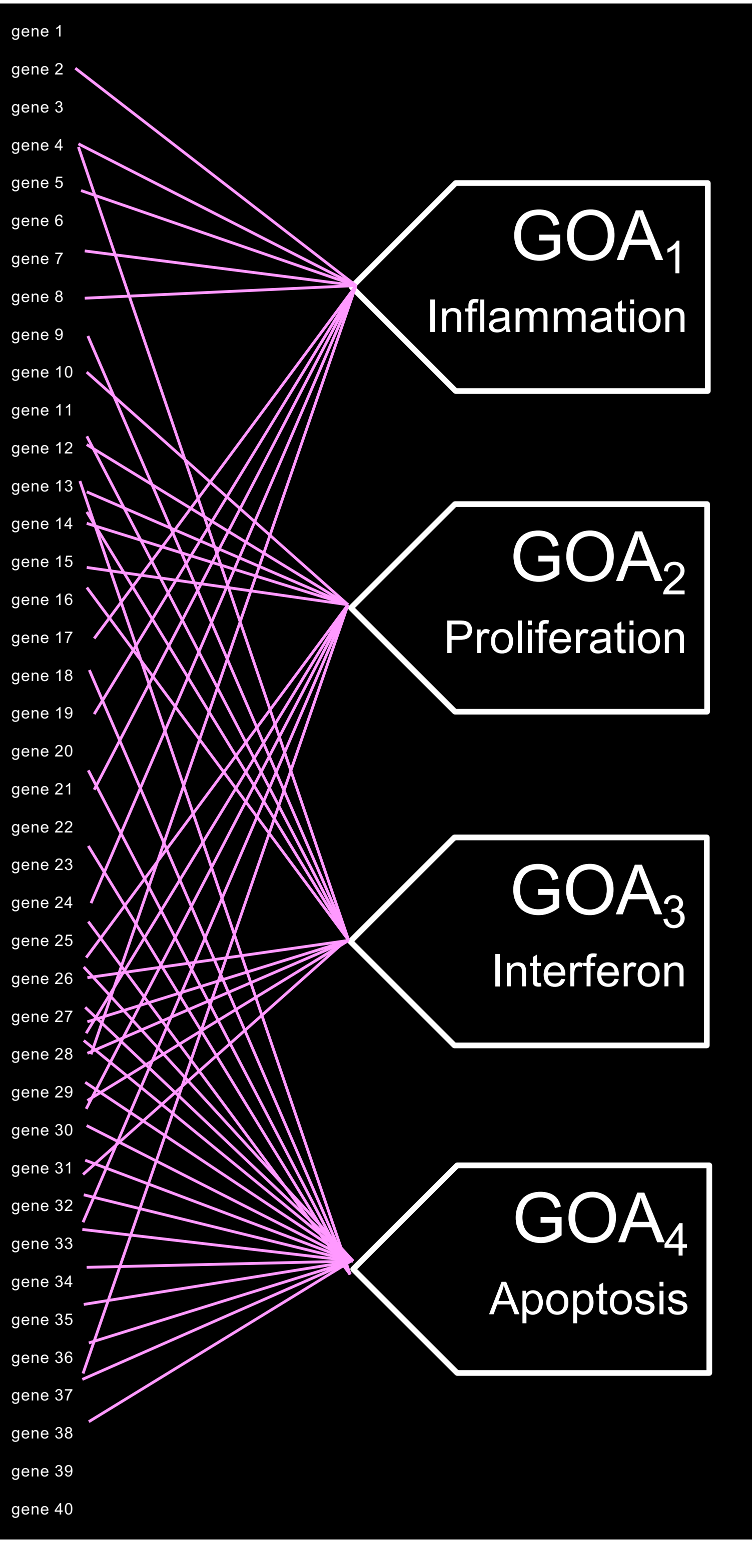
GOA_2
Proliferation

GOA_3
Interferon

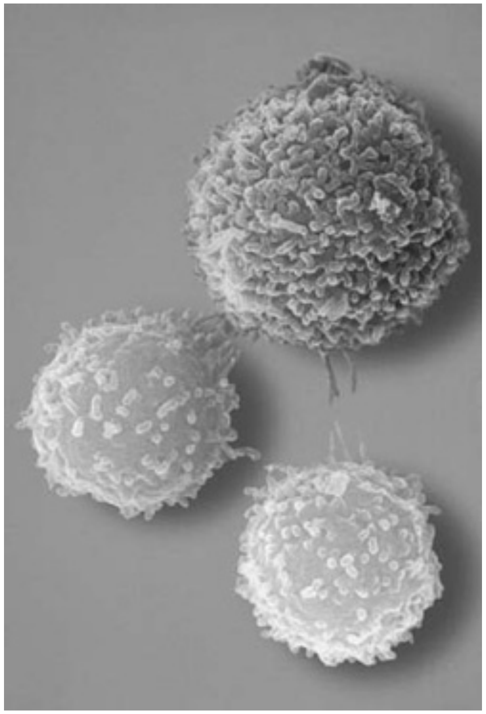
Interpreting results within an unbiased “expert annotation” reference space

A statistical caveat: what’s the right reference point for a null hypothesis?

- Genome-wide baseline?
- Expressed transcriptome?
- Up- vs. down-regulated gene sets?



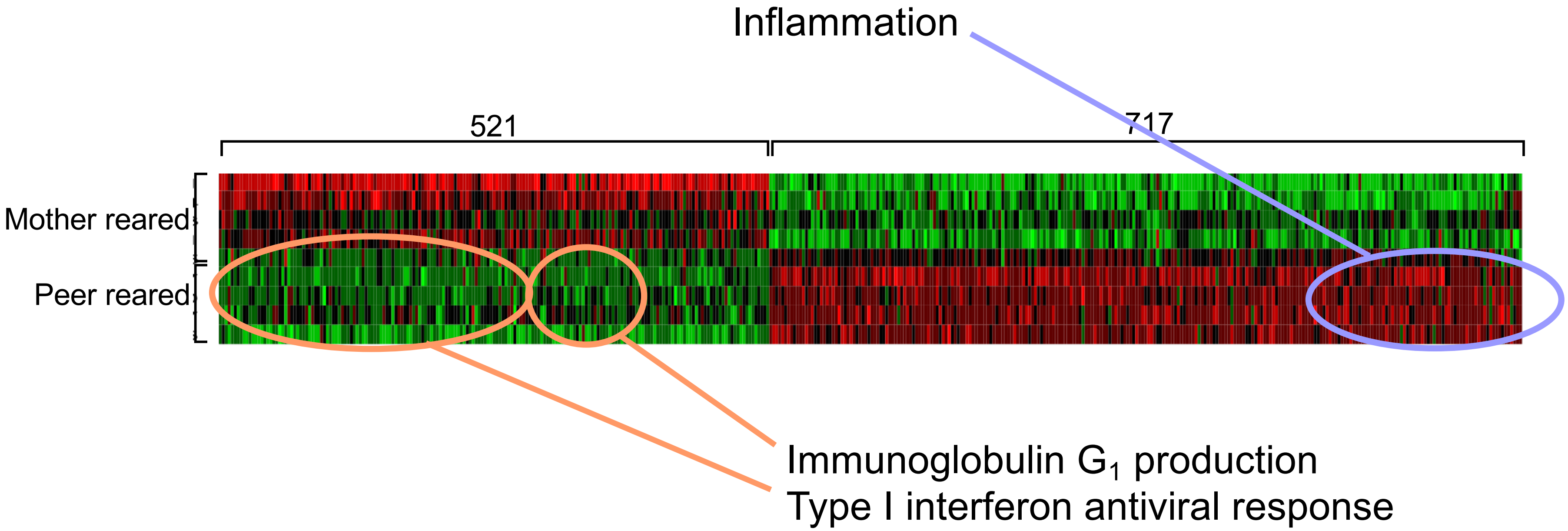
Social instability



Cole et al (2012) PNAS

CTRA – conserved transcriptional response to adversity

Social instability



Inflammation

Immunoglobulin G_1 production

Type I interferon antiviral response

Cole et al (2012) PNAS

Low SES

Social loss / bereavement

Post-traumatic stress

Cancer diagnosis

Social threat

Loneliness

Social instability

Chronic stress

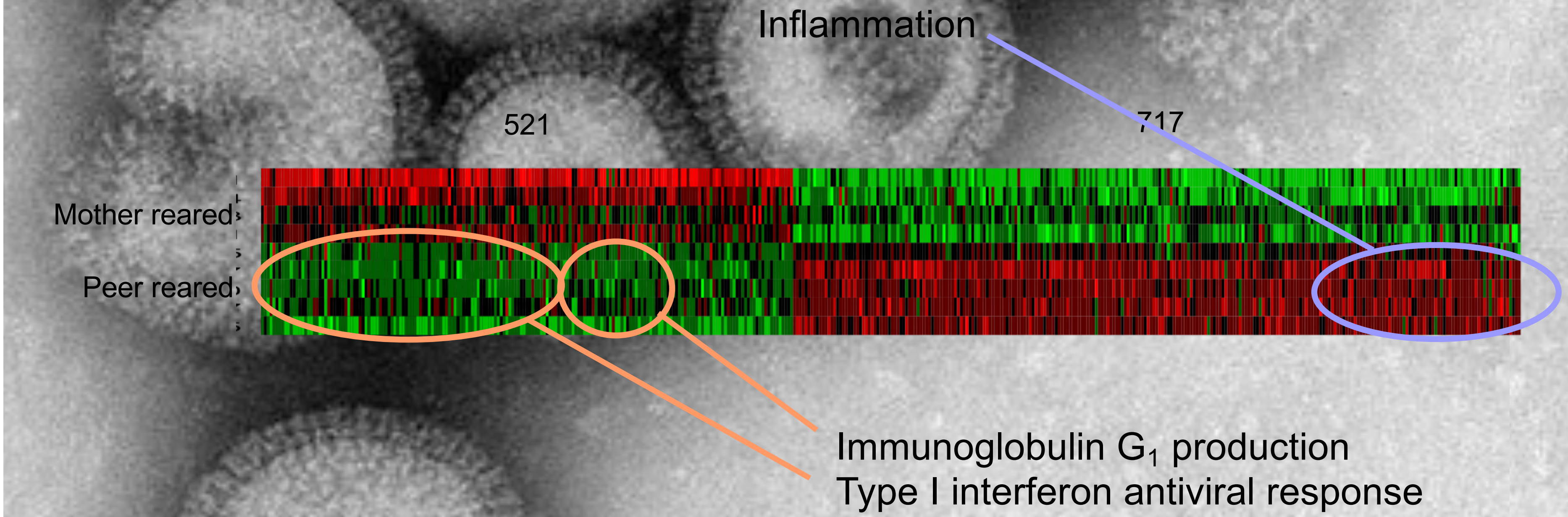
Low social rank

Caregiving for seriously ill

Anxiety

Early life adversity

CTRA – conserved transcriptional response to adversity



Immunoglobulin G_1 production
Type I interferon antiviral response

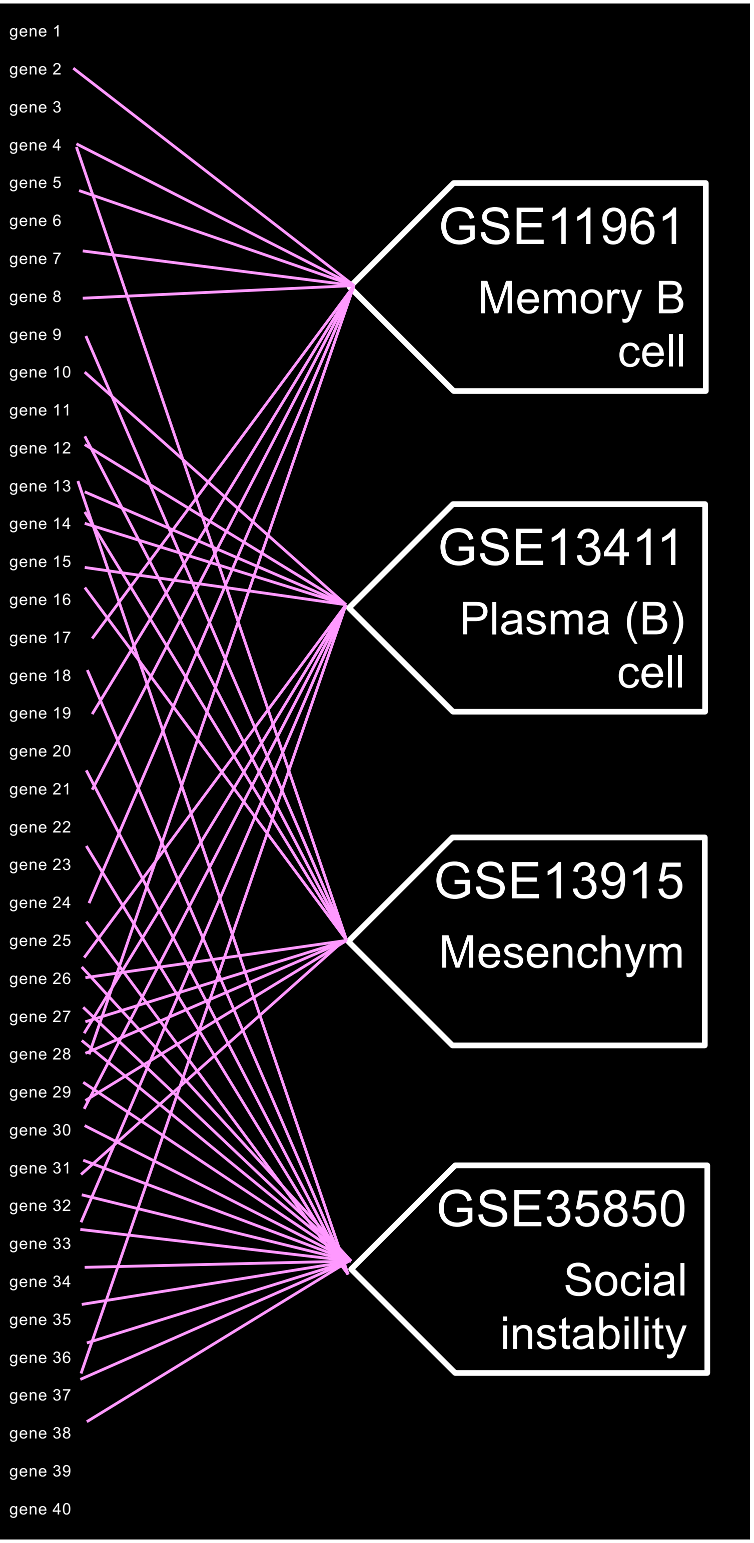
Irwin & Cole, Nature Reviews Immunology 2011
Cole, PLoS Genetics 2014
Nunn & Altizer, Infectious Diseases in Primates 2006

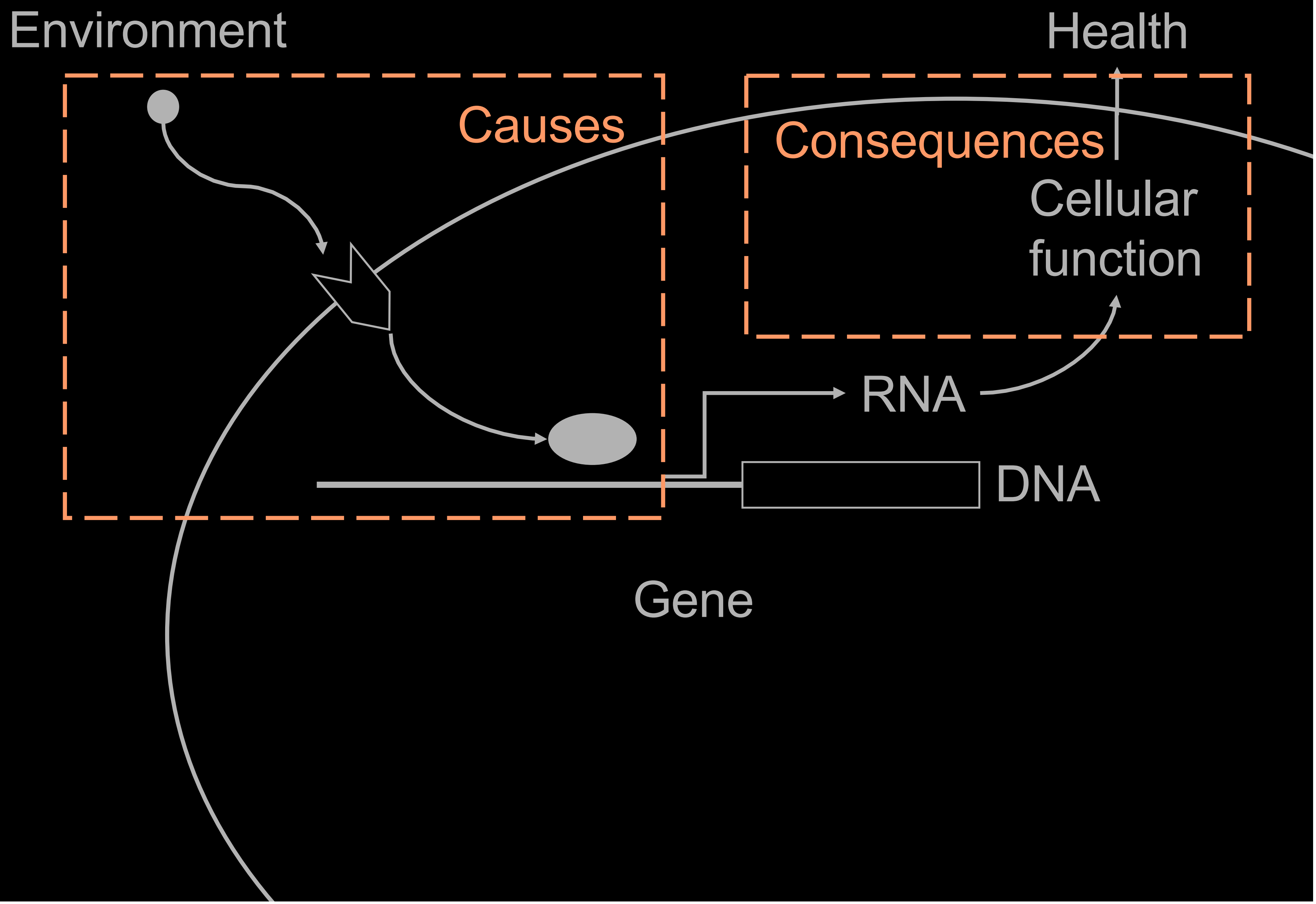
Interpreting results within an empirical transcriptome reference space
(vs. "expert" GO tags)

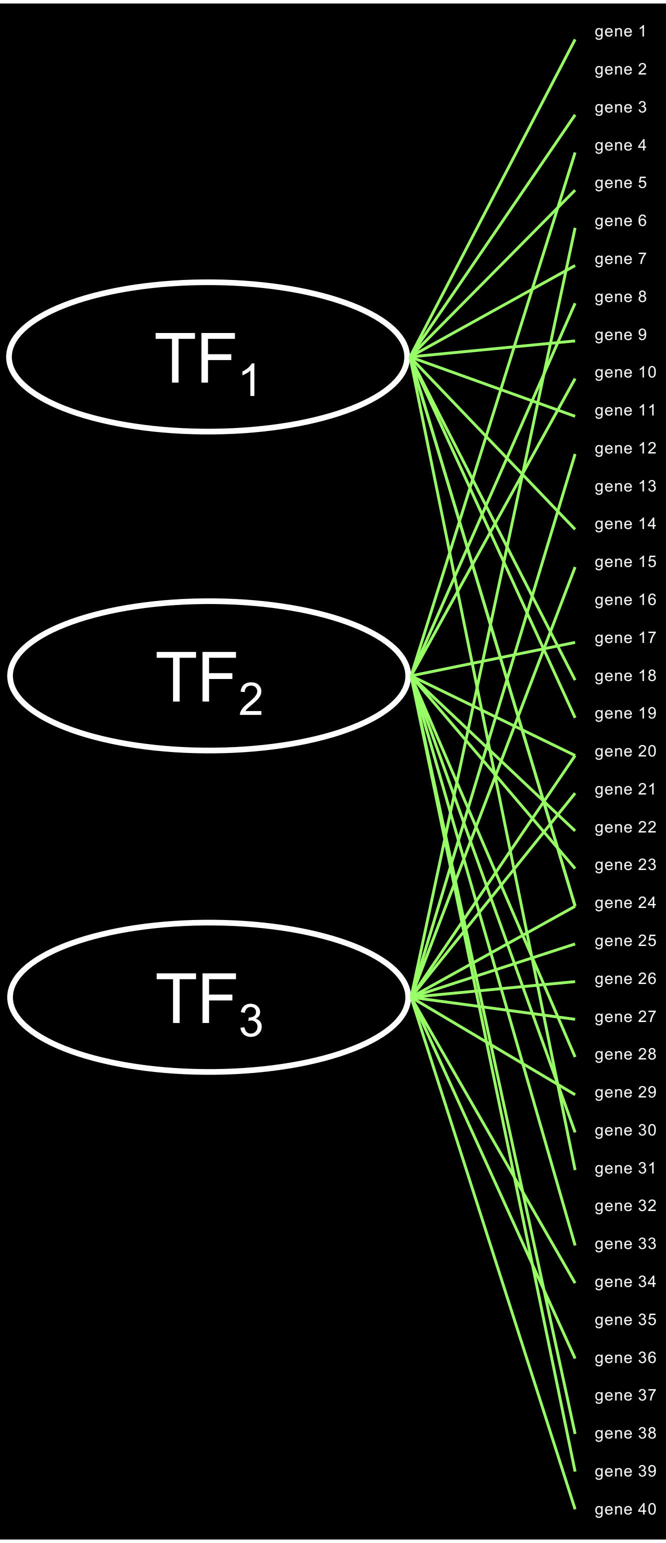
1. GEO search
(or ArrayExpress, GSEA, etc.)

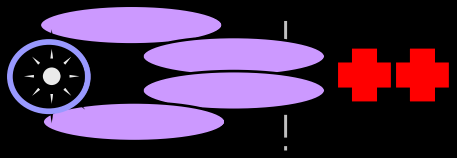
2. Grab the data & compute diagnosticity scores for their experimental "poles"

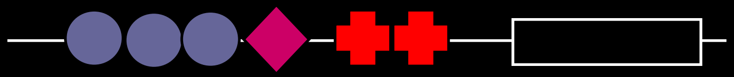
3. Test for gene set differential expression in your data (either them|you or you|them)

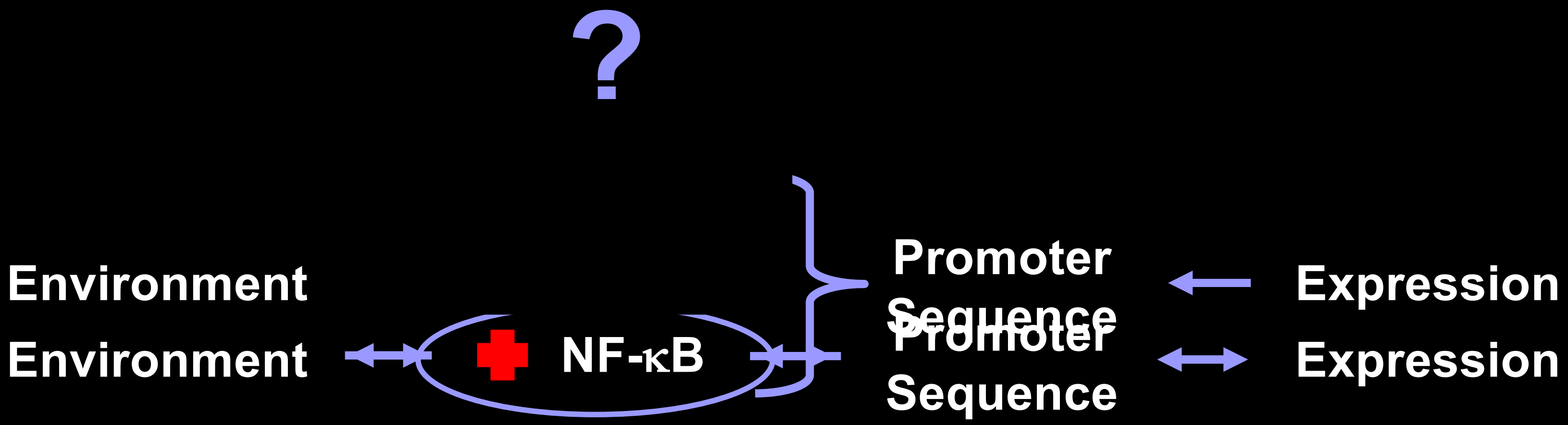



Environment
Health
Causes
Consequences
Cellular function
RNA
DNA
Gene


TF_1
TF_2
TF_3
gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40







?

Environment

Promoter Sequence

Expression

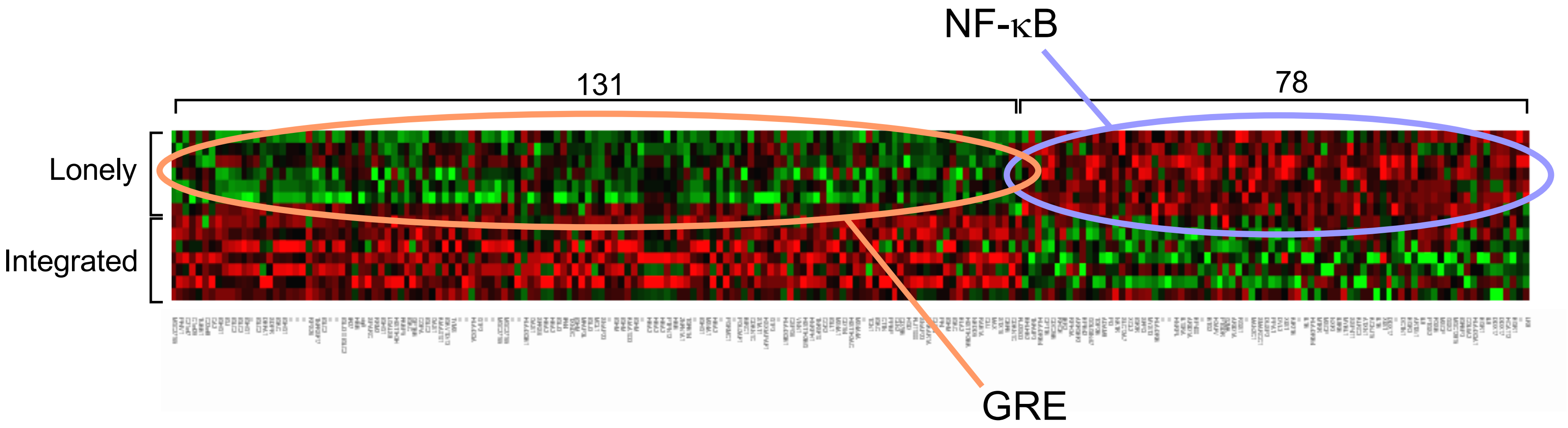
Environment

NF-κB

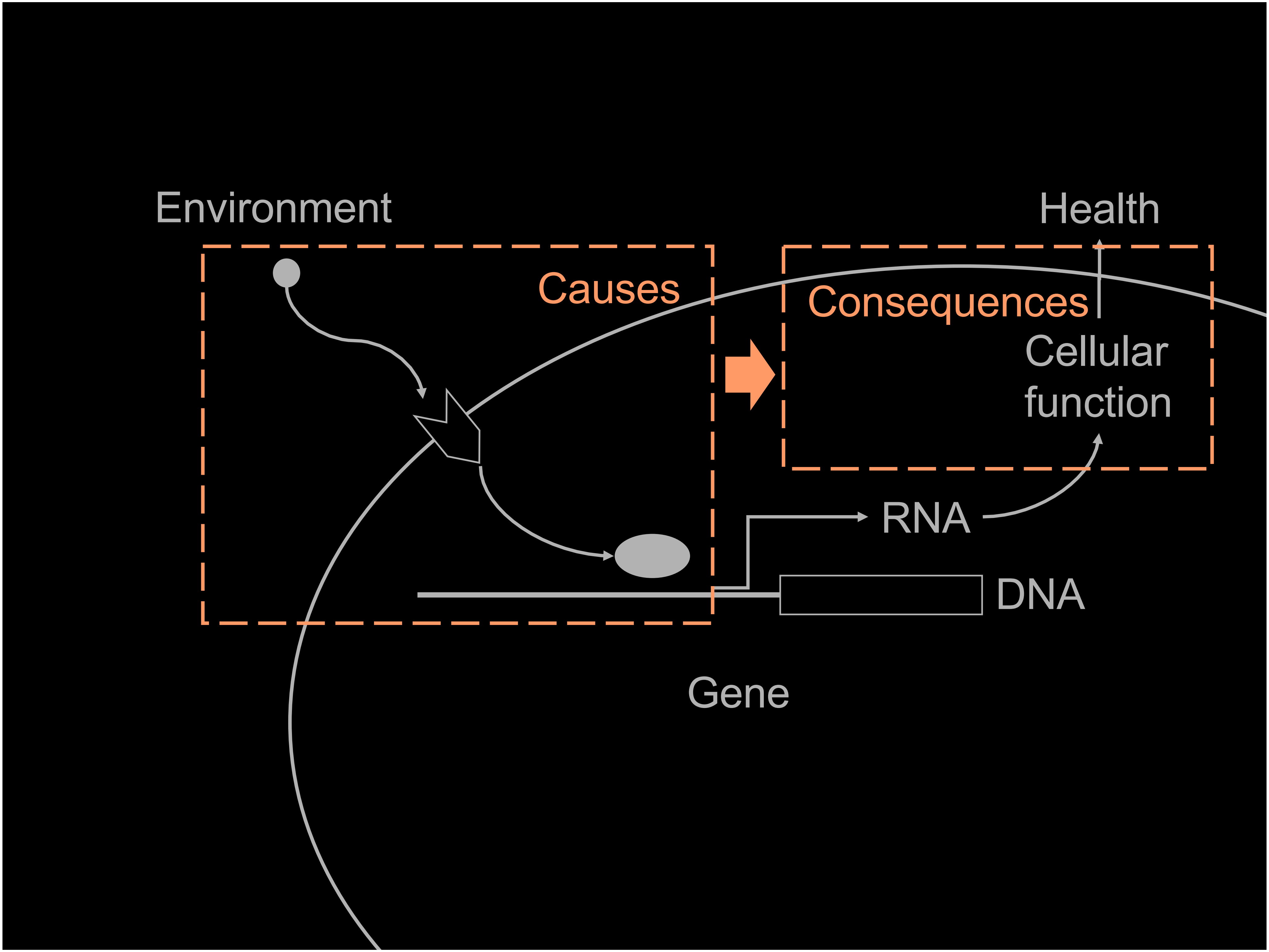
Promoter Sequence

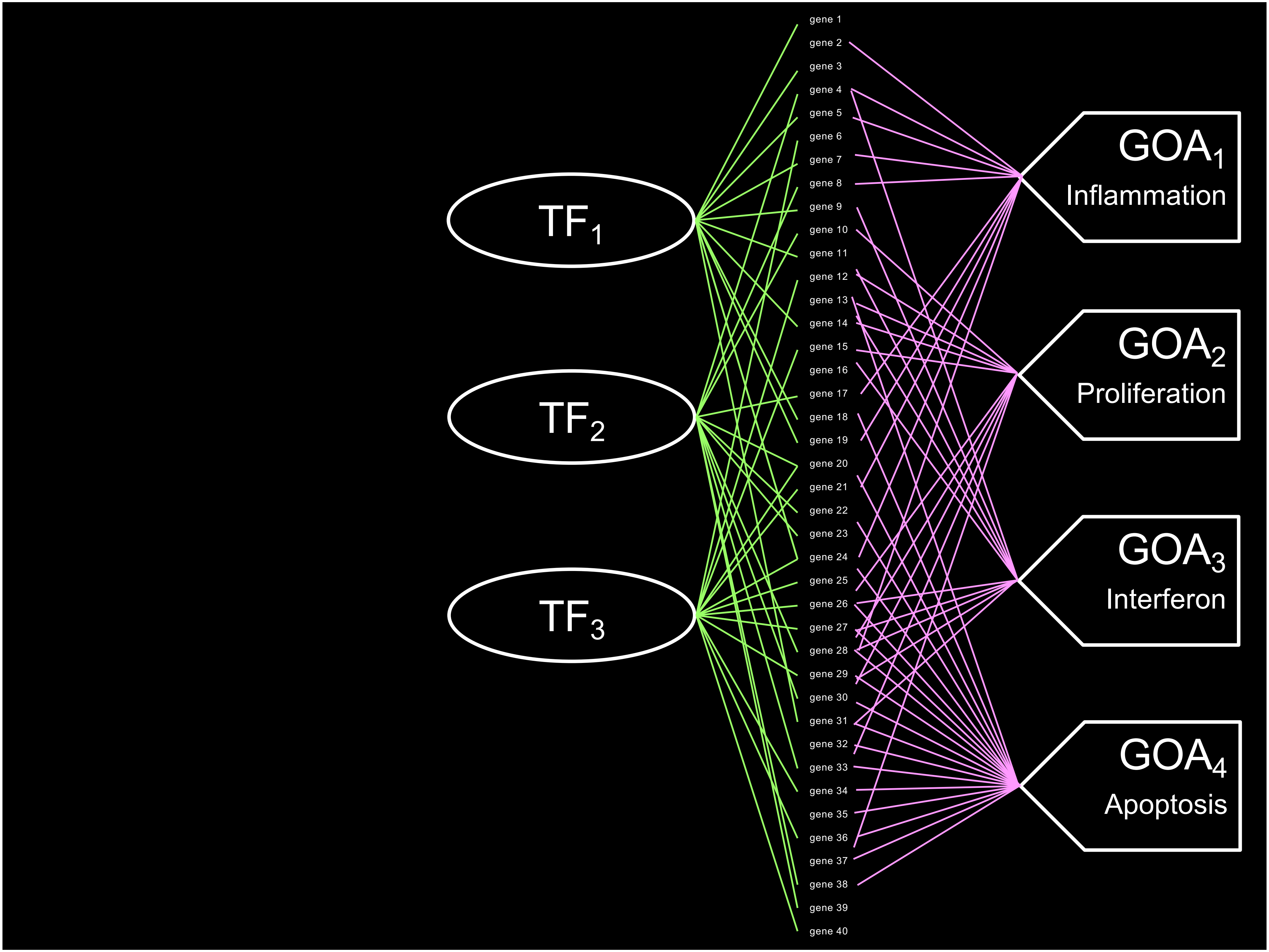
Expression

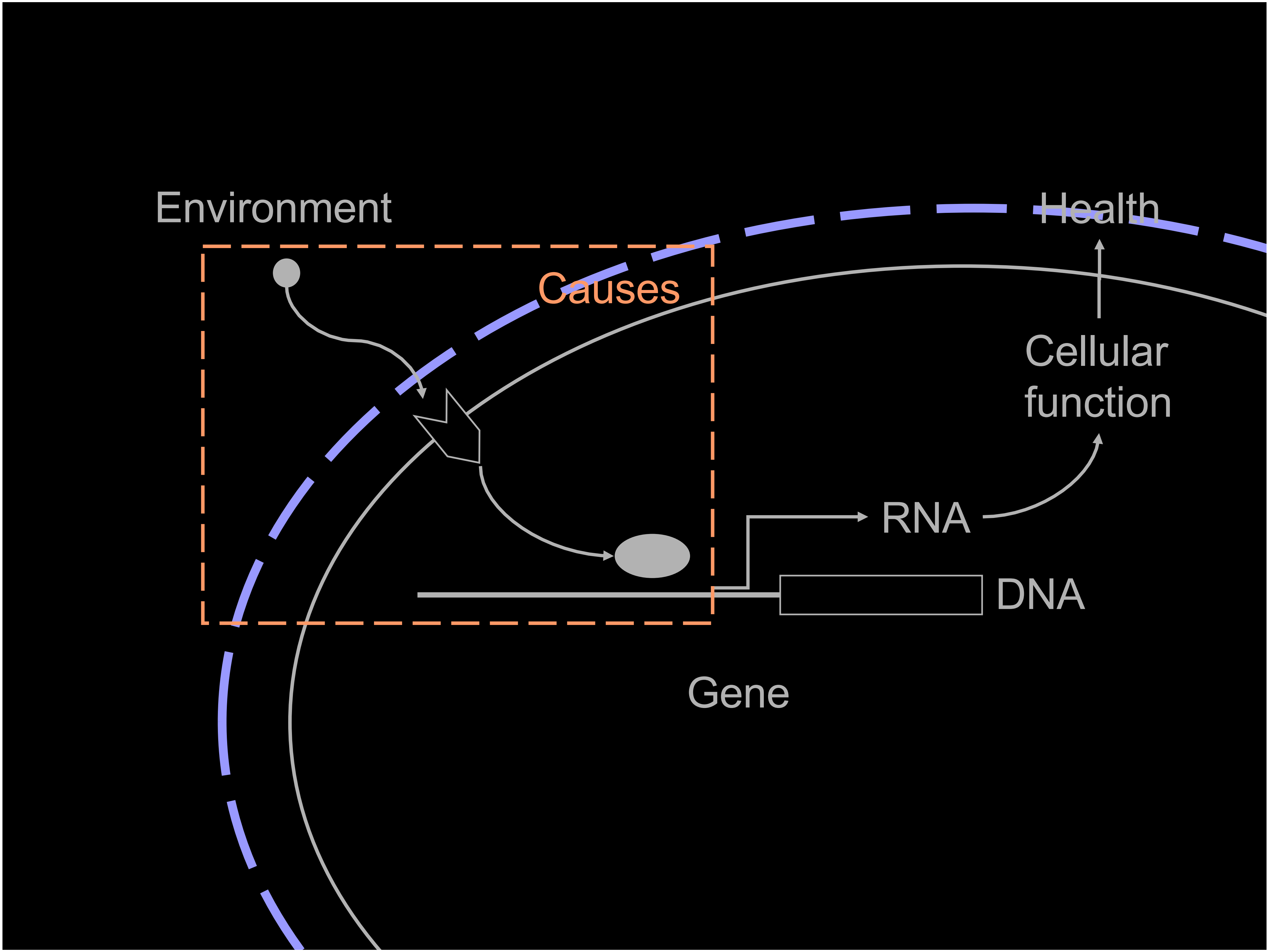
CTRA – Conserved Transcriptional Response to Adversity

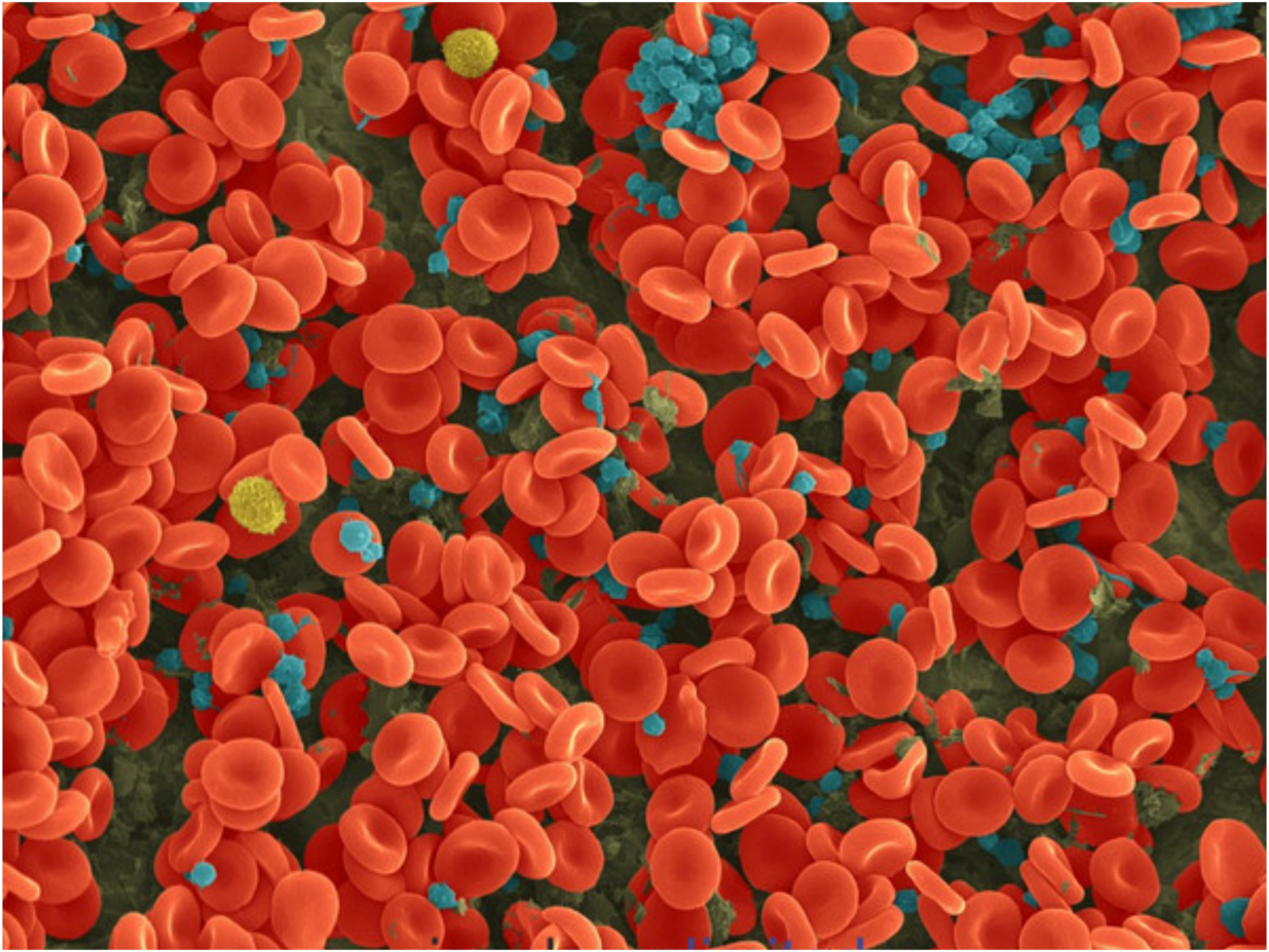


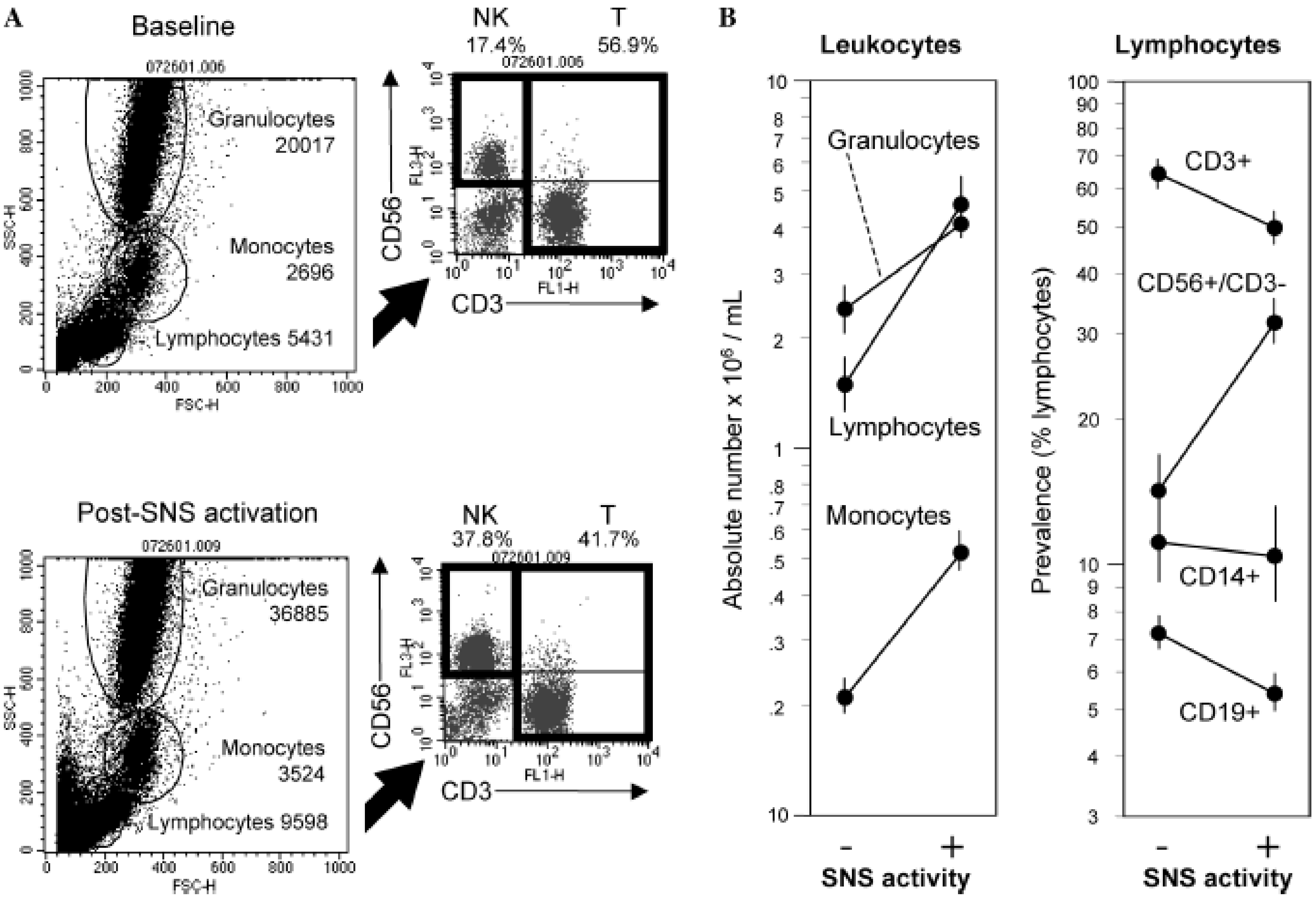
Iriwin & Cole (2011) Nature Reviews Immunology


Environment
Health
Causes
Consequences
Cellular function
RNA
DNA
Gene


TF$_1$
TF$_2$
TF$_3$
gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40
GOA$_1$
Inflammation
GOA$_2$
Proliferation
GOA$_3$
Interferon
GOA$_4$
Apoptosis

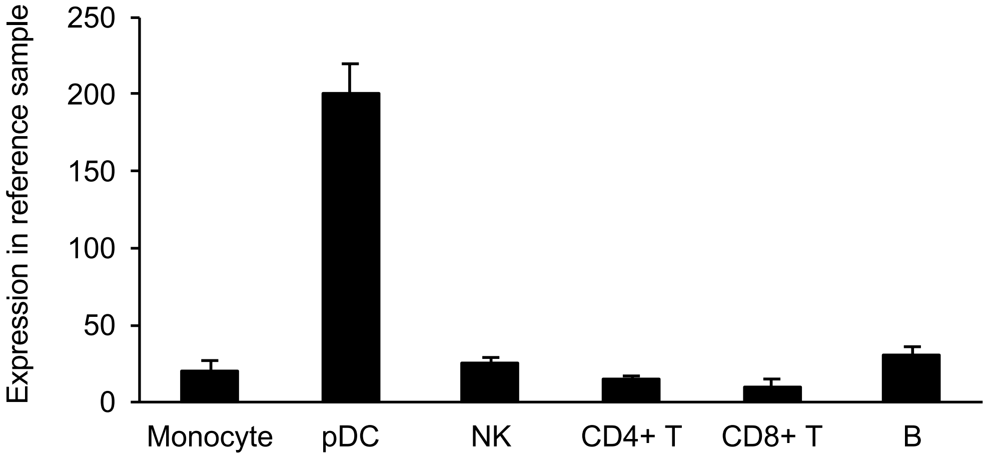

Environment
Health
Causes
Cellular
function
RNA
DNA
Gene





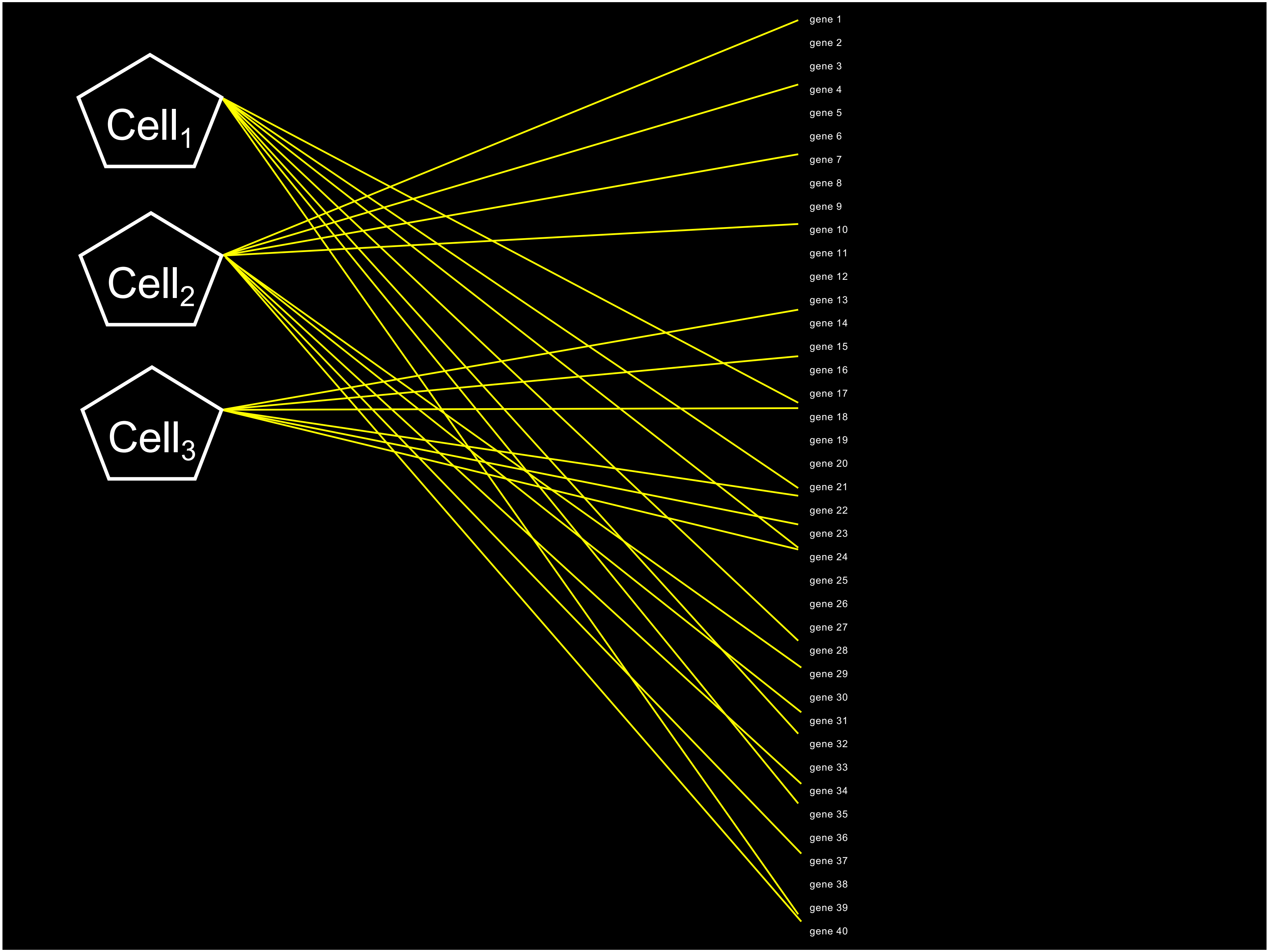
Richlin et al. Brain, Behavior & Immunity (2004)

Transcript origin analysis

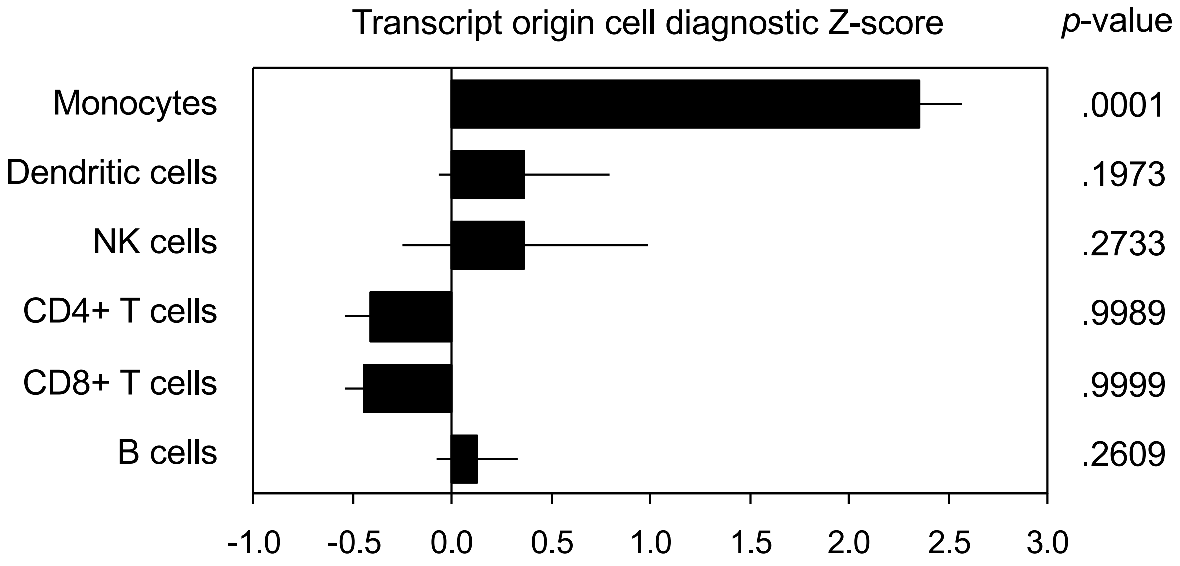


TOA *Z*-score $_{gc}$ = -0.4 4.9 -0.3 -0.5 -0.6 -0.3

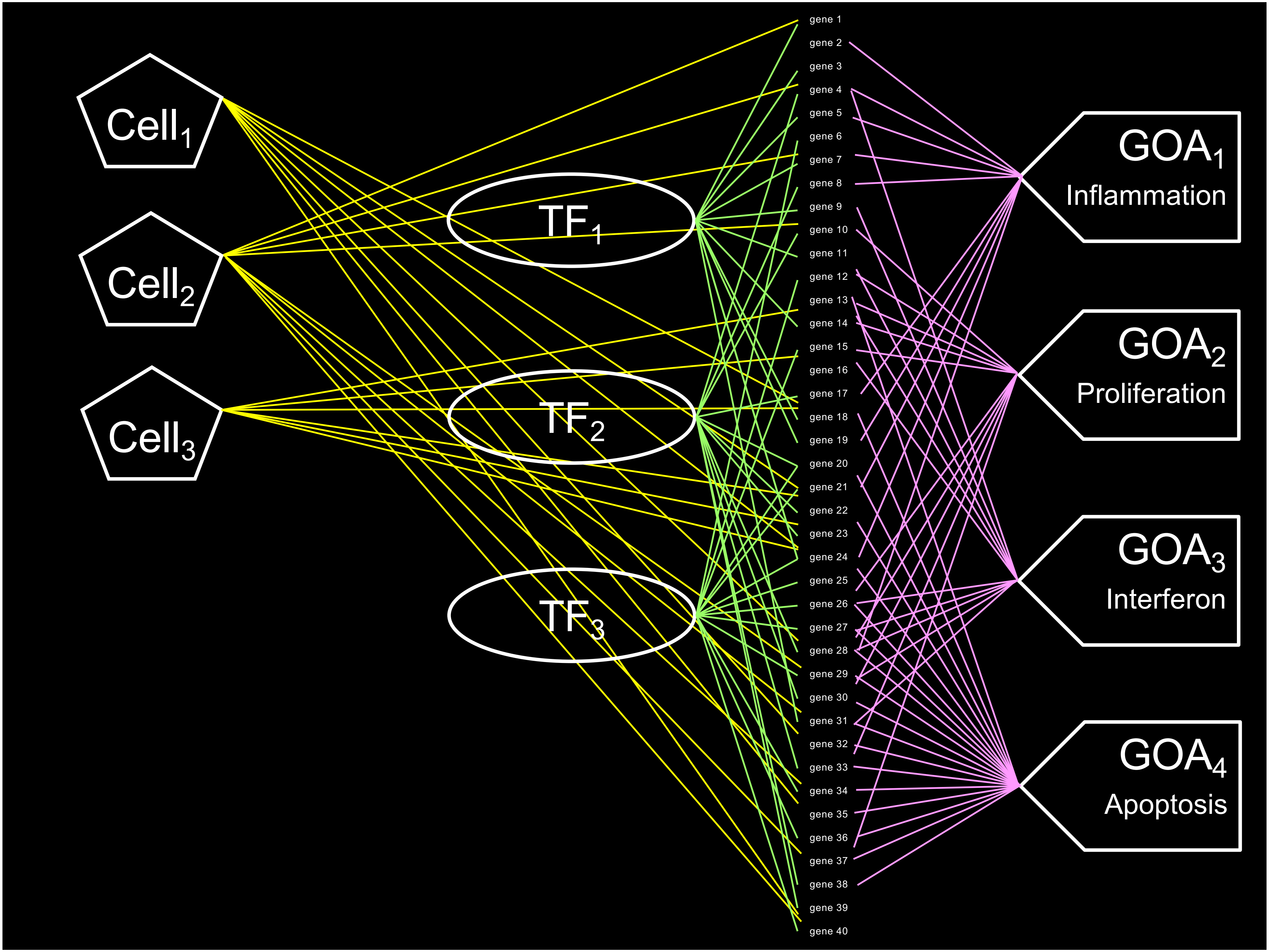
Cole et al (2011) PNAS


$Cell_1$
$Cell_2$
$Cell_3$
gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40

Cellular target of social adversity

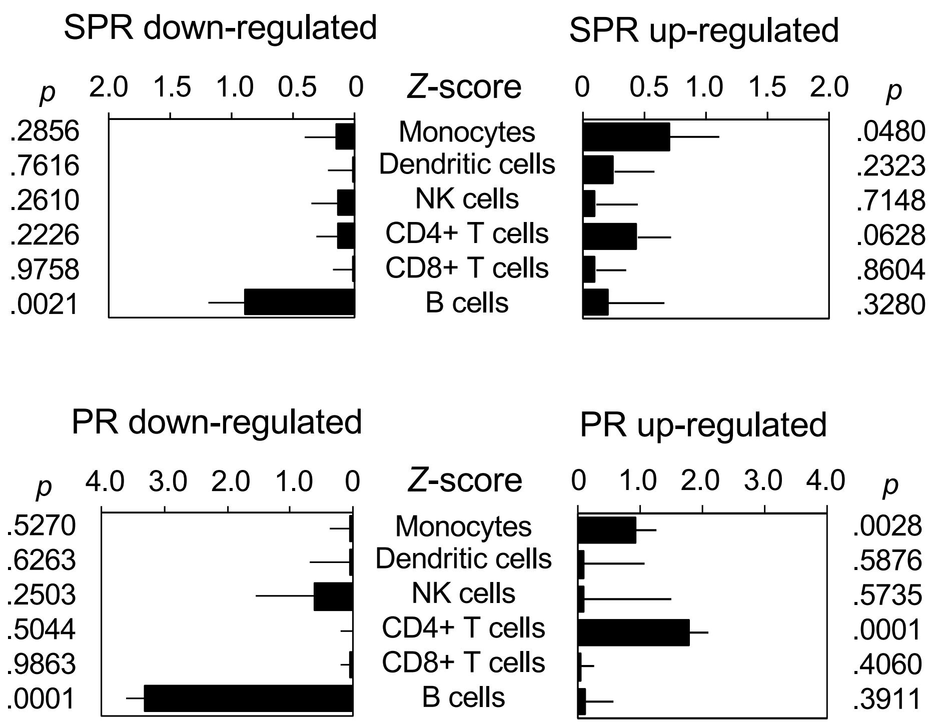


Cole et al (2011) PNAS

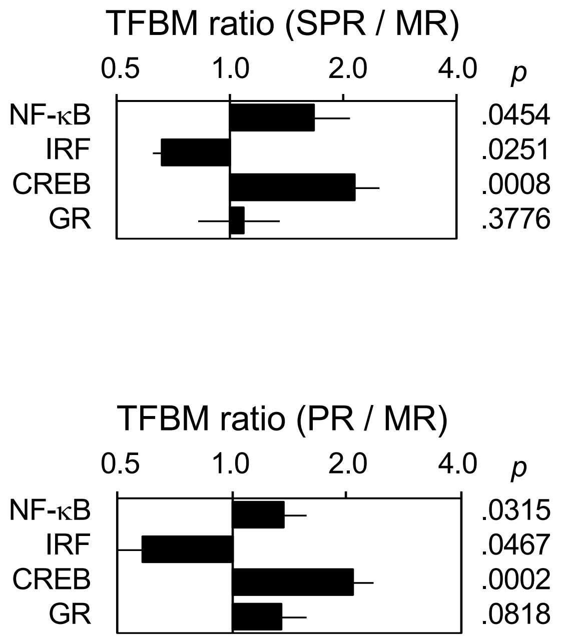

Cell$_1$
Cell$_2$
Cell$_3$
TF$_1$
TF$_2$
TF$_3$
gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40
GOA$_1$
Inflammation
GOA$_2$
Proliferation
GOA$_3$
Interferon
GOA$_4$
Apoptosis

Social instability

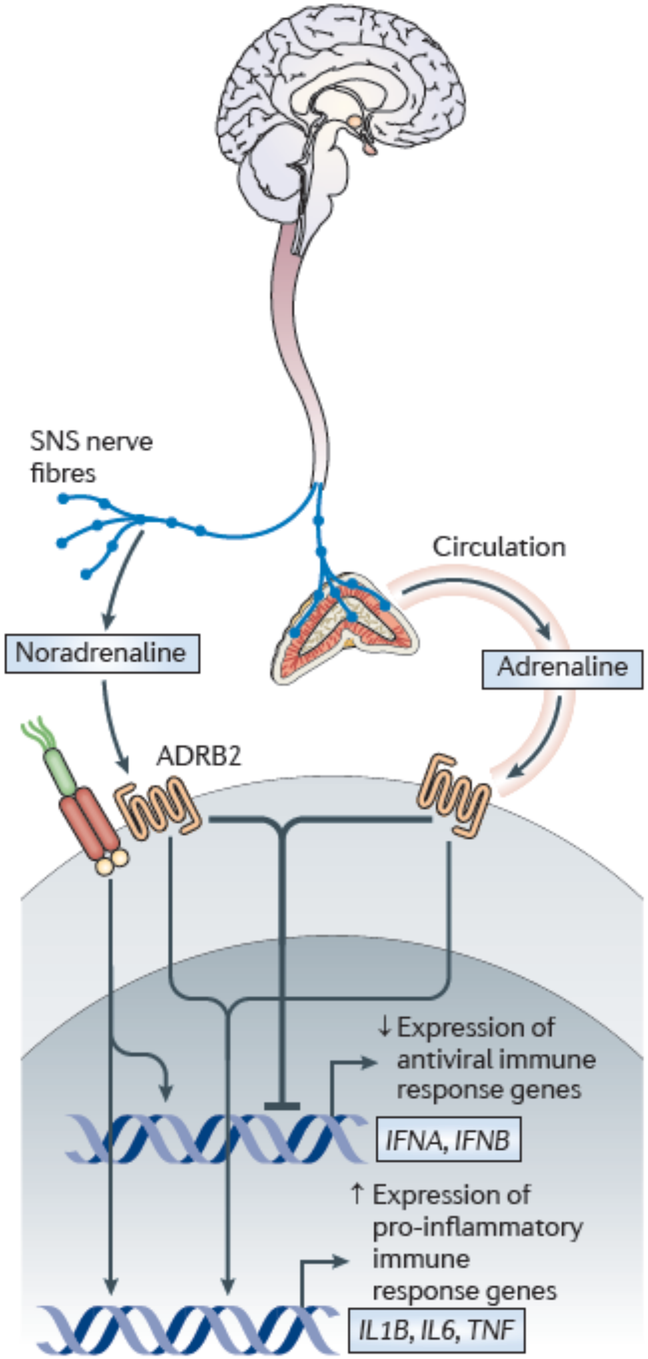
Cellular origin



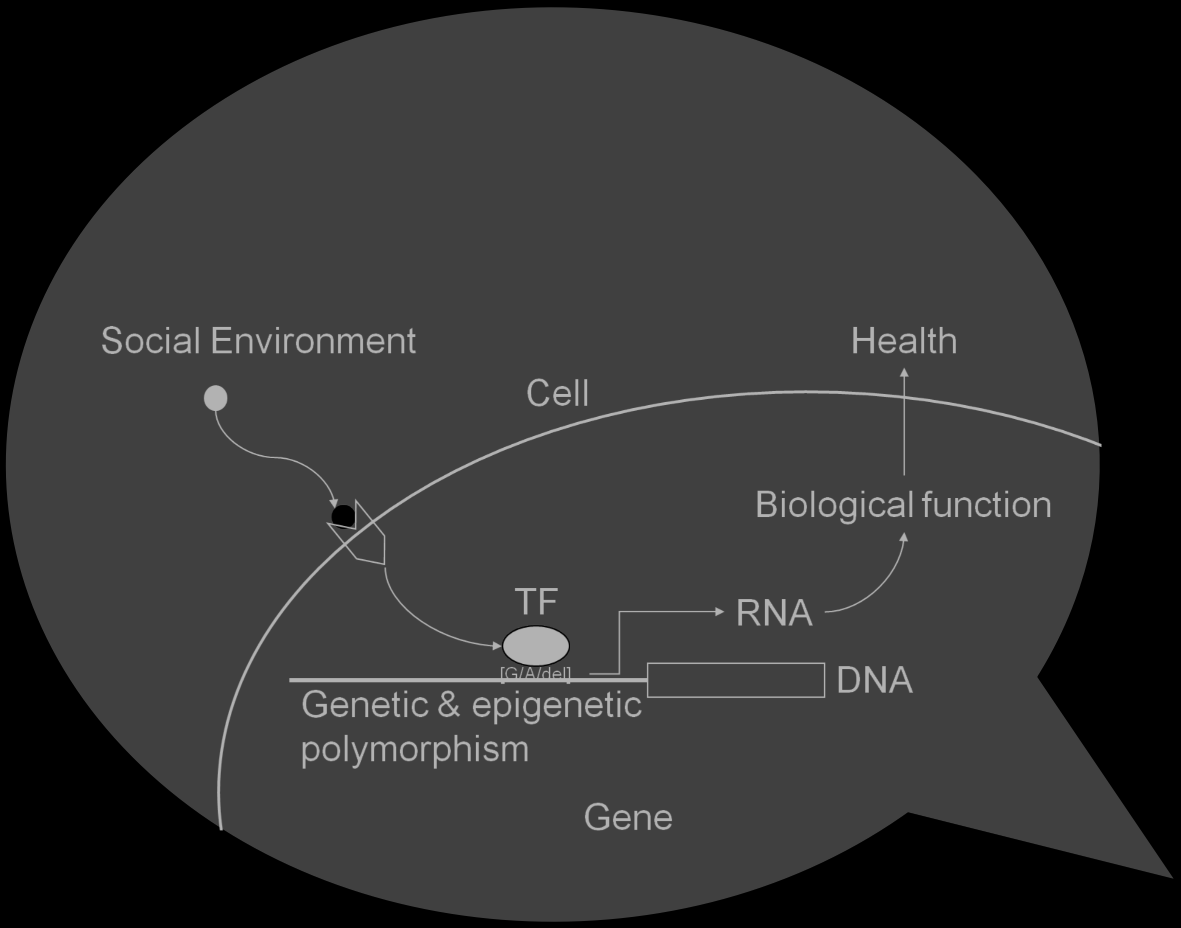
Transcription factor



Cole et al (2012) PNAS



SNS nerve fibres
Circulation
Noradrenaline
Adrenaline
ADRB2
↓ Expression of antiviral immune response genes
IFNA, IFNB
↑ Expression of pro-inflammatory immune response genes
IL1B, IL6, TNF



Social Environment

Cell

Health

Biological function

TF

RNA

[G/A/del]

DNA

Genetic & epigenetic polymorphism

Gene

gene 1

gene 2

gene 3

gene 4

gene 5

gene 6

gene 7

gene 8

gene 9

gene 10

gene 11

gene 12

gene 13

gene 14

gene 15

gene 16

gene 17

gene 18

gene 19

gene 20

gene 21

gene 22

gene 23

gene 24

gene 25

gene 26

gene 27

gene 28

gene 29

gene 30

gene 31

gene 32

gene 33

gene 34

gene 35

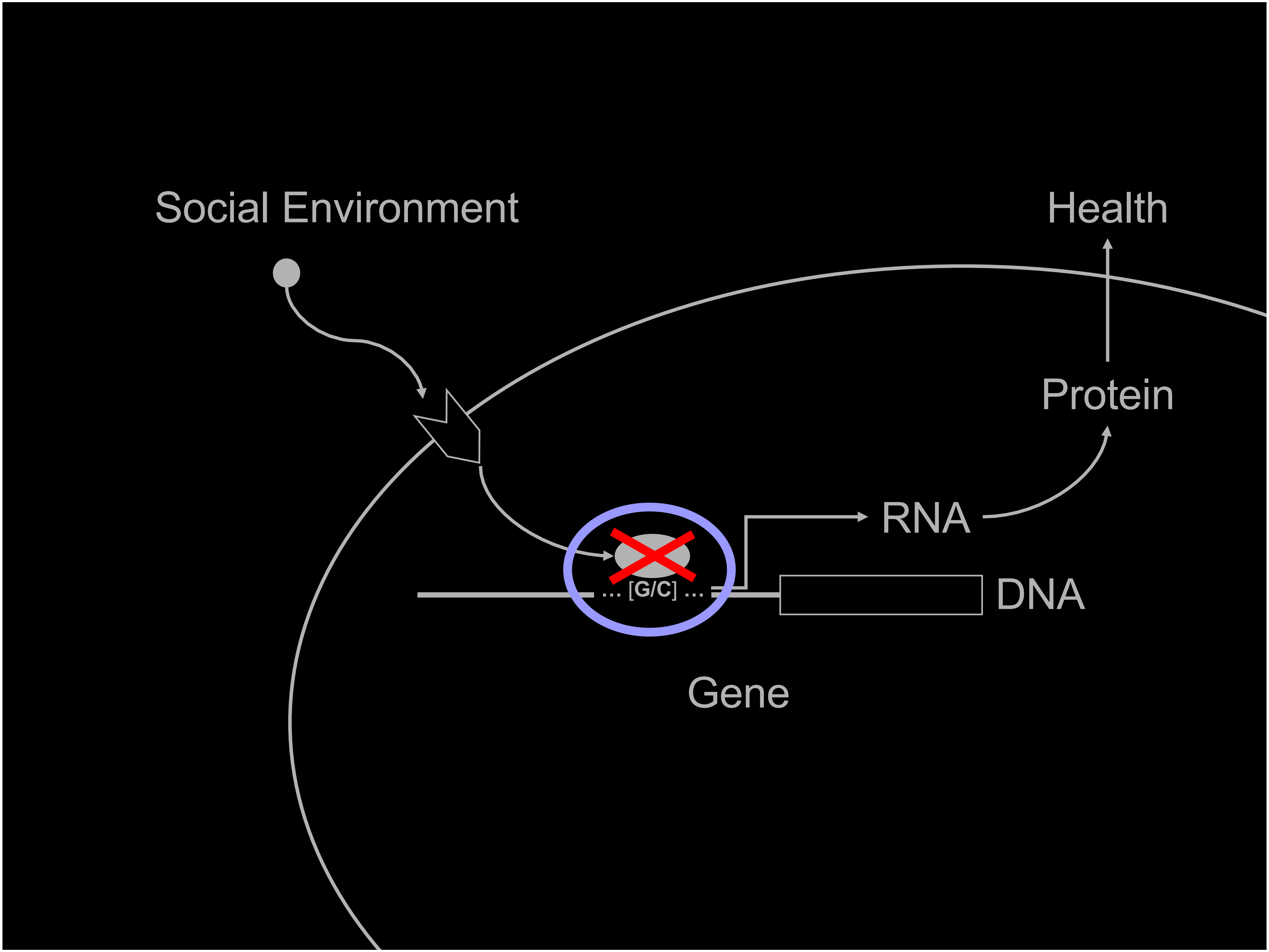
gene 36

gene 37

gene 38

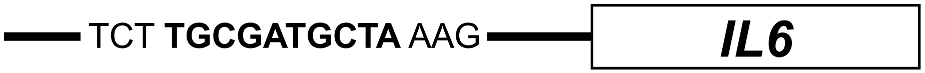
gene 39

gene 40


Social Environment
Health
Protein
RNA
... [G/C] ...
DNA
Gene

Gene x Environment Interaction

In silico

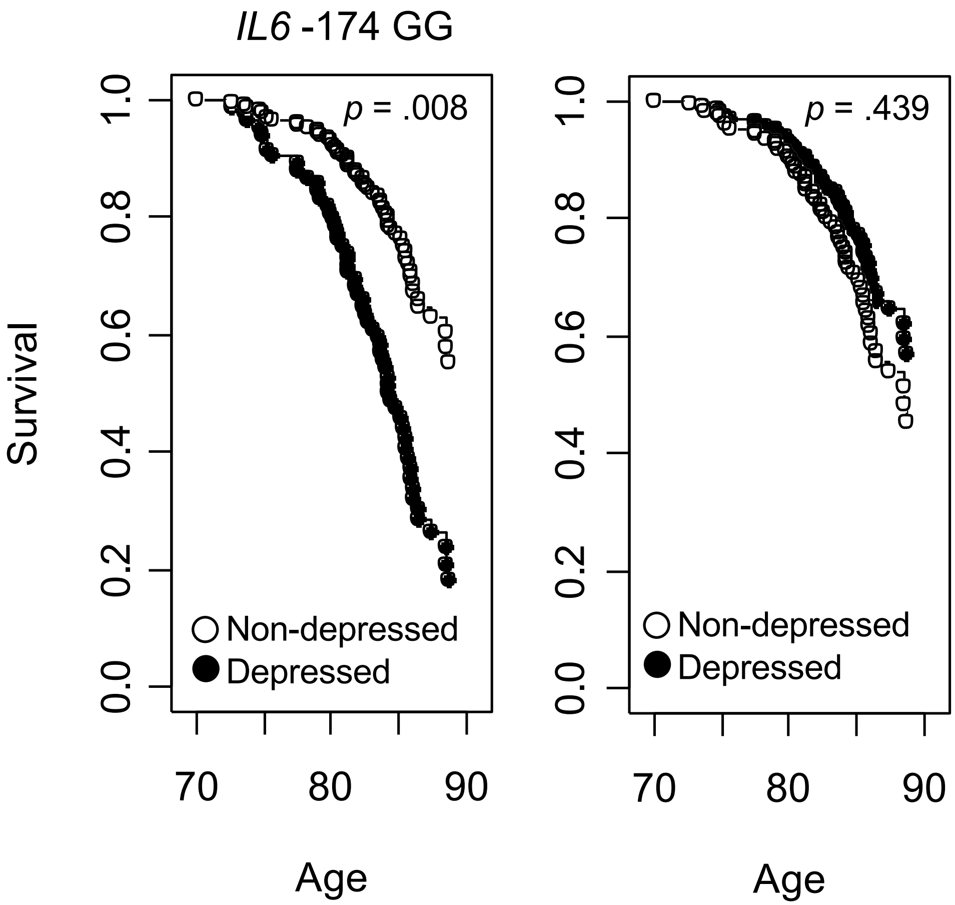


TCT **TGCGATGCTA** AAG

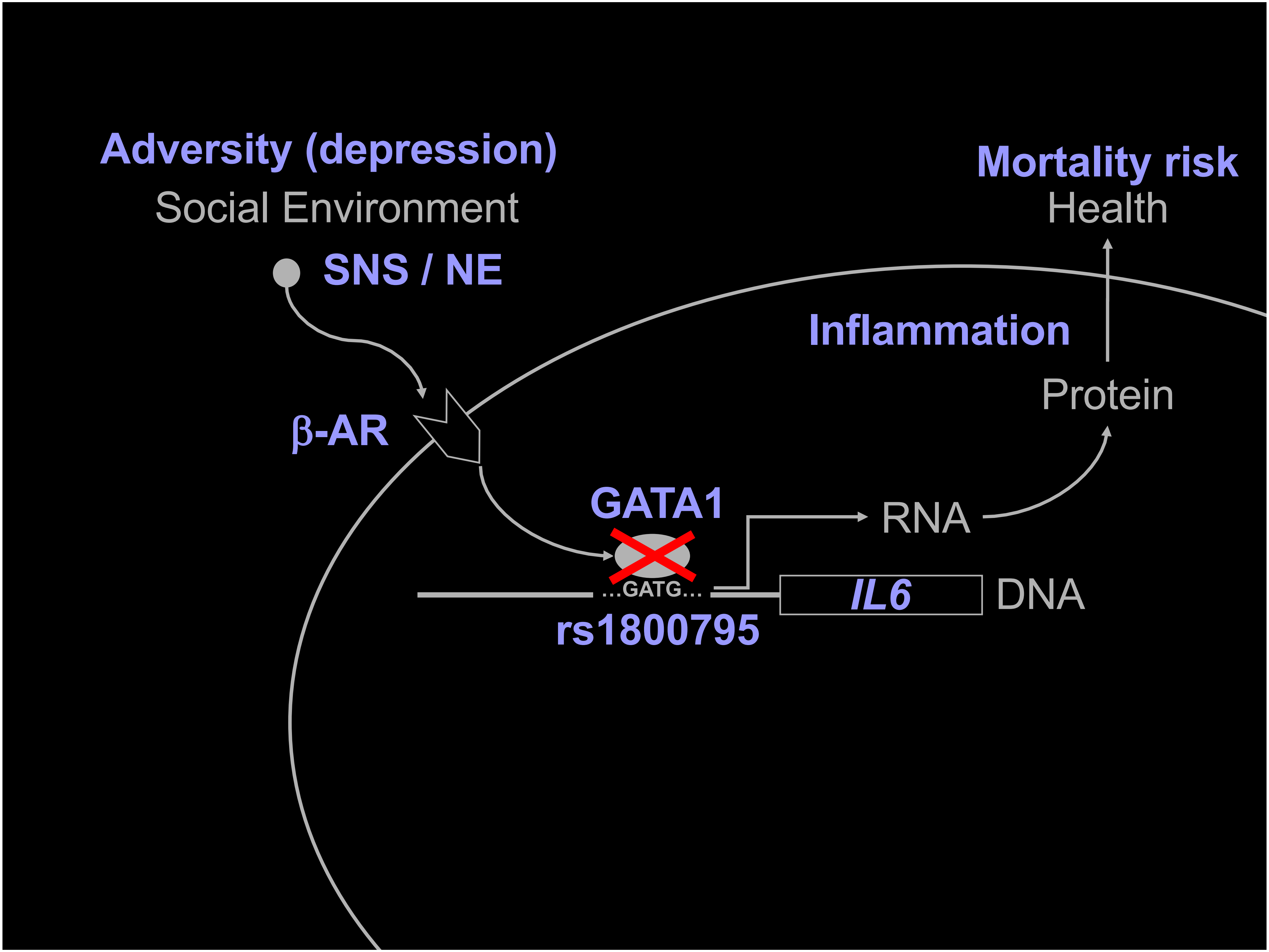
IL6

Cole et al. (2010) PNAS

Gene x Environment Interaction



Cole et al. (2010) PNAS



Adversity (depression)

Social Environment

SNS / NE

Mortality risk

Health

Inflammation

β-AR

Protein

GATA1

RNA

…GATG…

IL6

DNA

rs1800795

1205 GRE-modifying SNPs

RHCE -292
RHCE -292
RHCE -292
RHCE -292
LOC440576 -934
SOC -39
SOC -49
SOC -26
UNQ6122 -877
LAPTM5 -728
PHC2 -168
PHC2 -16
ITGB3BP -311
FLJ20331 -994
ZNF265 -663
ZNF265 -663
FUBP1 -778
LOC388650 -392
LOC388654 -957
PDE4DIP -175
COAS2 -435
LOC199882 -474
LOC440689 -692
LOC440689 -16
LOC441906 -496
FLG -17
LEP3 -631
RAB13 -310
LOC91181 -956
LOC91181 -956
LOC126669 -407
LOC440693 -399
PKLR -118
PKLR -597
FCRH1 -580
SPTA1 -163
SLAMF9 -256
KCNJ10 -383
ITLN1 -760
ITLN1 -760
F11R -798
F11R -798
LMX1A -85
SELP -144
LOC400796 -263
F13B -881
F13B -881
MYOG -951
LOC440712 -956
LGTN -331
FLJ10874 -676
GPATC2 -556
LOC440721 -625
AGT 1
FLJ10359 -367
LOC441927 -406
LOC440741 -564
MGC12466 -863
KIAA1720 -894
LOC388578 -522
LOC391205 -430
MIG-6 -618
MIG-6 -638
MIG-6 -678
LOC441870 -731
LOC440561 -255
LOC401940 -500
LOC401940 -564
LOC401940 -606
LOC339553 -400
LOC440753 -695
LOC388789 -593
FLJ38374 -686
LOC391241 -81
LOC388794 -28
C20orf70 -431
STK4 -122
PIGT -910
DNTTIP1 -479
C20orf67 -1
MMP9 -875
CEBPB -978
RNPC1 -370
RNPC1 -370
TH1L -26
TH1L -26
LOC400849 -714
LOC400849 -382
CGI-09 -309
FKHL18 -608
C20orf172 -118
TGM2 -220
TGM2 -220
LOC388798 -828
Kua-UEV -465
Kua-UEV -561
Kua -465
BTBD4 -590
C21orf99 -772
C21orf99 -13
KRTAP15-1 -566
B3GALT5 -889
B3GALT5 -889
B3GALT5 -889
B3GALT5 -889
B3GALT5 -889
LOC441955 -824
LOC441955 -824
LOC400858 -624
CLDN8 -17
KRTAP19-7 -127
DSCR1 -620
C21orf84 -232

LOC150221 -939
LOC91219 -352
LOC150236 -666
GSTT1 -141
SEC14L4 -746
SSTR3 -705
FLJ22582 -372
DIA1 -749
ATP5L2 -328
A4GALT -825
SULT4A1 -729
SULT4A1 -729
C2orf15 -882
LOC129521 -477
LOC440892 -918
IL1RL1 -332
MRPS9 -970
LOC442037 -839
IL1F7 -978
IL1F7 -978
IL1F7 -978
IL1F7 -978
MGC52000 -273
MGC52000 -466
MGC52057 -404
MAP1D -120
COL3A1 -310
SLC39A10 -921
LOC200726 -220
IL8RB -447
TUBA4 -643
FLJ25955 -24
ALPPL2 -296
UGT1A9 -651
UGT1A7 -351
UGT1A6 -224
UGT1A6 -402
TRPM8 -170
ASB1 -723
GCKR -204
LOC388938 -212
FLJ38348 -606
MSH2 -376
MSH2 -976
MSH2 -376
MSH2 -376
SBLF -59
LOC151443 -85
LOC391387 -134
SEMA4F -751
RBM29 -1
LOC339562 -621
LOC339562 -641
LOC200493 -245
TXNDC9 -714
FLJ40629 -946
LOC401005 -12
LOC389050 -170
ORC4L -16
ORC4L -16
ORC4L -16
ORC4L -16
ARL5 -895
ARL5 -895
NR4A2 -527
NR4A2 -527
NR4A2 -527
NR4A2 -527
ATP5G3 -55
ZNF533 -598
ZSWIM2 -772
PGAP1 -821
PGAP1 -827
SF3B1 -138
ORC2L -786
LOC391475 -413
CRYGC -765
PECR -942
SLC23A3 -412
LOC442070 -877
LOC129607 -488
LOC339789 -268
LOC130502 -558
ALK -710
BCL11A -615
BCL11A -615
BCL11A -615
BCL11A -615
PAP -438
PAP -438
PAP -531
CNTN4 -809
PPARG -584
PPARG -914
LOC401054 -926
GALNTL2 -427
FBXL2 -107
APRG1 -269
APRG1 -347
LOC440951 -20
LOC389123 -140
LOC285194 -808
NR1I2 -769
STXBP5L -480
LOC442092 -880
MRPS22 -897
KCNAB1 -793
LOC402146 -134
LOC90133 -2
NLGN1 -541
FLJ20522 -803
ATP2B2 -593

IDSP -819
MGC48628 -101
NDST3 -902
LOC401149 -733
LOC401038 -837
FLJ35630 -291
CYP4V2 -117
LOC401164 -978
LOC391727 -934
LOC399917 -840
ZAR1 -106
LOC401132 -18
PF4 -819
EIF4E -716
ADH7 -557
TACR3 -957
AGXT2L1 -631
PLA2G12A -795
PITX2 -411
PITX2 -411
LOC401155 -72
CDHJ -652
FGA -110
FGA -110
PPID -384
LOC441049 -368
GPM6A -203
LOC389833 -878
LOC389833 -288
LOC389833 -288
LOC389833 -878
LOC442102 -418
FGFBP1 -290
LOC441013 -188
FLJ00310 -289
FLJ00310 -881
FLJ00310 -289
FLJ00310 -289
FLJ00310 -289
FLJ00310 -289
FLJ00310 -289
LOC442127 -287
SRD5A1 -631
LOC345711 -877
LOC389281 -225
MGC42105 -669
PELO -938
BDP1 -918
DKFZp564C0469 -378
LOC134505 -63
TSLP -331
LOC340069 -755
SNCAIP -671
LOC441106 -646
SLC27A6 -484
CDC42SE2 -384
PHF15 -52
LOC389331 -27
PCDHA4 -26
PCDHA4 -26
PCDHB3 -623
PCDHB6 -212
PCDHB16 -609
ABLIM3 -474
LARP -716
LOC134541 -868
FGFR4 -472
FGFR4 -472
FGFR4 -745
FGFR4 -745
LOC442145 -7
LOC442146 -856
LOC345462 -604
LOC345462 -609
LOC442148 -595
OR2V2 -340
OR2V2 -901
TPPP -454
MYO10 -583
LOC441066 -463
GDNF -36
LOC345643 -568
FOXD1 -990
ARSB -493
DHFR -473
SPATA9 -748
CHD1 -581
STK22D -863
LOC389316 -227
CDO1 -360
FLJ33977 -166
LOC391824 -129
ALDH7A1 -920
CAMK2A -429
CAMK2A -429
C5orf4 -657
LOC345430 -332
DUSP1 -361
LOC285770 -132
NQO2 -705
MRS2L -22
HIST1H2BA -960
HIST1H2BD -597
HIST1H2BD -597
HIST1H2BH -618
HIST1H4I -283
HLA-H -477
MRPS18B -207
LOC401250 -26
LOC401250 -497
NFKBIL1 -305
LY6G5B -359
C6orf25 -413

LOC442279 -858
LOC401289 -82
LOC285766 -472
SERPINB6 -657
OFCC1 -367
LOC441129 -714
SMA3 -762
LOC222699 -719
LOC441138 -870
OR12D3 -872
LOC346171 -389
HCG4P6 -80
HCG4P6 -50
PSORS1C2 -78
PSORS1C2 -78
HLA-C -512
HLA-B -594
HLA-DRB1 -469
HLA-DRB1 -821
HLA-DQB2 0
HLA-DQB2 -333
HLA-DQB2 0
HLA-DOB -500
MLN -740
LRFN2 -452
C6orf108 -907
C6orf108 -907
PLA2G7 -227
CRISP1 -236
CRISP1 -236
IL17F -733
HMGCLL1 -759
LOC442226 -67
C6orf66 -832
DJ467N11.1 -34
RTN4IP1 -207
SLC22A16 -869
LOC442254 -307
DEADC1 -509
FLJ44955 -391
SYNE1 -484
SYNE1 -126
LOC389435 -451
LOC389435 -565
PIP3-E -457
T -9
T -3
LOC442280 -112
DKFZp434J154 -615
LOC401303 -632
LOC441198 -739
GHRHR -646
ADCYAP1R1 -60
C7orf16 -842
LOC441209 -41
GPR154 -435
GPR154 -435
C7orf36 -707
BLVRA -400
BLVRA -400
LOC51619 -311
WBSCR19 -38
LOC136288 -523
LOC392030 -632
FZD9 -485
LOC85865 -255
LOC442341 -390
AKR1D1 -159
LOC93432 -126
OR2F1 -160
OR2A5 -927
LOC441184 -336
LOC441186 -584
LOC441187 -654
LOC389831 -914
LOC389831 -914
LOC222967 -338
LOC222967 -338
LOC340267 -244
ICA1 -699
AGR2 -65
LOC389472 -184
LOC401316 -837
CRHR2 -610
PDE1C -20
LOC441210 -361
LOC222052 -77
LOC441224 -287
LOC441230 -143
LOC441245 -127
LOC441259 -954
CCL26 -441
SEMA3C -385
C7orf23 -761
PON1 -785
GATS -36
ACHE -715
ACHE -224
ACHE -715
ACHE -224
ORC5L -990
ORC5L -990
CHCHD3 -793
MGC5242 -861
LOC392997 -596
LOC392997 -596
FLJ44186 -168
HIPK2 -70
ZC3HDC1 -407
LOC402301 -14
BAGE4 -100
BAGE4 -648
MCPH1 -520

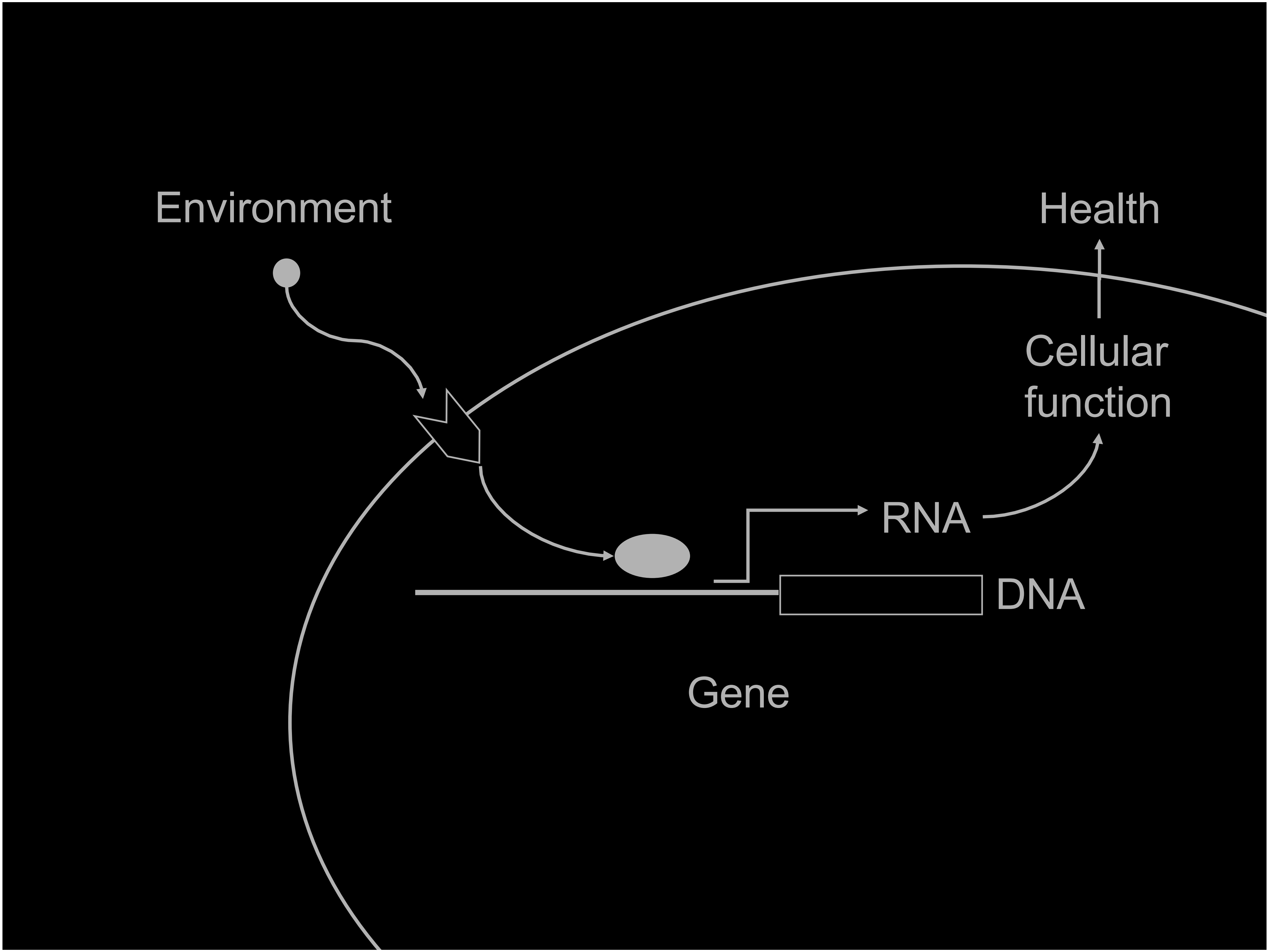
SPAG11 -622
SPAG11 -622
SPAG11 -971
DEFB104 -132
LOC389633 -370
ASAH1 -702
ASAH1 -882
FLJ22494 -242
FLJ22494 -781
SNAI2 -728
CPA6 -613
FSBP -393
[illegible]
[illegible]
LOC442399 -126
TOP1MT -477
LOC286126 -887
LOC340393 -922
DOCK8 -109
LOC441386 -327
C9orf93 -708
SH3GL2 -702
C9orf98 -376
LOC340501 -32
LOC441417 -394
DKFZP434M131 -944
SECISBP2 -404
LOC441453 -821
PHF2 -646
PHF2 -648
LOC441457 -742
LOC441457 -802
PRG-3 -971
RAD23B -998
SLC31A2 -380
OR1N2 -646
C9orf54 -2
C9orf54 -2
LAMC3 -895
LOC441473 -825
LOC441473 -825
LOC441473 -825
DBH -768
OBP2A -732
EGFL7 -330
EGFL7 -335
TRAF2 -32
LOC441408 -394
LOC389702 -288
C9orf46 -353
SLC24A2 -265
IFNA10 -138
IFNA14 -85
C9orf11 -311
C9orf24 -905
C9orf24 -905
UNQ470 -31
STOML2 -420
LOC392334 -904
LOC286327 -215
HNRPK -86
LOC441452 -955
DIRAS2 -896
LOC286359 -774
TXNDC4 -690
TXN -239
OR1L8 -459
DYT1 -561
ABO -790
ABO -789
ABO -790
XPMC2H -374
LOC4414474 -921
LOC389734 -489
LOC389734 -223
FCN1 -673
FCN1 -709
LOC441410 -990
GAGE1 -21
RRAGB -788
RRAGB -788
LOC340527 -194
SH3BGRL -944
DIAPH2 -921
DIAPH2 -921
HSU24186 -145
NXF2 -89
PLP1 -918
PLP1 -918
LOC286436 -713
SLC6A14 -962
LOC392529 -73
FLJ25735 -992
MAGEB4 -834
MAGEB4 -834
LOC389844 -822
LOC389844 -814
UBE1 -964
LOC203604 -16
LOC441481 -796
DMD -923
RPGR 3
ZNF21 -828
PRKY -308
LOC441537 -223
LOC441539 -222
LOC441535 -225
LOC441536 -223
LOC338588 -51
UCN3 -368
NET1 -14
MAPK8 -856

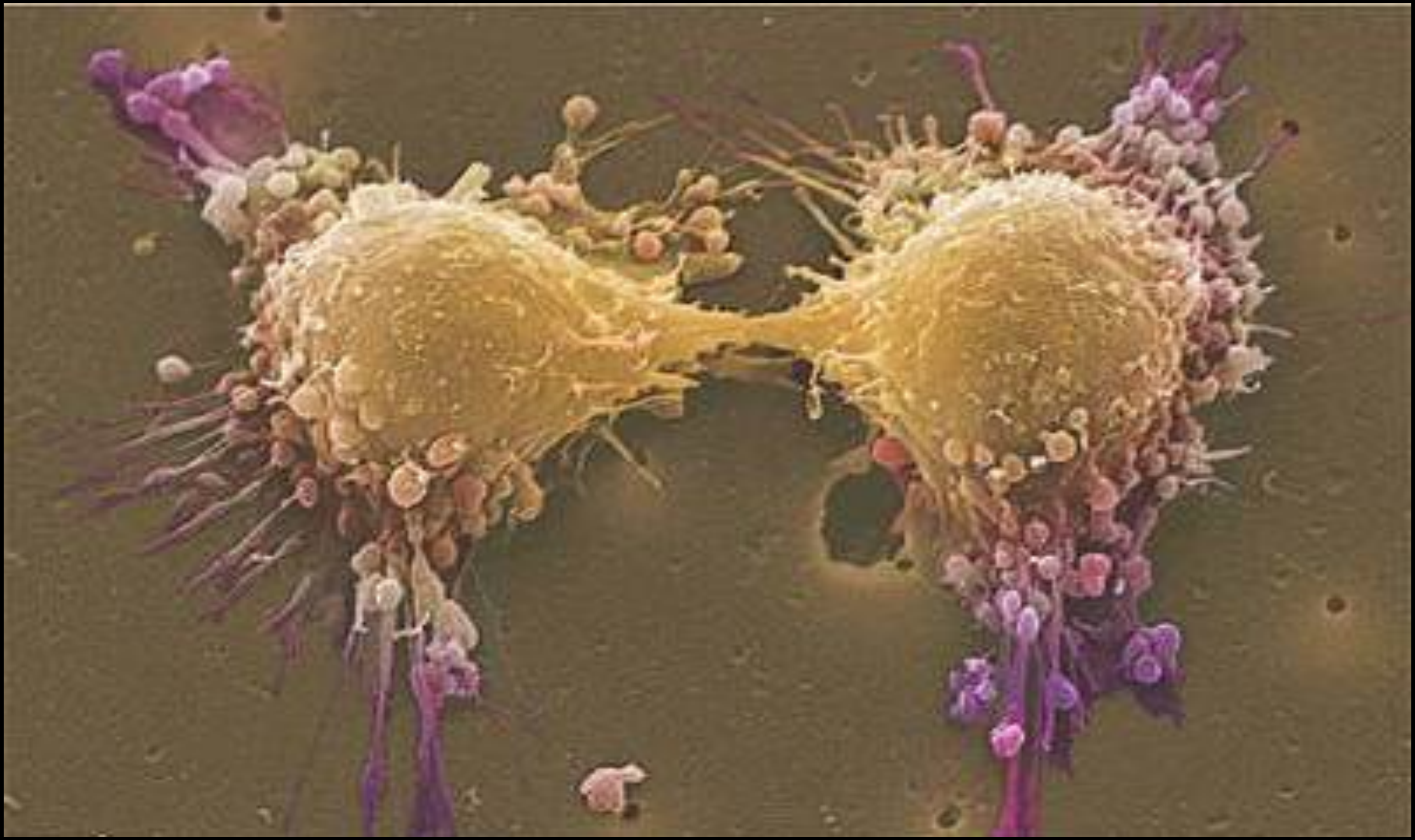
FANK1 3
TAF3 -544
LOC441547 9
LOC220998 -941
TPRT -277
C10orf68 -817
C10orf9 -269
ZNF33A -477
ZNF33A -477
LOC399744 -202
LOC399744 -202
PPYR1 -8[illegible]
PPYR1 -8[illegible]
LOC3994[illegible]
AKR1C2 -641
AKR1C2 -641
LOC441560 -504
LOC439975 -618
NEUROG3 6
AMID -452
PPP3CB -854
LOC439983 -240
LOC389988 -68
MMS19L -221
C10orf69 -121
GPR10 -555
C10orf93 -42
ASB13 -506
IL15RA -222
IL15RA -827
USP6NL -573
C10orf45 -181
NMT2 -912
SIAT8F -676
NEBL -727
C10orf52 -163
LOC439953 -879
LOC399737 -608
CTGLF1 -504
LOC439963 -500
KCNQ1 -40
LOC387746 -61
OR51F2 -640
TRIM34 -105
OR10A2 -851
SAA1 -721
SAA1 -722
LOC441593 -126
PDHX -845
TRIM44 -24
LOC90139 -660
NDUFS3 -929
LOC196346 -885
OR5T3 -97
CTNND1 -133
CTNND1 -116
CNTF -149
ROM1 -515
MARK2 -375
MARK2 -375
RAB1B -75
GSTP1 -841
GSTP1 -841
LOC440056 -824
USP35 -148
LOC390231 -471
OR4D5 -465
OR8G5 -809
MGC39545 -867
LOC399969 -328
LOC219797 -216
NUP98 -651
NUP98 -651
NUP98 -651
NUP98 -651
KIAA0409 -533
LOC283299 -427
LOC440026 -69
LOC440030 -675
LOC387754 -159
LOC144100 -631
HPS5 -917
HPS5 -917
HPS5 -917
LOC387764 -149
LOC440041 -221
FLJ31393 -362
OR8H1 -161
AGTRL1 -809
PRG2 -899
TCN1 -716
RAB3IL1 -976
KIAA0404 -771
CHRDL2 -754
KCTD14 -94
MRE11A -879
MRE11A -982
MMP7 -853
CRYAB -175
ZNF202 -527
LOC387820 -553
LOC387823 -178
CCND2 -350
NDUFA9 -485
KCNA5 -805
FLJ10665 -245
FLJ10665 -576
LOC285407 -743
LOC390299 -771
FLJ10652 -491
LOC144245 -455
PFKM -838

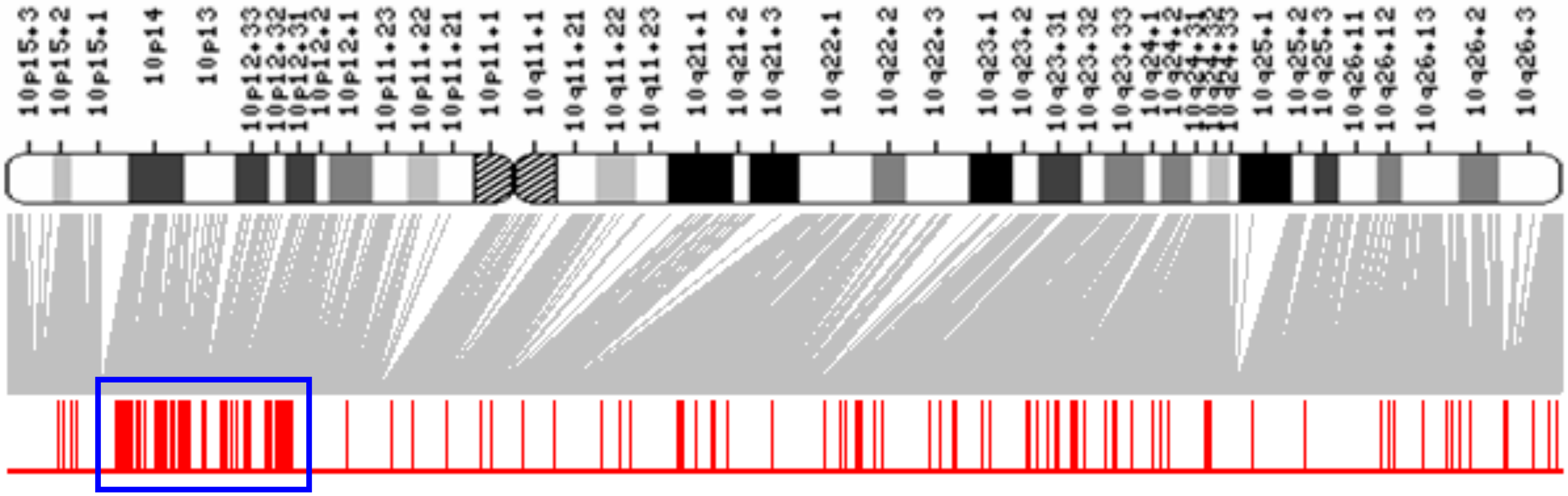
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
CLECSF12 -885
KLRK1 -349
PRB1 -589
PRB1 -589
PRB1 -589
ADAMTS20 -965
ADAMTS20 -965
SLC38A2 -638
K-ALPHA-1 -27
KIAA1602 -262
RACGAP1 -620
K6IRS3 -708
KRT4 -83
NPFF -777
STAT2 -94
FLJ32949 -500
IFNG -795
MGC26598 -498
HAL -358
DKFZp434M0331 -920
LOC400070 -223
TSC -785
GPR109B -392
EPIM -568
EPIM -568
GALNT9 -798
LOC440122 -169
LOC221140 -342
LOC440128 -877
LOC387912 -279
LOC341784 -327
NURIT -947
RB1 -525
DKFZP434K1172 -595
DKFZP434K1172 -595
LOC144983 -906
LOC144983 -892
LOC144983 -896
LOC400144 -807
PROZ -865
PROZ -865
CRYL1 -768
POSTN -32
LOC440134 -367
EBPL -973
GUCY1B2 -832
LOC338862 -918
LOC404785 -818
OR11H6 -269
C14orf92 -234
PSMA6 -219
KTN1 -222
C14orf166B -786
EVL -28
CCNB1IP1 -868
CCNB1IP1 -868
NEDD8 -143
BAZ1A -508
BAZ1A -508
NFKBIA -963
LOC283551 -302
CDKL1 -902
LOC400214 -138
RTN1 -974
LOC390488 -457
PLEK2 -465
PIGH -153
RDH11 -251
FLJ39779 -161
KIAA1509 -179
SERPINA2 -559
SERPINA2 -559
SERPINA2 -559
SERPINA9 -856
LOC390529 -204
LOC388073 -112
LOC400307 -332
LOC283694 -71
LOC400320 -443
FLJ35785 -414
LOC440249 -92
HH114 -991
PLA2G4B -483
CAPN3 -318
CAPN3 -318
CAPN3 -318
LOC400368 -320
SLC28A2 -275
DUT -32
SCG3 -739
LIPC -853
OSTbeta -781
LOC440289 -446
COMMD4 -790
LOC400433 -496
LOC390637 -55
FLJ11175 -113
LOC440224 -815
LOC283804 -112
CHSY1 -876
LOC440315 -303
LOC440315 -303
LOC400470 -62
LOC388076 -715

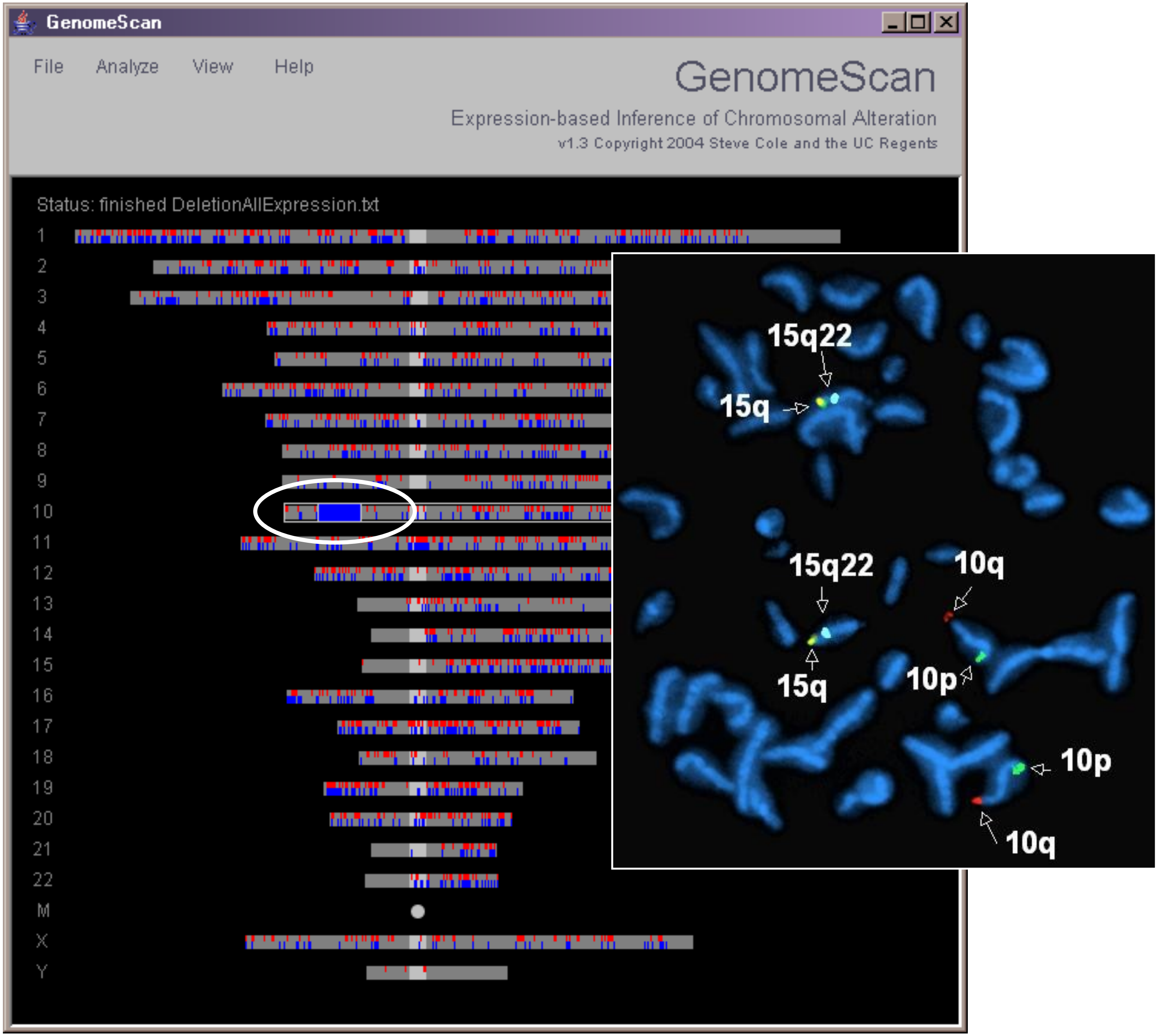
TNFRSF12A -968
DNAJA3 -24
ALG1 -464
ALG1 -464
FLJ12363 -773
LOC92017 -711
TMC7 -412
MGC16824 -271
RBBP6 -795
RBBP6 -795
RBBP6 -795
ITGAX -504
ERAF -510
LOC388248 -649
FLJ38101 -981
CES4 -221
MT1H -280
GAN -839
PLCG2 -534
CDH13 -906
HSBP1 -425
MLYCD -917
FLJ45121 -772
DPEP1 -765
FLJ32252 -288
FLJ32252 -346
MGC35212 -360
FLJ25410 -280
LOC400506 -715
LOC94431 -77
DOC2A -265
LOC441761 -889
LOC57019 -375
ZNF319 -360
DNCLI2 -857
DNCLI2 -857
DKFZP434A1319 -236
LOC439920 -70
CHST5 -601
CHST5 -756
LOC390748 -242
DPH2L1 -42
LOC388323 -892
MAP2K4 -128
MAP2K4 -128
KRTAP4-12 -78
JJAZ1 -789
CCL2 -912
PSMB3 -889
LOC440440 -1
FLJ25168 -244
SP2 -57
LOC388406 -800
TBX4 -465
DDX42 -212
DDX42 -212
LOC90799 -734
DKFZP586L0724 -829
SSTR2 -874
MRPS7 -822
MRPS7 -719
LOC388429 -804
NARF -669
NARF -669
GEMIN4 -911
OR1D2 -376
ALOX15 -267
SLC16A11 -346
CLECSF14 -596
CLECSF14 -640
FLJ40217 -393
RCV1 -761
CDRT1 -618
NOS2A -287
NOS2A -287
KRT25D -828
KRT12 -585
HUMGT198A -797
HUMGT198A -690
FLJ31222 -769
LOC284058 -524
GIP -957
LOC400619 -823
UNC13D -695
LOC339162 -685
LOC388462 -43
SEH1L -801
LOC284232 -988
LOC284232 -845
CABLES1 -281
CABYR -908
CABYR -908
CABYR -908
CABYR -908
CABYR -908
DSG3 -367
SLC14A1 -333
DCC -386
RAB27B -713
ZCCHC2 -249
LOC342808 -306
LOC284276 -397
MYOM1 -232
MC2R -113
LOC441817 -600
KIAA1632 -405
FBXO15 -123
FBXO15 -192
LOC390865 -489
TXNL4 -33
CDC34 -270
GZMM -678
C19orf21 -573

PSMC4 -215
PSMC4 -215
EGLN2 -452
LOC388549 -412
SYNGR4 -825
RPL13A -816
LOC402665 -925
FLJ46385 -176
LOC91661 -13
LAIR2 -705
LAIR2 -705
KIR2DL1 -763
KIR3DL2 3
ZNF583 -867
ZNF71 -861
MGC4728 -490
ZNF211 -76
ZNF211 -76
LOC401895 -957
APBA3 -13
FUT5 -174
TNFSF7 8
SH2D3A -273
8D6A -950
EIF3S4 -547
RAB3D -852
MGC20983 -338
MGC20983 -338
MGC20983 -338
NDUFB7 -741
LOC339377 -660
IL12RB1 -56
IL12RB1 -56
IL12RB1 -56
IL12RB1 -56
LOC148198 -361
CEBPA -564
UNQ467 -521
FLJ22573 -941
CLC -823
DYRK1B -849
DYRK1B -849
DYRK1B -849
PSG11 -297
PSG11 -297
PSG4 -299
PSG4 -299
PSG9 -435
FLJ34222 -415
ERCC2 -123
DMPK -988
PGLYRP1 -212
LIG1 -806
FLJ32926 -288
CGB8 -202
TEAD2 -546
FLJ20643 -895
LOC400712 -236
SIGLEC6 -972
SIGLEC6 -972
SIGLEC6 -972
ZNF577 -582
ZNF611 -148
ZNF600 -716
ZNF600 -37
NALP9 -489
PRDM2 -762
PRDM2 -762
LOC400743 -400
PADI1 -598
FLJ44952 -494
DJ462O23.2 -973
PPP1R8 5
PPP1R8 5
PPP1R8 5
ATPIF1 -766
ATPIF1 -766
ATPIF1 -766
LOC400581 -793
CGI-94 -384
FLJ14351 -753
UROD -715
LOC441885 -810
DKFZp761D221 -478
DKFZp761D221 -221
IL23R -322
CTH -6
CTH -6
AK5 -966
DNAJB4 -987
CDC7 -604
LOC388649 -426
DCLRE1B -406
LOC440610 -739
LOC440610 -584
LOC440610 -652
LOC441903 -538
LOC440673 -482
BNIPL -420
BNIPL -419
SPRR1B -826
SPRR1B -826
IL6R -110
IL6R -110
CKS1B -983
SYT11 -785
PMF1 -223
LOC164118 -75
FY -397
NCSTN -809
HSPA6 -839
HSPA6 -611


Environment
Health
Cellular function
RNA
DNA
Gene




10p15.3
10p15.2
10p15.1
10p14
10p13
10p12.33
10p12.32
10p12.31
10p12.2
10p12.1
10p11.23
10p11.22
10p11.21
10p11.1
10q11.1
10q11.21
10q11.22
10q11.23
10q21.1
10q21.2
10q21.3
10q22.1
10q22.2
10q22.3
10q23.1
10q23.2
10q23.31
10q23.32
10q23.33
10q24.1
10q24.2
10q24.31
10q24.32
10q24.33
10q25.1
10q25.2
10q25.3
10q26.11
10q26.12
10q26.13
10q26.2
10q26.3



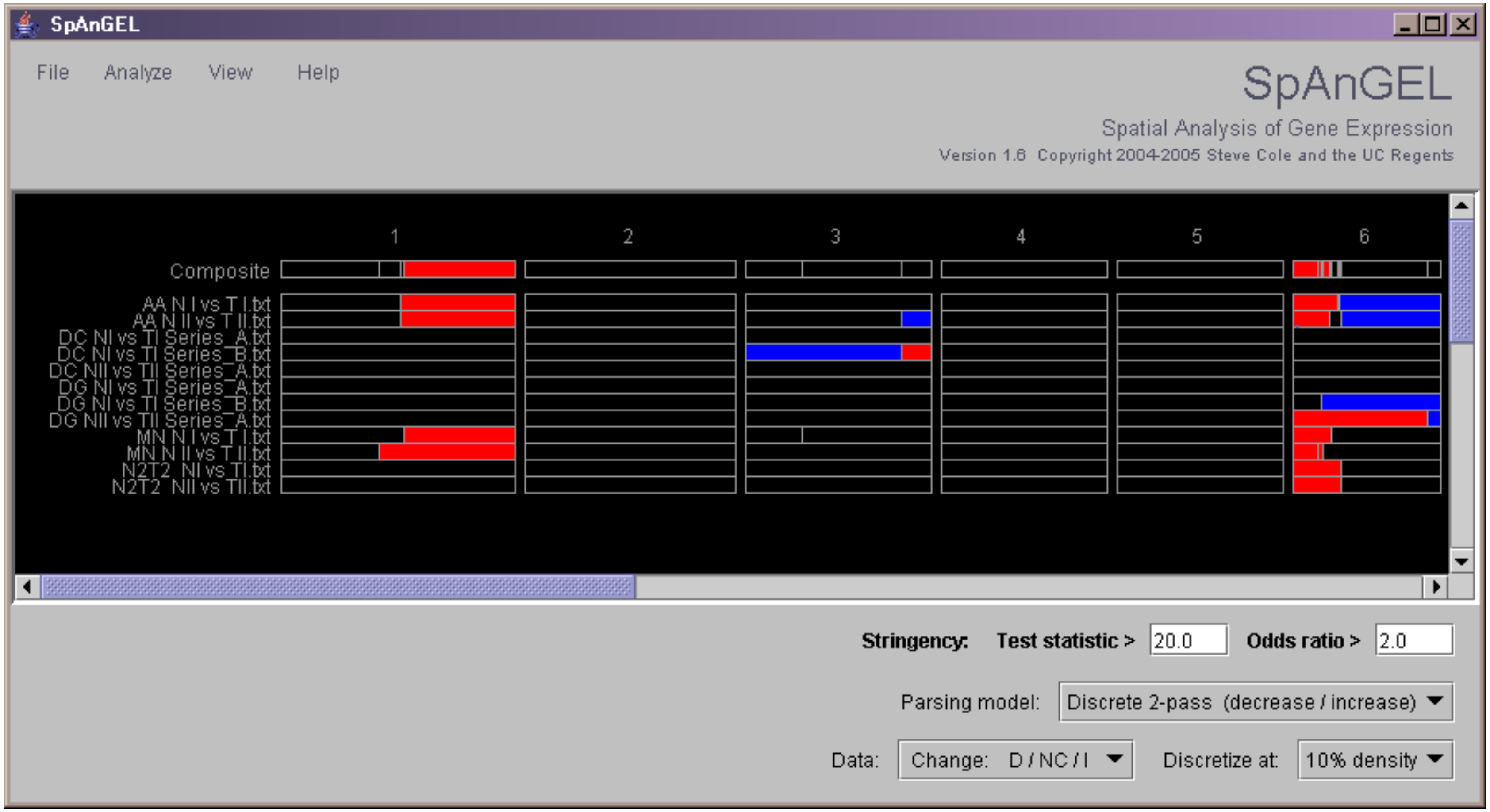
GenomeScan

Expression-based Inference of Chromosomal Alteration

v1.3 Copyright 2004 Steve Cole and the UC Regents

Status: finished DeletionAllExpression.txt

Chromosomal aberration profiles

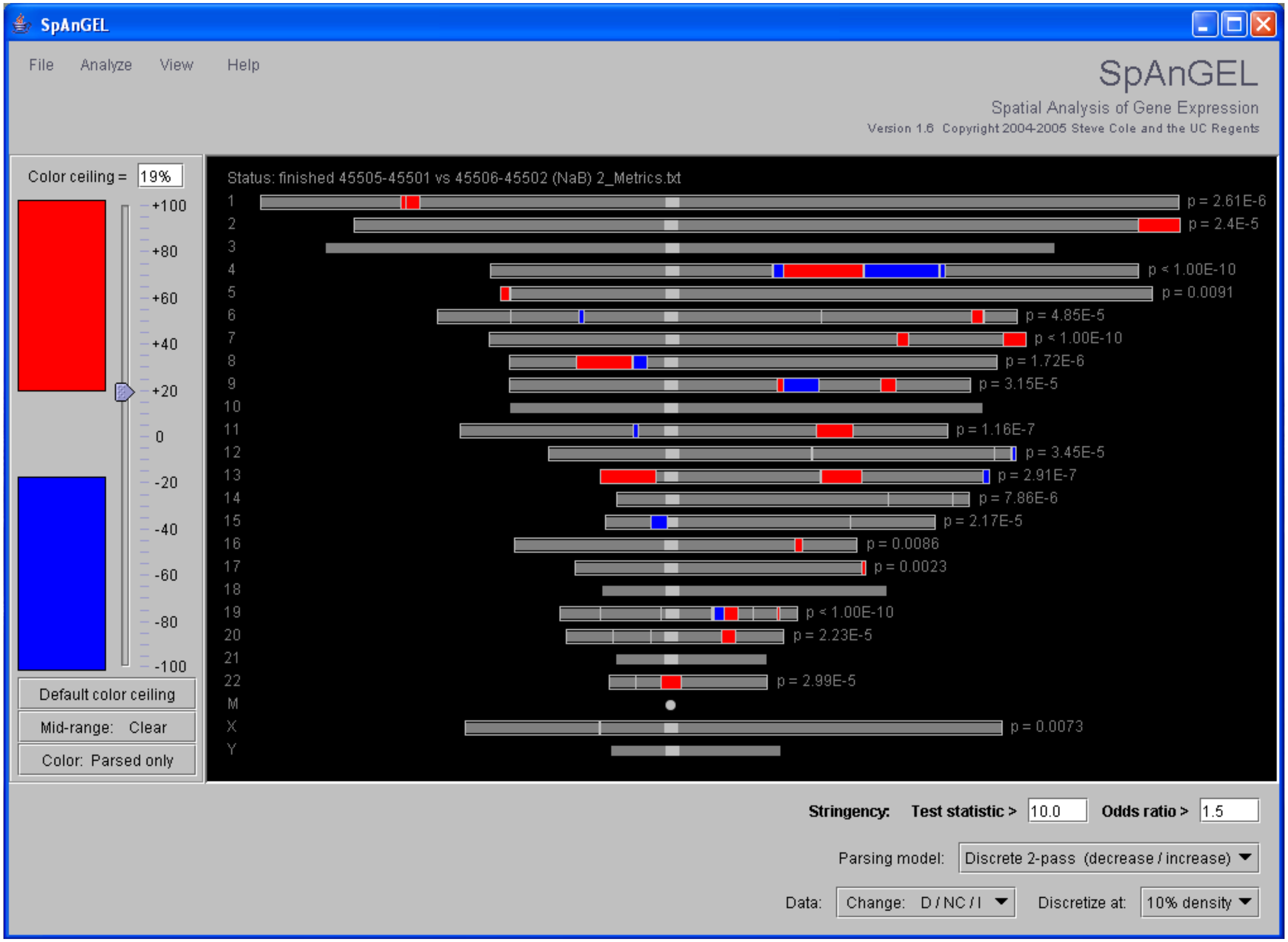


Stringency: Test statistic > 20.0 Odds ratio > 2.0

Parsing model: Discrete 2-pass (decrease / increase)

Data: Change: D / NC / I Discretize at: 10% density

Epigenetics, chromatin accessibility, & RIDGEs



Color ceiling = 19%

Mid-range: Clear

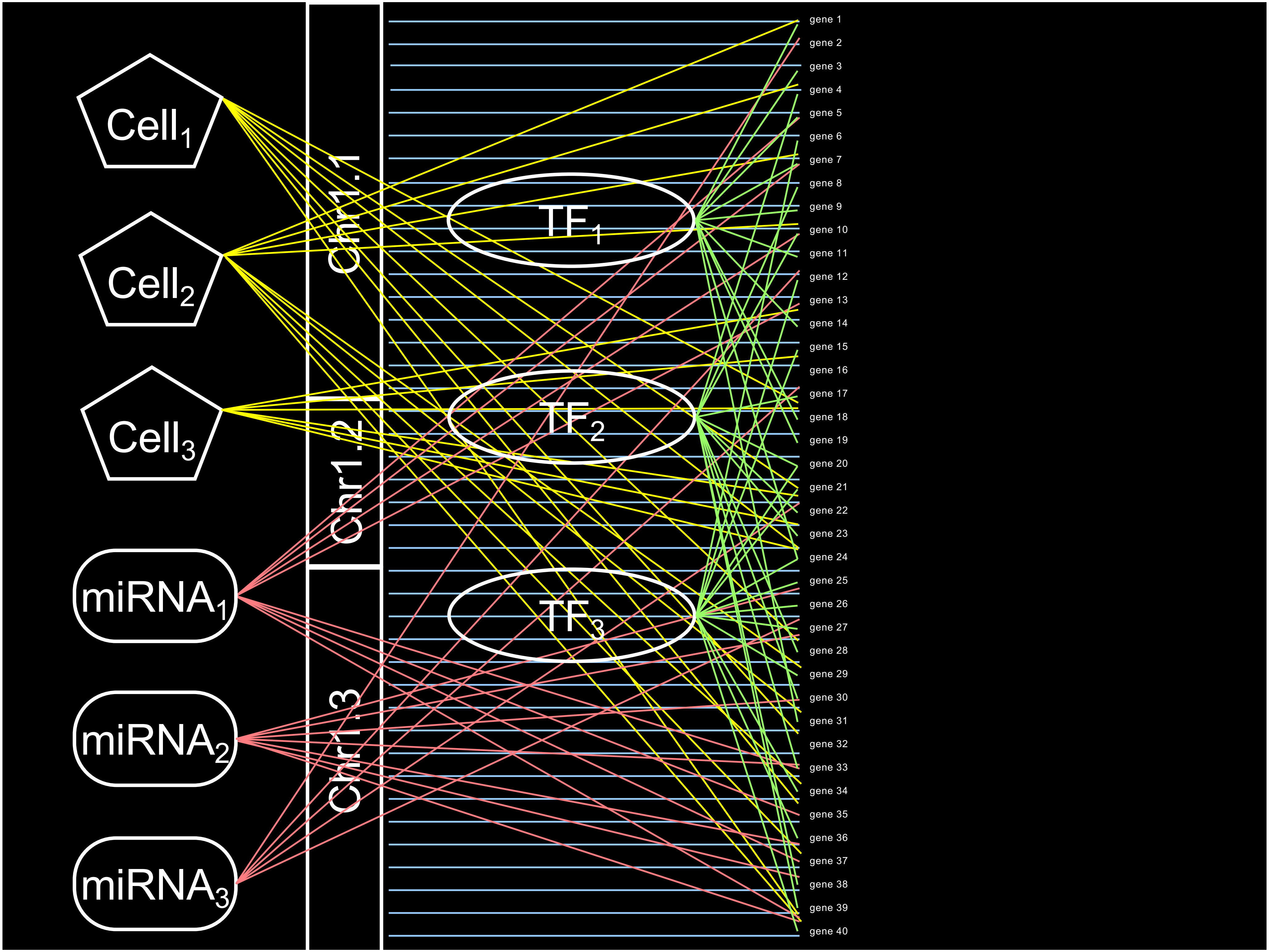
Color: Parsed only

Status: finished 45505-45501 vs 45506-45502 (NaB) 2_Metrics.txt

Stringency: **Test statistic >** 10.0 **Odds ratio >** 1.5

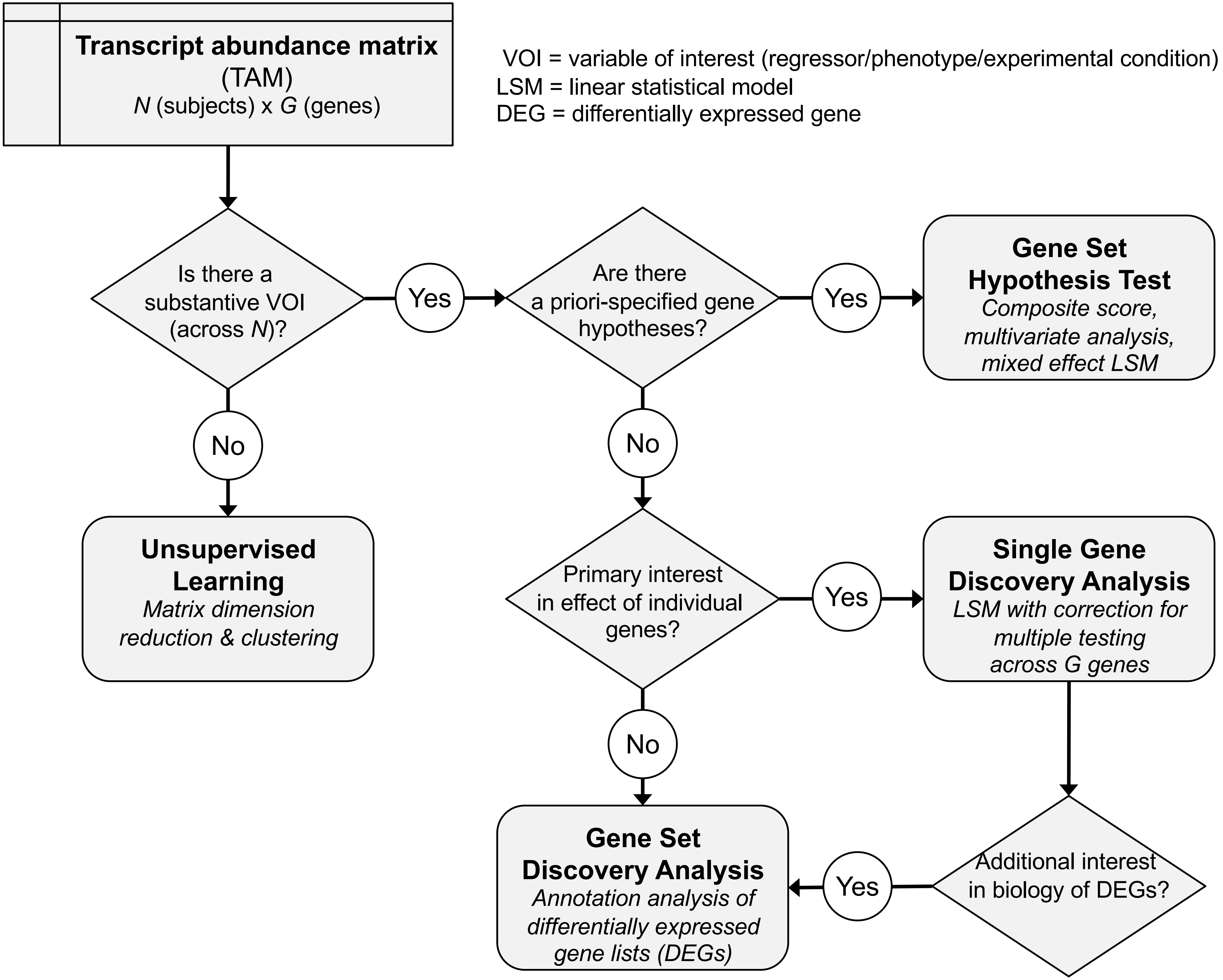
Parsing model: Discrete 2-pass (decrease / increase)

Data: Change: D / NC / I Discretize at: 10% density



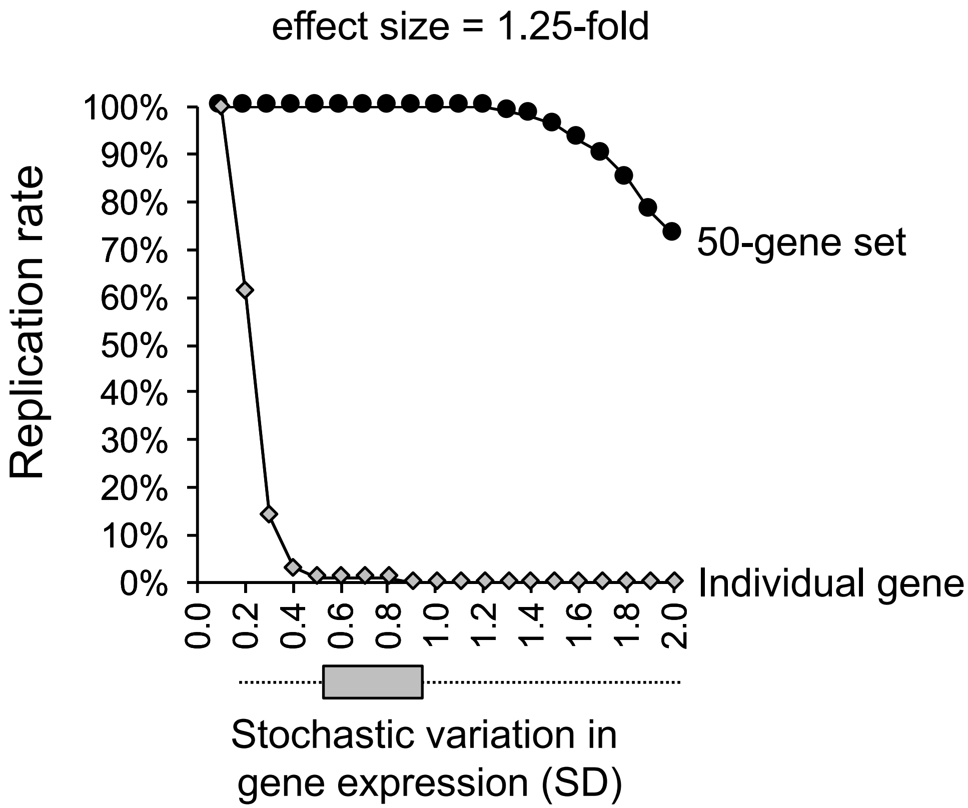
Cell$_1$
Cell$_2$
Cell$_3$
miRNA$_1$
miRNA$_2$
miRNA$_3$
Chr1.1
Chr1.2
Chr1.3
TF$_1$
TF$_2$
TF$_3$
gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40

Some notes on statistical testing for gene list analysis


Transcript abundance matrix
(TAM)
N (subjects) x G (genes)
VOI = variable of interest (regressor/phenotype/experimental condition)
LSM = linear statistical model
DEG = differentially expressed gene
Is there a substantive VOI (across N)?
Yes
No
Unsupervised Learning
Matrix dimension reduction & clustering
Are there a priori-specified gene hypotheses?
Yes
Gene Set Hypothesis Test
Composite score, multivariate analysis, mixed effect LSM
No
Primary interest in effect of individual genes?
Yes
Single Gene Discovery Analysis
LSM with correction for multiple testing across G genes
No
Additional interest in biology of DEGs?
Yes
Gene Set Discovery Analysis
Annotation analysis of differentially expressed gene lists (DEGs)

Painful lesson #1: If you want replicable results, focus on functionally related ***sets*** of genes (not individual gene transcripts)

Differential expression of an a priori gene set

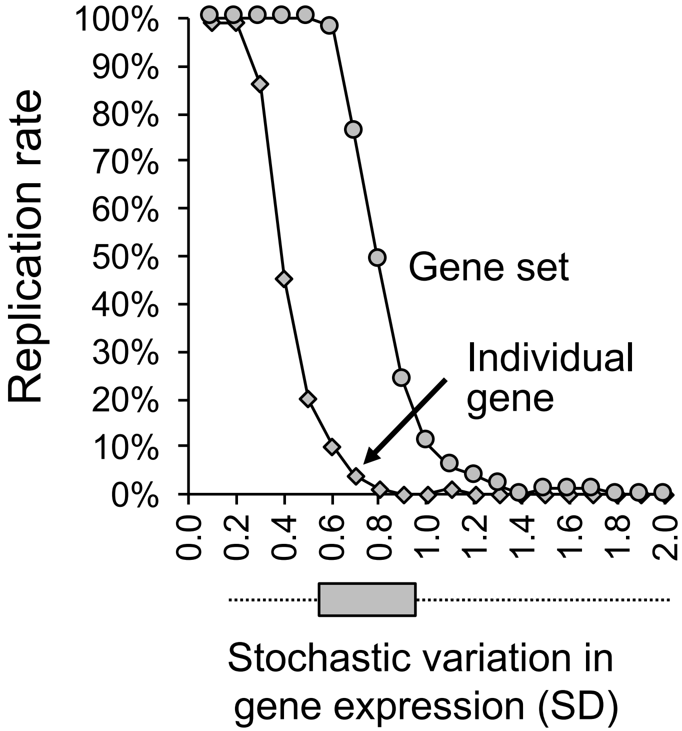


Fredrickson et al (2013) PNAS - SI

Painful lesson #2: If you want replicable results, conduct 2nd stage bioinformatics on gene lists derived from point estimates of ***effect size*** (not *p*/*q*-values)

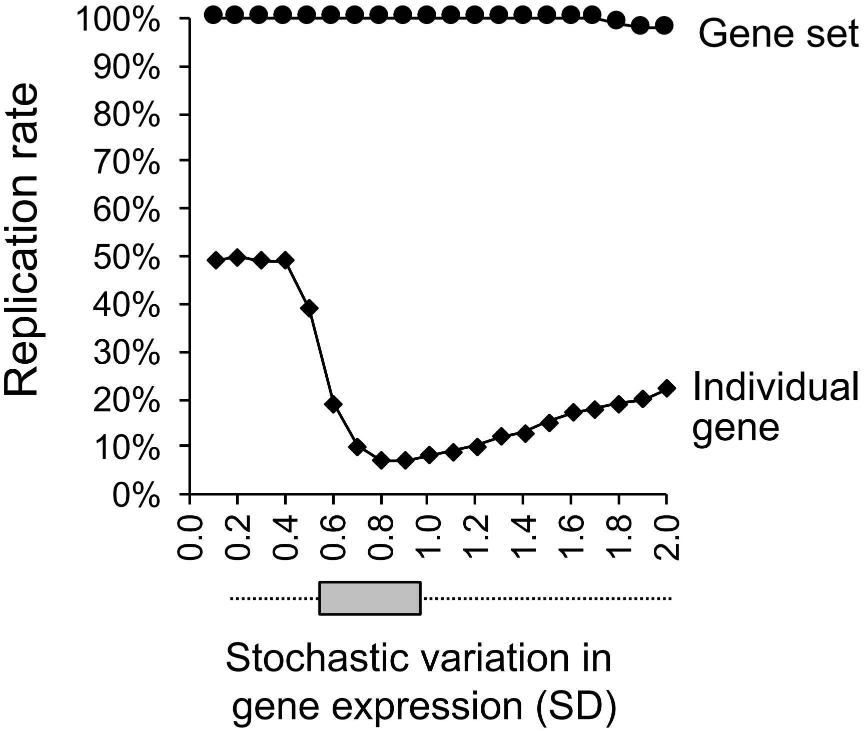
Gene set enrichment analysis

Low level input: *p*-value gene list

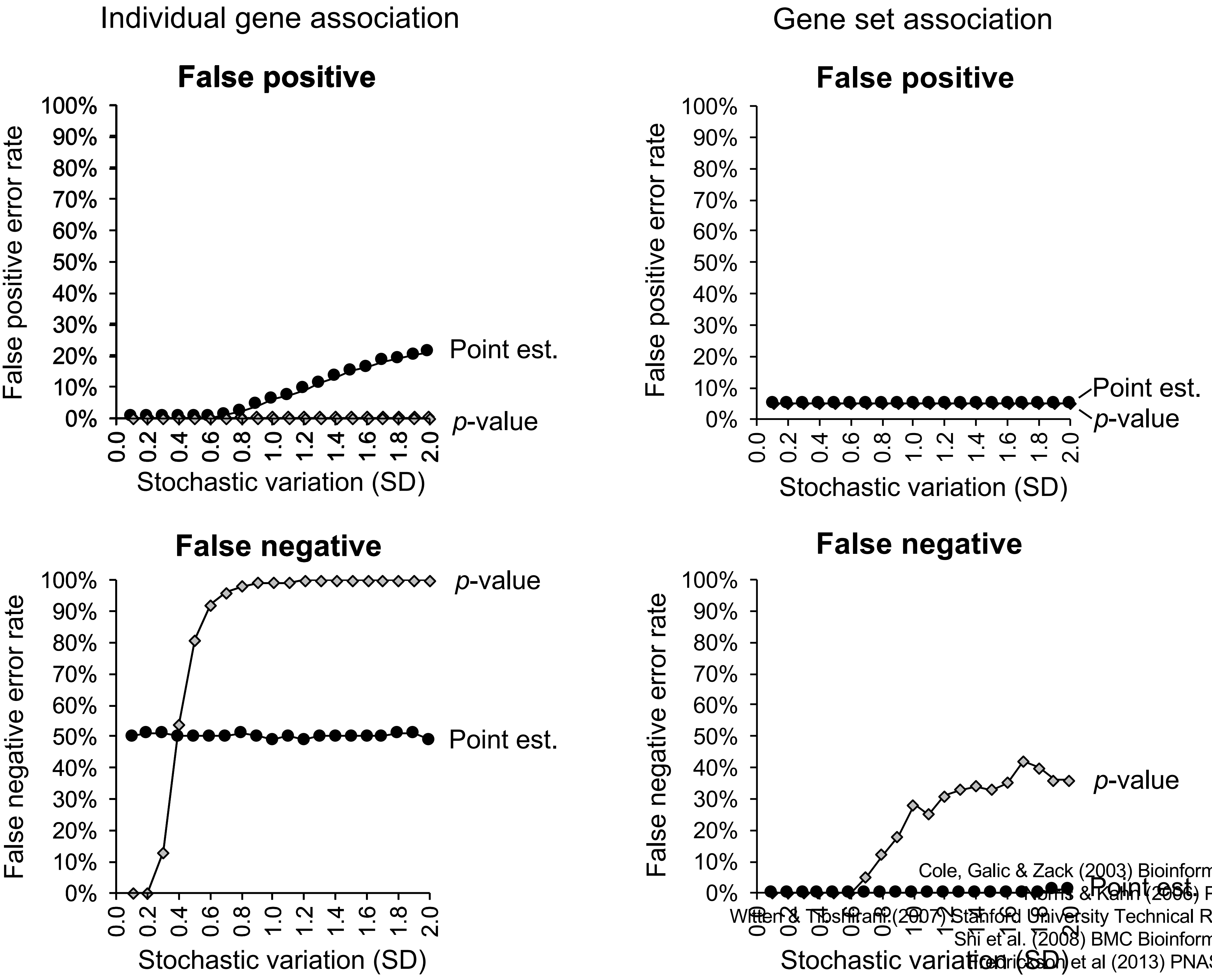


Gene set enrichment analysis

Low level input: point estimate gene list

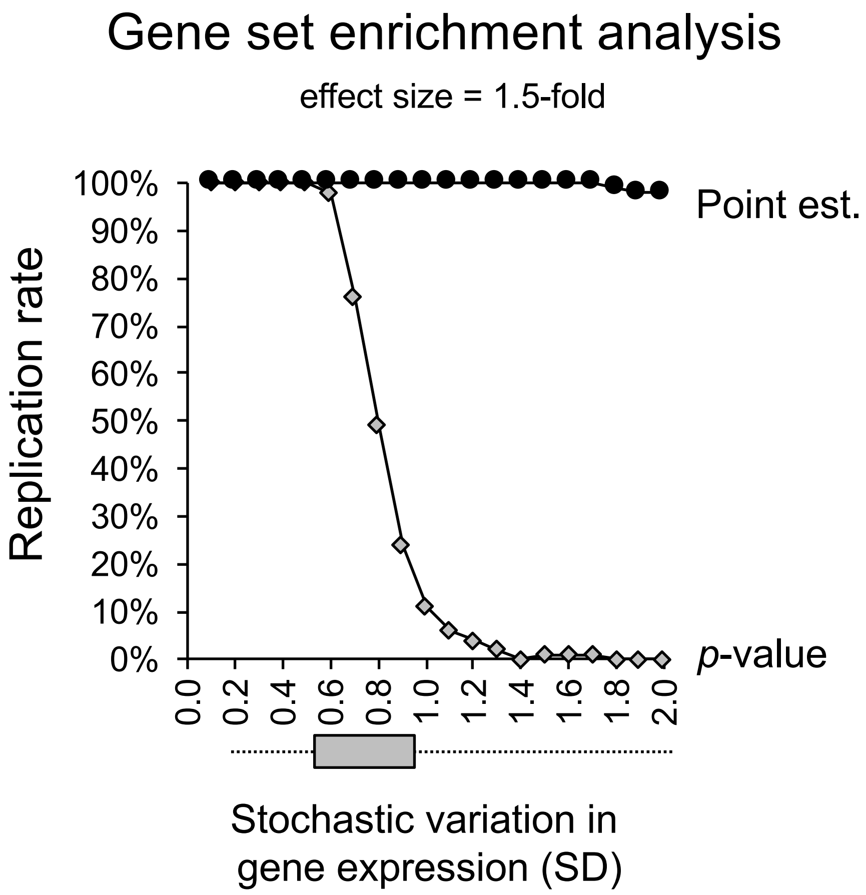


Fredrickson et al (2013) PNAS - SI



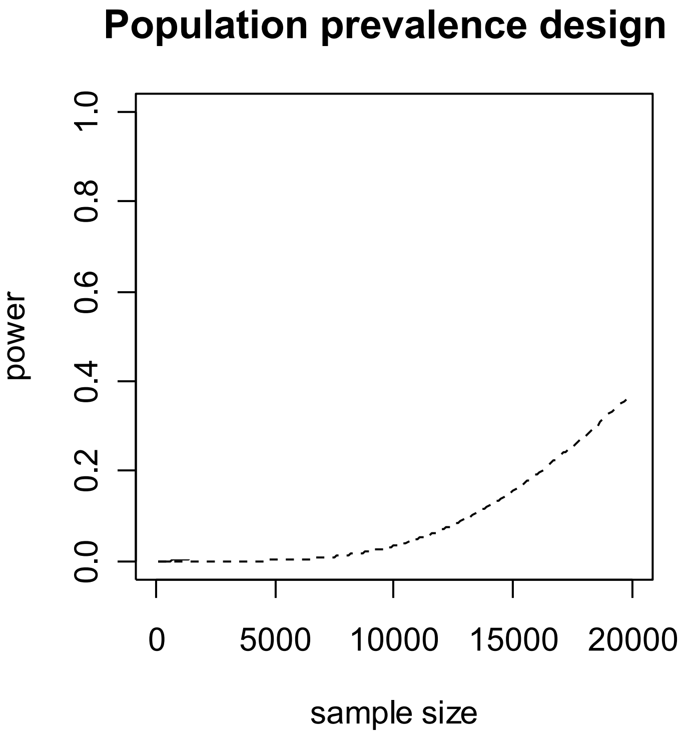
Cole, Galic & Zack (2003) Bioinformatics
Norris & Kahn (2006) PNAS
Witten & Tibshirani (2007) Stanford University Technical Report
Shi et al. (2008) BMC Bioinformatics
Fredriksson et al (2013) PNAS - SI

Implication of false negative errors for result replication



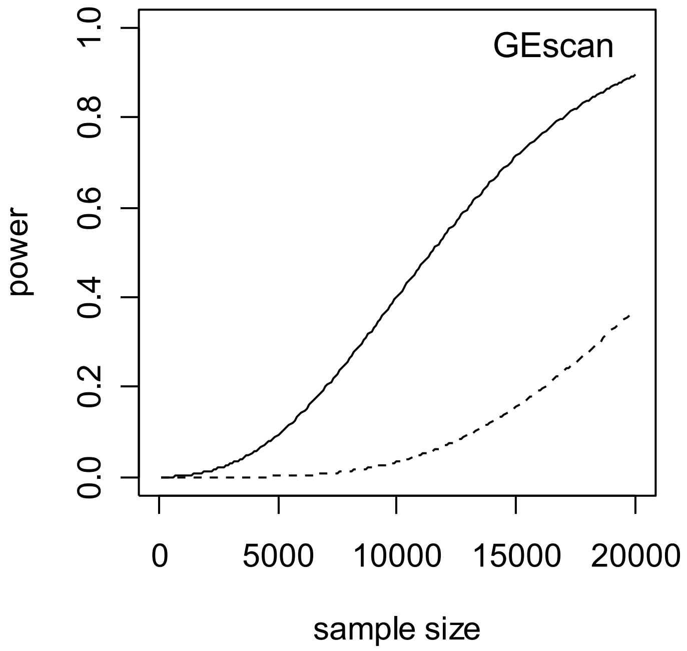
Cole, Galic & Zack (2003) Bioinformatics
Norris & Kahn (2006) PNAS
Witten & Tibshirani (2007) Stanford University Technical Report
Shi et al (2008) BMC Bioinformatics
Fredrickson et al (2013) PNAS - SI

Painful lesson #3: Hypotheses are ***extremely*** valuable for statistical power in genomics. Don't pretend you know nothing (unless you really do).

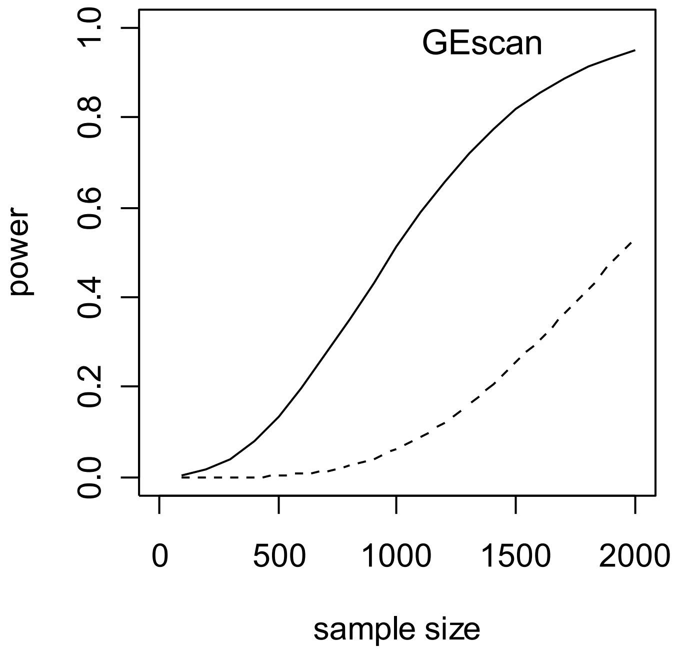


Painful lesson #3: Hypotheses are ***extremely*** valuable for statistical power in genomics. Don’t pretend you know nothing (unless you really do).

Population prevalence design

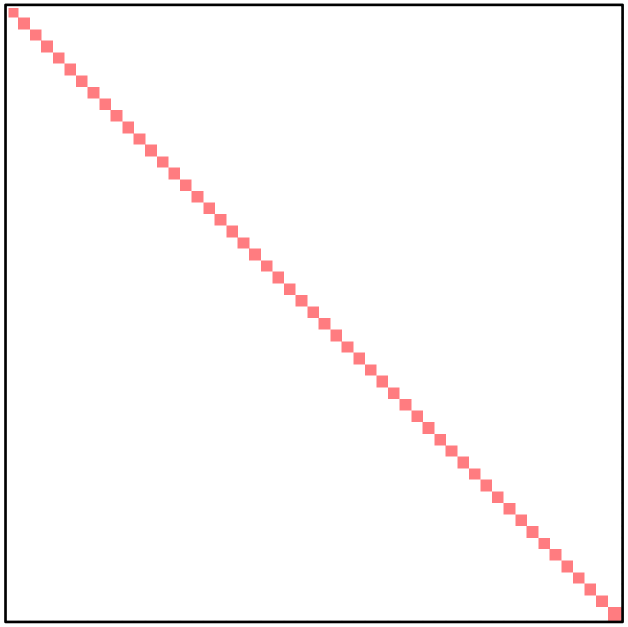


Outcome-stratified design

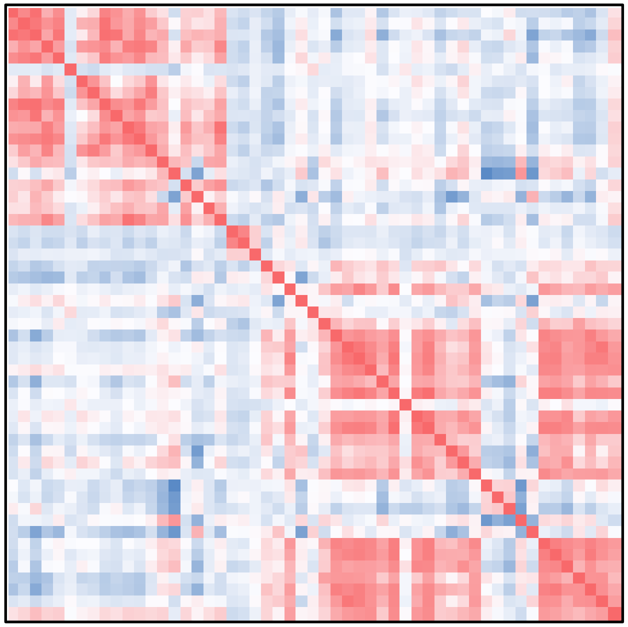


Painful lesson #4: Most *p*- and *q*-values are wrong (due to non-independence). Don’t trust them.

Inter-gene correlation assuming independence



Empirical inter-gene correlation



Efron (2010) J Am Stat Assoc

Painful lesson #1: If you want replicable results, analyze functionally related ***sets*** of genes (not individual gene transcripts)

Sets derive from a biological model of mass action

(i.e., environmental conditions, pathogens, distinct cell types, differentiation states, TFs, chromosomal & epigenetic alterations, genetic polymorphisms, etc.)

Painful lesson #2: If you want replicable results, conduct interpretive bioinformatics on gene lists derived from point estimates of ***effect size*** (not *p*/*q*-values)

Effect sizes (not SNRs) reflect mass biological processes

(i.e., environmental conditions, pathogens, cells, differentiation states, TFs, chromosomal & epigenetic alterations, genetic polymorphisms, etc.)

Painful lesson #3: Hypotheses are ***extremely*** valuable for statistical power, and that is the limiting resource in genomics. Don’t pretend you know nothing.

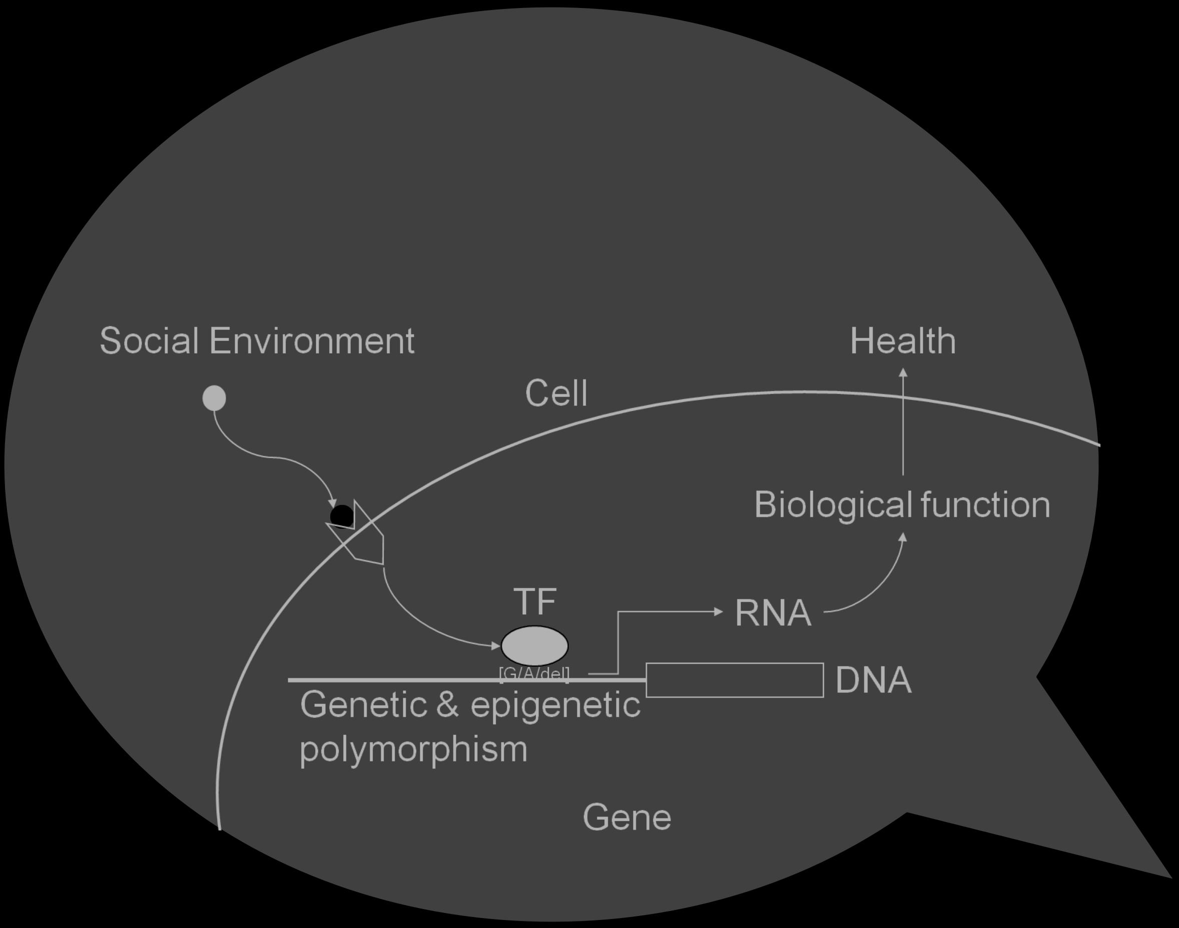
Hypotheses derive from a biological model

(i.e., or perhaps at least a set of ~100 imaginable models << 20,000 genes)

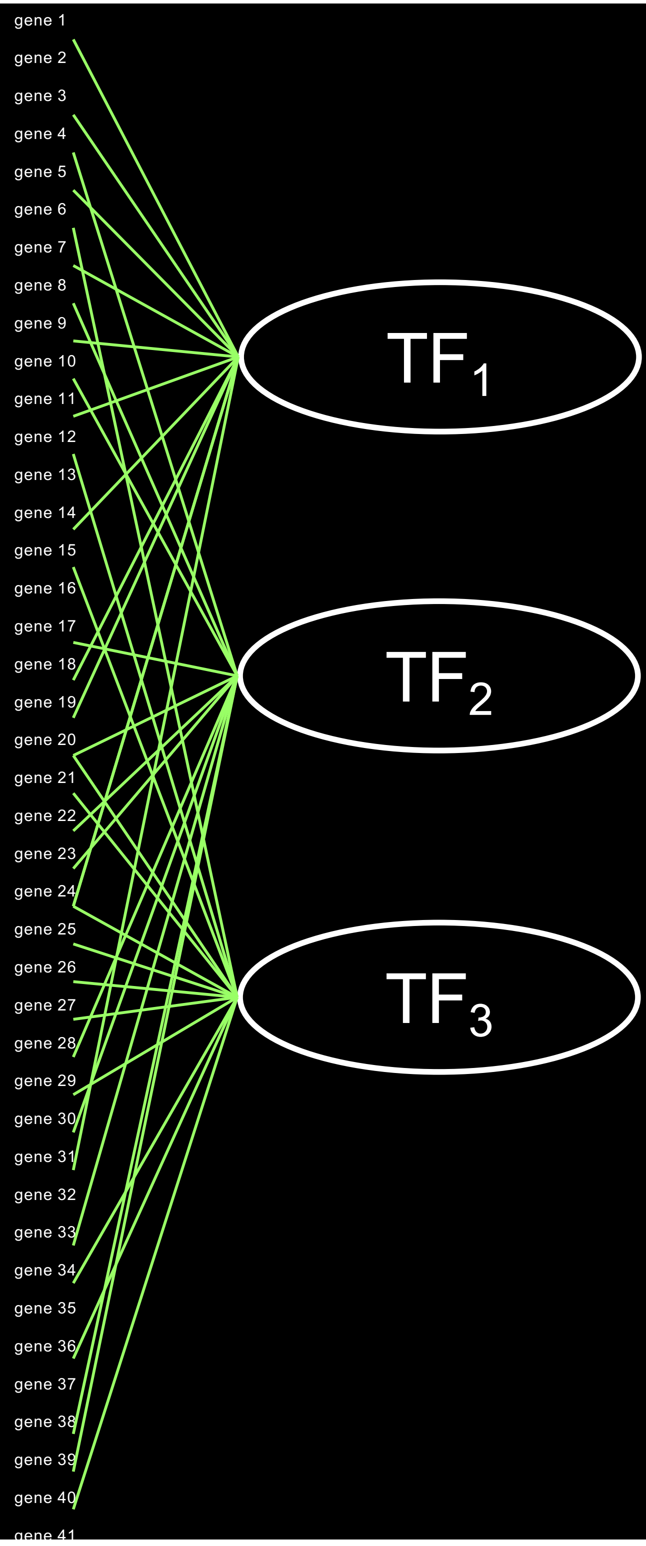
Painful lesson #4: *p/q*-values are inaccurate due to mass action correlation.

Model the biological processes that create the correlation

(work with the gene sets, not against them)



gene 1
gene 2
gene 3
gene 4
gene 5
gene 6
gene 7
gene 8
gene 9
gene 10
gene 11
gene 12
gene 13
gene 14
gene 15
gene 16
gene 17
gene 18
gene 19
gene 20
gene 21
gene 22
gene 23
gene 24
gene 25
gene 26
gene 27
gene 28
gene 29
gene 30
gene 31
gene 32
gene 33
gene 34
gene 35
gene 36
gene 37
gene 38
gene 39
gene 40



Statistical power

100x narrowed search space

Increased reliability

100x indicators

Substantive understanding

Move from *what happened* to *why*