SUPPLEMENTAL INFORMATION

Optimization of reverse transfection

A GFP expression vector was used to assess the performance of various transfection protocols. We tested multiple transfection reagents including Lipofectamine 2000, Fugene 6, Fugene HD, TransIT-Neural and TransIT-2020, and a wide range of the reagent to DNA ratios, and cell densities. We measured GFP expression 48 hours post-transfection. Cells were maintained in media without phenol red. The starting cell number appeared to have less impact on GFP expression for Lipofectamine transfection than for Fugene HD transfection. We thus opted for Lipofectamine 2000 for our screens. We also monitored the kinetics of GFP expression after Lipofectamine transfection. GFP expression increased steadily over time until 40 hours post-transfection and started to plateau at 48 hours (see Supplementary Fig. 1B).

Data acquisition and background subtraction

Image acquisition must be rapid enough to minimize changes in GFP/RFP during the scan of a whole library. The speed of scanning is inversely correlated with the pixel size of an acquired image. An optimal scan would take the shortest time but still produce an image resolution of sufficient data quality. To examine the effect of pixel size on data quality, we measured the consistency of signal intensities acquired at different pixel sizes. We randomly plated different numbers of pflareG-exon 18 cells per well to generate a range of fluorescence intensities, and measured GFP and RFP at different resolutions. The whole well fluorescence signals (i.e. the sum of pixel intensities within a well) from scanning at 200 µm pixels showed almost perfect linear correlation with signals from a 100 µm scan (Supplementary Fig. 4A-B, R^2=0.999 and 0.997 for the GFP and RFP channels). The regression coefficients are 3.9248 for the GFP channel and 4.0054 for the RFP channel, consistent with a 100 µm scan producing four times more pixels per well than a 200 µm scan. At a pixel size of 200 µm, the whole 44-plate MGC library can be scanned within 100 minutes for both channels. Scanning at larger pixel sizes (e.g. 500 µm) greatly increased data variance, particularly at low fluorescence intensities. This higher variance is presumably due to the lower number of pixels per well, as a single well from a 384-well plate is covered by 49 500-µm-pixels, and by 256 200-µm-pixels.

To test the uniformity of fluorescence signals, we added uniform media to a 384-well plate and measured well intensities (i.e. background signals for the HTS) at different positions on the scanner. Individual well intensities varied by 4-7% at different positions. This small variation can be further diminished by plate-to-plate normalization.

We next examined potential sources of fluorescent noise and variation to allow proper background correction in the HTS. We found that the cellular autofluorescence of the parental N2a cells was about 1.9% the level of the media fluorescence in the GFP channel and 6.6% the media fluorescence in the RFP channel. Thus, relative to the media fluorescence, the native cellular autofluorescence is negligible. However, we found that media fluorescence was subject to two forms of variation. The average

absolute signal intensity varied from plate to plate, and intensities varied between the individual wells within a plate.

We found that fluorescence intensities vary substantially between individual wells of a 384-well plate containing equal volumes of media in each well (Supplementary Fig. 4C-D). Most obviously, wells on the edge of a plate had higher signals in both the GFP and RFP channels than inner wells (up to 30%). These well-to-well differences increased as plates were incubated for longer times, probably due to evaporation of the media along the edges. These well effects were very reproducible from plate to plate, such that fluorescent signals of individual wells in one plate correlated strongly with the signals of corresponding wells in another equivalent plate (Supplementary Fig. 4E-F). This indicates that well-specific media background can be calculated from a media-only plate and used for subtraction of the media background across the multiple plates of the library. To test this, we seeded equal numbers of pflareG-exon 18 cells into each well of a plate, except wells A23, A24, B23 and B24, which contained media alone. We also generated a background plate containing media alone in all wells. To correct the inter-plate variation, the intensity of each of the four media-alone wells for the cell plate was divided by the intensity of the corresponding well from the background plate. These four values were then averaged to derive a scaling factor for the cell plate. Each well-specific media fluorescence from the background plate was multiplied by the scaling factor and subtracted from the total fluorescence of the corresponding well in the cell plate to allow inter-plate normalization and well-by-well background subtraction. We found that if the averaged signals from media alone wells from each cell plate were subtracted as a single background value for that plate, the cell plates exhibited significant well-to-well variation, with well intensities ranging from 20% below to 40% above the mean intensity (Supplementary Fig. 4G-H). The CV for corrected well intensities was 9.5% for the GFP channel and 7.4% for the RFP channel. In contrast, the intensities were far more uniform across the plate after the inter-plate well-by-well background correction (Supplementary Fig. 4I-J). The CV dropped to 4.5% for the GFP channel and 4.4% for the RFP channel. We applied the inter-plate background subtraction in quantifying the well intensities measured in all plates of the screen.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Optimization of reverse transfection. (A) N2a cells were plated in 384 well plates on top of a previously deposited GFP expressing plasmid under various conditions. Transfection efficiency, gauged by GFP expression level, varied with the transfection reagent, its ratio to the DNA plasmids and the cell number. Transfection reagent to DNA ratios are shown on X axis. L2k: Lipofectamine 2000. F6: Fugene 6. FHD: Fugene HD. Neu: TransIT-Neural. 2020: TransIT-202. Cell number is color coded and shown on the right. (B) N2a cells were reverse transfected with a GFP expressing plasmid using lipofectamine 2000. GFP expression was measured over time and plotted.

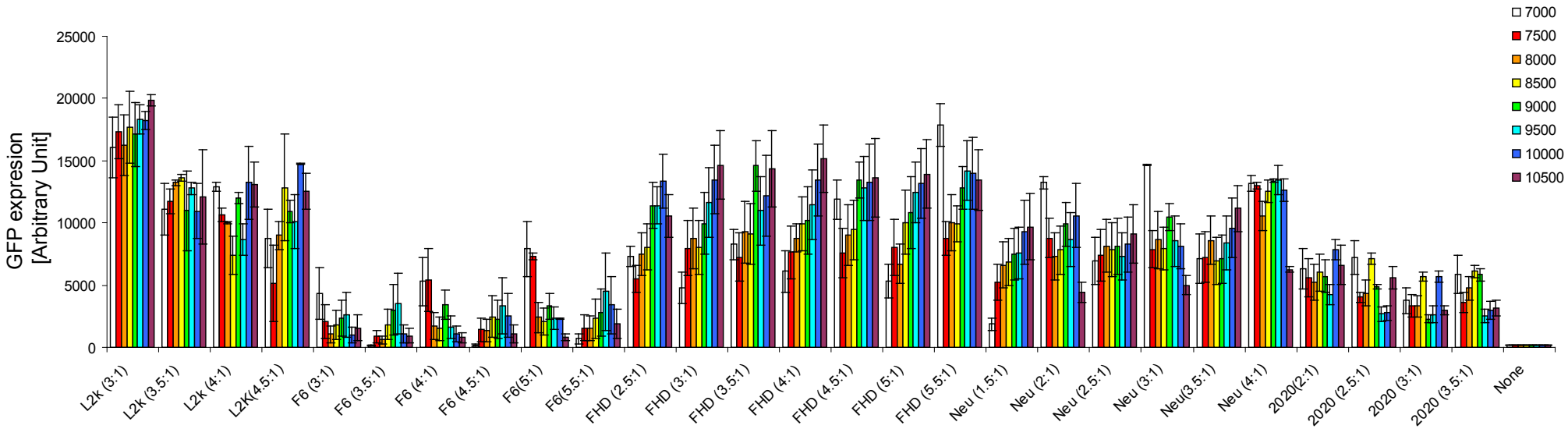
Supplementary Figure 2. Splicing of endogenous *Dlg4* exon 18 in various reporter cell clones.

Supplementary Figure 3. RT-PCR validation of changes in pflareG-exon18 splicing after transient transfection of the identified genes. The percent spliced in (PSI) values are indicated below the gels.

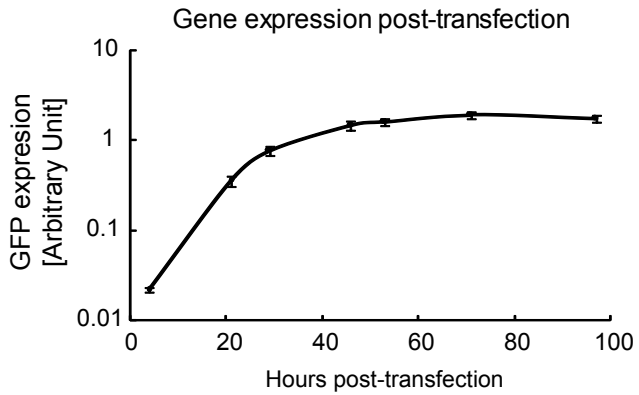
Supplementary Figure 4. Data acquisition and well-specific background subtraction. Data quality of well-based fluorescence signals was preserved when the scanning pixel size increases from 100 µm to 200 µm for both GFP (*A*) and RFP (*B*) channels. Heat maps of GFP (*C*) and RFP (*D*) signals indicated well-to-well background variation within a 384-well plate containing equal media volumes in each well. The value of each well was the ratio of its signal intensity to the mean intensities of all wells within the plate. The well-specific backgrounds were consistent from plate to plate for GFP (*E*) and RFP (*F*), as the well signals of one plate (e.g. plate #2) were correlated with the corresponding well signals of another plate (e.g. plate #1). This allowed for interplate background subtraction. Heat maps of the GFP (*G* and *I*) and RFP (*H* and *J*) signals from a uniform cell plate showed higher well-to-well variation with intraplate background correction (*G* and *H*) than with interplate background correction (*I* and *J*). Well value was normalized by the mean intensities of all wells.

Supplementary figure 1

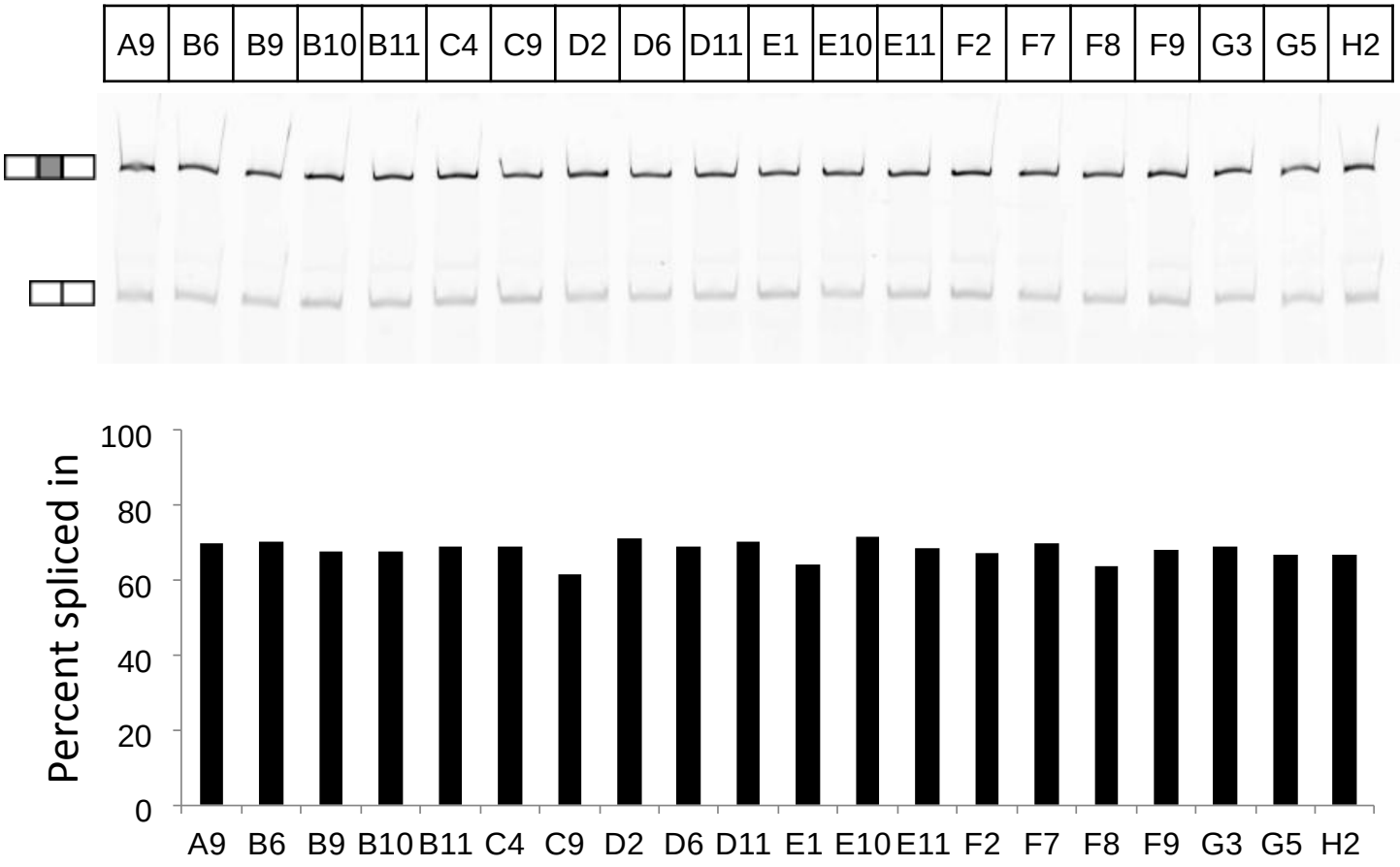
A



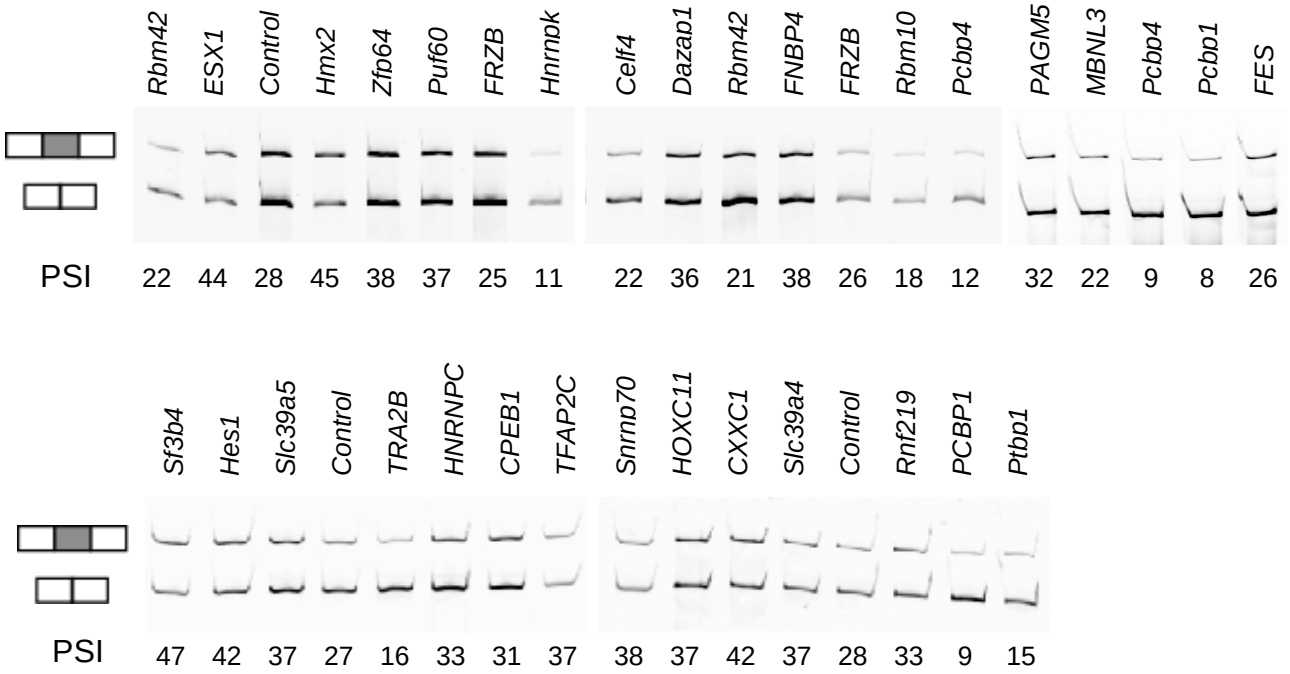
B



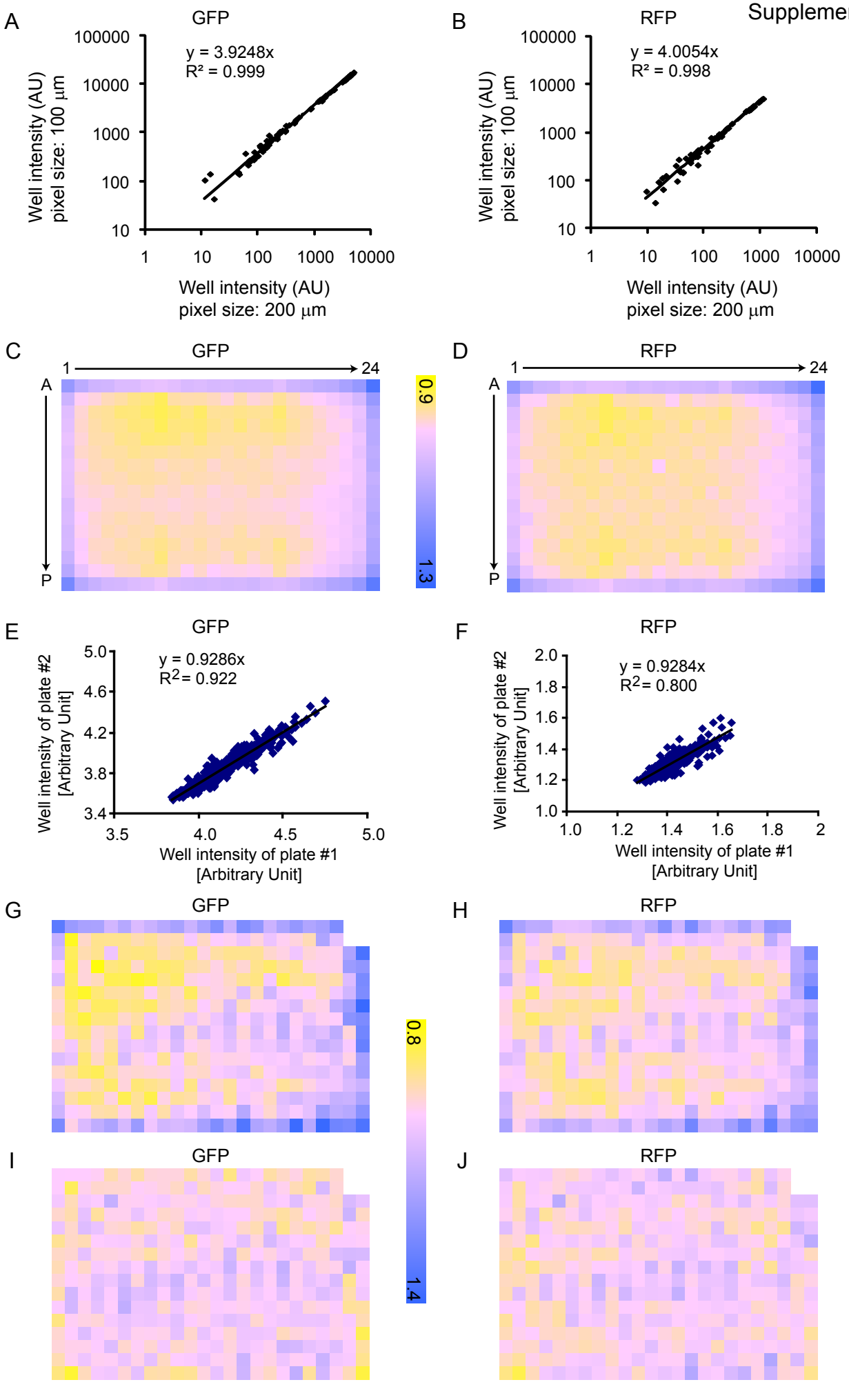
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary table 1. Enriched terms of overlapping hits

Category	Term	Count of genes	%	Fold Enrichment	Bonferroni corrected p-value	Genes
GOTERM_BP_FAT	mRNA processing	14	35.0	13.96	7.97E-10	*HNRNPC, Snrnp70, SF1, PUF60, Celf4, SCAF1, SRSF3, Sf3b4, Ptbp1, CPEB1, PCBP1, TRA2B, Ptbp3, Hnrnpk*
GOTERM_BP_FAT	mRNA metabolic process	14	35.0	12.44	3.47E-09	*HNRNPC, Snrnp70, SF1, PUF60, Celf4, SCAF1, SRSF3, Sf3b4, Ptbp1, CPEB1, PCBP1, TRA2B, Ptbp3, Hnrnpk*
GOTERM_BP_FAT	RNA splicing	12	30.0	14.80	3.42E-08	*HNRNPC, Snrnp70, SF1, PUF60, Celf4, SCAF1, SRSF3, Sf3b4, Ptbp1, PCBP1, TRA2B, Hnrnpk*
GOTERM_BP_FAT	RNA processing	14	35.0	8.32	5.17E-07	*HNRNPC, Snrnp70, SF1, PUF60, Celf4, SCAF1, SRSF3, Sf3b4, Ptbp1, CPEB1, PCBP1, TRA2B, Ptbp3, Hnrnpk*
GOTERM_BP_FAT	RNA splicing, via transesterification reactions	7	17.5	24.48	7.16E-05	*HNRNPC, Snrnp70, SF1, Celf4, SRSF3, PCBP1, TRA2B*
GOTERM_BP_FAT	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	7	17.5	24.48	7.16E-05	*HNRNPC, Snrnp70, SF1, Celf4, SRSF3, PCBP1, TRA2B*
GOTERM_BP_FAT	nuclear mRNA splicing, via spliceosome	7	17.5	24.48	7.16E-05	*HNRNPC, Snrnp70, SF1, Celf4, SRSF3, PCBP1, TRA2B*
GOTERM_CC_FAT	ribonucleoprotein complex	10	25.0	8.35	3.74E-05	*Sf3b4, HNRNPC, CPEB1, SF1, Pcbp1, Snrnp70, Pcbp4, Dazap1, PUF60, Hnrnpk*
GOTERM_CC_FAT	spliceosome	5	12.5	15.82	1.18E-02	*Sf3b4, HNRNPC, SF1, Snrnp70, Hnrnpk*
GOTERM_MF_FAT	RNA binding	18	45.0	7.15	2.03E-09	*HNRNPC, Snrnp70, SF1, PUF60, Celf4, SCAF1, SRSF3, Sf3b4, Ptbp1, CPEB1, PCBP1, TRA2B, Pcbp4, Ptbp3, Dazap1, Rbm42, Rbm10, Hnrnpk*
KEGG_PATHWAY	Spliceosome	6	15.0	32.12	8.87E-06	*SRSF3, HNRNPC, PCBP1, Snrnp70, PUF60, TRA2B*
INTERPRO	RNA recognition motif, RNP-1	13	32.5	21.99	1.29E-11	*HNRNPC, Snrnp70, PUF60, Celf4, SRSF3, Sf3b4, Ptbp1, CPEB1, TRA2B, Ptbp3, Dazap1, Rbm42, Rbm10*
INTERPRO	Nucleotide-binding, alpha-beta plait	12	30.0	19.97	4.81E-10	*HNRNPC, Snrnp70, PUF60, Celf4, SRSF3, Sf3b4, Ptbp1, TRA2B, Ptbp3, Dazap1, Rbm42, Rbm10*
INTERPRO	K Homology, type 1, subgroup	4	10.0	53.24	3.51E-03	*SF1, PCBP1, Pcbp4, Hnrnpk*
INTERPRO	K Homology, type 1	4	10.0	55.66	3.06E-03	*SF1, PCBP1, Pcbp4, Hnrnpk*

Note that both human and mouse genes are listed, with human genes indicated in all capital letters.

Supplementary table 2. Sequences of siRNAs

Gene	siRNA	Sequence
Celf4	1	UUCAUGGGAAUGGUCGACGGG
	2	UAAUCUGACCAUAGGCAGCAG
Dazap1	1	UCAUAGAUCAUAACCACCUCT
	2	UCUUUAAACUUGACAAAUCCA
Pcbp4	1	UUGAGGCCGAUGAAGUUGGAG
	2	UCACUGCUCCAUACUGACCCT
Puf60	1	UAUUCGAUAAAACCAUAGCCC
	2	UUUCGGAGAACCAUCACUGTA
Rbm10	1	UUGGACACCAUGGGACUGCAG
	2	ACAGAGGCUGUAGAUAUGCCA
Rbm42	1	ACCUGCAUCGCUUCUACCGTG
	2	UUCCGGUCCUUCCACAUGCTT
Sf3b4	1	UACAUUAGCCUUGACAAGCTA
	2	UUUCCAUAGAGUUUGAUCATG
Snrnp70	1	UUCAUGCUCAUACUCGAUGAA
	2	AUUGGGAUCAUUGUGAGGGTC