***N**etworks for **TR**anscriptional **A**ctivity **CE**ll a**R**rays (NTRACER)*

NTRACER aims to identify the dynamics of signaling processes that control an observed phenotype using dynamic measurements of TFr activity[1]. NTRACER uses a combination of prior knowledge and an ensemble of inference methods to determine the possible relationships between the given cellular inputs and TFrs. NTRACER employs normalized activity data from the significant TFrs as input. The computational pipeline involves three main steps: i) statistical analysis to identify significant changes in the TFr activity data, ii) generation of an initial network topology, and iii) network identification. Overall, the envisioned computational pipeline was developed to identify highly robust and consistent results in the final networks by protecting against the erroneous identification of edges that could result from noisy data. Robustness and consistency were accomplished by a combination of data pre-filtering, structure optimization and bootstrapping techniques.

The dynamic network was originally modeled as a three-level Boolean paradigm[1]. Herein, we present an improved methodology to avoid data discretization. The most likely connections present at each time are established later by minimizing the difference between the experimental data and the simulation, where the relationships between the nodes are calculated as linear or quadratic regression models. Additionally, we have expanded the libraries of available dynamic inference methods, we have automated the determination of the shortest paths between each TFr and the applied extracellular stimuli, and we have allowed the inclusion of self-loops, where a single TFr was allowed to act on itself from the TRACER experimental data only. Solely new modifications added to NTRACER since our original publication[1] are discussed here.

Inference methods

Multiple inference methods were incorporated into NTRACER to establish new possible connections between extracellular cues (i.e., RGD and stiffness) and measured TFrs not previously reported in the scientific literature[1]. A combination of modified methods to account for dynamics, either linear or non-linear, was summarized into a unique inference network. Linear methods included PLSR[2], similarity index, SI[3], and linear ordinary differential equations (ODE) based on TIGRESS[4]. Non-linear methods included newer strategies, such as dynamic mutual information methods (ARACNE[5], CLR[6], MRNET[7], dynamic random forest[8]), as well as well-established dynamic methods, such as dynamic Bayesian networks[9]. 500 bootstrapping samples from the normalized TFr data were employed to determine connections using the above methods. If a connection appeared in any method in more than 65% of the bootstrapping cases or it appeared greater than 700 times across all methods, it was deemed significant. Those cut-offs were selected based on the frequency distributions of the bootstrapping results. The selected cut-offs coincided with initial frequency of the second distribution of the bimodal bootstrapping results (**Fig. S9**).

*Dynamic PLSR***:** Dynamic PLSR was employed to infer connections between the stimuli or inputs and the TFrs using the *pls* package[10]. The first two interpolated time points (3 and 4.5 hrs) for each TFr, $X_{j,t<ti}$, were employed to regress them against the different conditions (i.e., stiffness or RGD concentration), Y_i (Eq. 4). Connections were considered significant if their loads for their first component were greater than 0.15. The directionality of the interaction is given by the sign of the loading.

$$Y_i = \sum_{j=1}^{n} B_j X_{j,t<t_i} \quad \text{(Eq.4)}$$

Identification of TFrs that most likely affected other constructs was based on the differences between scaled activities acquired at two consecutive time points for the same TFr. TFr activities were regressed with respect to activity at the previous time point of the constructs, which is an approximation to the first derivative over time of the given TFr. Connections were considered significant if their loads for their first component were greater than 0.3.

$$X_{i,t=t+1} - \square_{i,t=t} = \sum_{j=1}^{n} B_j X_{j,t=t} \text{ (Eq. 5)}$$

Dynamic Similarity Index (SI): The SI is defined as the scalar product of the dynamic trajectories of the average activities of two TFrs over time. Therefore, if the dynamic trends of two TFrs were similar, the SI is close to 1, and if they were similar but in completely opposite directions (anti-correlated), the SI value would be -1. A SI index close to 0 indicates that there is no correlation between the dynamic trends of the two TFrs. Here, we calculated the SI of two dynamic trajectories, but where one was delayed with respect to the other, so that we could infer the directionality and sign of the observed correlation in the following manner:

$$SI = \frac{(average(X_{j,t+1})-1)(average(X_{i\in n,t})-1)}{\sqrt{\sum_{k=1}^{m}(average(X_{j,k})-1)^2}\sqrt{\sum_{k=1}^{m}(average(X_{j,k})-1)^2}} \text{ (Eq. 6)}$$

Similarly, we have employed the original definition to calculate the relationships between extracellular conditions or stimuli and TFrs by only employing the first two interpolation times. All the connections that have an abs(SI) ≥ 0.95 were considered significant and 0.9 in the case of edges between stimuli and TFrs.

ODE-TIGRESS: Lasso regression with feature selection stabilization has been successfully applied to infer biological connections[4]. Here, we presented a modification of the procedure, ODE-TIGRESS, where an approximation of the first derivative over time for a given TFr is regressed with respect to all the other TFrs and stimuli present in the system. Lasso regression was performed using the *lars* package, with a regularization penalty, λ, equal to unity, aiming to minimize L:

$$L = \frac{X_{i,t=t+1} - X_{i,t=t}}{t_{t+1} - t_t} - \sum_{j=1}^{n} \beta_j \, X_{j,t=t} + \lambda \sum_{j=1}^{n} \beta_j \quad \text{(Eq. 7)}$$

1000 samples were generated from the original data by randomly multiplying each value by a factor between 0 to 1. An interaction was deemed significant if it was present in at least 99% of the iterations. Directionality and sign were granted by the regression parameters.

Dynamic mutual information: Mutual information (MI) methods were not only considered to determine interactions between stimuli and TFrs, as in the original version of NTRACER, but also between TFrs. The mutual information matrices (MIM) for relationships between inputs and TFrs were constructed as for the dynamic PLSR case. The sign of each interaction between a stimulus and a given TFr was determined by the initial slope over time for each stimulus. For interactions between TFrs, MIM was merged from two matrices: one that contained all the data except the last time point and another that contained all the data points except the first time point. This method provided the MI between the different TFrs with directionality, representing changes between immediately successive time points. The *minet* package[11] was selected to assess the MIM with the Schurmann-Grassberger estimate of the entropy[12] by equal frequency for discretization of the data for ARACNE, CLR and MRNET. Inference networks were created from interactions between each TFr at an initial time point versus all TFrs at the following time points with values greater than 0, as found using any of the above methods. Default parameters were used otherwise.

Dynamic Bayesian Networks: Dynamic Bayesian networks were obtained assuming that all the data were not independent, due to the short experimental frequency used, and no prior knowledge was provided to BANJO[13], http://www.cs.duke.edu/~amink/software/banjo/. No parents were allowed for any of the stimuli, and all the data were discretized into three intervals for each type of extracellular stimulus. Simulated annealing with random local moves was the choice for the searching strategy with the default parameters and a maximum parent size of 5.

Banjo was run 500 times, and interactions were obtained from the top network for each run. Interaction signs were given by the influence score.

Dynamic random forest: For the dynamic random forest version, concepts from GENIE3[14] were incorporated, but with modifications to permit handling time-series data by the non-linear random forest approach. The approximation to the first derivative over time was calculated as above. A total of 1000 random trees were created using the data for all the TFrs and treatments for each time point employing the *randomForest* package[15]. The square root of the total number of all the TFrs and conditions was used to select the number of random TFrs to start populating the trees. The importance of a node was measured by the reduction in the residuals. Edges were considered significant if they appeared in the top 10% ranked weights. Directionality was guaranteed by the temporal order.

Consensus inference network: A total of 500 bootstrap samples were generated using the weights described above and the inference methods listed above applied to each bootstrapping sample. To combine all bootstrap samples, edges were deemed significant if there were present in more than 65% of the runs for at least one inference method or if the number of the times that was significant by some of the investigated inference methods exceeded the 700 counts (140% of the 500 bootstrap samples). These cut-offs were selected based on the bimodal frequency distribution for each method alone and all methods combined. Specifically, they were selected to coincide with the start of the second distribution of the bimodal graph (**Fig. S9**)

Determination of TFr networks evolution over time upon chemical and physical alterations of the extracellular environment

The initial network topology originated from an equally weighted number of prior knowledge sources and inference methods. Prior knowledge and inference networks were combined into a unique structure that served as a combined initial knowledge network model for the modified version of CellNOptR[16] in NTRACER. The improved NTRACER (NTRACER v2.0) was employed to identify the most likely connections present at each time point, penalizing network complexity. First, the initial network was simplified by removing all connections that did not include edges between the external stimuli (i.e., adhesion peptide concentration and gel stiffness) and TFrs or between TFrs.

NTRACER v2.0 was adapted from the three-level Boolean to a continuous paradigm, where edges represent linear and non-linear interactions between the nodes. This modification allowed accommodating continuous variable levels (i.e., stiffness and RGD concentration). These features were required in order to capture the cellular biphasic response upon chemical and physical environmental cues. The prediction of the output from the model was obtained from a regression model that accounts for the contributions of all the input nodes to a given TFr activity. Initially the regression model was assumed linear. However, lack of fit to a linear model was estimated with the rainbow test[17] (p-value≤0.1), and a quadratic term was added to model the non-linear effects.

Assume that the following reactions are active in the random structure *i*:

$$A \rightarrow B$$

$$B \dashv C$$

$$A \rightarrow C$$

NTRACER v2.0 fits a linear model for each of the output nodes, in this case, B and C, as a function of their input nodes:

$$B_{t=t+1} = \alpha_1 A_{t=t} \quad \text{(Eq.8)}$$

$$C_{t=t+1} = \alpha_1 A_{t=t} - \alpha_2 B_{t=t} \quad \text{(Eq.9)}$$

Note that NTRACER v2.0 aims to predict the next temporal response of a given node, in this case, B and C, based on the previous temporal values of A and B. In addition, for each of the models and coefficients, NTRACER v2.0 determines the lack of fit to a linear model using the rainbow test [17] from the *lmtest* package[18]. If the alternative hypothesis is significant (p-value≤0.1), in other words, if the relationship is not linear, an additional squared term is added to the model. Assume that ifα_2 were not significant, then NTRACER v2.0 will fit the following model:

$$C_{t=t+1} = \alpha_1 A_{t=t} - \alpha_2 B_{t=t} + \alpha_3 B_{t=t}^2 \quad \text{(Eq.10)}$$

Another addition to NTRACER v2.0 is the manner in which TFrs are allowed to participate in self-loop edges. Here, we incorporated a penalty for self-loop edges and avoided models with only auto-regressive edges.

$$Score = \frac{1}{N}\left(\sum_{i=1}^{NC}(x_M - x_i)^2 + 0.1(N - NC)\right) + \frac{1}{NInp}\left(si\square e_{Pen} NSig + Stim_{Pen} SP^{(OrdT-1)} NStim + InhM_{Pen} size_{Pen} NInhM + sl_{Pen} NSl\right) \quad \text{(Eq. 11)}$$

Here, N is the total number of experimental observations; NC is the total number of simulations in which the model converged; x_M represents the simulation results from the model; x_i denotes the discretized experimental results; NA_{Pen}, $size_{Pen}$, $Stim_{Pen}$, $InhM_{Pen}$,and sl_{Pen} are the penalties assigned to the size of non-converged simulation results, number of edges from TFrs, stimuli, InhM, and self-loops, respectively; *NInp, NSig, NStim,NInhM* and *NSl* are the size of the total number of edges, number of edges originated from TFrs, stimuli, InhM and self-loops respectively; *SP* is the stimuli policy increased to penalize the appearance of long-term stimuli edges, and *OrdT* indicates the order of the experimental time whose structure is being optimized.

Only TFrs with significantly different activities among treatments in at least one time point (meta-analysis false discovery rate (fdr)-corrected *p*-value ≤ 0.02) were subsequently studied. In order to reduce the computational time, a two-level factorial design with a central point was conducted to determine the parameters that yielded the lowest score for the same number of iterations (**Table S1**).

References

1. M. Weiss, B. P. Bernabé, S. Shin, S. Asztalos, S. Dubbury, M. Mui, A. Bellis, D. Bluver, D. Tonetti and J. Saez-Rodriguez, *Integrat Biol*, 2014.
2. S. Wold, M. Sjöström and L. Eriksson, *Chemometrics Intelligent Lab Sys*, 2001, **58**, 109-130.
3. A. Siletz, M. Schnabel, E. Kniazeva, A. J. Schumacher, S. Shin, J. S. Jeruss and L. D. Shea, *PLoS ONE*, 2013, **8**, e57180.
4. A.-C. Haury, F. Mordelet, P. Vera-Licona and J.-P. Vert, *BMC Sys Biol*, 2012, **6**, 145.
5. A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera and A. Califano, *BMC Bioinformatics*, 2006, **7**, S7.
6. J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins and T. S. Gardner, *PLoS biology*, 2007, **5**, e8.
7. P. E. Meyer, K. Kontos, F. Lafitte and G. Bontempi, *EURASIP journal on bioinformatics and systems biology*, 2007, **2007**.
8. L. Breiman, *Machine learning*, 2001, **45**, 5-32.
9. J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink and E. D. Jarvis, *Bioinformatics*, 2004, **20**, 3594-3603.
10. V. A. Smith, J. Yu, T. V. Smulders, A. J. Hartemink and E. D. Jarvis, *Plos Comput Biol*, 2006, **2**, 1436-1449.
11. P. E. Meyer, F. Lafitte and G. Bontempi, *BMC Bioinformatics*, 2008, **9**, 461.
12. T. Schurmann and P. Grassberger, *Chaos*, 1996, **6**, 414-427.
13. J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink and E. D. Jarvis, *Bioinformatics*, 2004, **20**, 3594-3603.
14. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel and P. Geurts, *PLoS One*, 2010, **5**.
15. A. Liaw and M. Wiener, *R News*, 2002, **2**, 18-22.
16. J. Saez-Rodriguez, L. G. Alexopoulos, J. Epperlein, R. Samaga, D. A. Lauffenburger, S. Klamt and P. K. Sorger, *Mol Sys Biol*, 2009, **5**.
17. J. M. Utts, *Commun Stat-Theory Meth*, 1982, **11**, 2801-2815.
18. A. Zeileis and T. Hothorn, *Diagnostic checking in regression relationships*, R News, 2002.

Supplemental Figures and Tables

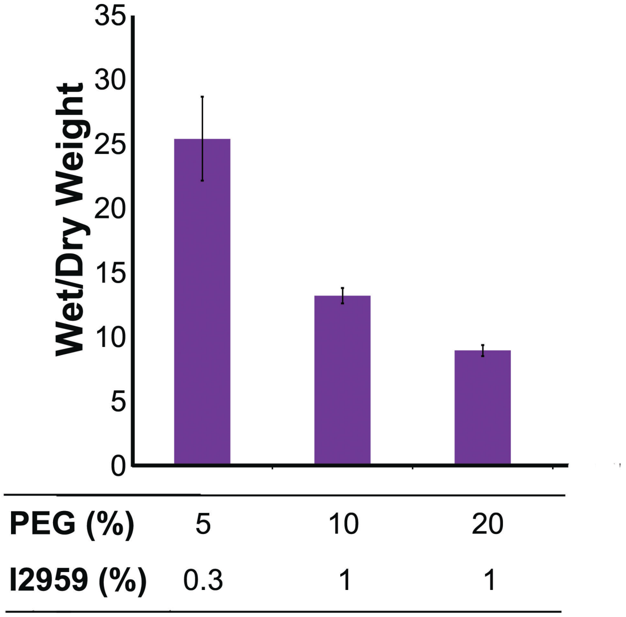

Figure S1. Swelling rations of human foreskin fibroblasts cultured on PEG hydrogels with varying modulus or RGD concentration. Swelling ratios were calculated by comparing the weight of hydrogels swollen to equilibrium (>12 hrs at room temperature) in PBS, pH 7.4 and after lyophilization: swelling ratio = wet weight/dry weight.

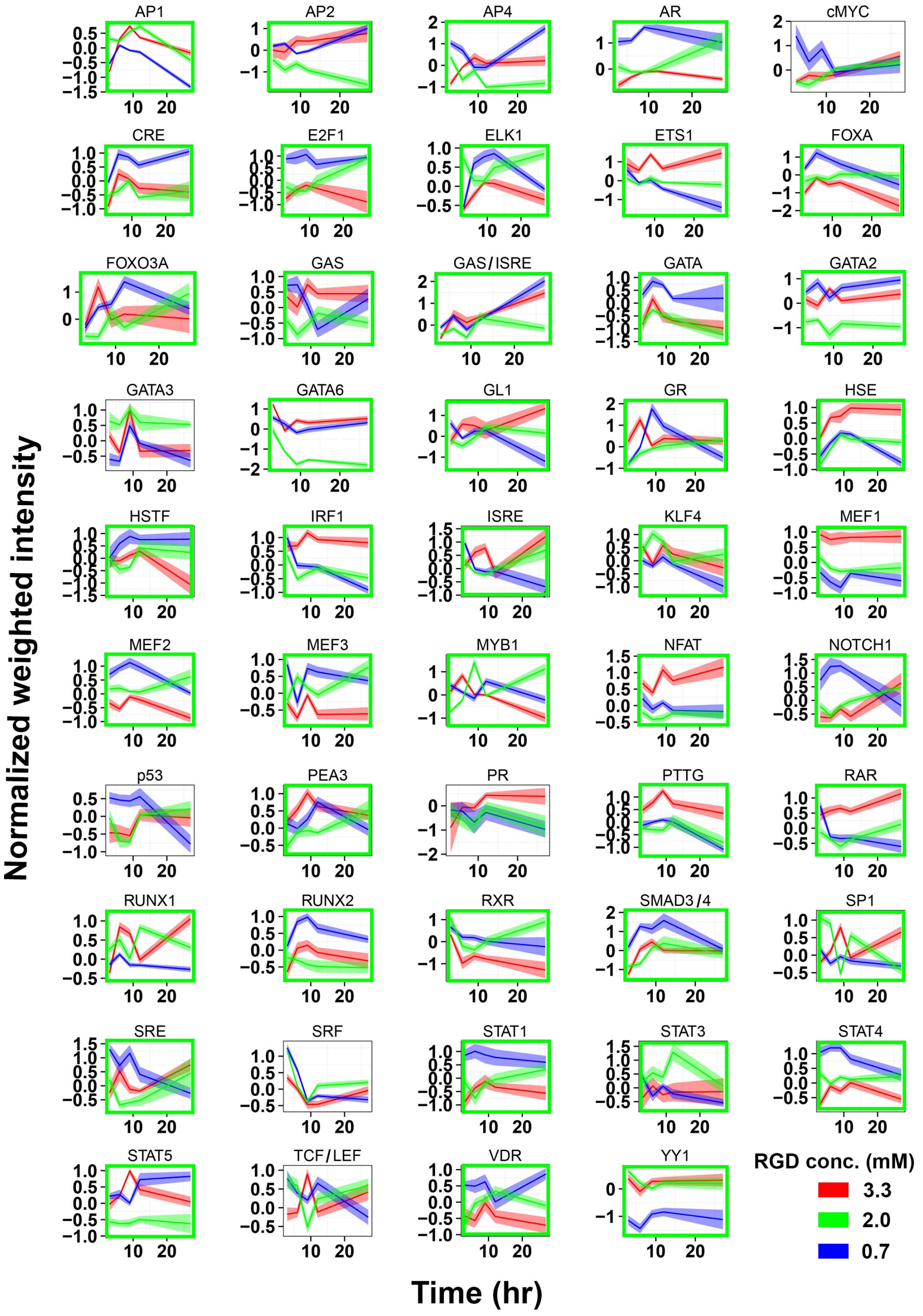

Figure S2. Dynamic TFr activity trends different levels of adhesion motif (RGD) concentration. Mean weighted normalized and 95% confidence intervals. Green squares indicate that there is at least there is one significant difference between the RGD concentrations at one of the measured times (meta-analysis fdr-corrected p-value ≤0.02). Gel stiffness 1.5 kPa.

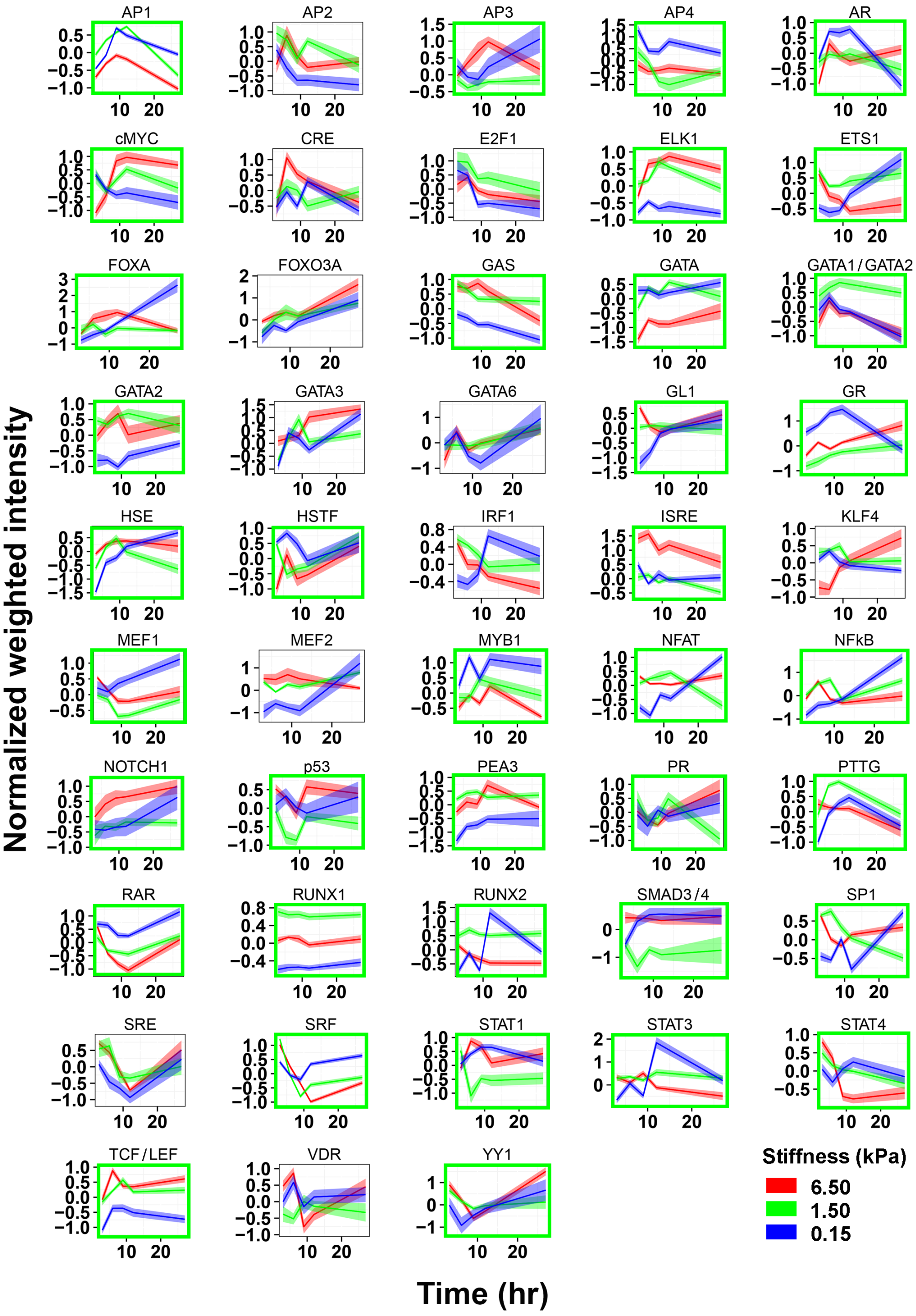

Figure S3. Dynamic TFr activity trends different levels of gel stiffness. Mean weighted normalized and 95% confidence intervals. Green squares indicate that there is at least there is one significant difference between the upon variation of the gel stiffness in at least one of the measured times (meta-analysis fdr-corrected p-value ≤0.02). Adhesion concentration 2.0 mM.

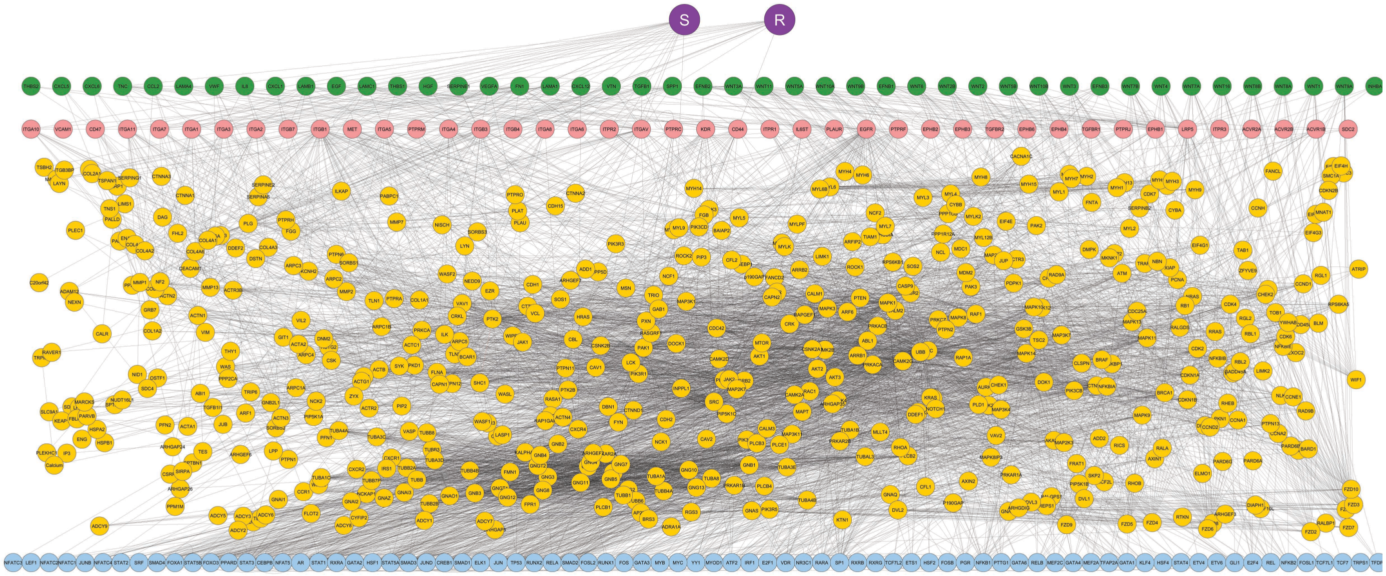

Figure S4. Mechanotransduction signaling network. Purple nodes are the two mechanotransduction explored variables, S for gel stiffness and R for RGD concentration; green nodes are ligands such as fibronectin or collagen; red nodes are membrane proteins such as receptors, integrins or cadherins; yellow nodes represent cytosolic proteins (i.e, kinases, phosphatases); blue nodes are the transcription factors whose consensus sequences was employed to generate the TFr employed in the study. Connections were obtained from the GENEGO database. The initial experimental network employed for NTRACER that incorporate the connections between the ECM and TFs also contained experimentally determined connections from the adhesome database (see methods for more details).

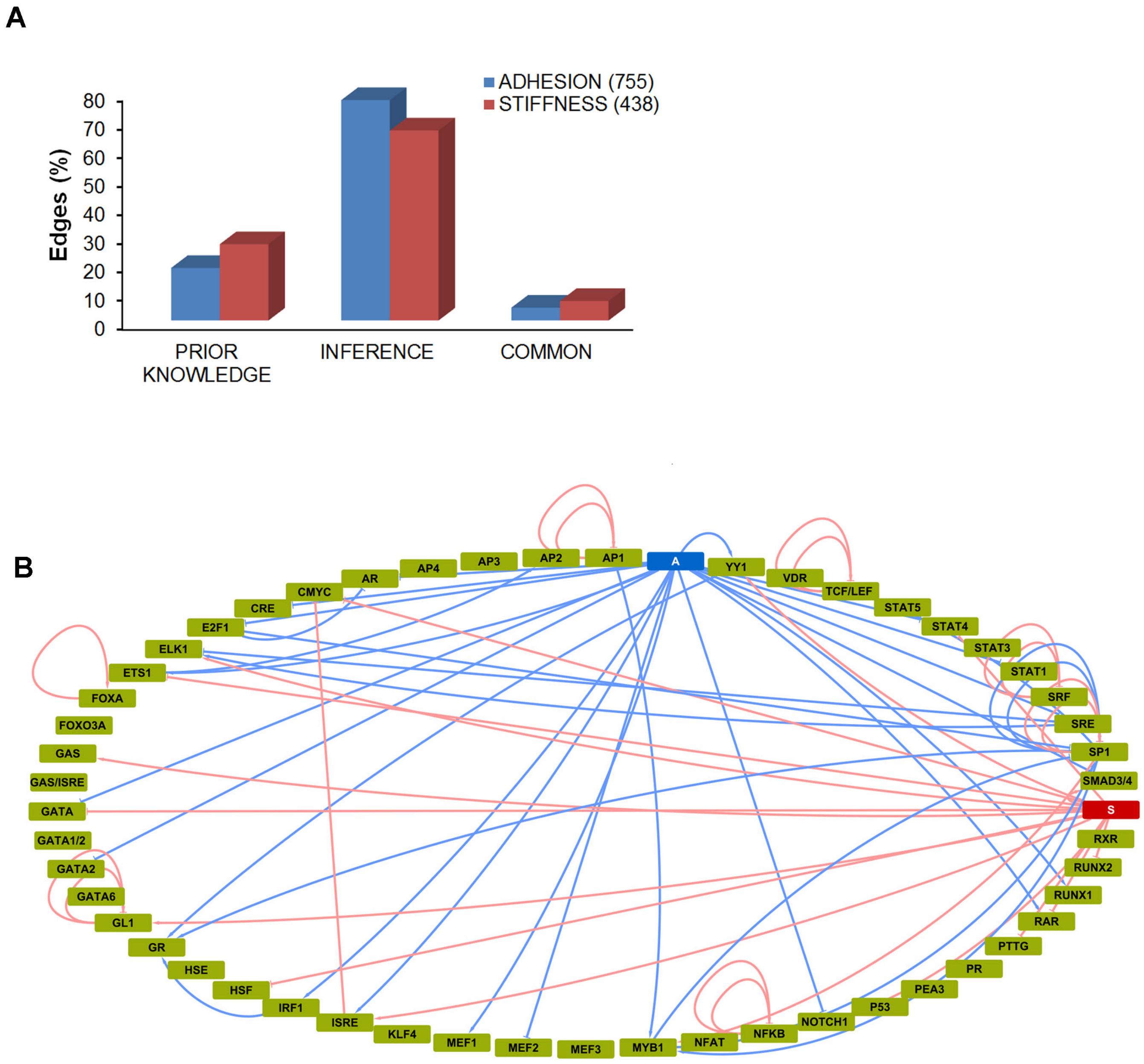

Figure S5 Initial networks for NTRACER for A) variations in RGD concentration; B) variation in gel stiffness. A) Number of edges originated only from literature curation (i.e., prior knowledge), or only from inference methods and those that were common between both approaches. Total number of edges is indicated between parenthesis. B) The two mechanotransduction explored variables (adhesion and stiffness) correspond to the blue and red nodes, respectively; Green nodes are transcription factors reporters. Only edges that are common between the two sources, prior knowledge and inference methods, are represented. Adhesion common edges are indicated in light blue and stiffness common edges are represented in light red.

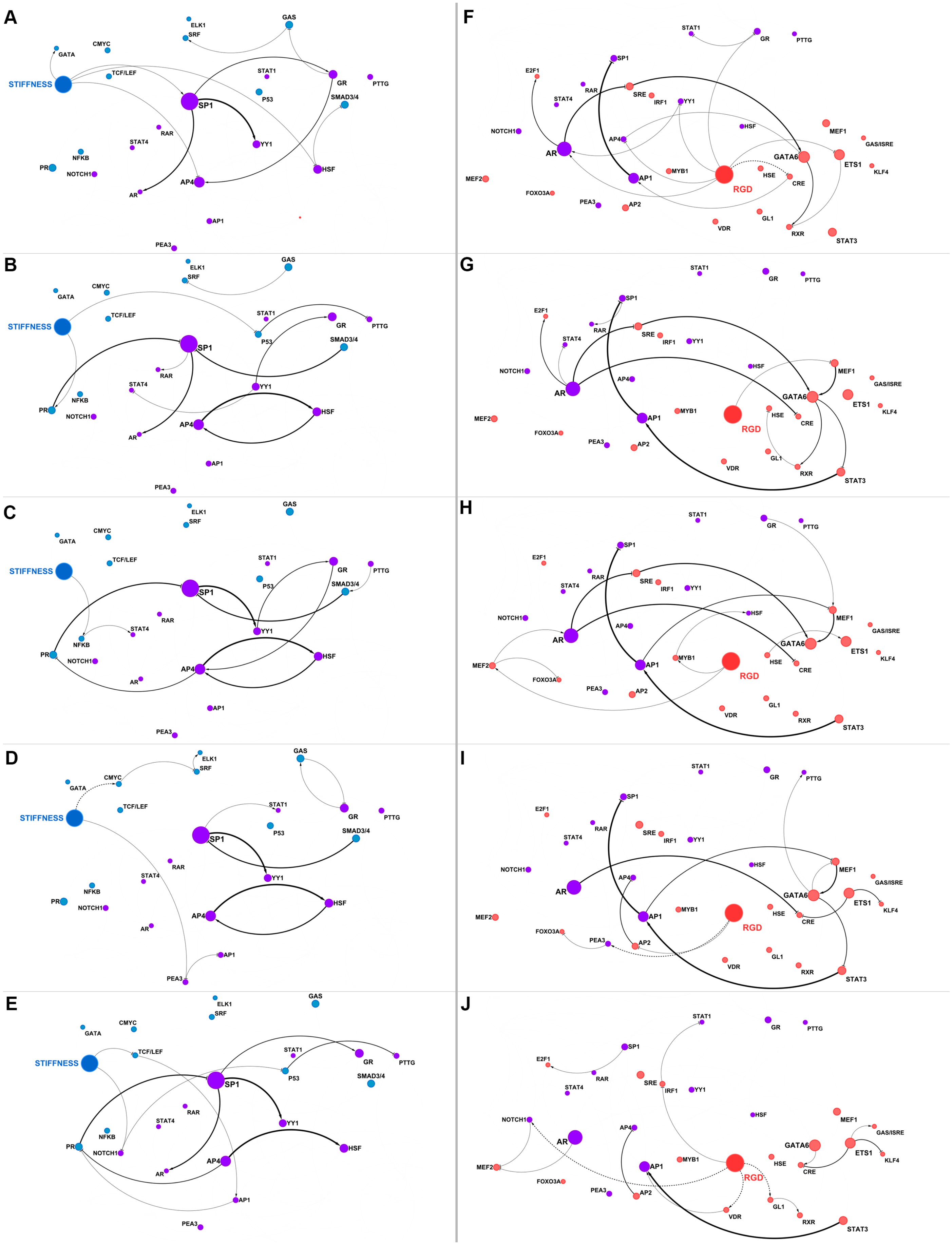
A
GAS
ELK1
SRF
CMYC
GATA
STIFFNESS
TCF/LEF
STAT1
GR
PTTG
P53
SP1
SMAD3/4
RAR
STAT4
YY1
NFKB
PR
HSF
NOTCH1
AP4
AR
AP1
PEA3
B
ELK1
GAS
SRF
CMYC
GATA
TCF/LEF
STAT1
STIFFNESS
GR
PTTG
P53
SP1
SMAD3/4
RAR
YY1
STAT4
NFKB
PR
HSF
NOTCH1
AP4
AR
AP1
PEA3
C
GAS
ELK1
SRF
CMYC
GATA
TCF/LEF
STAT1
STIFFNESS
GR
PTTG
SP1
P53
SMAD3/4
RAR
STAT4
YY1
NFKB
PR
HSF
NOTCH1
AP4
AR
AP1
PEA3
D
ELK1
GAS
SRF
CMYC
GATA
TCF/LEF
STAT1
STIFFNESS
GR
PTTG
SP1
P53
SMAD3/4
RAR
STAT4
YY1
NFKB
PR
HSF
NOTCH1
AP4
AR
AP1
PEA3
E
GAS
ELK1
SRF
CMYC
GATA
TCF/LEF
STAT1
STIFFNESS
GR
PTTG
SP1
P53
SMAD3/4
RAR
STAT4
YY1
NFKB
PR
HSF
NOTCH1
AP4
AR
AP1
PEA3
F
STAT1
GR
PTTG
SP1
E2F1
RAR
SRE
IRF1
YY1
STAT4
HSF
MEF1
NOTCH1
AP4
GAS/ISRE
GATA6
AR
ETS1
KLF4
MYB1
HSE
MEF2
AP1
CRE
RGD
FOXO3A
PEA3
AP2
GL1
VDR
RXR
STAT3
G
STAT1
GR
PTTG
SP1
E2F1
RAR
SRE
STAT4
IRF1
YY1
HSF
MEF1
NOTCH1
AP4
GAS/ISRE
AR
GATA6
ETS1
KLF4
MYB1
HSE
CRE
AP1
MEF2
RGD
FOXO3A
PEA3
AP2
GL1
VDR
RXR
STAT3
H
STAT1
GR
PTTG
SP1
E2F1
RAR
SRE
STAT4
IRF1
YY1
HSF
NOTCH1
MEF1
AP4
GAS/ISRE
AR
GATA6
ETS1
KLF4
MYB1
HSE
AP1
MEF2
CRE
RGD
FOXO3A
PEA3
AP2
VDR
GL1
RXR
STAT3
I
STAT1
GR
PTTG
SP1
E2F1
RAR
SRE
STAT4
IRF1
YY1
HSF
NOTCH1
MEF1
AP4
GAS/ISRE
AR
GATA6
ETS1
KLF4
MYB1
HSE
CRE
AP1
MEF2
RGD
FOXO3A
PEA3
AP2
VDR
GL1
RXR
STAT3
J
STAT1
GR
PTTG
SP1
E2F1
RAR
SRE
IRF1
YY1
STAT4
HSF
MEF1
NOTCH1
AP4
GAS/ISRE
AR
GATA6
ETS1
KLF4
MYB1
HSE
CRE
MEF2
AP1
RGD
FOXO3A
PEA3
AP2
GL1
VDR
RXR
STAT3

Fig. S6: Dynamic TF activity networks for changes in stiffness (A-E) and RGD levels (F-J) in PEG hydrogels. Hydrogel conditions and TFrs are represented as nodes, while the connections between them are represented by directed edges. Only edges active at each temporal step (e.g., 0-3 hrs, A and F; 3-6 hrs, B and G; 6-9 hrs, C and H; 9-12 hrs, D and I; and 12-27 hrs, E and J) are represented. Nodes affected by changes in both RGD and stiffness levels are represented in purple. Nodes only affected by RGD changes or only by stiffness changes are colored in red and aqua, respectively. Edges corresponding to linear relationships between nodes are represented by continuous lines. Edges corresponding to non-linear relationships are represented with dashes. Node size is proportional to the number of nodes that can potentially alter the TFr activity level. Similarly, edge thickness is proportional to the number of times that are activated during the measured experimental times. Activation or inhibitory effects on the downstream nodes is represented by an arrow or a T respectively.

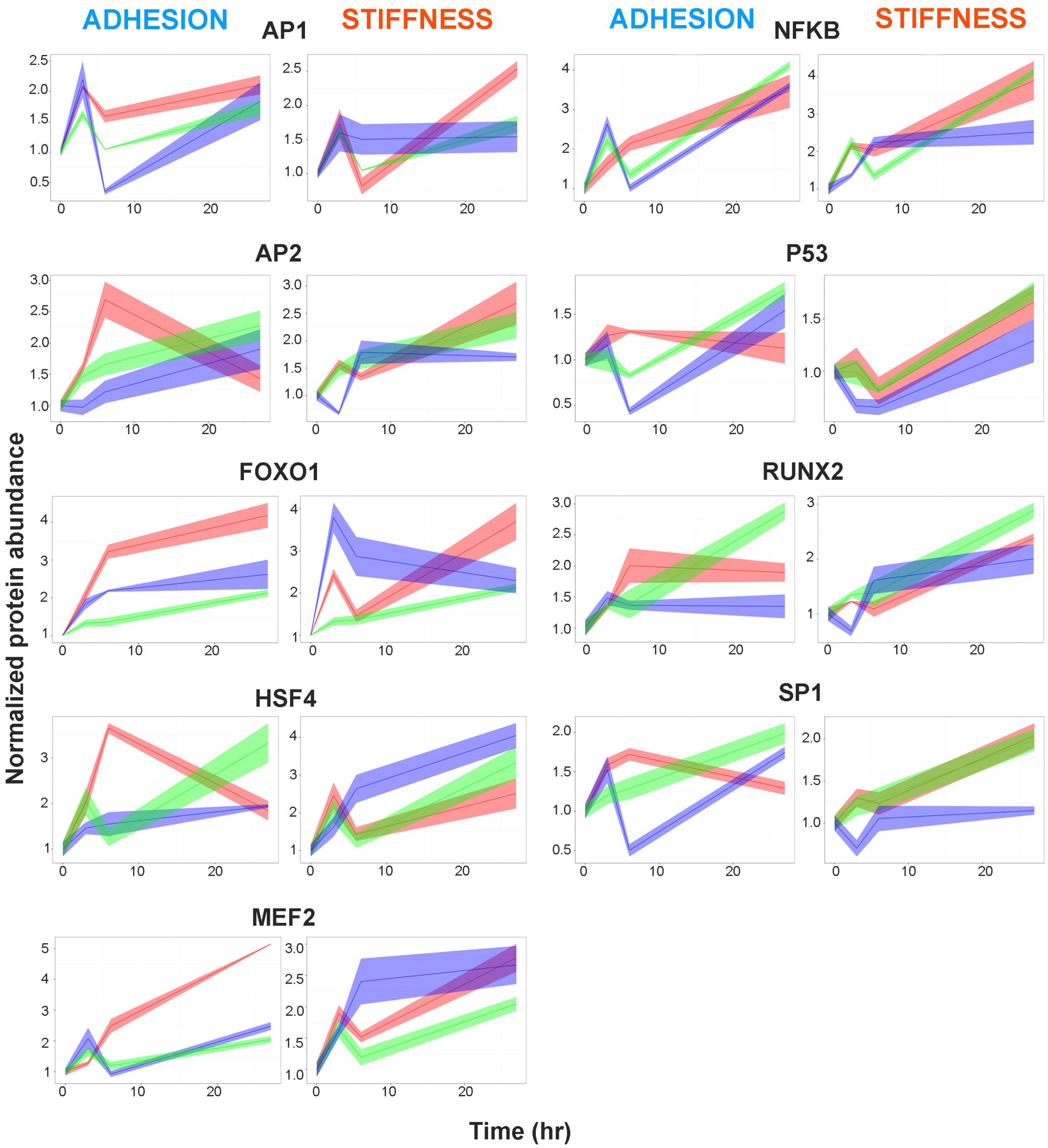

Fig S7. Dynamic TF abundance trends for different levels of adhesion motif (RGD) concentration and PEG gel stiffness levels. Mean normalized protein abundance and 95% confidence intervals. The colors of each trends are the same as Fig. S2 and Fig. S3

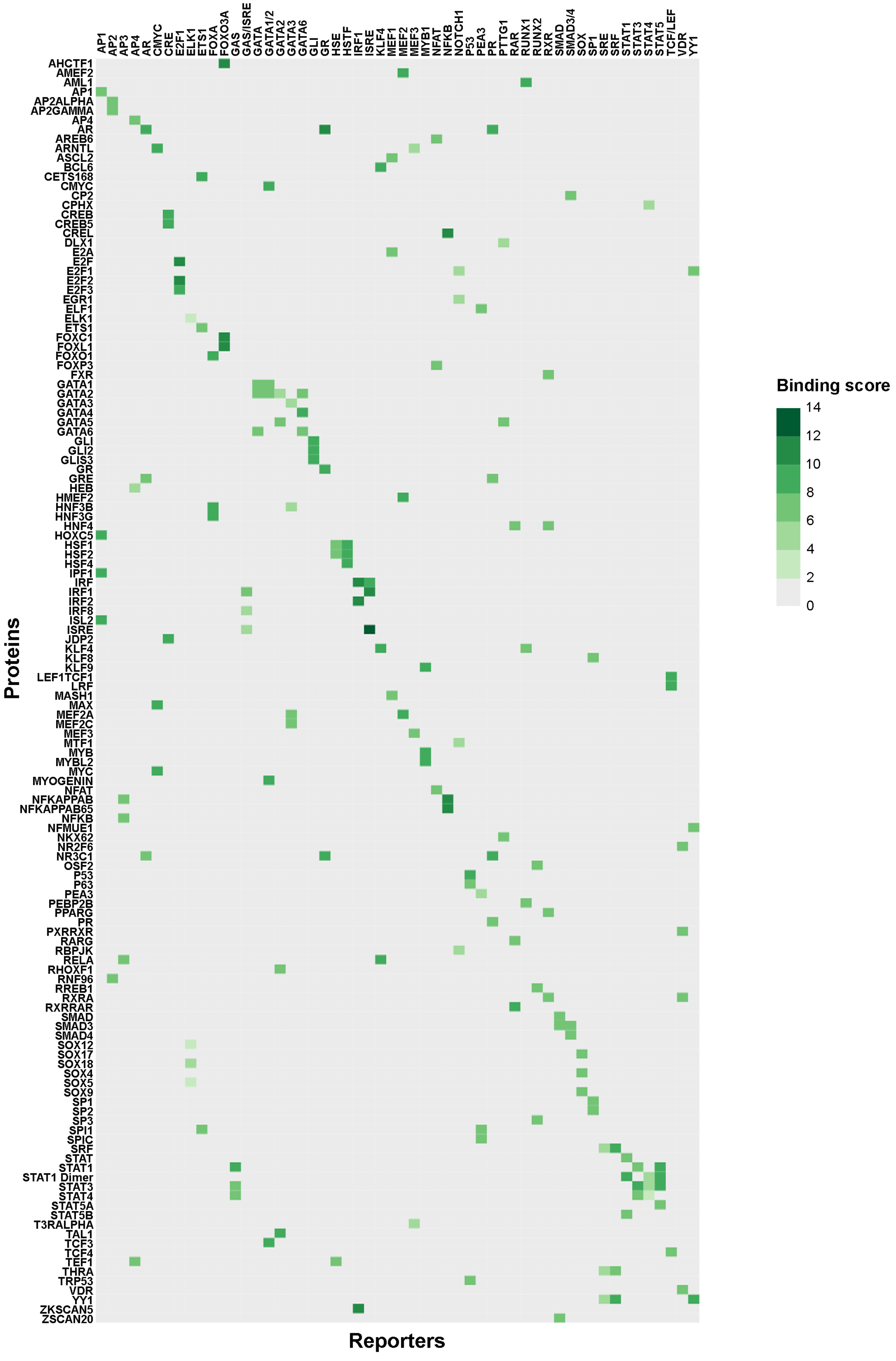
AP1
AP2
AP3
AP4
AR
CMYC
CRE
E2F1
ELK1
ETS1
FOXA
FOXO3A
GAS
GAS/ISRE
GATA
GATA1/2
GATA2
GATA3
GATA6
GLI
GR
HSE
HSTF
IRF1
ISRE
KLF4
MEF1
MEF2
MEF3
MYB1
NFAT
NFKB
NOTCH1
P53
PEA3
PR
PTTG1
RAR
RUNX1
RUNX2
RXR
SMAD
SMAD3/4
SOX
SP1
SRE
SRF
STAT1
STAT3
STAT4
STAT5
TCF/LEF
VDR
YY1
AHCTF1
AMEF2
AML1
AP1
AP2ALPHA
AP2GAMMA
AP4
AR
AREB6
ARNTL
ASCL2
BCL6
CETS168
CMYC
CP2
CPHX
CREB
CREB5
CREL
DLX1
E2A
E2F
E2F1
E2F2
E2F3
EGR1
ELF1
ELK1
ETS1
FOXC1
FOXL1
FOXO1
FOXP3
FXR
GATA1
GATA2
GATA3
GATA4
GATA5
GATA6
GLI
GLI2
GLIS3
GR
GRE
HEB
HMEF2
HNF3B
HNF3G
HNF4
HOXC5
HSF1
HSF2
HSF4
IPF1
IRF
IRF1
IRF2
IRF8
ISL2
ISRE
JDP2
KLF4
KLF8
KLF9
LEF1TCF1
LRF
MASH1
MAX
MEF2A
MEF2C
MEF3
MTF1
MYB
MYBL2
MYC
MYOGENIN
NFAT
NFKAPPAB
NFKAPPAB65
NFKB
NFMUE1
NKX62
NR2F6
NR3C1
OSF2
P53
P63
PEA3
PEBP2B
PPARG
PR
PXRRXR
RARG
RBPJK
RELA
RHOXF1
RNF96
RREB1
RXRA
RXRRAR
SMAD
SMAD3
SMAD4
SOX12
SOX17
SOX18
SOX4
SOX5
SOX9
SP1
SP2
SP3
SPI1
SPIC
SRF
STAT
STAT1
STAT1 Dimer
STAT3
STAT4
STAT5A
STAT5B
T3RALPHA
TAL1
TCF3
TCF4
TEF1
THRA
TRP53
VDR
YY1
ZKSCAN5
ZSCAN20
Proteins
Reporters
Binding score
14
12
10
8
6
4
2
0

Figure S8.Possible TFs binding to each of the studied TF reporters. Each sequences for all the TFr employed in the study were scanned to determine the most likely TFs whose consensus binding sequences were highly similar to TFr sequence itself. The binding score represents the likelihood of a given TF to bind a given reporter, accounting for sequence similarity and the non-overlapping motifs. Only the top 3 rank TF for each TFr are represented as well as the TF whose consensus binding sequence was employed for the design of each reporter (i.e., for AP1 reporter, the AP1 consensus sequence was employed).

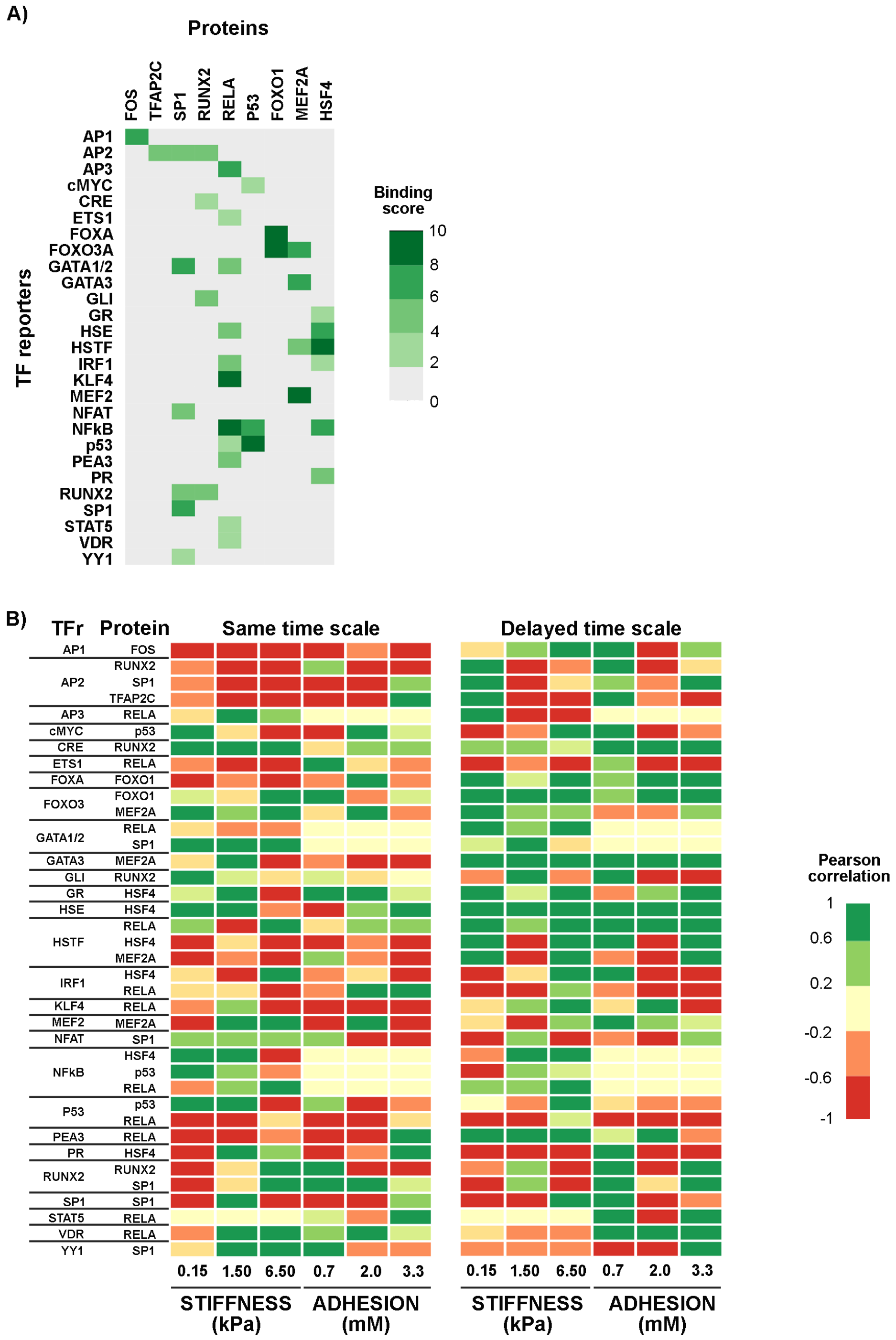

Figure S9. Validation of TRACER measurements using microWesterns arrays (MWA). A) Possible binding sites of the proteins whose abundance was measured by MWA. Using FIMO, we identified the most likely reporters that could bind to each of the studied TFrs. We limited the list to the top 3 proteins as well as the protein that was employed for the design of the reporter (i.e., for AP1 reporter, the AP1 consensus sequence was employed). B) Most likely TF reporters that selected proteins that were analyzed by MWA arrays could bind, using the same time scale (left panel) and delayed time scale (right panel).

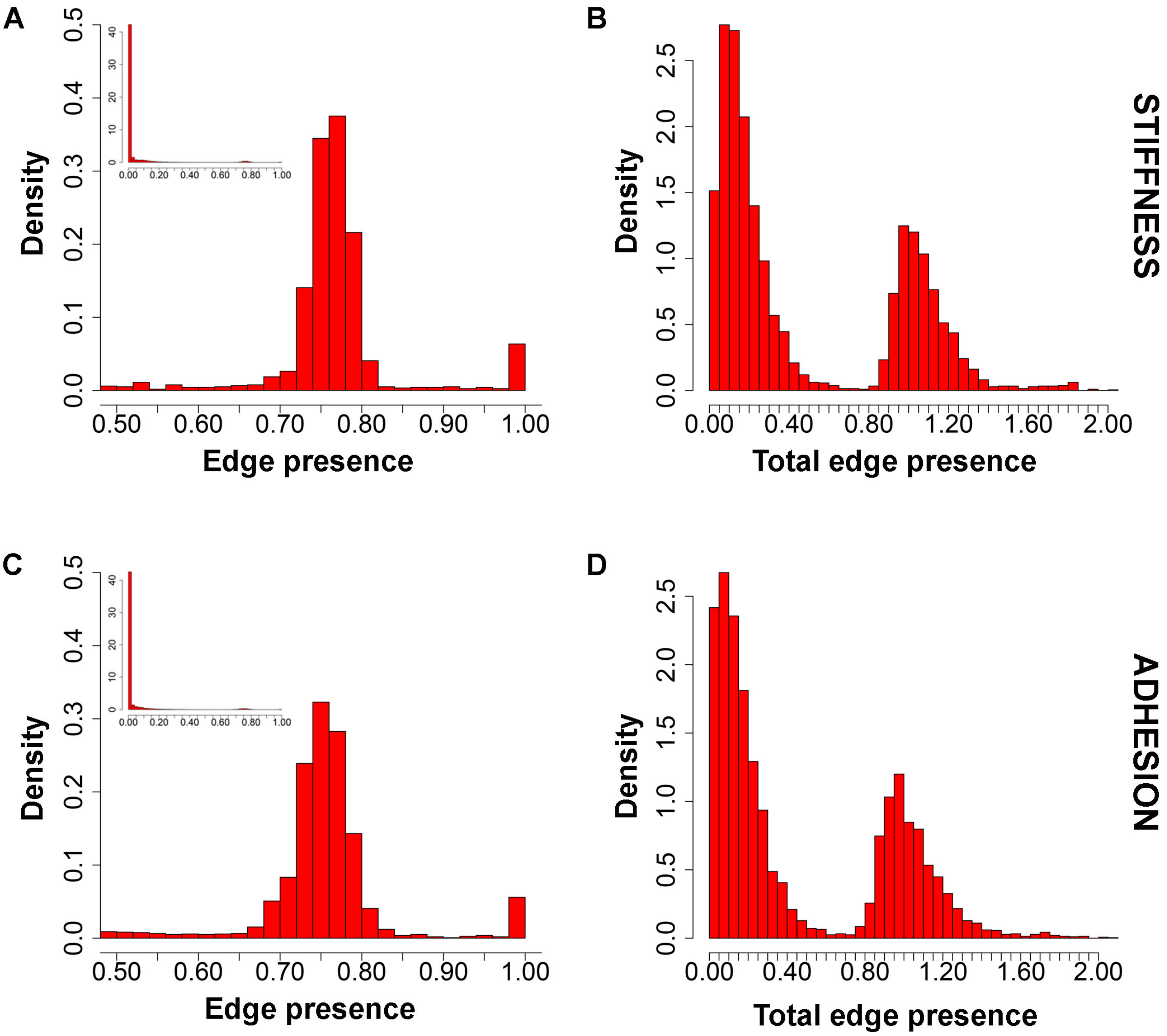

Figure S10. Histograms of edge presence for the different explored inference methods. A) Histogram forthe presence of an edge in a given method (e.g., TD-PLSR, TD-MI) in the 500 bootstrapping runs for the experiments in which stiffness was altered; B) Histogram for the total summation of the presence of each edge independently of the inference method runs for the experiments in which stiffness was altered. C and D panels represents the same histograms for the experiments in which adhesion was altered.

Table S1: Optimized parameters employed for NTRACER

Parameters	Values
Population size	50
Percentage of non-present edges for random start	0.5
Elitism	5
Probability of mutations	0.001
Selective pressure	3
Deactivation mechanism factor penalty	48
Edge penalty	2
Self-loop penalty	6
Stimuli penalty	2

Table S2: List of the TF antibodies employed in the microwestern arrays

Antibody	Company	Catalog number
TFAP2C/AP2-gamma	Aviva	ARP38284_T100
FOS	Santa Cruz Biotechnology	sc-52
FOXO1	Cell Signaling Technologies	9462
HSF4	Aviva	ARP32652
MEF2A	Cell Signaling Technologies	9736
P53	Cell Signaling Technologies	9282
RELA	Aviva	P100779
RUNX2	Cell Signaling Technologies	8486
SP1	Abcam	ab13370
Lamin A+C	AbCam	ab8984

Table S3: List of microarrays employed for the identification of overexpressed TF gene targets in mechanotransduction related transcriptomic measurements

Group	Array Express ID	Cells/Tissue	Variables	FC	p-value	fdr corrected
Stiffness	E-GEOD-22011	Human lung fibroblasts	Different matrix stiffness	1.3	0.005	no
	E-GEOD-33603	Young patient quadriceps	Massage therapy after exercise	1.3	0.01	no
	E-GEOD-10125	Human dermal fibroblast cells	3 hours of cycle mechanical loading	1.3	0.01	yes
RGD	E-MEXP-1273	Human mesenchymal stem cells from adipose tissue	Monolayer or LVG or RGD alginate	1.3	0.05	no
Both	E-GEOD-6432	Human fibroblasts	Culture in petri dish or attached to a tissue engineered scaffold	1.3	0.01	yes
	E-GEOD-44811	Adipose stromal cells	2D or 3D collagen culture	1.3	0.01	yes
	E-GEOD-39475	Human foreskin fibroblasts	Attached versus released 3D collagen matrix	1.3	0.01	yes
	E-GEOD-3003	Human CD34+ hematopoietic cells	Suspension culture or collagen I matrix	1.2	0.05	yes
Fibrosis	E-GEOD-17978	Non-culture pulmonary fibroblasts from idiopathic pulmonary fibrosis (IPF)	Patients versus normal control donors	1.3	0.01	yes