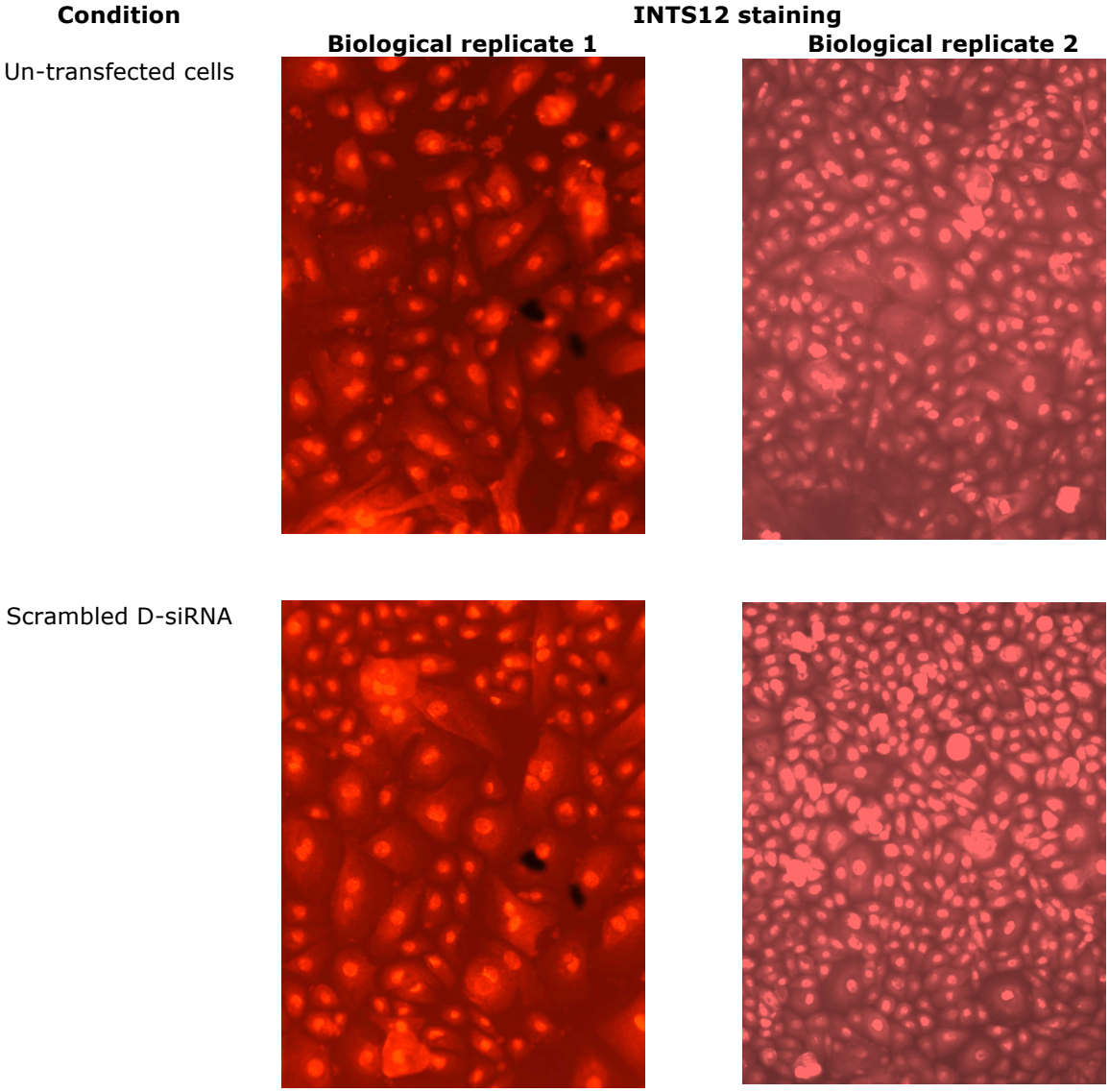
Supplemental Figures

Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways

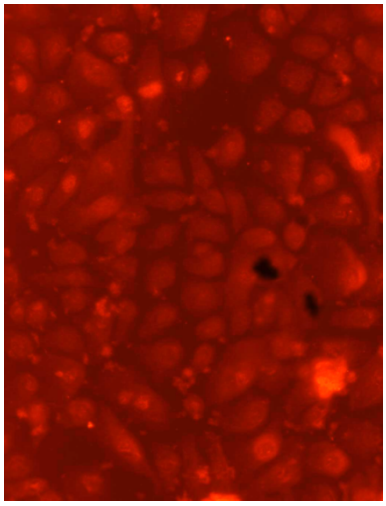
Kheirallah et al.

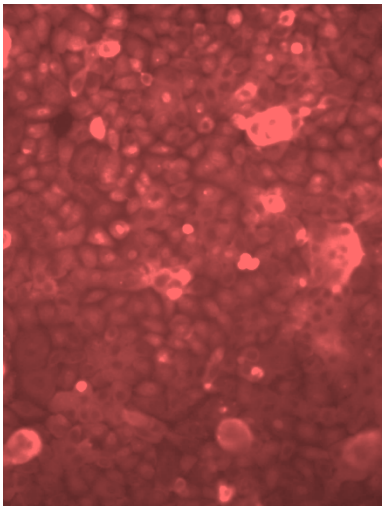
Figure S1: Qualitative comparison of biological replicates of INTS12 protein knockdown.

Representative images from the first experiment are shown in the first column and are compared to the images from the independent experiment shown in the second column. Based on these observations it is possible to say that D-siRNA treatment resulted in INTS12 protein depletion and indicates the specificity of used antibody as there is a notable decrease of staining among cells in which RNAi was initiated. In agreement with previous reports, INTS12 appears to have a nuclear sub-cellular localization.

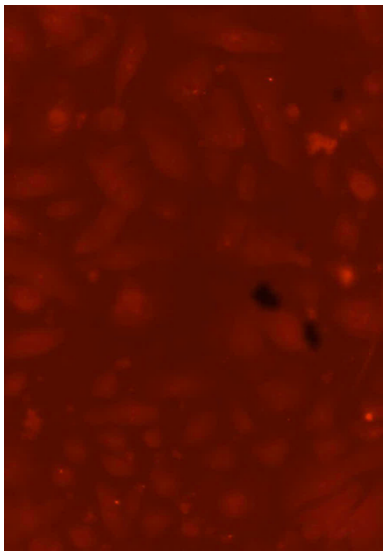


D-siRNA INTS12 knockdown





Isotype control



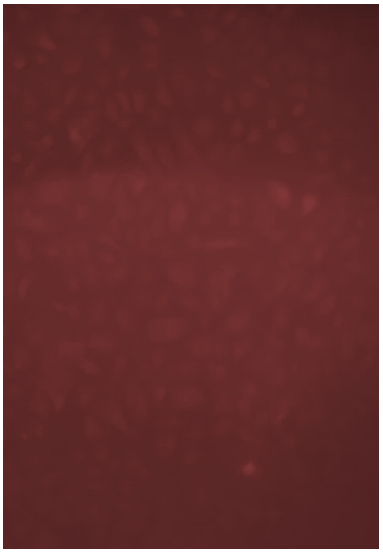


Figure S2: qPCR expression profiling of LEP expression in additional donor cells.

LEP is significantly upregulated in validation donor HBECs depleted of INTS12. Statistical tests were performed comparing to scrambled D-siRNA control: *P<0.05, ****P<0.0001. Individual ΔΔCt gene expressions are *GAPDH* normalized and relative to the mean of the scrambled D-siRNA condition. No significant difference was observed between un-transfected and scrambled D-siRNA transfected cells.

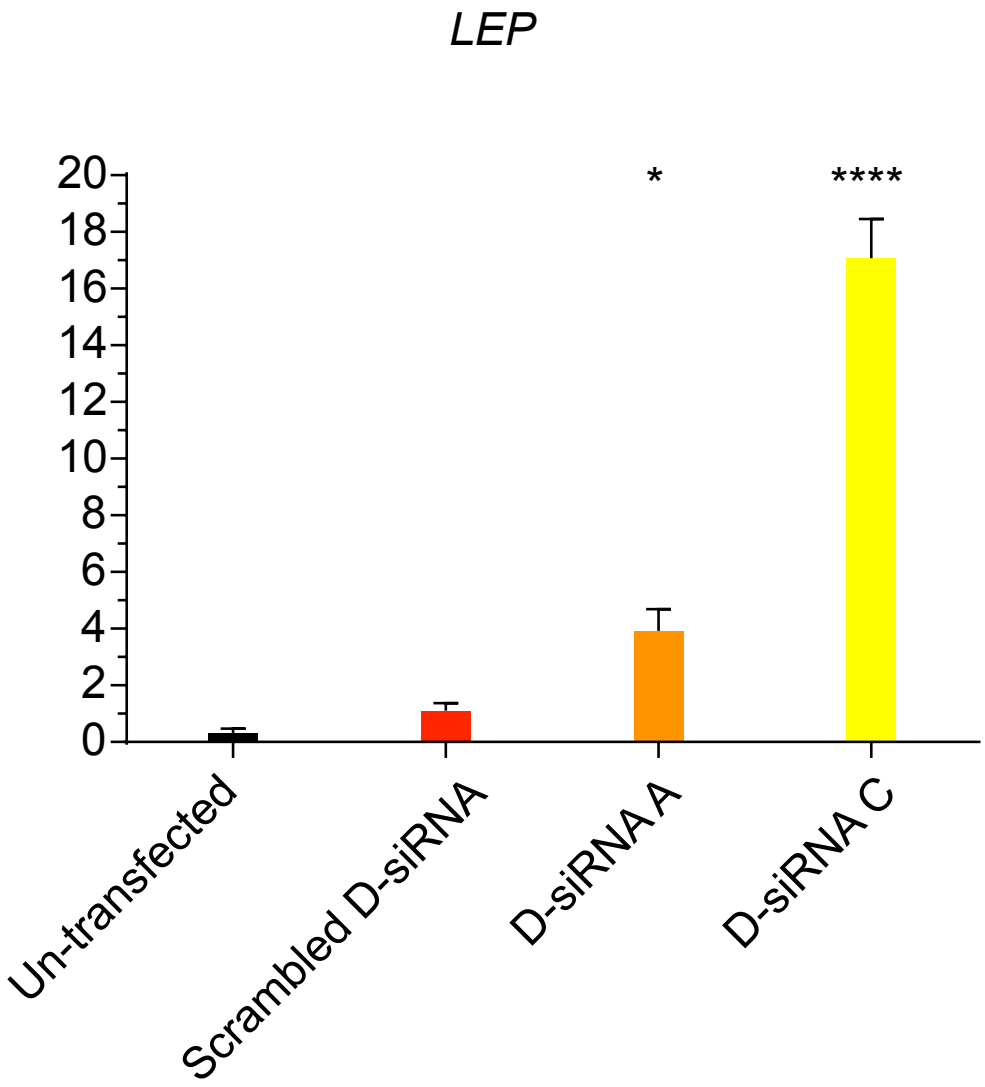


Figure S3: Box plots representing log_{10} of RNAseq FPKM expression values of genes belonging to the top dysregulated pathways.

Stars indicate the GSEA-derived significance of pathway dysregulation.

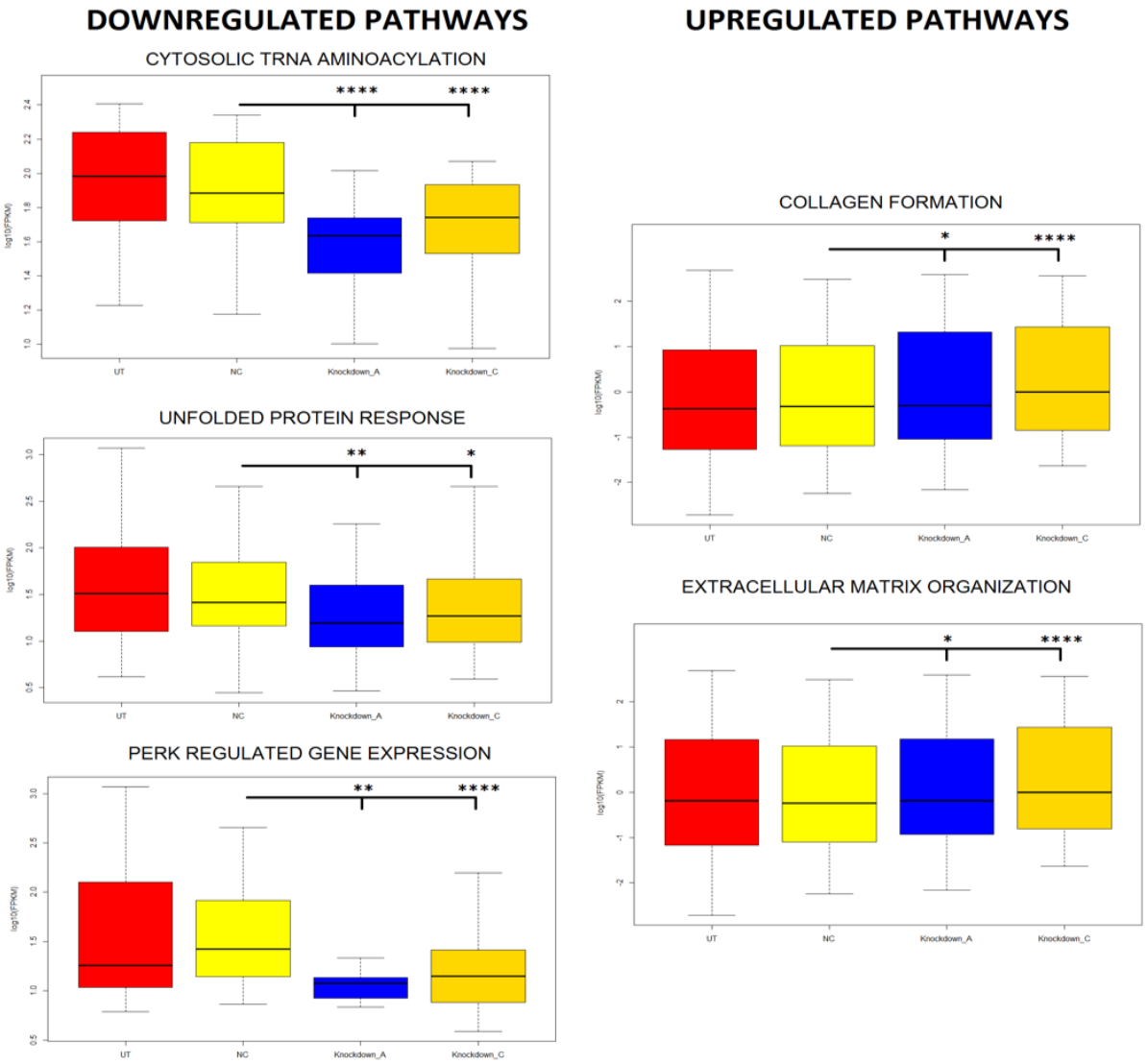
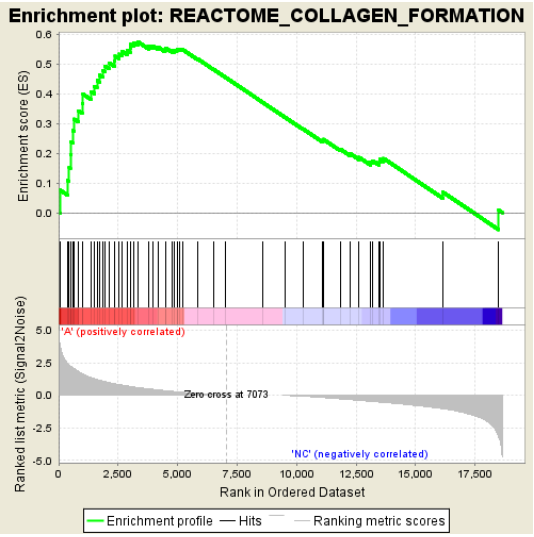


Figure S4: Enrichment plots of pathways upregulated by INTS12 knockdown.

Enrichment plots of reproducibly upregulated pathways in D-siRNA A and C analyses are shown with indicated statistical significance and normalized enrichment scores of their respective upregulations. The FDR and normalized enrichment score values were rounded up to one and three significant figures respectively.

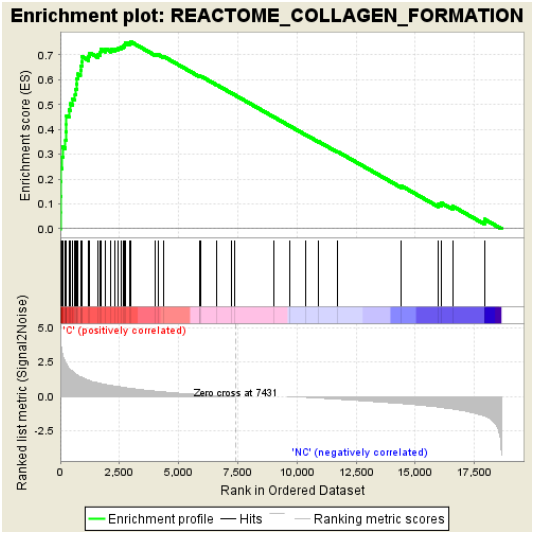
D-siRNA A analysis



FDR = 0.01

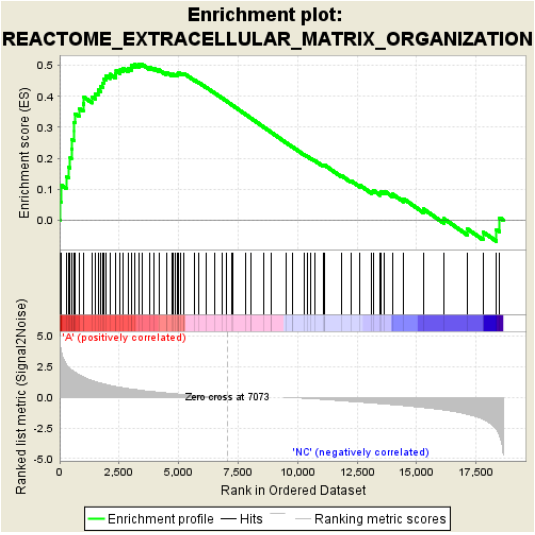
Normalized enrichment score = 1.87

D-siRNA C analysis



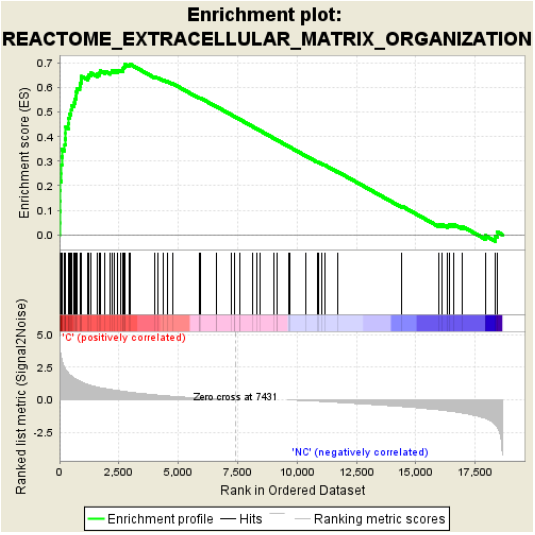
FDR < 0.00001

Normalized enrichment score = 2.42



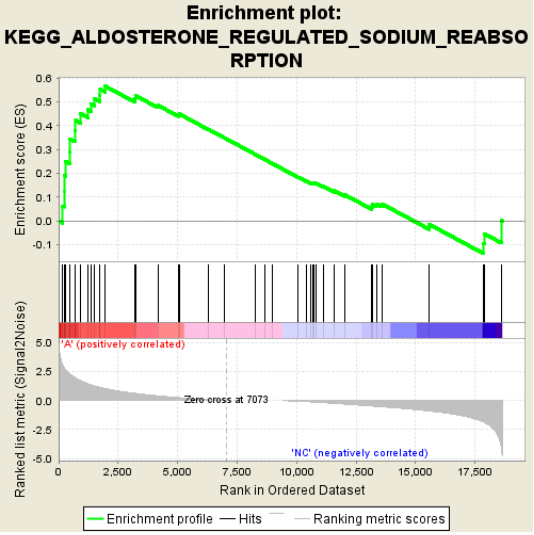
FDR = 0.03

Normalized enrichment score = 1.78



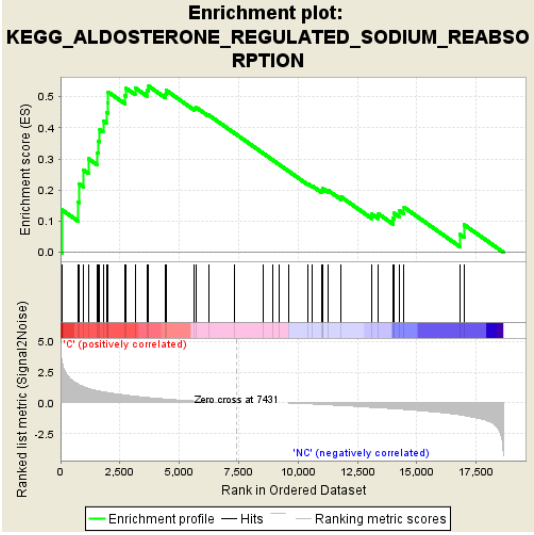
FDR < 0.00001

Normalized enrichment score = 2.40



FDR = 0.04

Normalized enrichment score = 1.74



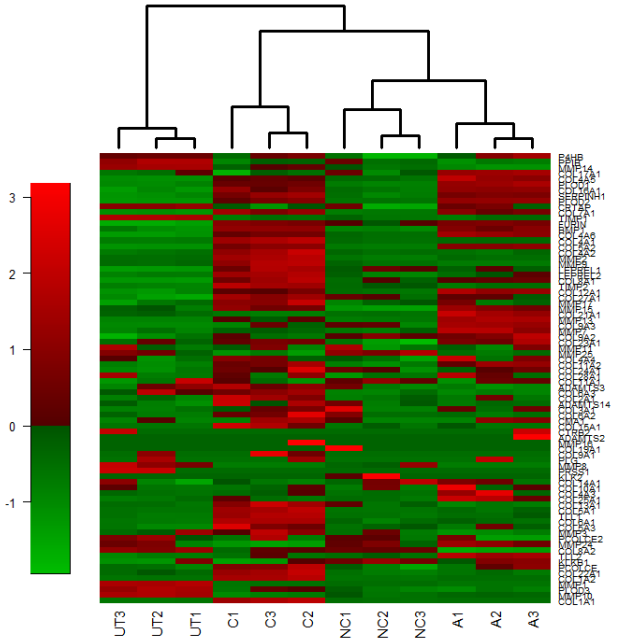
FDR = 0.05

Normalized enrichment score = 1.64

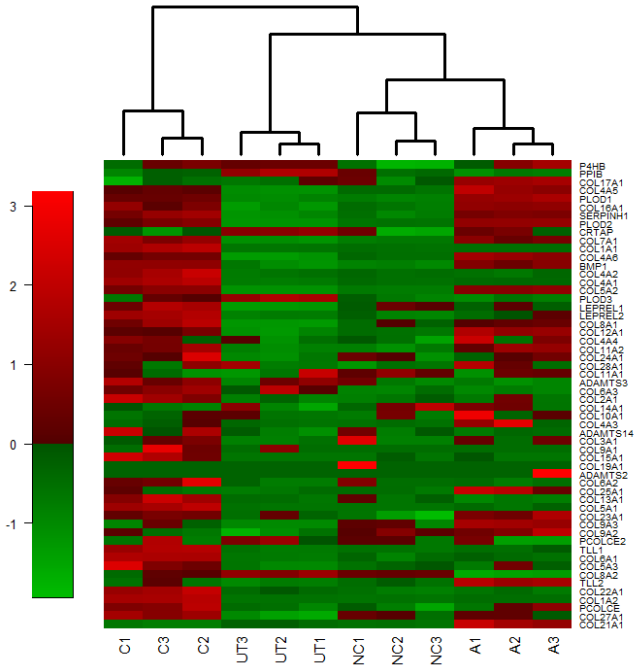
Figure S5: Gene expression heatmaps of genes belonging to reproducibly upregulated pathways.

Green and red colours on the Z-scale indicate lower and higher expression respectively. Samples were clustered by unsupervised hierarchical clustering and resulted in clustering of three biological replicate samples of each of the four conditions: un-transfected cells (UT), cells transfected with scrambled D-siRNA control (NC), cells transfected with anti-INTS12 D-siRNA A (A) and cells transfected with anti-INTS12 D-siRNA C (C).

Extracellular matrix organization (REACTOME)



Collagen formation (REACTOME)



Aldosterone regulated sodium re-absorption (KEGG)

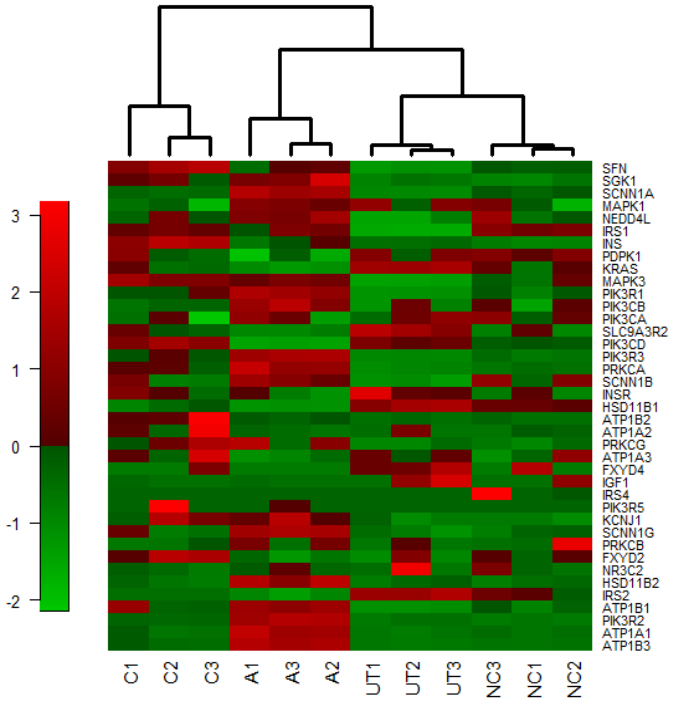
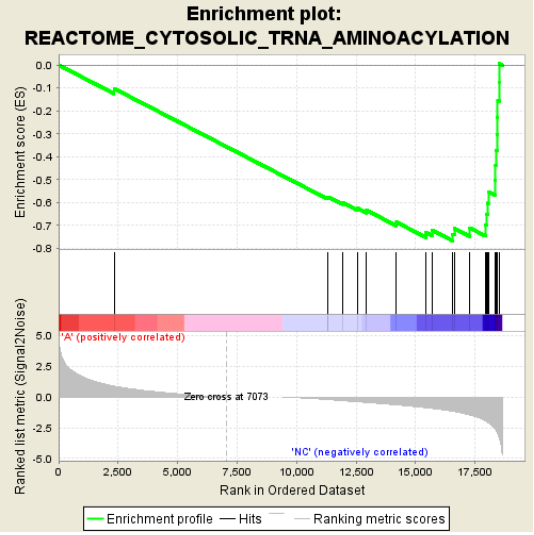


Figure S6: Enrichment plots of pathways downregulated by INTS12 knockdown.

Enrichment plots of reproducibly downregulated pathways in D-siRNA A and C analyses, except cytosolic tRNA aminoacetylation (REACTOME) and PERK regulated gene expression (REACTOME), are shown with indicated statistical significance and normalized enrichment score of their respective downregulations. The FDR and normalized enrichment score values were rounded up to one and three significant figures respectively.

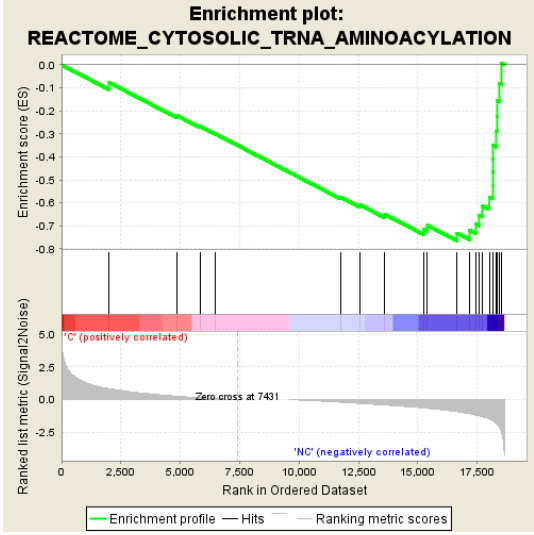
D-siRNA A analysis



FDR = 0.0004

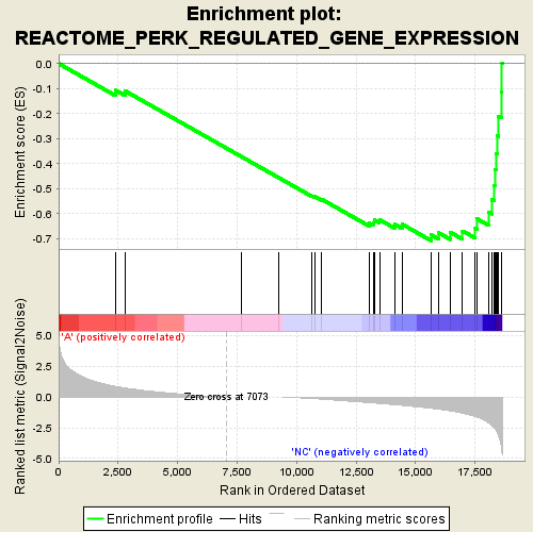
Normalized enrichment score = -2.05

D-siRNA C analysis



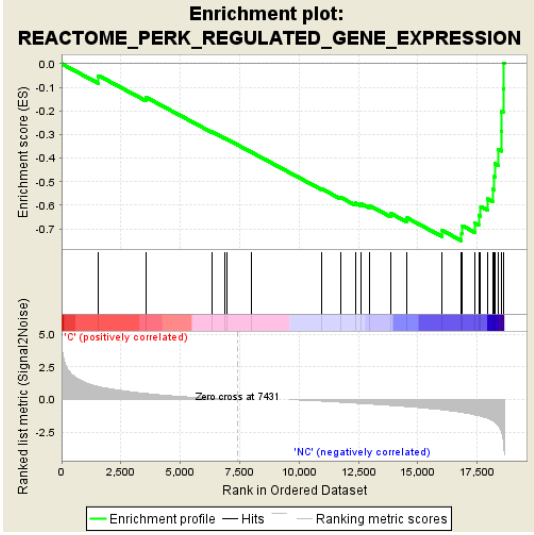
FDR = 0.00009

Normalized enrichment score = -2.10



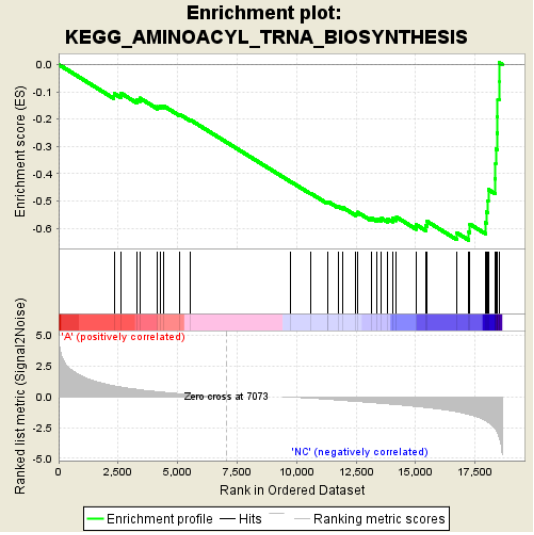
FDR = 0.002

Normalized enrichment score = -1.95



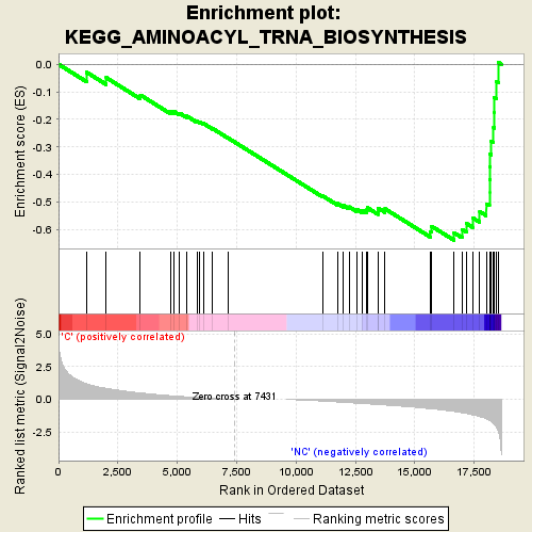
FDR = 0.00006

Normalized enrichment score = -2.12



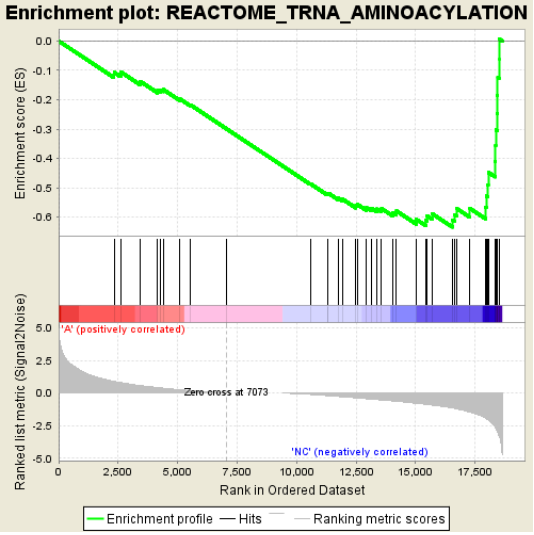
FDR = 0.003

Normalized enrichment score = -1.90



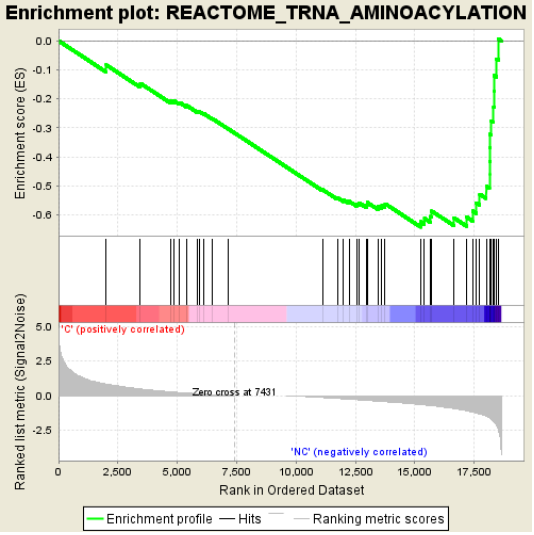
FDR = 0.004

Normalized enrichment score = -1.92



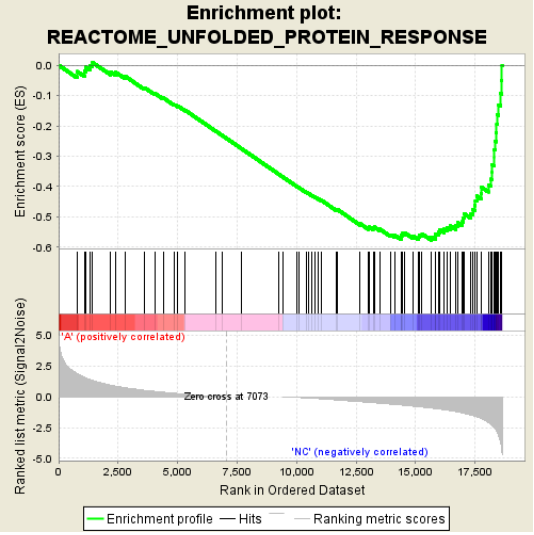
FDR = 0.004

Normalized enrichment score = -1.89



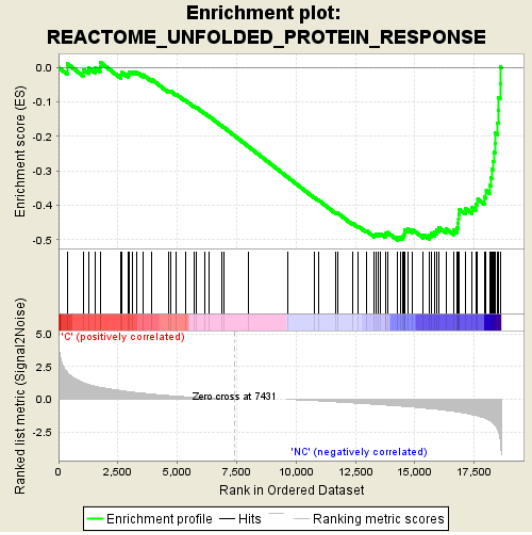
FDR = 0.003

Normalized enrichment score = -1.94



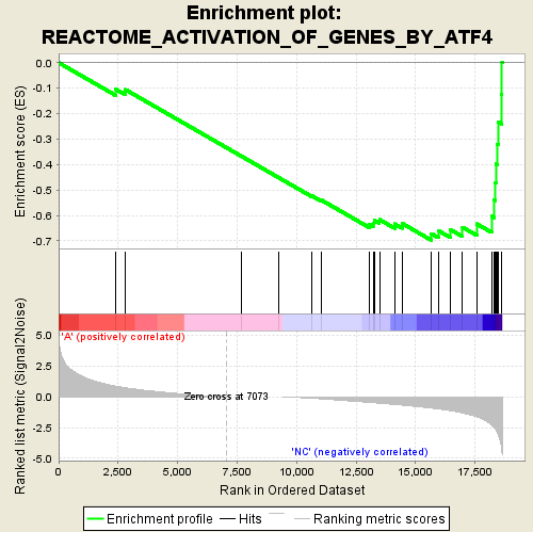
FDR = 0.003

Normalized enrichment score = -1.91



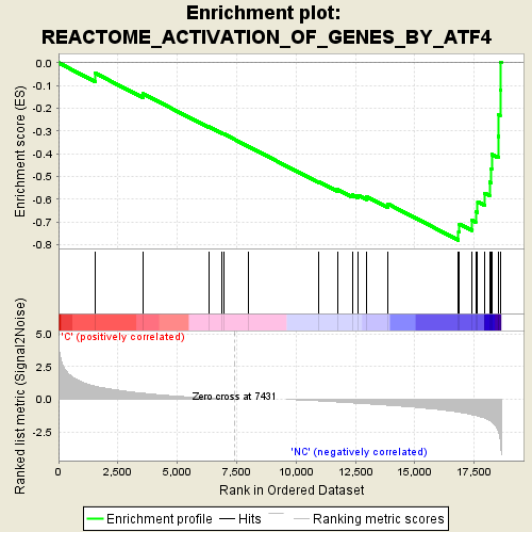
FDR = 0.03

Normalized enrichment score = -1.72



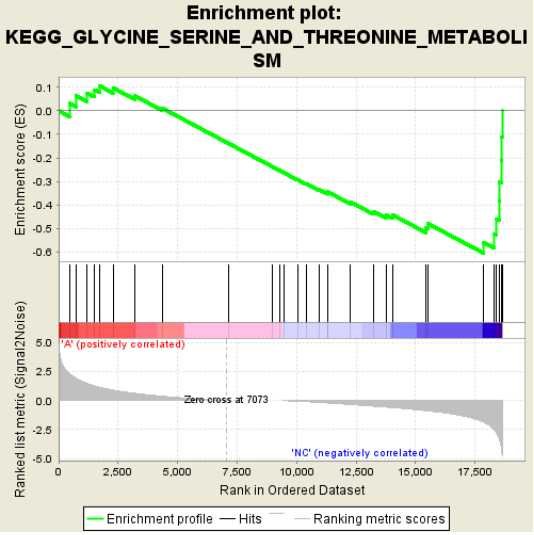
FDR = 0.006

Normalized enrichment score = -1.86



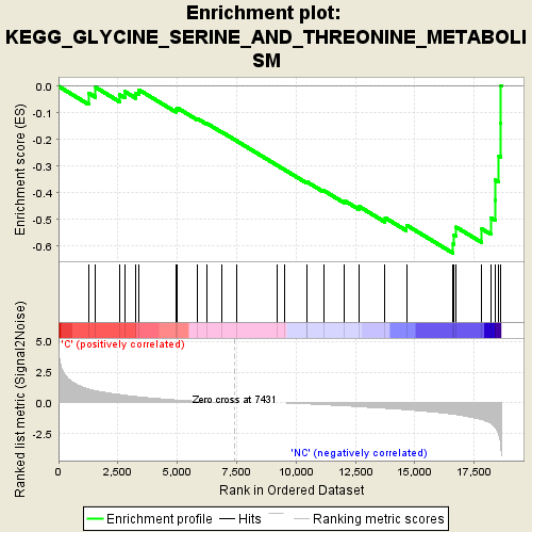
FDR = 0.0001

Normalized enrichment score = -2.10



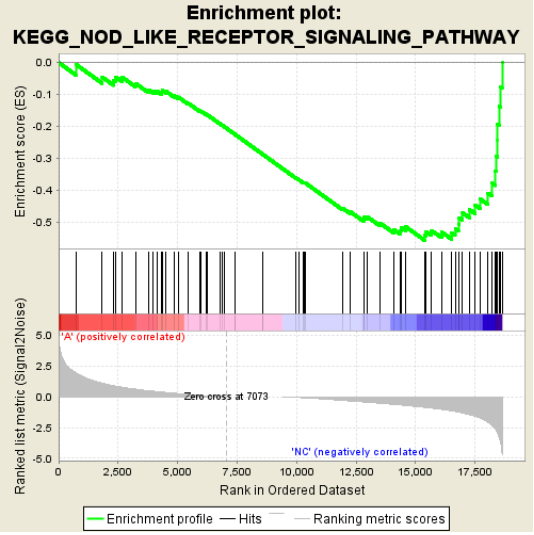
FDR = 0.03

Normalized enrichment score = -1.72



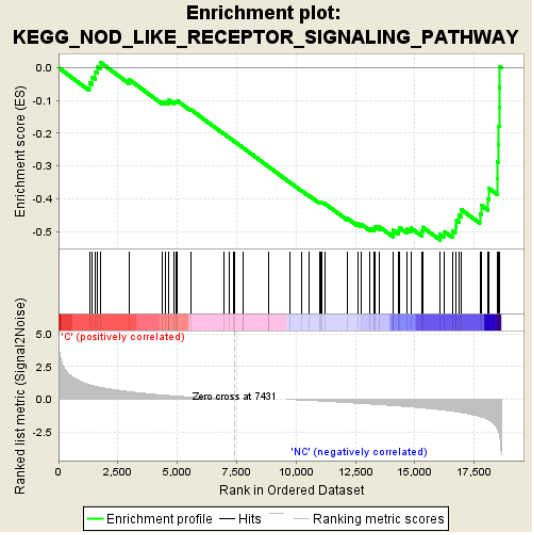
FDR = 0.02

Normalized enrichment score = -1.79



FDR = 0.01

Normalized enrichment score = -1.80



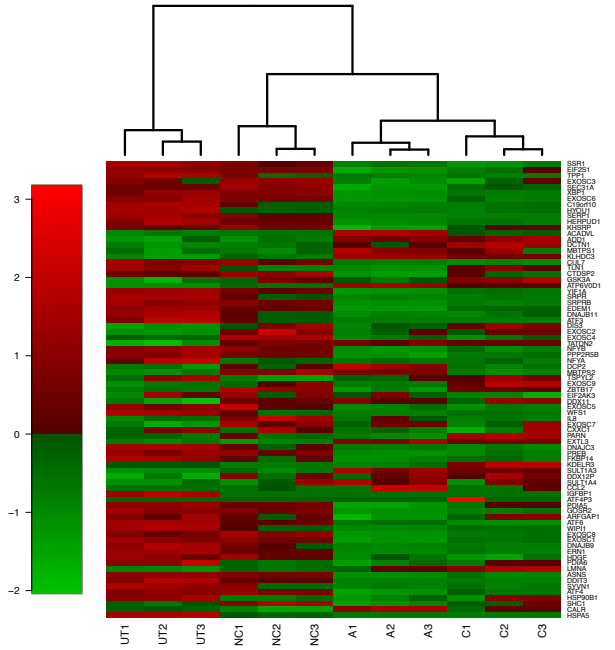
FDR = 0.03

Normalized enrichment score = -1.74

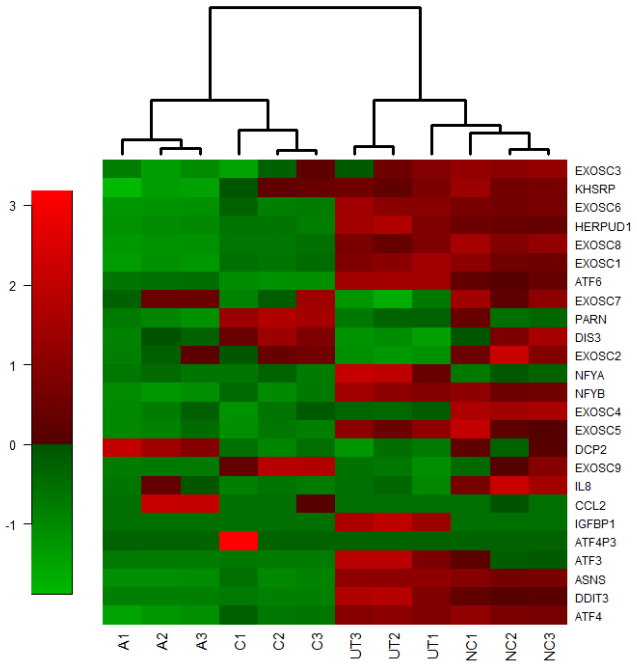
Figure S7 Gene expression heatmaps of genes belonging to reproducibly downregulated pathways.

Green and red colours on the Z-scale indicate lower and higher expression respectively. Samples were clustered by unsupervised hierarchical clustering and resulted in clustering of three biological replicate samples of each of the four conditions: un-transfected cells (UT), cells transfected with scrambled D-siRNA control (NC), cells transfected with anti-*INTS12* D-siRNA A (A) and cells transfected with anti-*INTS12* D-siRNA C (C).

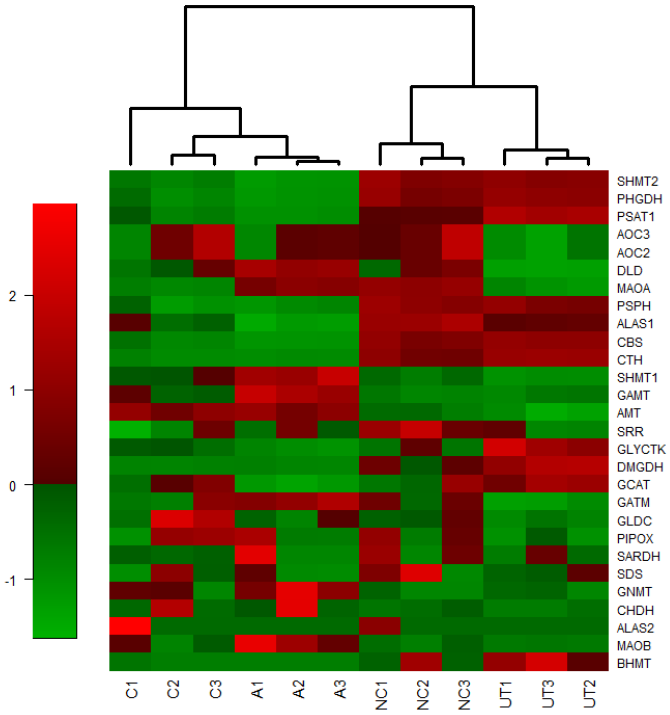
Unfolded protein response (REACTOME)



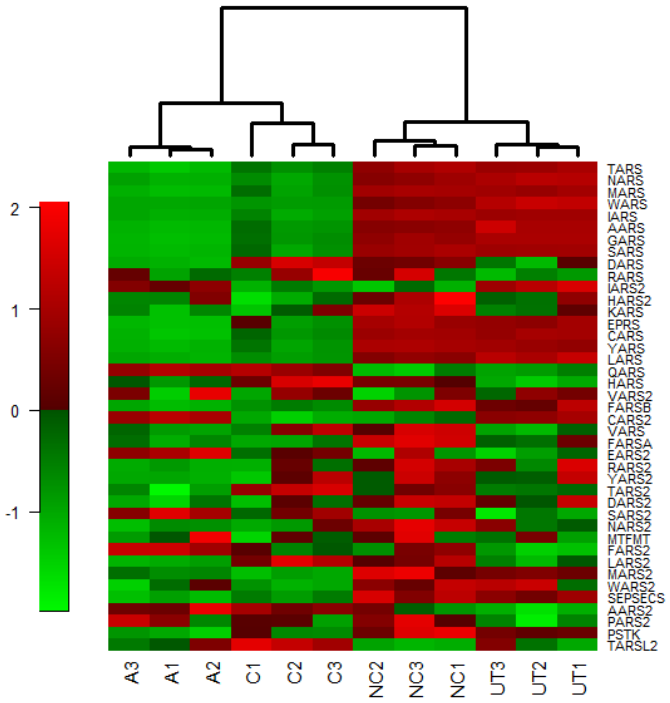
Activation of genes by ATF4 (REACTOME)



Glycine, serine and threonine metabolism (KEGG)



Aminoacyl tRNA biosynthesis (KEGG)



NOD like receptor signalling (KEGG)

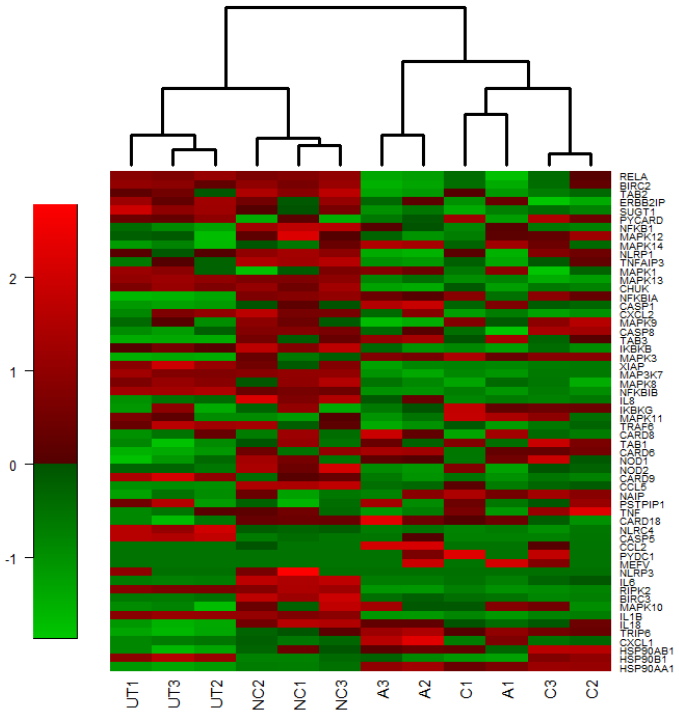


Figure S8: Biological reproducibility of INTS12 knockdown.

Correlation of ChIPseq signals in donor replicates of active regions defined by the start coordinate of the most upstream interval and the end coordinate of the most downstream interval (a union of donor 1 and donor 2 intervals) revealed a Pearson's correlation of 0.85 (P<0.0001).

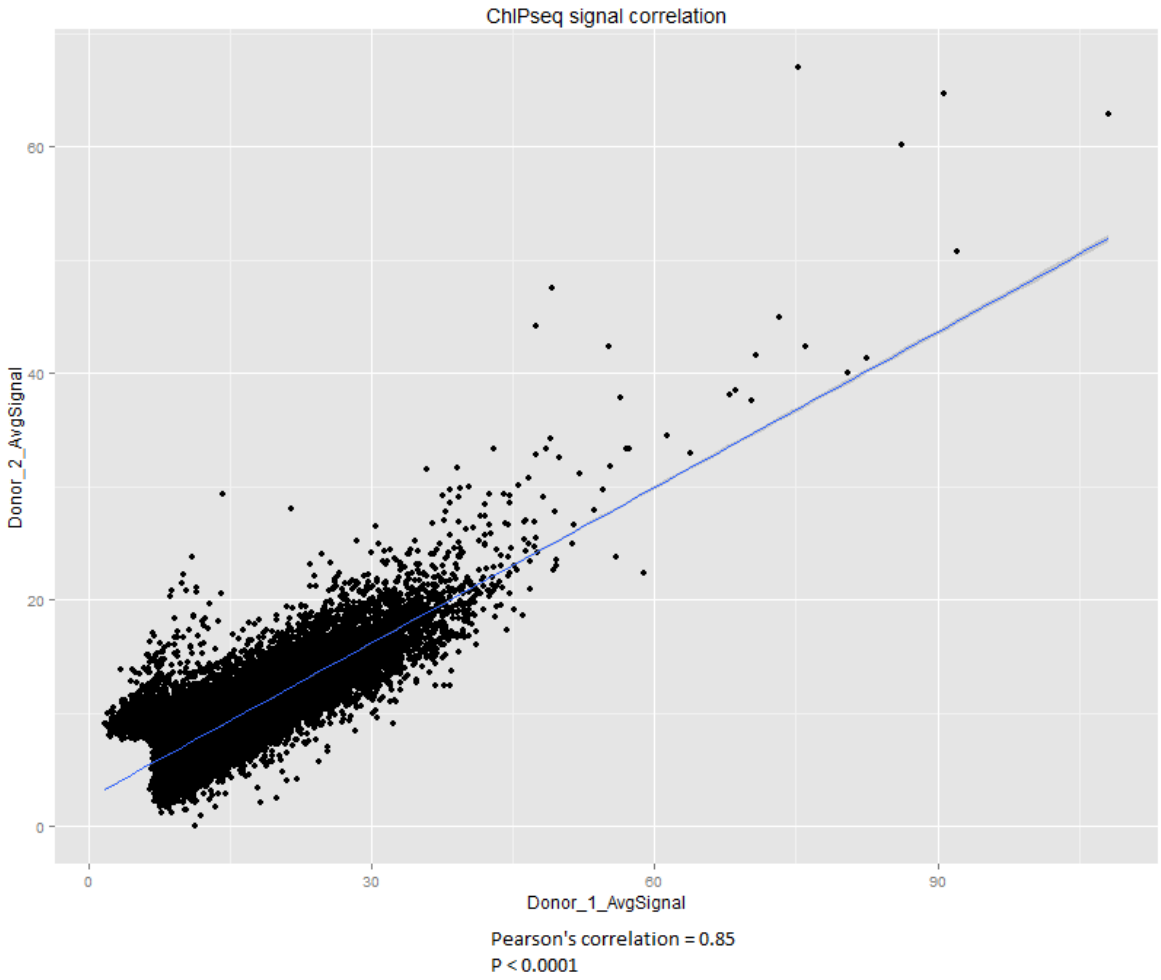


Figure S9: ChIP-PCR validation of ChIPseq findings.

Three ChIPseq positive binding sites (*POR*, *ACTB*, *NBPF1*) shown in green boxes and one negative binding site (*Untr12*) shown in blue box were selected for ChIP-PCR testing to determine the number of binding events detected per thousand donor 1 (D1) and donor 2 (D2) cells. ChIP-PCR results corresponded well with ChIPseq data as seen on the genome browser.

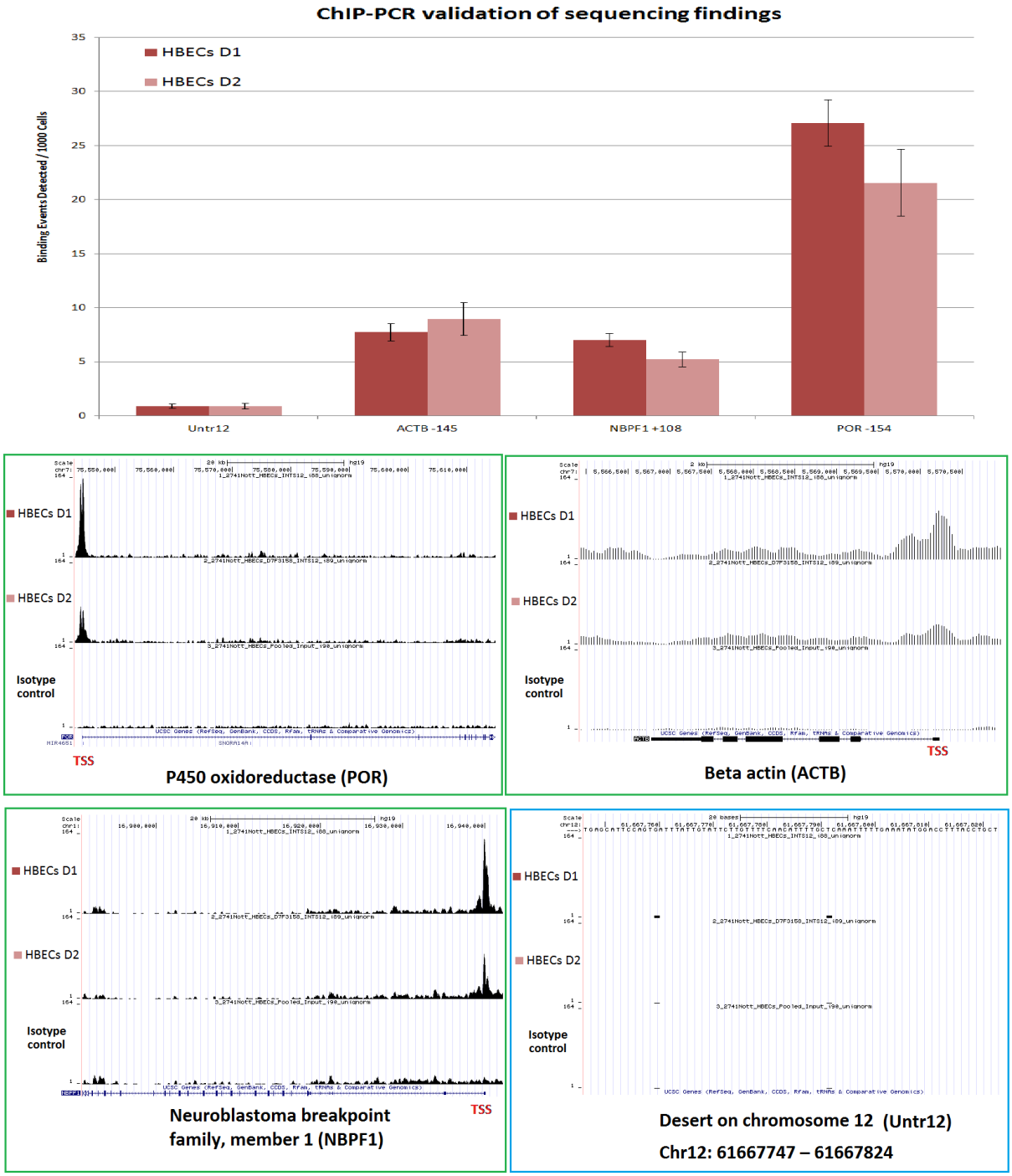


Figure S10: INTS12 ChIPseq peaks over the human genome in donor 1 and donor 2 cells.

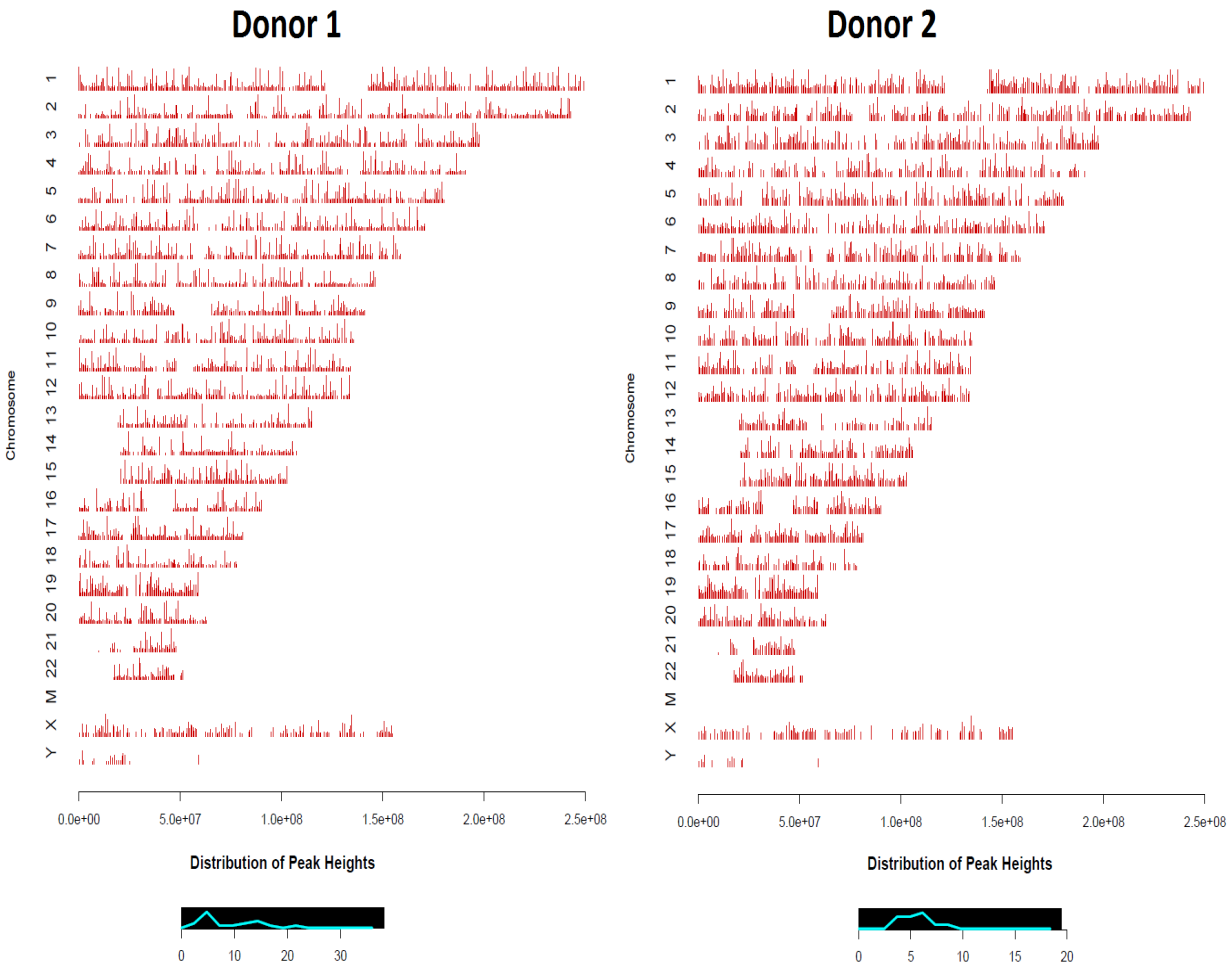
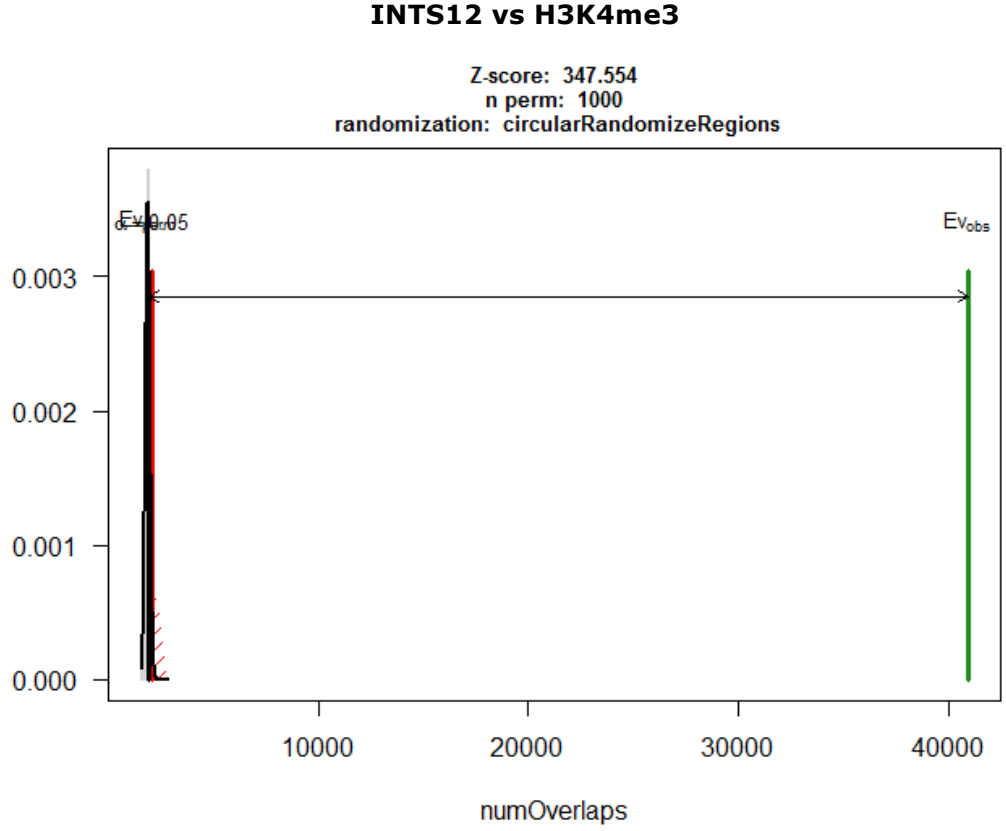
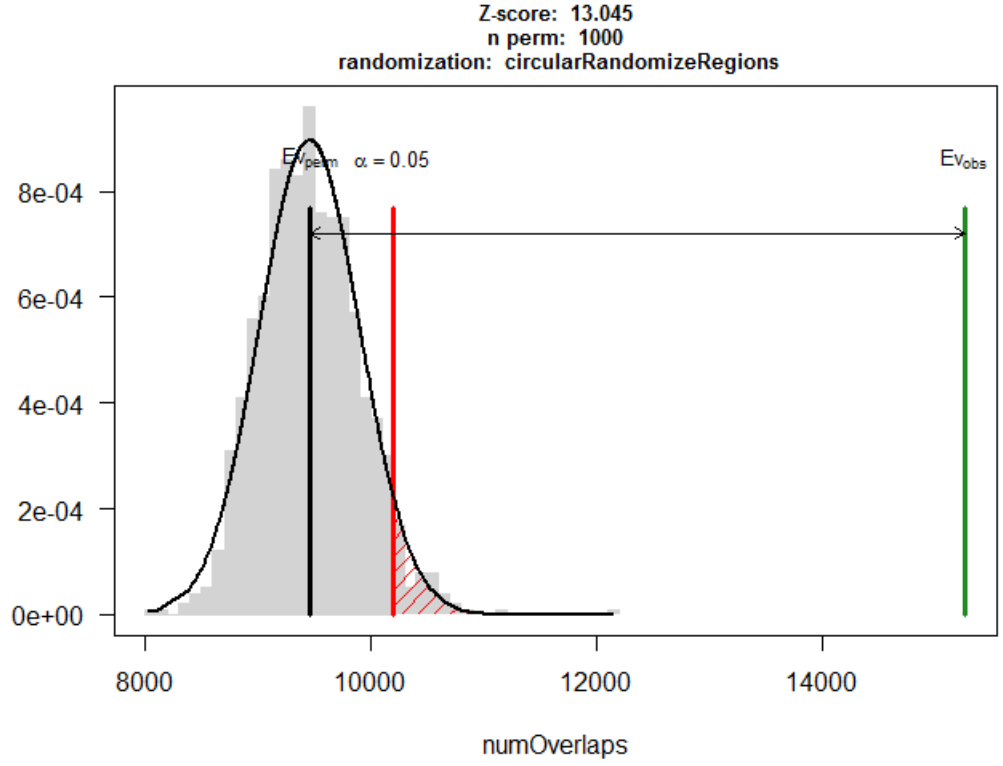


Figure S11: The distance between the average of distribution of intersection in random shuffling of target sites and the observed number of overlaps with INTS12 binding.

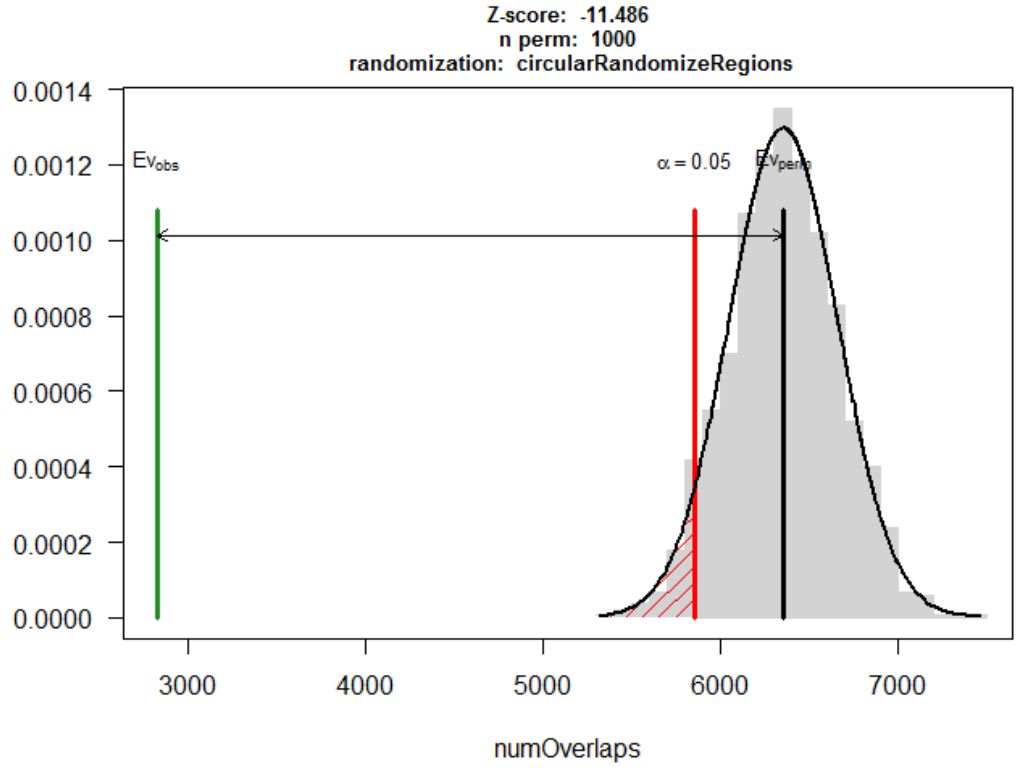
Random walk represents the frequency distribution of overlaps between INTS12 and test regions generated by shuffling H3K4me3, H3K36me3, H3K27me3, DNaseI and CTCF sites in thousand times permutation test. The larger the Z-score distance between the observed and permuted distribution of intersection, the less likely it is to have occurred by chance. Negative Z-score indicates that the observed connection is less than expected by chance.



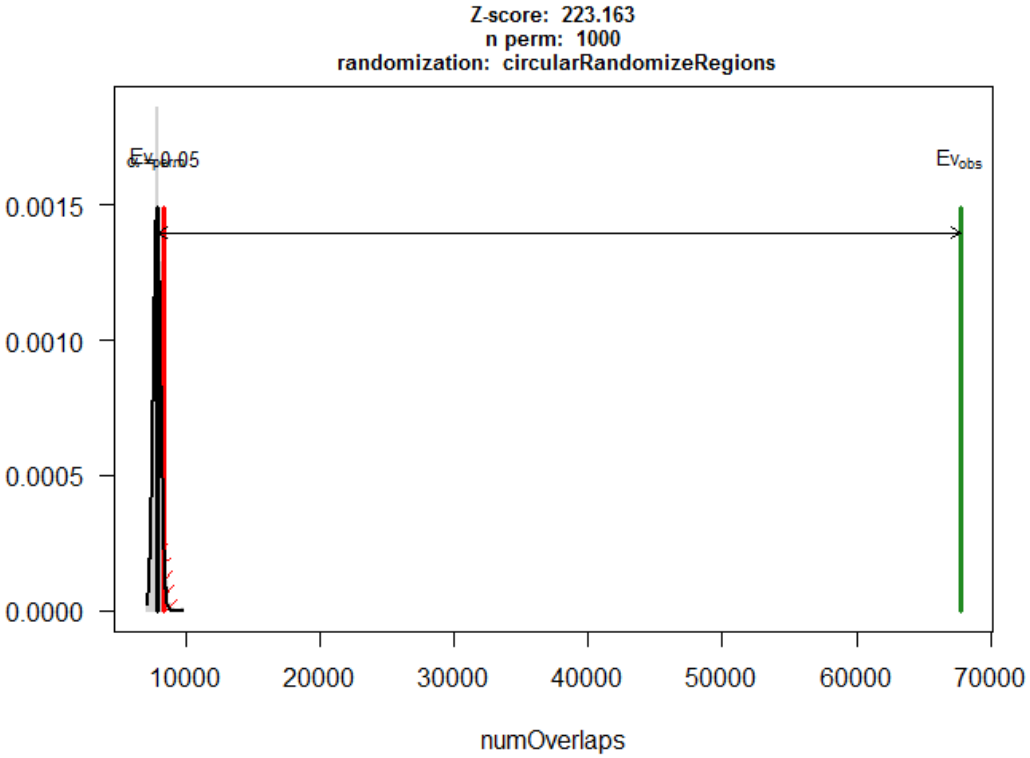
INTS12 vs H3K36me3



INTS12 vs H3K27me3



INTS12 vs DNaseI



INTS12 vs CTCF

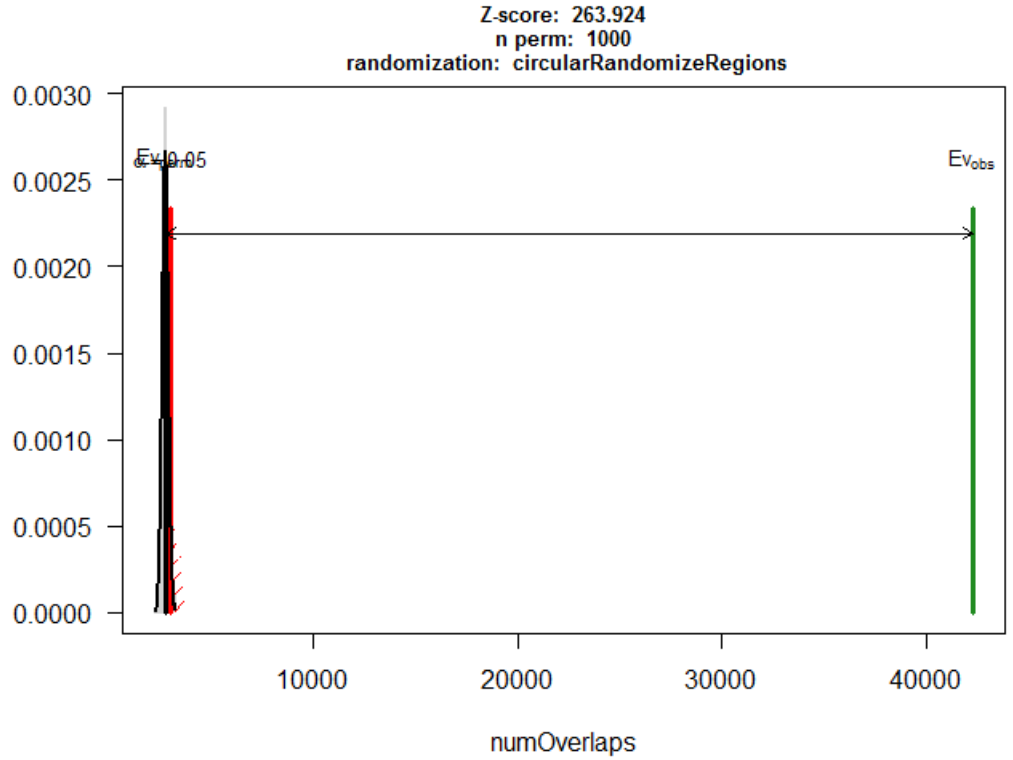
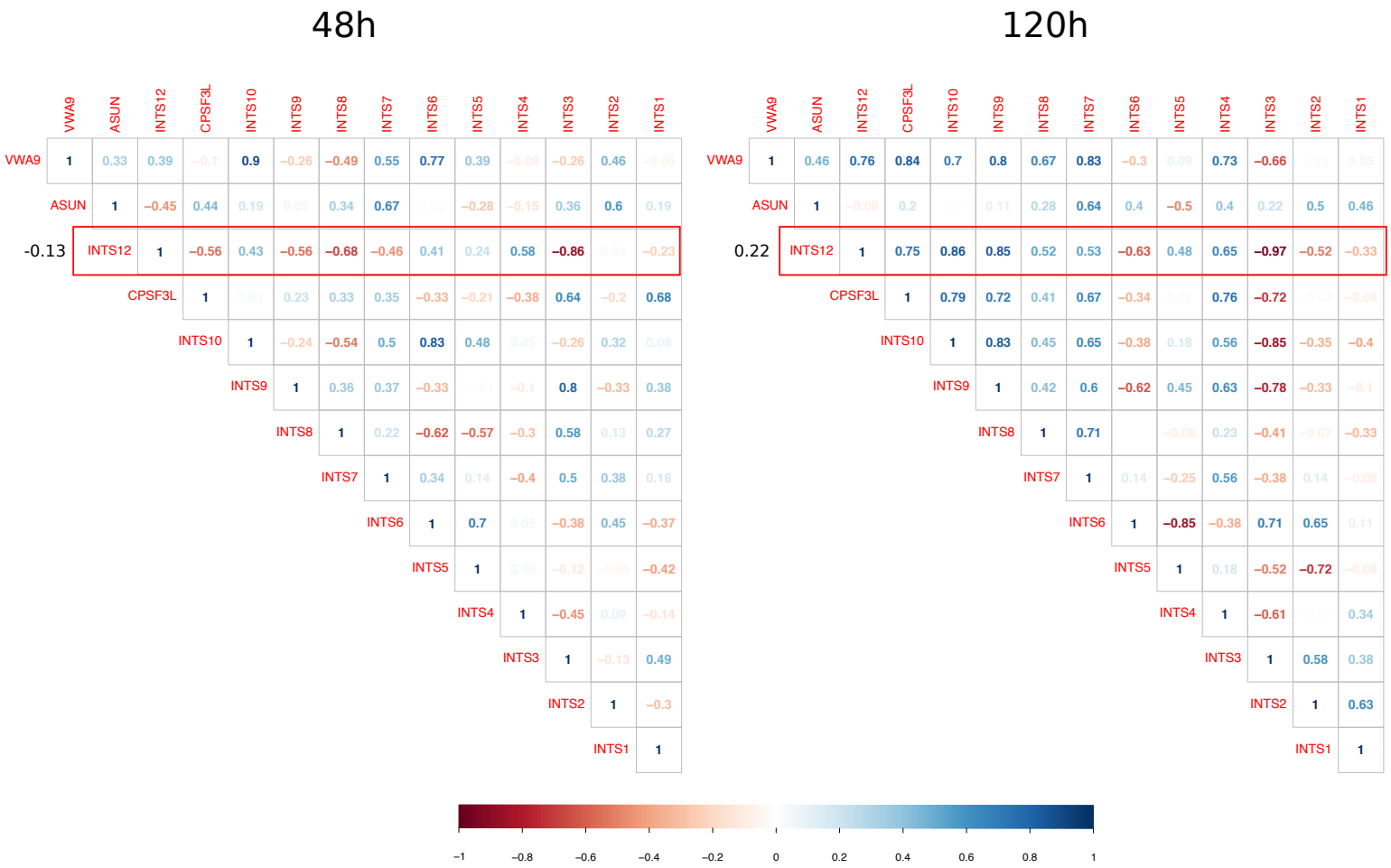


Figure S12: Correlation of INTScom members at 48h and 120h. Numbers and colours are indicative of Pearson's correlation coefficients. INTS12 column is highlighted in red box next to the average of coefficients.

INTS12 appears to have poor correlation with other INTScom members in HBECs suggesting its functional independence from the rest of the complex.



Supplemental Tables

Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways

Kheirallah et al.

Table S1: Hits of the INTS12 BLASTP search against a database of human proteins.

BLASTP search identified significant sequence similarity between canonical full length human INTS12 protein (NP_001135943.1) and PHD finger family.

BLASTP SEARCH HITS						
Description	**Max score**	**Total score**	**Query cover**	**E value**	**Ident**	**Accession**
PHD finger protein 1 isoform b	50.8	50.8	11%	3.00E-06	39%	NP_077084.1
PHD finger protein 1 isoform a	50.4	50.4	11%	4.00E-06	39%	NP_002627.1
PHD finger protein 21A isoform a	45.4	45.4	11%	2.00E-04	40%	NP_001095272.1
PHD finger protein 21A isoform b	44.7	44.7	14%	3.00E-04	37%	NP_057705.3
sp110 nuclear body protein isoform a	42.7	42.7	12%	0.001	40%	NP_004500.3
histone-lysine N-methyltransferase 2A isoform 1 precursor	43.1	43.1	11%	0.001	38%	NP_001184033.1
histone-lysine N-methyltransferase 2A isoform 2 precursor	43.1	43.1	11%	0.001	38%	NP_005924.2
sp110 nuclear body protein isoform c	42.7	42.7	12%	0.001	40%	NP_536349.2
metal-response element-binding transcription factor 2	41.6	41.6	11%	0.002	30%	NP_031384.1

isoform a						
metal-response element-binding transcription factor 2 isoform c	40.8	40.8	11%	0.004	30%	NP_001157863.1
metal-response element-binding transcription factor 2 isoform b	40.8	40.8	11%	0.005	30%	NP_001157864.1
bromodomain adjacent to zinc finger domain protein 2B isoform a	38.1	38.1	27%	0.034	25%	NP_038478.2
bromodomain adjacent to zinc finger domain protein 2B isoform b	38.1	38.1	27%	0.038	25%	NP_001276904.1

Table S2: Summary of proteins with sequence similarity to INTS12.

Summary of proteins showing sequence similarity to human INTS12 provide evidence for putative chromatin and gene regulation roles.

INTS12-SIMILAR PROTEIN SUMMARIES	
PHD finger protein 1 isoform a and b	This gene encodes a Polycomb group protein. The protein is a component of a histone H3 lysine-27 (H3K27)-specific methyltransferase complex, and functions in transcriptional repression of homeotic genes. The protein is also recruited to double-strand breaks, and reduced protein levels results in X-ray sensitivity and increased homologous recombination. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, May 2009]
PHD finger protein 21A isoform a and b	The PHF21A gene encodes BHC80, a component of a BRAF35 (MIM 605535)/histone deacetylase (HDAC; see MIM 601241) complex (BHC) that mediates repression of neuron-specific genes through the cis-regulatory element known as repressor element-1 (RE1) or neural restrictive silencer (NRS) (Hakimi et al., 2002 [PubMed 12032298]).[supplied by OMIM, Nov 2010].
sp110 nuclear body protein isoform a and c	The nuclear body is a multiprotein complex that may have a role in the regulation of gene transcription. This gene is a member of the SP100/SP140 family of nuclear body proteins and encodes a leukocyte-specific nuclear body component. The protein can function as an activator of gene transcription and may serve as a nuclear hormone receptor coactivator. In addition, it has been suggested that the protein may play a role in ribosome biogenesis and in the induction of myeloid cell differentiation. Alternative splicing has been observed for this gene and three transcript variants, encoding distinct isoforms, have been identified. [provided by RefSeq, Jul 2008]
histone-lysine N-methyltransferase 2A isoform 1 and 2 precursor	This gene encodes a transcriptional coactivator that plays an essential role in regulating gene expression during early development and hematopoiesis. The encoded protein contains multiple conserved functional domains. One of these domains, the SET domain, is responsible for its histone H3 lysine 4 (H3K4) methyltransferase activity which mediates chromatin modifications associated with epigenetic transcriptional activation. This protein is processed by the enzyme Taspase 1 into two fragments, MLL-C and MLL-N. These fragments reassociate and further assemble into different multiprotein complexes that regulate the transcription of specific target genes, including many of the HOX genes. Multiple chromosomal translocations involving this gene are the cause of certain acute lymphoid leukemias and acute myeloid leukemias. Alternate splicing results in multiple transcript variants.[provided by RefSeq, Oct 2010]

metal-response element-binding transcription factor 2 isoform a, b and c	No description available
bromodomain adjacent to zinc finger domain protein 2B isoform a	No description available

Table S3: Sequences of D-siRNAs used for INTS12 knockdown.

Oligo	Sequence
D-siRNA #A	5’-GGAAUGGAAAUAGUGGAACAUCAGG-3’
D-siRNA #B	5’-GGCAAUCAAUUAGUAGAAUGUCAGG-3’
D-siRNA #C	5’-GCGUUUAAGAGAACAGAAGUCAAGA-3’

Table S4: Sequences of forward and reverse primers/probes used in snRNA processing and gene expression qPCR assays.

SYBR Green		
Target	**Oligo**	**Sequence**
Immature U1	Forward primer	5′-GATGTGCTGACCCCTGCGATTTC-3′
	Reverse primer	5′-GTCTGTTTTTGAAACTCCAGAAAGTC-3′
Immature U2	Forward primer	5′-TTGCAGTACCTCCAGGAACGG-3′
	Reverse primer	5′-CAGGGAAGCAGTTAAGTCAAGCC-3′
Immature U4	Forward primer	5′-AGCTTTGCGCAGTGGCAGTATCG-3′
	Reverse primer	5′-AGCTTTGCGCAGTGGCAGTATCG-3′
Immature U5	Forward primer	5′-TACTCTGGTTTCTCTTCAGATCGC-3′
	Reverse primer	5′-TTCTATTGTTGGATTACCAC-3′
MARS	Forward primer	5′-TACCCATTACTGCAAGATCC-3′
	Reverse primer	5′-CTTGCTGTTTCAGTACAGTC-3′
GARS	Forward primer	5′-GTGTTAGTGGTCTGTATGAC-3′
	Reverse primer	5′-GTCTTTAAAACTGGCTCAGG-3′
ASNS	Forward primer	5′-GATTGGCTGCCTTTTATCAG-3′
	Reverse primer	5′-AATTGCAAATGTCTGGAGAG-3′
ATF4	Forward primer	5′-CCTAGGTCTCTTAGATGATTACC-3′
	Reverse primer	5′-CAAGTCGAACTCCTTCAAATC-3′
LEP	Forward primer	5′-TCAATGACATTTCACACACG-3′
	Reverse primer	5′-TCCATCTTGGATAAGGTCAG-3′
TaqMan		
INTS12	Forward primer	5′-CTCCAGCTGTCAAAGATCCATT-3′
	Reverse primer	5′-GAGAGCTGCTGGATTCTGAAGT-3′
	Probe	5′-TGGCTGCAAAAGCTGCCCATCCAG-3′

Table S5: Sequences of forward and reverse primers used in ChIP-PCR assays of INTS12 binding to the indicated sites.

Target gene/primer	Primer sequence	Site (hg19)	Type of binding
POR forward	5′-CAGGGTCCGAGCTGTAGAAG-3′	TSS-145	Positive
POR reverse	5′-CCGGCAGAGAAATGAAAGTG-3′		
NBPF1 forward	5′-CACCTACGCCTCCCAGTACC-3′	TSS+108	Positive
NBPF1 reverse	5′-GCCTTGGGTTATCCTGACAC-3′		
ACTB forward	5′-AACTCTCCCTCCTCCTCTTCC-3′	TSS-154	Positive
ACTB reverse	5′-CCTCTCCCCTCCTTTTGC-3′		
Untr12 forward	5′-TGAGCATTCCAGTGATTTATTG-3′	Chr12:61667747–61667824	Negative
Untr12 reverse	5′-AAGCAGGTAAAGGTCCATATTTC-3′		

Table S6: A summary of the survey of scientific literature that investigated the effect of INTScom members on snRNA processing. Ezzedine et al. 2011, Chen et al. 2012 and Chen et al. 2013 studies were undertaken on fly S2 cells while Baillat et al. 2005 study was undertaken on human HeLa cells.

Study	INTScom protein targeted	Effect on snRNA processing
Ezzedine et al. 2011	INTS1	Moderate
Ezzedine et al. 2011	INTS2	Strong
Ezzedine et al. 2011	INTS3	Very weak
Ezzedine et al. 2011	INTS4	Very strong
Ezzedine et al. 2011	INTS5	Moderate
Ezzedine et al. 2011	INTS6	Weak
Ezzedine et al. 2011	INTS7	Weak
Ezzedine et al. 2011	INTS8	Moderate
Ezzedine et al. 2011	INTS9	Very strong (two experiments)
Ezzedine et al. 2011	INTS10	Very weak
Ezzedine et al. 2011	INTS11	Strong
Ezzedine et al. 2011	INTS12	Very weak (one experiment) and weak (another experiment)
Ezzedine et al. 2011	INTS1	Moderate
Ezzedine et al. 2011	INTS2	Strong
Ezzedine et al. 2011	INTS3	Very weak
Ezzedine et al. 2011	INTS4	Very strong
Ezzedine et al. 2011	INTS5	Moderate
Ezzedine et al. 2011	INTS6	Weak
Ezzedine et al. 2011	INTS7	Weak
Ezzedine et al. 2011	INTS8	Moderate
Ezzedine et al. 2011	INTS9	Very strong (two experiments)
Ezzedine et al. 2011	INTS10	Very weak
Ezzedine et al. 2011	INTS11	Strong
Ezzedine et al. 2011	INTS12	Very weak (one experiment) and weak (another experiment)
Ezzedine et al. 2011	INTS1	Above 30 and below 100 fold relative to control

Ezzedine et al. 2011	INTS2	Below 3 fold relative to control
Ezzedine et al. 2011	INTS3	Less than 1 fold relative to control
Ezzedine et al. 2011	INTS4	300 fold relative to control
Ezzedine et al. 2011	INTS5	Below 3 fold relative to control
Ezzedine et al. 2011	INTS6	Below 3 fold relative to control
Ezzedine et al. 2011	INTS7	10 fold relative to control
Ezzedine et al. 2011	INTS8	Below 3 fold relative to control
Ezzedine et al. 2011	INTS9	Above 100 but below 300 fold relative to control
Ezzedine et al. 2011	INTS10	Below 3 fold relative to control
Ezzedine et al. 2011	INTS11	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS12	Below 3 fold relative to control
Ezzedine et al. 2011	INTS1	Above 30 and below 100 fold relative to control
Ezzedine et al. 2011	INTS2	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS3	Less than 1 fold relative to control
Ezzedine et al. 2011	INTS4	Above 300 and below 1000 fold relative to control
Ezzedine et al. 2011	INTS5	Above 10 fold relative to control
Ezzedine et al. 2011	INTS6	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS7	10 fold relative to control
Ezzedine et al. 2011	INTS8	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS9	Above 100 but below 300 fold relative to control
Ezzedine et al. 2011	INTS10	Below 3 fold relative to control
Ezzedine et al. 2011	INTS11	Above 10 but below 30 fold relative to control
Ezzedine et al. 2011	INTS12	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS1	30 fold relative to control

Ezzedine et al. 2011	INTS2	Below 3 fold relative to control
Ezzedine et al. 2011	INTS3	Less than 1 fold relative to control
Ezzedine et al. 2011	INTS4	Above 30 and below 100 fold relative to control
Ezzedine et al. 2011	INTS5	3 fold relative to control
Ezzedine et al. 2011	INTS6	Below 3 fold relative to control
Ezzedine et al. 2011	INTS7	10 fold relative to control
Ezzedine et al. 2011	INTS8	Below 3 fold relative to control
Ezzedine et al. 2011	INTS9	30 fold relative to control
Ezzedine et al. 2011	INTS10	Below 3 fold relative to control
Ezzedine et al. 2011	INTS11	3 fold relative to control
Ezzedine et al. 2011	INTS12	Below 3 fold relative to control
Ezzedine et al. 2011	INTS1	Above 30 and below 100 fold relative to control
Ezzedine et al. 2011	INTS2	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS3	Below 3 fold relative to control
Ezzedine et al. 2011	INTS4	300 fold relative to control
Ezzedine et al. 2011	INTS5	Above 10 fold relative to control
Ezzedine et al. 2011	INTS6	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS7	Above 10 but below 30 fold relative to control
Ezzedine et al. 2011	INTS8	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS9	100 fold relative to control
Ezzedine et al. 2011	INTS10	Below 3 fold relative to control
Ezzedine et al. 2011	INTS11	Above 3 but below 10 fold relative to control
Ezzedine et al. 2011	INTS12	3 fold relative to control
Chen et al 2012	INTS12	Moderate
Chen et al 2012	INTS12	Moderate
Chen et al 2012	INTS9	Above 30 but below 100 fold relative to

		control
Chen et al 2012	INTS9	Above 10 but below 30 fold relative to control
Chen et al 2012	INTS12	Above 1 but below 3 fold relative to control
Chen et al 2012	INTS12	Above 1 but below 3 fold relative to control
Chen et al 2013	INTS12	Moderate
Chen et al 2013	INTS12	Moderate
Chen et al 2013	INTS12	Between 5 to 7 fold relative to control
Chen et al 2013	INTS12	Between 1 to 3 fold relative to control
Chen et al 2013	INTS12	Between 3 to 5 fold relative to control
Chen et al 2013	INTS12	Between 3 to 5 fold relative to control
Chen et al 2013	INTS12	No effect
Baillat et al. 2005	INTS11	3 fold relative to control
Baillat et al. 2005	INTS11	4 fold relative to control
Baillat et al. 2005	INTS1	2 fold relative to control
Baillat et al. 2005	INTS1	4 fold relative to control

Supplemental Methods

Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways

Kheirallah et al.

Cell Culture

Human bronchial epithelial cells (HBEC) were purchased from Clonetics-Biowhittaker (MD, USA). Cells were cultured in HBEC basal medium (BEGM) from Lonza (Berkshire, UK; Product code CC-2540) prepared by addition of all the recommended supplements per manufacturer specifications excluding gentamicin. All laboratory experiments were performed using passage 3 cells. Prior to experiments cells were grown at 37^{o}C with 5% CO_2 until ~95% confluent with BEGM media change every ~48h.

RNAi

Interferin (Polyplus Transfection) was used for gene knockdown optimizations. INTS12 silencing efficiency was tested using D-siRNAs A, B and C (OriGene, Table S3). Subsequently D-siRNAs A and C were tested at 0.1nM, 1nM and 10nM concentrations and a concentration of 1nM was chosen for optimal silencing efficiency. Two D-siRNAs were used in the experiments to account for off-target effects and thus to internally validate our observations. For main RNAseq and functional experiments the effects of INTS12 depletion were assessed 120h after initiation of interference. To ensure appropriate knockdown D-siRNA transfections were administered on two occasions at days zero and three of the experiment. To compare the acute and chronic transcriptomic responses to knockdown, RNAseq profiling was also performed 48h after the initiation of interference. In all experiments there were four experimental conditions: un-transfected cells, cells transfected with scrambled D-siRNA control, and cells transfected with D-siRNAs A and C. Each experimental condition was performed in three independent biological replicates.

RNAseq

Total RNA was extracted using a mammalian total RNA prep kit with on-column DNaseI digestion (Sigma-Aldrich). Sequencing samples were ensured to have RNA integrity number scores greater or equal to 8 (Agilent Technologies). The sequencing libraries were prepared with Illumina TruSeq RNA Sample Prep Kit v2. mRNA was poly-A selected by capturing total RNA samples with oligo-dT coated magnetic beads. The mRNA was then fragmented and randomly primed. cDNA was synthesised using random primers. Finally, a ready-for-sequencing library was prepared by end-repair, phosphorylation, A-tailing, adapter ligation and PCR amplification. Paired-end sequencing was performed on the HiSeq2000 platform (Illumina) using TruSeq v3 chemistry over 100 cycles yielding approx. 40 million reads per sample.

qPCR

Cultured cells were lysed and RNA was extracted using silica-membrane columns (Sigma-Aldrich). 1µg of total RNA was converted to cDNA using the SuperScript synthesis system leveraging random hexamer priming (Invitrogen). Prior to reverse transcription, RNA was treated with DNaseI for a second time to ensure complete removal of any remaining traces of genomic DNA (gDNA). Each reverse transcriptase positive sample had equivalent reverse transcriptase negative control sample. For TaqMan assays (Applied Biosystems) the final volume of qPCR mix per single well was 20µl consisting of 2µl of cDNA template, 6.4µl of DNase and RNase free water, 0.3µM of forward primer, 0.3µM of reverse primer, 0.1µM of probe, and 10µl of x2 TaqMan master mix (Applied Biosystems). For SYBR Green assays the final volume of qPCR mix per single well was 25µl consisting of 5µl of cDNA template, 6.4µl of DNase and RNase free water, 0.25µM of forward primer, 0.25µM of reverse primer, and 12.5µl of x2 Brilliant III Ultra-Fast SYBR Green master mix (Agilent). Reverse transcriptase positive samples were run in triplicate while reverse transcriptase negative samples were run in duplicates. Every qPCR ran had a water only control. qPCR oligo sequences are shown in the Table S4. Housekeeping GAPDH expression was run using pre-developed assay reagents (Life Technologies). QPCR-derived relative to GAPDH and control gene expression was analysed using ΔΔCt method [62]. QPCR technical validation of RNAseq findings was performed using at least three biological cDNA replicates derived from total RNA used in sequencing thus were upon the same donor cells. QPCR biological validation of target genes was performed upon different donor cells with at least three biological cDNA replicates.

RNAseq and Pathway Data Analysis

The quality of raw fastq files was assessed on fastqc. Tuxedo analysis pipeline was used for RNAseq analysis [63]: (1) TopHat read alignment was performed upon hg19 build, (2) Cufflinks transcriptome assembly was performed on individual sample basis and merged by Cuffmerge using reference-based assembly, (3) Cuffdiff differential gene expression was performed using Cuffmerge-predicted annotation. Loci with Benjamin-Hochberg corrected P value [60] below 0.05 were considered significant. Transcriptomic comparisons were performed comparing scrambled D-siRNA to each anti-INTS12 D-siRNA and comparing un-transfected cells with scrambled D-siRNA transfected cells in order to account for off-target and mere transfection effects respectively.

In order to perform pathway analyses, fragments per kilobase per million reads (FPKM) expression values were obtained for each gene per individual RNAseq sample using Cuffnorm. Loci containing multiple amalgamated genes were separated into individual genes and had assigned the equivalent expression values, while genes occurring multiple times on the dataset had their expression values summated using in-house written python script. Scripts can be accessed on GitHub repository (https://github.com/msxakk89/dataset_preperation_scripts). Gene set enrichment analysis using 4722 curated gene sets including 1320 canonical pathway definitions from the Molecular Signatures Database [35] was used, comparing scrambled D-siRNA to each anti-INTS12 D-siRNA and comparing un-transfected cells with scrambled D-siRNA transfected cells. Pathways with Benjamin-Hochberg corrected P value below 0.05 were considered significant. Pathways reproducibly dysregulated by the two different D-siRNA treatments were considered further. Top candidate pathways with the highest enrichment score in both D-siRNAs were chosen for further functional analysis. Results of the pathway analysis were displayed in a Cleveland's plot using ggplot2 R package while pathway heatmaps were drawn using heatplus R package. Boxplots were drawn using build-in R function. Pearson's correlations of gene expression were calculated using hmisc R package and drawn using ggplot2.

Comparison of acute and chronic transcriptomic responses to INTS12 knockdown aimed at identifying core subset genes significantly differentially expressed in 48h and 120h time points respectively. The rational of the analysis was similar to pathway analysis, i.e. genes were shortlisted if were reproducibly dysregulated in both anti-INTS12 D-siRNAs but not in scrambled D-siRNA. Genes that were dysregulated in both anti-INTS12 D-siRNAs in a given direction while in the opposite direction in the scrambled D-siRNA sample were also included.

Core subset of genes was identified by determining the common genes between the 48h and 120h significant gene lists. Enrichment of lung biology relevant gene set was performed via Fisher's exact over-representation analysis using the background of protein coding genes. Correlation of INTS12 with INTScom was calculated by averaged Pearson's correlation over all the complex members.

Protein synthesis by ^{35}S-Methionine incorporation assay

Rates of protein synthesis were measured using EasyTag ^{35}S protein labelling for 10 minutes in labelling medium, followed by lysis of cells in passive lysis buffer (Promega) and TCA precipitation on filter paper as described previously for NIH3T3 cells [64]. Three biological replicates with four technical replicates each were performed. In parallel, the same samples were assayed for total protein using 200µl Coomassie Protein Assay Reagent (Thermo) with 10µl of lysate in microtitre plates and a Synergy HT plate reader (Biotek) at 595 nm. Background for lysis buffer alone was subtracted. For each replicate, the radioactive incorporation was divided by the protein assay measurements thus yielding a measure of incorporation per amount of total protein. Statistical significance of difference in protein synthesis in INTS12 depleted cells was determined by one-way ANOVA analysis of variance followed by Fisher's Least Significant Difference test.

Assessment of proliferative capacity by cell counts

Proliferative capacity was assessed by comparing total cell counts at the beginning and at end of the knockdown, i.e. at the beginning of experiment cells were seeded at the same density in all the conditions. At the end HBECs were washed with PBS, treated with trypsin/EDTA at 37^{o}C for 10min to allow all the cells to detach and were re-suspended in 1ml of culture media. Samples were coded and mixed to perform counting without knowledge of the condition and conditions were decoded later. Cell counts were performed on haemocytometer in technical triplicate per each condition, averaged and total cell count estimates derived accordingly. Experiment was performed in four biological replicates.

ChIPseq

HBECs from two different donors were fixed with formaldehyde solution for 15 min. Formaldehyde solution contained 11% formaldehyde (Sigma), 0.1M sodium chloride (Sigma), 1mM EDTA (Sigma), 50mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Applichem). Fixation was quenched with 0.125 M glycine (Sigma). Chromatin was isolated by the addition of lysis buffer (Active Motif), followed by disruption with a Dounce homogenizer (Active Motif) to allow for efficient chromatin preparation. Lysates were sonicated and the DNA sheared to an average length of 300-500bp. Genomic DNA for each replicate sample was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation (Active Motif). Pellets were re-suspended and the resulting DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield. 30µg chromatin of each sample was precleared with protein A agarose beads (Invitrogen). Unprecipitated genomic DNA (i.e. input control) was prepared from a pool of equal aliquots of the two donor samples. Genomic DNA regions of interest were isolated using 4µg of antibody against INTS12 (Sigma cat. num. HPA03577) following manufacturer's specifications (Active Motif). Complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65^{o}C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Initially a pilot experiment was conducted where DNA libraries obtained from single donor were sequenced on NextSeq 500 sequencing machine (Illumina) yielding ~8 million single-ended 75bp reads in order to assess the success of ChIPseq. For the definite experiment, sequencing libraries (Illumina) were prepared from the both ChIP and input DNAs and the resulting libraries were sequenced yielding ~40 million reads per two ChIP samples from each donor cells and one input control of both donors.

ChIP-PCR

INTS12 peak regions used for qPCR validation were prioritized based on ChIPseq signals observed on the genome browser. Three positive regions and one negative region were chosen for ChIP-PCR validation. PCR primers were designed to span these regions (Table S5). qPCR reactions were carried out in triplicate upon 12.5ng of gDNA from each donor and input control using SYBR Green assay (Bio-Rad). Ct values were converted into the number of binding events detected per 1000 cells according to the manufacturers of ChIP-PCR kit specifications (Active Motif).

ChIPseq Data Analysis

Reads were BWA aligned [65] to hg19 using default settings. Artefactual read duplicates were removed using samtool prior to further analyses. MACS INTS12 peak calling was run on each donor separately comparing ChIPseq samples to input control [66]. Calling was performed with a multiple comparisons corrected P value of less than 0.05 considered as significant. Generated fragment pileup signal was normalized to library size. Fragment pileup was converted to wig files based on fold enrichment above input background for each donor. To compare peak metrics between two donor samples, overlapping intervals were grouped into active regions, which were defined by the start coordinate of the most upstream interval and the end coordinate of the most downstream interval. In locations where only one sample had an interval, this interval defined the active region. ChIP signal at these active regions was compared between the two donor samples and correlation drawn and calculated by ggplot2 and rcmdr R packages respectively. Intervals were annotated, percentage of total INTS12 binding sites falling on the fixed annotated genomic features and enrichment over meta-gene body determined using CEAS package [67]. The proportion of binding proximal to TSS was calculated by dividing the number of significant peaks close the TSS (TSS±1000bp) by the number of significant peaks falling within the broader region surrounding the TSS (TSS±3000). Enrichment over various gene classes, expressed/not expressed, or differentially expressed genes was drawn using ngs.plot [68]. Gene classes were retrieved using Ensembl's BioMart tool. HOMER and MEME were used for *de novo* identification of enriched DNA motif at INTS12 binding sites [51, 52]. TomTom was used to compare *de novo* identified motif to a set of currently known motifs [53]. BETA was used to predict INTS12 regulatory function [49].

ENCODE data retrieval and analysis

Airway epithelial cells specific epigenetic and CTCF ChIPseq datasets were obtained from ENCODE data repository (ENCBS417ENC; www.encodeproject.org) and analysed as INTS12 ChIPseq datasets with the only difference that broad region calling was used for the epigenetic marks. Percent of overlap between INTS12 intervals and ENCODE intervals and its statistical significance was determined using regioneR R package with random permutation test. Correlation of ChIPseq signals and conservation of binding analyses were performed using cistrome [69].

Immunofluorescence

Cells were grown on 8-chamber glass slides seeding 8000 cells onto each chamber and were left un-treated or were transfected with anti-INTS12 and scrambled D-siRNAs as described previously. Cells were fixed using 4% formaldehyde and blocked/permeabilized with PBS, 10% goat serum, 1% BSA, and 0.15% Triton-X. Cells were incubated with antibody against INTS12 (Sigma cat. num. HPA03577) at 4°C overnight and rhodamine-TRITC labelled secondary for 1 hour at room temperature. Controls were incubated with primary isotype control (Abcam) antibody followed by secondary antibody. Cells were visualized epifluorescently and exposures were kept constant across the conditions to avoid artefactual differences in the observed fluorescence intensity.