[Frontiers in Bioscience 11, 1998-2006, May 1, 2006]

Micro-RNA-like effects of complete intronic sequences

Aubrey E. Hill [2,4], Jeong S. Hong [1,3,5], Hui Wen [1,3,5], LiHong Teng [1,2], David T. McPherson [2], Sylvia A. McPherson [6], Dana N. Levasseur [7] and Eric J. Sorscher [1,3,5]

[1]Department of Medicine, [2] Pittman General Clinical Research Center, [3] Physiology and Biophysics, [4] Department of Computer and Information Sciences, [5]The Gregory Fleming James Cystic Fibrosis Center, [6] Center for AIDS Research , at the University of Alabama at Birmingham,1918 University Blvd., 796 MCLM, Birmingham, Alabama 35294-0005 and [7] Department of Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Mayer 656, Boston, MA 02115

TABLE OF CONTENTS

1. ABSTRACT

MicroRNAs (miRNAs) have been suggested as suppressors of numerous target genes in human cells. In this report, we present gene chip array data indicating that in the absence of miRNA sequences, complete human introns are similarly capable of coordinating expression of large numbers of gene products at spatially diverse sites within the genome. The expression of selected intronic sequences (6a, 14b and 23) derived from the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene caused extensive and specific transcriptional changes in epithelial cells (HeLa) that do not normally express this gene product. Each intron initiated a distinctive pattern of gene transcription. Affected genes such as FOXF1, sucrase-isomaltase, collagen, interferon, complement and thrombospondin 1 have previously been linked to CFTR function or are known to contribute to the related processes of epithelial differentiation and repair. A possible regulatory function of this nature has not been demonstrated previously for non-coding sequences within eukaryotic DNA. The results are consistent with the observation that splicesomal introns are found only in eukaryotic organisms and that the number of introns increases with phylogenetic complexity.

2. INTRODUCTION

Although traditionally viewed as evolutionary debris (1-13), functions ascribed to intronic DNA (mRNA export, transcript stability, nonsense mediated decay, etc.) that act in cis have become increasingly apparent (14-25). In addition, very recent evolutionary and developmental models have speculated that intronic DNA might somehow govern genomic networks that underlie human differentiation and development (2-4). Small non-coding RNAs (microRNAs, or miRNAs) are often found within intronic DNA, and regulate mRNA transcript stability in plants and animals. A specific miRNA was recently shown to alter expression of approximately 100 different mRNAs in the HeLa (cervical cancer) cell line (5) and to shift gene expression towards a profile suggestive of brain tissue. The specificity of these findings for 22nt RNA sequences (as opposed to larger, non-coding segments of RNA) is not known. Intronic RNA is transcribed from DNA following expression of the vast majority of human genes. Nevertheless, the genomic consequences of expressing a eukaryotic intron in heterologous cells (i.e. that otherwise do not express the associated gene) have not been examined previously. In the present experiments, we

Table 1. Primers and their sequences

Designation	Primer sequence 5′ → 3′
6a-F	GGATCCTCGAGTAGCAACCTATTTTCATAAC
6a-R	GGATCCTCGAGCTGTAAATCAATCAATCAATC
14b-F	GGATCCTCGAGTGAGTATTCCATGTCCTATTG
14b-R	GGATCCTCGAGCTGTAAAGCAAATAGGAAATC
23-F	GGATCCTCGAGTGAGTCTTTATAACTTTACTTAAG
23-R	GGATCCTCGAGCTAGGGAAATGGCTGTTAGTG

explored intron function by transcribing three sequences (normally spliced from the cystic fibrosis transmembrane conductance regulator, *CFTR*) in HeLa cells that have undetectable endogenous *CFTR* mRNA.

3. MATERIALS AND METHODS

3.1. Amplification of CFTR introns 6a, 14b and 23

Intronic sequence information was obtained from the University of California, Santa Clara Genome Bioinformatics web site (26). Introns 6a, 14b, and 23 from the CFTR genomic sequence (Figure 1) were PCR amplified using primers shown in Table 1.

PCR products were cloned into pCR4-Blunt vector (Invitrogen, Carlsbad, CA), and then subcloned into XhoI digested pHR′ CMVLucWSin-18 (see below). Correct PCR amplification was confirmed in each case by double stranded DNA sequencing.

3.2. Lentivirus generation

Lentiviral vector production as described by Trono and colleagues (27) was based on a lentiviral packaging system consisting of three plasmids; pMD.G (envelope plasmid), pCMVDR8.91 (packaging plasmid), and pHR'CMVWSIN-18 (transfer vector). 293T cells (2.5 x 10^6) were seeded in 10-cm plastic dishes containing Dulbecco modified Eagle medium (DMEM) with 10% fetal bovine serum (FBS) 24 hours prior to transfection. Forty micrograms plasmid DNA cocktail containing 5 µg pMD.G, 15 µg pCMVDR8.91 (which expresses Gag, Pol, Tat, and Rev) and 20 µg SIN transfer vector plasmid were used to transfect each 10-cm dish. Transfection medium was removed after 14 to 16 hours and replaced with DMEM/F12 without phenol red (Invitrogen, Carlsbad, CA) and containing 2% FBS. Viral supernatant was collected after an additional 24 hours, cleared by low-speed centrifugation, and filtered through a low protein-binding 0.22-µm polyethersulfone filter (Millipore, Beford, MA). The virus was concentrated 1000-fold by centrifugation at 26,000 rpm for 90 minutes at 8°C using an SW-28 rotor (Beckman, Palo Alto, CA), resuspended into serum-free cell growth medium and allowed to incubate on ice for 2 hours before storage at -80°C. Viral titer was approximately 1x 10^9 TU/ml.

3.3 Clonal cell line preparation

One hundred thousand HeLa cells were split into 6-well plates. Two microliters of lenitvirus encoding each intron were added to the medium and incubated 24 hours. The cells were split into 96 wells at a density of 0.3 cells/well. Cells were observed daily and only single clones were expanded and screened for intron expression by RT-PCR. RNA from clonal cells was isolated using RNeasy Mini Kit (Qiagen, Valencia, CA) and analyzed using One-Step RT-PCR Kit (Qiagen, Valencia, CA) with CFTR intron specific primers. As expected (28), CFTR intronic mRNA was not detectable in RNA from parental (non-transduced) HeLa cells. For each intron, 5 to 10 distinct clones were isolated, grown in separate wells and purified separately prior to DNA microarray or other analysis.

3.4. DNA microarray analysis

A sample of cellular mRNA from individually derived cell clones was used to synthesize double-stranded cDNA by reverse transcription using One Cycle cDNA Synthesis Kit (Affymetrix, San Clara, CA). Final labeling of cRNA with biotin-UTP was performed with GeneChip IVT Labeling Kit (Affymetrix, San Clara, CA), according to manufacturer protocol. Hybridization, washing, staining and scanning were performed as described (Affymetrix GeneChip Analysis Manual). cRNA samples were initially hybridized to Affymetrix Test 3 arrays for quality control and subsequently to Affymetrix Human Genome U133 Plus 2.0 arrays. Hybridizations were carried out at 45°C for 16 h in a rotisserie oven set at 60 r.p.m. Following hybridization, arrays were washed and stained with streptavidin-phycoerythrin (Affymetrix Fluidics Station 450) using the standard antibody amplification protocol. Arrays were scanned with the Affymetrix GeneChip Scanner 3000 at 532 nm and 2.5 microns per pixel. Expression values and fold changes were calculated using GeneChip Operating Software (GCOS). Data were normalized by scaling the target intensity to 150.

4. RESULTS AND DISCUSSION

4.1. Consequences of expression of intronic sequences in human cells

Transcription of each of three *CFTR* introns (Figure 1) revealed a unique pattern of gene expression compared to (parental, non-transduced) HeLa controls. These results are presented in Tables 2-6, and indicate transcripts affected by a single intron. Only those transcripts that consistently changed by at least 3-fold with a signal intensity $\geq$ 50 (a unit-less number) were deemed up or down regulated specifically by intron expression. Two replicates of intron 6a clone 1 and one replicate of intron 6a clone 2 are presented, as well as two independent clones each for introns 14b and 23. For comparison, the fold change data for all three introns are also provided in each table.

The observed fluorescence (total putative positives) on the Affymetrix chip is the sum of those transcripts that are actually present in the biological sample (true positives) and random background fluorescence (false positives). Assuming the worst case in which all positives are false, the probability of any single transcript appearing as positive is the quotient of the total putative positives divided by the total number of transcripts on the chip. In the case of intron 6a clone 1 (Table 2), the total apparent positive rate is 77/47,000 (1.6 x 10^{-3}) for replicate one and 66/47,000 (1.4 x 10^{-3}) for replicate two. In the absence of any biological causality, the probability of a single gene appearing in both technical replicates from 47,000 possible

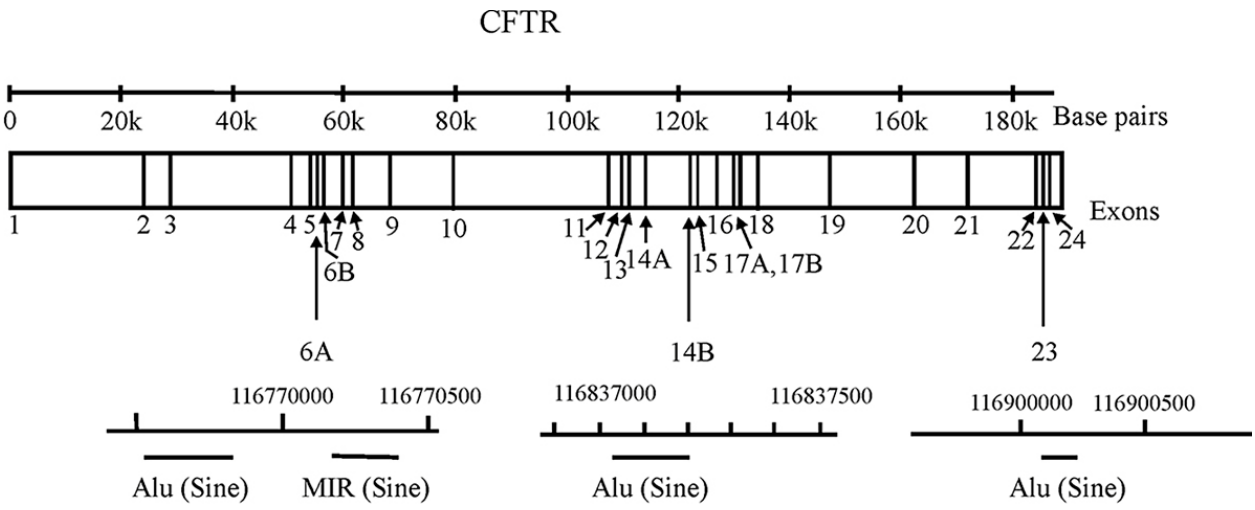

Figure 1. Exon-intron Structure of CFTR (including Alu and MIR Sequences of introns 6a, 14b and 23).

Table 2. Transcripts up-regulated by CFTR Intron 6a

Probe	Accession	Description	Intron 6 Clone 1 replicate 1 Fold change	Intron 6 Clone 1 replicate 2 Fold change	Intron 6 Clone 2 replicate 1 Fold change	Intron 14 Clone 1 replicate 1 Fold change	Intron 14 Clone 2 replicate 1 Fold change	Intron 23 Clone 1 replicate 1 Fold change	Intron 23 Clone 2 replicate 1 Fold change
205935_at	NM_001451	forkhead box F1 (FOXF1)	55.7	22.6	8.6	NC	NC	NC	3.5
206664_at	NM_001041	sucrase-isomaltase	14.9	8.0	84.5	5.7	NC	NC	NC
206100_at	NM_001874	carboxypeptidase M (CPM)	13.0	6.1	16.0	5.3	7.0	-2.5	2.5
233413_at	AU156421	placenta	13.0	55.7	7.0	NC	NC	NC	NC
236081_at	AI929792	fetal brain	13.0	3.2	6.5	NC	NC	NC	2.3
209781_s_at	AF069681	T-Star	8.0	5.3	6.1	1.9	NC	-1.9	2.1
207076_s_at	NM_000050	argininosuccinate synthetase	7.5	4.9	5.7	4.9	2.8	-3.0	6.5
235019_at	BE878495	lung large cell carcinoma	7.5	7.0	11.3	4.3	5.3	NC	1.6
229947_at	AI088609	pregnant_uterus	7.0	3.2	4.0	NC	NC	NC	NC
207030_s_at	NM_001321	cysteine and glycine-rich protein 2 (CSRP2)	6.1	6.5	7.0	NC	1.3	1.7	1.7
207173_x_at	D21254	OB-cadherin-1	5.7	5.3	9.9	NC	NC	NC	24.3
211126_s_at	U46006	smooth muscle LIM protein (h-SmLIM)	5.7	5.7	7.0	NC	NC	1.6	1.6
235706_at	AW663908	fetal lung, testis, B-cell	5.7	5.7	9.2	3.5	5.3	NC	2.3
204466_s_at	BG260394	synuclein, alpha (non A4 component of amyloid precursor)	5.3	4.6	5.3	2.1	2.1	NC	1.4
211546_x_at	LG36674	synuclein (NACP) alpha (non A4 component of amyloid precursor)	4.9	5.7	4.9	1.6	1.5	NC	NC
201278_at	N21202	disabled (Drosophila) homolog 2 (mitogen-responsive phosphoprotein)	4.6	3.2	6.5	2.1	2.0	1.6	4.3
207827_x_at	L36675	synuclein (NACP)	4.6	4.9	5.7	2.0	1.5	NC	NC
227176_at	AL565362	fetal brain	4.6	4.3	5.7	3.0	4.6	NC	2.5
201280_s_at	NM_001343	disabled (Drosophila) homolog 2 (mitogen-responsive phosphoprotein) (DAB2)	4.3	4.9	8.0	1.9	2.0	1.5	3.7
208396_s_at	NM_005019	phosphodiesterase 1A, calmodulin-dependent	4.0	6.1	9.9	4.9	2.8	NC	5.3
210757_x_at	AF188298	disabled (Drosophila) homolog 2 (mitogen-responsive phosphoprotein)	4.0	3.0	5.7	1.9	1.7	1.2	3.5
227290_at	AU157881	placenta	4.0	3.2	7.0	2.8	NC	NC	1.5
241765_at	AI469884	gi:4331974	4.0	9.8	5.3	3.2	3.5	NC	NC
201279_s_at	BC003064	disabled (Drosophila) homolog 2 (mitogen-responsive phosphoprotein)	3.7	3.7	6.5	2.0	2.1	1.3	4.3
225727_at	AB033026	KIAA1200 protein, partial cds gi:6330406	3.7	26.0	18.4	NC	NC	NC	NC
227051_at	AU157716	placenta	3.7	4.0	6.1	NC	NC	NC	-1.5
229139_at	AI202201	adult prostate	3.7	4.6	4.0	2.5	2.0	-13.9	1.7
1552695_a_at	NM_052885	solute carrier family 2 (facilitated glucose transporter)	3.2	5.7	5.7	NC	NC	NC	NC
204467_s_at	NM_000345	synuclein, alpha (non A4 component of amyloid precursor)	3.2	4.9	5.3	1.6	1.3	1.5	1.3
233547_x_at	N53248	phosphodiesterase 1A, calmodulin-dependent	3.2	4.6	8.0	2.6	NC	NC	2.8
226271_at	N46350	multiple sclerosis lesions	3.0	3.0	7.5	NC	NC	-10.6	NC

Note: only those transcripts increased in 2 of 2 (both replicates of clone 1) distinct and separately derived clones are shown, NC - no change, Clone 1, replicate 1, total of 77 transcripts increased by >= 3 fold, Clone 1, replicate 2, total of 66 transcripts increased by >= 3 fold, Clone 2, replicate 1, total of 126 transcripts increased by >= 3 fold

sequences is the product of the individual probabilities ($p = 0.0000023$) or conversely the probability of the same transcript not appearing in both replicates is $1 - 2.3 \times 10^{-6}$ or 0.999998. Likewise, the probability of a single gene appearing to be positive in both technical replicates of clone 1 and the single replicate of clone 2 (a biological replicate with 126 total putative positives) is $p = 0.00000000617$.

Similar values of significance were observed for 65 genes up-regulated by both independently derived clones expressing intron 14b (65/47,000 or < 0.14% of the

Table 3. Transcripts down-regulated by CFTR Intron 6a

Probe	Accession	Description	Intron 6 Clone 1 replicate 1 Fold change	Intron 6 Clone 1 replicate 2 Fold change	Intron 6 Clone 2 replicate 1 Fold change	Intron 14 Clone 1 replicate 1 Fold change	Intron 14 Clone 2 replicate 1 Fold change	Intron 23 Clone 1 replicate 1 Fold change	Intron 23 Clone 2 replicate 1 Fold change
225540_at	BF342661	glioblastoma	-13.0	-3.7	-5.3	-2	-5.3	-1.4	-2.8
209101_at	M92934	Human connective tissue growth factor	-11.3	-4.8	-3.3	-1.7	-4.3	-3.5	-1.4
201110_s_at	NM_003246	thrombospondin 1 (THBS1)	-10.6	-9.9	-7.5	-2.8	-3.7	-2.5	-1.9
222108_at	AC004010	lymphoblastoid	-6.5	-15.5	-5.7	-1.6	-1.5	-1.9	-1.4
203474_at	NM_006633.1	IQ motif containing GTPase activating protein 2 (IQGAP2)	-6.5	-8.2	-6.1	-1.5	NC	-4.0	-2.3
212912_at	AI992251	anaplastic oligodendroglioma	-6.5	-3.8	-4.6	-10.6[1]	-1.3	-1.5	+1.3
231577_s_at	AW014593	guanylate binding protein 1, interferon-inducible	-6.1	-5.7	-3.0	NC	-2.5	+1.5	-2.5
201324_at	NM_001423.1	epithelial membrane protein 1 (EMP1)	-6.1	-4.6	-3.3	-1.6	-1.9	+1.5	NC
215785_s_at	AL161999.1	amygdala	-5.7	-8.5	-6.5	-14.9[1]	-2.5	-19.7[1]	+2.3
202270_at	NM_002053.1	guanylate binding protein 1, interferon-inducible, 67kD (GBP1)	-4.6	-5.9	-4.0	NC	-2.3	+1.6	-2.1
202269_x_at	BC002666.1	guanylate binding protein 1, interferon-inducible, 67kD	-4.6	-3.6	-3.0	NC	-1.6	+1.6	-2.5
203060_s_at	AF074331.1	PAPS synthetase-2 (PAPSS2)	-4.0	-4.9	-4.6	-1.6	-1.6	+1.7	NC
212097_at	AU147399	mammary gland	-4.0	-3.8	-3.0	NC	-2.0	+1.9	-2.1
225604_s_at	AA284532	tyrosine 3-monooxygenasetryptophan 5-monooxygenase activation protein	-4.0	-3.5	-3.7	-4.0	-3.5	-1.2	-2.0
203963_at	NM_001218.2	carbonic anhydrase XII (CA12)	-3.7	-3.5	-3.7	-1.7	-3.3	NC	-4.0
226873_at	AI631210	pooled germ cell tumors	-3.5	-3.4	-3.3	-2.0	NC	NC	NC
215867_x_at	AL050025.1	Consensus includes Hs.5344 adaptor-related protein complex 1, gamma 1 subunit	-3.2	-3.5	-4.3	-2.1	-2.3	NC	-4.0
214164_x_at	BF752277	Consensus includes adaptor-related protein complex 1, gamma 1 subunit	-3.0	-3.8	-3.3	-2.1	-2.3	NC	-4.0
201430_s_at	W72516	Consensus includes dihydropyrimidinase-like 3	-3.0	-3.2	-4.0	-2.1	-2.5	-1.3	-2.6
226789_at	W84421	fetal_heart	-3.0	-3.4	-3.3	NC	-5.3[1]	+2.1	NC

Note: only those transcripts decreased in 2 of 2 (both replicates of clone 1) distinct and separately derived clones are shown, [1] signal intensity < 50, NC - no change, Clone 1, replicate 1, total of 105 transcripts decreased by >= 3 fold, Clone 1, replicate 2, total of 66 transcripts decreased by >= 3 fold, Clone 2, replicate 1, total of 61 transcripts decreased by >= 3 fold

transcripts available on the Affymetrix HG-U133 Plus 2 chip) ($p=0.0000265$) and the 9 concordant gene products in two independent clones expressing intron 23 ($p=0.00000373$). These results indicate a nonstochastic expression pattern attributable to transduced cell lines independently constructed, expanded, and analyzed for mRNA expression. The criteria that transcripts were only considered to be differentially expressed if they were altered by at least 3-fold in both the treatment replicates versus the controls (with a signal intensity $\geq$ 50) are restrictive, but they also provide confidence that the results are due to biological causality.

For comparison, we measured endogenous levels of each of three introns in intestinal cell lines (T84 and HT29) and airway cells (Calu3) by RT-PCR. In all three of these CFTR-expressing cell lines, the level of intron 23 was comparable to that of intron 23 transduced HeLa cells. Intron 14b was expressed at a much higher level in the HeLa cells containing the transduced intron than in the other three cell lines, whereas intron 6 was not detectable in non-transduced cells (Figure 2).

4.2. Interpreting genomic signalling by introns 6a, 14b, and 23

We have recently published results (29) demonstrating that certain exon-intron combinations (based on number of nucleotides in each) occur more frequently than would be expected from the random pairing of these genomic elements. We interpreted these findings to suggest a common structural function of introns, as opposed to the view of introns as random evolutionary debris. The results of the current study are compatible with this view. We have demonstrated that expression of a single intron in the absence of its gene of origin is sufficient to dramatically alter a cellular expression profile. While the mechanism by which introns induce these changes is unknown, it is obvious that the genomic effects are due to the presence of the intron, and not the gene, its protein product or a downstream metabolite, since CFTR is not expressed in HeLa cells.

The findings also establish that changes in cellular transcription mediated by intronic DNA are intron-specific. Each of three *CFTR* sequences displayed a unique and highly reproducible expression profile with only a few transcripts in common among the sequences examined. The three introns tested here were of different lengths and derived from different positions in *CFTR*. The experiments therefore implicate each expression pattern as a function of the specific intronic base sequence. Intron 14b is 668 bp in length. Introns 6a and 23 are 1136 bp and 1343 bp, respectively. If the expression profiles initiated by each of these were determined by length, rather than sequence, intron 6a and 23 might be expected to alter many of the same mRNAs. In fact, none of the transcripts up-regulated by 6a were also increased by intron 23. On the other hand, of 65 transcripts up-regulated by intron 14b, 7.7% (5/65), were also up-regulated by intron 6a and 4.6% (3/65) were up-regulated by intron 23. None of the intronically-signaled transcripts (Tables 2-5) were RNAses or other gene products that might be expected as part of a more general response to rid the cell of aberrant mRNA, further indicating specificity of the observed expression profiles.

The cell lines studied here (parental HeLa, intron 6a, intron 14b, intron 23) did not express CFTR as measured by the corresponding Affymetrix probe (205043_at). HeLa cells expressing intron 23 did exhibit a substantial increase in sequences derived from CFTR exons 23, 24, and 24a (Affymetrix probe 217026_at) (Table 6). This is consistent with the expression of intron 23, because exon 24a is the result of alternative splicing that incorporates part of the sequence of intron 23 to form exon 24a (30). The result is therefore a positive confirmation that the clone was expressing intron 23.

Table 4. Transcripts up-regulated by CFTR Intron 14b

			Intron 6 Clone 1 replicate 1	Intron 6 Clone 1 replicate 2	Intron 6 Clone 2 replicate 1	Intron 14 Clone 1 replicate 1	Intron 14 Clone 2 replicate 1	Intron 23 Clone 1 replicate 1	Intron 23 Clone 2 replicate 1
Probe	**Accession**	**Description**	**Fold change**	**Fold change**	**Fold change**	**Fold change**	**Fold change**	**Fold change**	**Fold change**
202411_at	NM_005532.1	interferon, alpha-inducible protein 27 (IFI27)	NC	NC	NC	137.2	48.5	NC	NC
1555229_a_at	BC007010.1	complement component 1, s subcomponent	NC	NC	NC	36.8	34.3	NC	NC
226757_at	AA131041	colon tumor	NC	-7.5	NC	36.8	8.0	NC	2.8
203153_at	NM_001548.1	interferon-induced protein with tetratricopeptide repeats	NC	-8.6[1]	NC	34.3	7.5	NC	1.9
214587_at	BE877796	collagen, type VIII, alpha	NC	NC	9.9	29.9	21.1	NC	21.1
230748_at	AI873273	fetal lung NbHL19W, testis NHT, and B-cell	NC	-1.4[1]	NC	27.9	5.3	NC	10.6
229450_at	AI075407	senescent fibroblast	NC	-3.0	NC	22.6	4.6	NC	NC
225645_at	AI763378	pyruvate dehydrogenase kinase 4 mRNA, 3 untranslated region, partial sequence	NC	NC	NC	21.1	27.9	NC	NC
1564511_a_at	AK055684.1	KIAA1061 protein weakly similar to follistatin-related	2.8	4.3[1]	5.3	16.0	10.6	NC	NC
204415_at	NM_022873.1	interferon, alpha-inducible protein transcript variant 3	NC	NC	NC	16.0	4.3	NC	NC
208747_s_at	M18767.1	complement subcomponent C1s, alpha- and beta-chains, complete cds	NC	-7.0 [1]	NC	13.9	9.2	NC	NC
219209_at	NM_022168.1	melanoma differentiation associated protein-5 (MDA5)	NC	NC	NC	13.0	3.2	NC	NC
204747_at	NM_001549.1	interferon-induced protein with tetratricopeptide repeats 4	NC	-1.4	NC	13.0	3.0	NC	NC
217767_at	NM_000064.1	complement component 3 (C3)	NC	-2.3	NC	12.1	10.6	3.5	NC
200795_at	NM_004684.1	SPARC-like 1	NC	-2.6[1]	1.5	12.1	4.9	-2.8[1]	NC
210797_s_at	AF063612.1	2-5oligoadenylate synthetase-related protein p30 (OASL) alternatively spliced	NC	-2.5	NC	11.3	3.7	NC	NC
219478_at	NM_021197.1	WAP four-disulfide core domain 1 (WFDC1)	9.9	1.2	9.2	10.6	36.8	-4.6[1]	2.6
222957_at	AK025617.1	glycoprotein beta-Gal 3-sulfotransferase	NC	NC	NC	9.8	9.8	NC	NC
225767_at	AI825833	Weakly similar to PIHUB6 salivary proline-rich protein precursor	NC	NC	NC	9.2	5.7	NC	NC
221577_x_at	AF003934.1	prostate differentiation factor mRNA	NC	NC	5.3	8.6	8.6	2.0	3.2
214022_s_at	AA749101	interferon induced transmembrane protein 1	1.5	1.5	2.1	8.0	6.1	NC	-1.5
205483_s_at	NM_005101.1	interferon-stimulated protein	NC	-3.0	1.4	8.0	3.0	-1.5	1.7
226237_at	AL359062.1	gi:8518189	2.5	NC	6.1	7.5	8.6	-1.7	4.0
1559263_s_at	BG397809	mRNA for FLJ00361 protein.	NC	NC	13.0	7.5	7.5	NC	NC
218872_at	NM_017899.1	Homo sapiens hypothetical protein FLJ20607	3.0	NC	NC	7.5	5.0	-9.2	NC
227410_at	AW264102	three pooled meningiomas	3.2	2.0	2.6	7.5	4.6	-1.9	4.3
204439_at	NM_006820.1	hypothetical protein, expressed in osteoblast (GS3686)	-3.2[1]	-9.2 [1]	-4.6	7.5	4.3	NC	-4.3[1]
210095_s_at	M31159.1	growth hormone-dependent insulin-like growth factor-binding protein	1.5	NC	NC	7.0	9.2	2	1.5
1554576_a_at	BC007242.1	E1A enhancer binding protein	NC	NC	NC	7.0	3.0	NC	4.3
206392_s_at	NM_002888.1	retinoic acid receptor responder (tazarotene induced)	NC	1.6	8.0	6.5	104	NC	NC
211603_s_at	U35622.2	EWS proteinE1A enhancer binding protein chimera mRNA	NC	NC	NC	6.5	3.0	NC	NC
212143_s_at	BF340228	insulin-like growth factor binding protein 3	NC	NC	NC	6.1	7.0	2.1	NC
201601_x_at	NM_003641.1	interferon induced transmembrane protein 1	1.4	1.5	2.5	6.1	5.3	NC	NC
212067_s_at	AL573058	complement component 1, r subcomponent	NC	NC	1.7	6.1	4.9	NC	NC
204748_at	NM_000963.1	prostaglandin-endoperoxide synthase 2	1.7	-2.0	1.4	5.7	5.7	-2.0[1]	2
211161_s_at	AF130082.1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	-1.9	-3.5	2.1	5.3	29.9	8.6	5.3
206100_at	NM_001874.1	carboxypeptidase M (CPM)	13.0	6.1	16.0	5.3	7.0	-2.5[1]	2.5
203882_at	NM_006084.1	interferon-stimulated transcription factor 3, gamma	NC	NC	NC	5.3	4.0	NC	NC
203477_at	NM_001855.1	collagen, type XV, alpha 1	3.0	1.6	2.5	4.6	9.2	-8.6[1]	2.3
201852_x_at	AI813758	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	-1.9	-2.6	1.9	4.3	32.0	4.9	4.3
215076_s_at	NM_000064.1	highly similar to procollagen alpha 1(III) chain precursor	-2.6	-2.8	1.6	4.3	10.6	5.7	4.9
235019_at	BE878495	large cell carcinoma	7.5	7.0	11.3	4.3	5.3	NC	1.6
1554997_a_at	AY151286.1	cyclooxygenase 2b, alternatively spliced	3.0	NC	NC	4.3	3.7	NC	NC
44783_s_at	R61374	infant brain	2.5	1.3	3.3	4.3	3.7	NC	2.5
203691_at	NM_002638.1	protease inhibitor 3, skin-derived (SKALP) (PI3)	2.6	NC	NC	4.3	3.5	NC	NC
209730_at	U38276.1	semaphorin III family homolog	NC	NC	NC	4.3	3.2	NC	NC
227819_at	AA524536	VTS20631 mRNA, g-protein coupled receptor family, partial cds	NC	NC	NC	4.0	9.8	NC	NC
225762_X_AT	AL531683	Weakly similar to PIHUB6 salivary proline-rich protein precursor	NC	NC	NC	4.0	6.1	NC	NC
204036_at	AW269335	endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor	NC	NC	NC	4.0	4.6	NC	NC
209676_at	J03225.1	lipoprotein-associated coagulation inhibitor	2.5	NC	2.1	4.0	4.3	NC	1.6
236795_at	AU150078	teratocarcinoma	NC	NC	NC	4.0	3.0	-2.1[1]	3.2
47560_at	AI525402	Prostate	NC	NC	NC	3.7	5.3	NC	NC
205302_at	NM_000596.1	insulin-like growth factor binding protein 1 (IGFBP1)	NC	NC	NC	3.7	4.6	NC	NC
226769_at	AI802391	total fetus	NC	NC	NC	3.7	4.0	NC	NC
205825_at	NM_000439.2	proprotein convertase subtilisinkexin type 1 (PCSK1)	2.5	1.7	3.0	3.7	3.2	-3.2[1]	1.6
210145_at	M68874.1	phosphatidylcholine 2-acylhydrolase (cPLA2)	2.8	1.3	3.3	3.7	3.0	NC	1.6
235706_at	AW663908	Moderately similar to ALU7_HUMAN ALU SUBFAMILY SQ SEQUENCE	5.7	5.7	9.2	3.5	5.3	NC	2.3
219888_at	NM_003116.1	sperm associated antigen 4 (SPAG4)	NC	NC		3.2	4.6	NC	NC
226388_at	AI675780	highly similar to HSTFIISH mRNA for transcription elongation factor TFIIS	NC	NC	NC	3.2	3.7	NC	NC
241765_at	AI469884	gi:4331974	4.0	9.8	5.3	3.2	3.5	NC	NC
AFFX-HUMRGE/M10098_5_at	M10098	18S rRNA gene	-3.2	1.4	NC	3.1	22.6	1.3	NC
202948_at	NM_000877.1	interleukin 1 receptor, type I	NC	-1.7	NC	3.0	5.3	-8.6	NC
227099_s_at	AW276078	Weakly similar to weak similarity to collagens C.elegans	NC	NC	NC	3.0	4.6	NC	NC
227176_at	AL565362	fetal brain	4.6	4.3	5.7	3.0	4.6	NC	2.5
222557_at	AL353715	chrom. 20 contains the 5 end of the GMEB2 gene for glucocorticoid modulatory element binding protein 2	2.0	NC	NC	3.0	3.7	-1.7	NC

Note: only those transcripts increased in 2 of 2 distinct and separately derived clones are shown, [1] signal intensity < 50, NC - no change, Clone 1, replicate 1, total of 179 transcripts increased by >= 3 fold, Clone 1, replicate 2, total of 327 transcripts increased by >= 3 fold

Table 5. Transcripts down-regulated by CFTR Intron 14b

Probe	Accession	Description	Intron 6 Clone 1 replicate 1 Fold change	Intron 6 Clone 1 replicate 2 Fold change	Intron 6 Clone 2 replicate 1 Fold change	Intron 14 Clone 1 replicate 1 Fold change	Intron 14 Clone 2 replicate 1 Fold change	Intron 23 Clone 1 replicate 1 Fold change	Intron 23 Clone 2 replicate 1 Fold change
234989_at	AV699657	Weakly similar to 1207289A reverse transcriptase related protein	-1.5	NC	-2.3	-8.0	-4.0	-2.5	1.9
222450_at	AL035541	soluble phosphoenolpyruvate carboxykinase 1	-3.0	-2.5	-2.8	-4.9	-3.7	NC	-1.5
227062_at	AU155361	plectin 1, intermediate filament binding protein, 500kD	NC	NC	-1.6	-4.3	-3.0	+5.3	+2.5
213478_at	AB028949.1	KIAA1026 gene	-2.5	-2.6	-3.5	-4.3	-3.2	-13.0[1]	NC
225604_s_at	AA284532	tyrosine 3-monooxygenase tryptophan 5-monooxygenase activation protein, eta polypeptide	-4.0	-3.5	-3.7	-4.0	-3.5	-1.2	-2
222449_at	AL035541	soluble phosphoenolpyruvate carboxykinase 1	-2.0	-2.6	-2.5	-3.7	-3.2	NC	-1.6
221011_s_at	NM_030915.1	Lbh, a novel conserved nuclear protein expressed during early limb and heart development	-3.5	-3.7	-2.3	-3.2	-4.0	+1.2	-2.1
229765_at	AW511239	fetal lung, testis, and B-cell	-2.8	NC	-3.3	-3.0	-3.0	+1.7	+1.5

Note: only those transcripts decreased in 2 of 2 distinct and separately derived clones are shown, [1] signal intensity < 50, NC - no change, Clone 1, replicate 1, total of 42 transcripts decreased by >= 3 fold, Clone 1, replicate 2, total of 38 transcripts decreased by >= 3 fold

Table 6. Transcripts up-regulated by CFTR Intron 23

Probe	Accession	Description	Intron 6 Clone 1 replicate 1 Fold change	Intron 6 Clone 1 replicate 2 Fold change	Intron 6 Clone 2 replicate 1 Fold change	Intron 14 Clone 1 replicate 1 Fold change	Intron 14 Clone 2 replicate 1 Fold change	Intron 23 Clone 1 replicate 1 Fold change	Intron 23 Clone 2 replicate 1 Fold change
217026_at	M96936	Homo sapiens cystic fibrosis transmembrane conductance regulator (CFTR) gene, exons 23, 24a, and 24	NC	NC	NC	NC	NC	19.7	9.8
232458_at	AU146808	FLJ11469 fis	NC	-1.3[1]	NC	NC	2.7	10.6	5.3
211161_s_at	AF130082.1	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	-1.9	-3.5	2.1	5.3	29.9	8.6	5.3
1554195_a_at	BC021680.1	Prostate, adenocarcinoma	NC	-4.0[1]	2.0	1.1	NC	6.5	3.2
215076_s_at	AU144167	highly similar to procollagen alpha 1(III) chain precursor	-2.6	-2.8	1.6	4.3	29.9	5.7	4.9
238320_at	AV659198	liver	1.6	NC	NC	NC	NC	5.3	3.5
201852_x_at	AI813758	collagen, type III, alpha 1 (Ehlers-Danlos syndrome type IV, autosomal dominant)	-1.9	-2.6	1.9	4.3	32.0	4.9	4.3
234118_at	AK000141.1	FLJ20134 fis, colon	NC	NC	NC	NC	NC	4.3	3.7
204529_s_at	AI961231	well-differentiated endometrial adenocarcinoma	4.0	NC	2.8	2.6	2.5	3.0	3.7

Note: only those transcripts increased in 2 of 2 distinct and separately derived clones are shown, [1] signal intensity < 50, NC - no change, Clone 1, replicate 1, total of 142 transcripts increased by >= 3 fold, Clone 1, replicate 2, a total of 58 transcripts increased by >= 3 fold

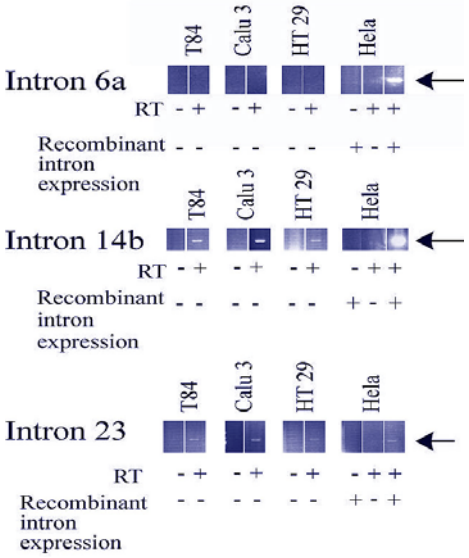

Figure 2. Cells expressing endogenous CFTR (Calu3: human airway serous glandular cells; HT-29 and T84: human colonic epithelium) were studied by RT-PCR using mRNA from approximately 100 ng template and primers as described in Methods. PCR settings were: 95° C x 15 min. initial PCR activation; 50° C x 30 min. reverse transcriptase; 94° C x 1 min. denaturation; 55° C x 1 min. annealing and 72° C x 1 min. extension. Controls without reverse transcriptase treatment are also shown.

4.3. Intronic signalling is independent of microRNA expression

MicroRNAs are 22 nucleotide RNA sequences known to suppress transcriptional activity in animals and plants, including human cells (5-7). Many of the > 300 miRNAs identified in the human genome are encoded within intronic DNA. Gene repression by miRNA sequences is not fully understood, but often involves interactions with 3' untranslated regions of targeted genes through partial complimentary. Recent findings suggest far reaching influence of miRNAs in genome-wide transcription. For example, transfection and expression of a single miRNA (miR-124) in HeLa cells downregulated numerous mRNAs and shifted the expression profile towards that of brain (5). The present experiments indicate that similar effects occur when entire introns are expressed in the HeLa cell line. Of note is that the *CFTR* introns 6a, 14b and 23 do not contain miRNAs (as judged by either the stem-loop search tool available from the miRNA registry website, or any known miRNAs listed on that site (31)).

4.4. Introns 6a, 14b, and 23 activate genes tied to CFTR function

The simultaneous changes of a number of the genes described here are suggestive of mRNA networks responsible for sustaining normal epithelial cell physiology. CFTR is expressed in lung, intestine, and other epithelium, where it acts as a chloride channel. Absence of CFTR leads to profound disruption of the normal tissue

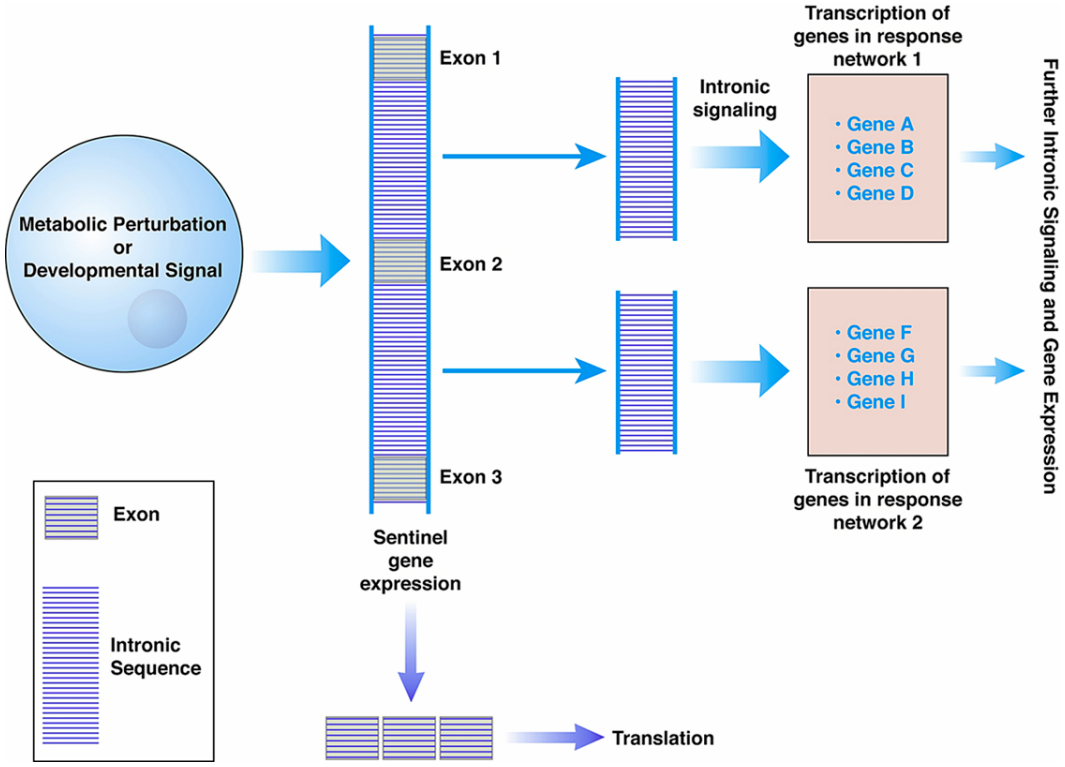

Figure 3. Model depicting intronic signalling and gene network notification. Activation of a cellular pathway (e.g. following a cellular insult or developmental cue) begins by expression of a sentinel gene. Intronic mRNA from the sentinel activates additional response networks (two are depicted), and further rounds of intronic signalling/notification.

architecture. Several genes found to be differentially expressed in response to CFTR intronic RNA seem more than coincidentally tied to CFTR function. For example, expression of intron 6a in HeLa cells caused an average 36-fold increase in sucrase-isomaltase (SI), an enzyme that is expressed only in small intestinal enterocytes (32) and fetal colon (33). The expression of this gene, which was also affected in one of the clones expressing intron 14b, has been shown to correlate with expression of CFTR during the differentiation of intestinal epithelial cells (34). The most significant down-regulated gene was thrombospondin-1, which was on average nine fold less in the cells expressing intron 6a as compared to parental HeLa cells. The level of thrombospondin has been shown to be associated with the degree of lung injury (35). A mouse model deleted for thrombospondin-1 (36) exhibited a number of abnormalities including persistent inflammation, vascular smooth muscle hyperplasia and mucinous metaplasia, which were noted to resemble a CF pulmonary phenotype. Several other transcripts affected by intron 14b and intron 23 are related to the collagen (7 instances), interferon (8 instances), and complement (4 instances) protein families crucial to normal epithelial physiology, or specifically implicated in the CF disease process. Expression of intron 6a resulted in an average 29-fold change in the level of Forkhead Box f1 (FOXf1). This gene product has been shown to cause activation of a number of lung-specific genes (37) and to be essential for lung repair (38).

The majority of mRNAs described in Tables 1 to 5 do not conform to a discrete epithelial attribute. However, based on a surprising number of gene products known to have crucial functions in major CF affected organs such as lung and intestine, we suggest that even in HeLa (cervical cancer) cells, CFTR introns may remain capable of shifting expression towards sequences important to epithelial and tissue physiology. The present experiments provide a means by which these observations can be better understood in the future.

4.5. Potential significance of the observation that intronic sequences modify gene expression in trans

These observations support a notion conceptualized previously (1-4) that intronic DNA may somehow provide a material signal contributing to the organization of certain genetic networks. In the paradigm suggested by the present findings, genes in a common metabolic pathway would influence expression of other genes in the same pathway. If transcription of a particular gene were activated (e.g. by a change in cellular metabolism) related genes would be notified of this circumstance by virtue of intronic RNA (Figure 3). The arrangement would coordinate waves of gene expression important to particular cellular processes, and quantitatively depend upon relative levels of spliced mRNAs. The signals would also persist for some period of time beyond the transcription of the originally activated gene.

The notion presented here is also consistent with the observation that the number of introns increases with the complexity of organisms (up the phylogenetic hierarchy). Splicesomal introns are only found in eukaryotes, with the lower eukaryotes having fewer. More complicated organisms would be expected to require complex transcriptional control to successfully complete development and differentiation. Increased complexity also necessitates adaptability (e.g. organ system and homeostatic responsiveness to infections, toxins, climate, resource availability, etc.), compounding the need for more sophisticated and systematic transcriptional regulation. Our findings predict that certain intronless (structural) genes such as histones, keratins and actin, might less actively participate in the complex regulatory responses necessary to manage environmental stimuli of this sort. Part and parcel of the model would be cascades of interrelated genes triggering expression of downstream gene networks, in a fashion fundamental to the processes of development, aging and epigenetic adaptation to the environment.

5. ACKNOWLEDGEMENTS

We would like to thank Gretchen Cloud and Grier Page for their helpful discussions and data analysis. The authors are indebted to Mrs. Jane Schell and Ms. Carolyn Cox for help preparing the manuscript. We also thank Mr. Eddie Walthall for initiating and refining some of the analysis described here.

6. REFERENCES

1. Mattick J.S.: RNA regulation: a new genetics? *Nature Reviews Genetics* 5, 316-323 (2004)

2. Mattick J.S. & J. Gagen: Accelerating networks in biology, engineering, and society. *Science* 307, 856-8 (2005)

3. Mattick J.S. & M.J. Gagen: The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol* 18, 1611-30 (2001)

4. Mattick J.S.: What makes a human? *The Scientist* 19(4), 32 (2005)

5. Lim L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, & D.P. Bartel: Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769-73 (2005)

6. Lim L.P., M.E. Glasner, S. Yekta, C.B. Burge, & D.P. Bartel: Vertebrate MicroRNA genes. *Science* 299, 1540 (2003)

7. Baskerville S. & D.P. Bartel: Microarray profiling of microRNAs reveals frequent coexperssion with neighbouring miRNAs and host genes. *RNA* 11(3), 241-7 (2005)

8. Aloni Y., R. Dhar, O. Laub, M. Horowitz, & G. Khoury: Novel mechanism for RNA maturation: the leader sequences of simian virus 40 mRNA are not transcribed adjacent to the coding sequences. *Proc Natl Acad Sci USA* 74, 3686-3690 (1977)

9. Berget S.M., C. Moore, & P. Sharp: Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* 74, 3171-3175 (1977)

10. Celma M.L., R. Dhar, J. Pan, & S.M. Weissman: Comparison of the nucleotide sequence of the messenger RNA for the major structural protein of SV40 with the DNA sequence encoding the amino acids of the protein. *Nucleic Acids Res* 4, 2549-50 (1977)

11. Chow L.T., R.E. Gelinas, T.R. Broker, & R.J. Roberts: An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8 (1977)

12. Dunn A.R. & J.A. Hassell: A novel method to map transcripts: evidence for homology between an adenovirus mRNA and discrete multiple regions of the viral genome. *Cell* 12, 23-36 (1977)

13. Klessig D.F.: Two adenovirus mRNAs have a common 5' terminal leader sequence encoded at least 10 kb upstream from their main coding regions. *Cell* 12, 9-21 (1977)

14. Forsdyke D.: Are introns in-series error-detecting sequences? *J Theor Biol* 93, 861-866 (1980)

15. Forsdyke D.: A stem-loop "kissing" model for the initiation of recombination and the origin of introns. *Mol Biol Evol* 12, 949-958 (1995)

16. Forsdyke D.: An alternative way of thinking about stem-loops in DNA. A case study of the human G0S2 gene. *J Theor Biol* 192, 489-504 (1998)

17. Barrette I.H., S. McKenna, D.R. Taylor, & D. Forsdyke: Introns resolve the conflict between base order-dependent stem-loop potential and the encoding of RNA or protein: further evidence from overlapping genes. *Gene* 270, 181-189 (2001)

18. Doyle G.G.: A general theory of chromosome pairing based on the palindromic DNA model of Sobell with modifications and amplifications. *J Theor Biol* 70, 171-184 (1978)

19. Ares M., L. Grate, & M.H. Pauling: A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* 5, 1138-1139 (1999)

20. Frederickson R.M., M.R. Micheau, A. Iwamoto, & N.G. Miyamoto: 5' flanking and first intron sequences of the human beta-actin gene required for efficient promoter activity. *Nucleic Acids Res* 17, 253-270 (1989)

21. Luo M. & R. Reed: Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci USA* 96, 14937-14942 (1999)

22. Fong Y.W., & Q. Zhou: Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414, 929-933 (2001)

23. Maniatis R. & R. Reed: An extensive network of coupling among gene expression machines. *Nature* 416, 499-506 (2002)

24. Hormuzdi S., R. Penttinen, R. Jaenisch, & P. Bornstein: A gene-targeting approach identifies a function for the first intron in expression of the alpha1(I) collagen gene. *Mol Cell Biol* 18, 3368-3375 (1998)

25. Beaton M.J. & T. Cavalier-Smith: Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Proc R Soc Lond B Biol SciI* 266, 2053-2059 (1999)

26. University of California, Santa Clara Genome Bioinformatics web site (http://genome.ucsc.edu/)

27. Trono Laboratory web site (http://www.tronolab.unige.ch/x_research2.htm)

28. Clancy J.P., F.E. Ruiz, & E.J. Sorscher: Adenosine and its nucleotides activate wild-type and R117H CFTR through an A2B receptor-coupled pathway. *Am J Physiol* 276(2 Pt 1), C361-9 (1999)

29. Hill A. & E. Sorscher: Common structural patterns in human genes. *Bioinformatics* 20, 1632-1635 (2004)

30. Yoshimura K., C.S. Chu, & R.G. Crystal: Alternative splicing of intron 23 of the human cystic fibrosis transmembrane conductance regulator gene resulting in a novel exon and transcript coding for a shortened intracytoplasmic C terminus. *J Biol Chem* 268, 686-690 (1993)

31. Griffiths-Jones S.: The microRNA registry. *Nucl Acid Res* 32, D109-11 (2004)

32. Suh E., & P.G. Traber: An intestine-specific homeobox gene regulates proliferation and differentiation. *Mol Cell Biol* 16, 619-625 (1996)

33. Wiltz O., C.J. O'Hara, & G.D. Steele Jr.: Mercurio AM: Expression of enzymatically active sucrase-isomaltase is a ubiquitous property of colon adenocarcinomas. *Gastroenterology* 100, 1266-78 (1991)

34. Sood R., C. Bear, W. Auerbach, E. Reyes, T. Jensen, N. Kartner, J.R. Riordan, & M. Buchwald: Regulation of CFTR expression and function during differentiation of intestinal epithelial cells. *EMBO J* 11, 2487-94 (1992)

35. Idell S., R. Maunder, A.M. Fein, H.I. Switalska, G.P. Tuszynski, J. McLarty, & S. Niewiarowski: Platelet-specific alpha-granule proteins and thrombospondin in bronchoalveolar lavage in the adult respiratory distress syndrome. *Chest* 96, 1125-32 (1989)

36. Crawford S.E., V. Stellmach, J.E. Murphy-Ullrich, S.M. Ribeiro, J. Lawler, R.O. Hynes, G.P. Boivin, & N. Bouck: Thrombospondin-1 is a major activator of TGF-beta1 in vivo. *Cell* 93, 1159-1170 (1998)

37. Hellqvist M., M. Mahlapuu, L. Samuelsson, S. Enerbaeck, & P. Carlsson: Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2. *J Biol Chem* 271, 4482-4490 (1996)

38. Kalinichenko V.V., Y. Zhou, B. Shin, D.B. Stolz, S.C. Watkins, J.A. Whitsett, & R.H. Costa: Wild-type levels of the mouse Forkhead Box f1 gene are essential for lung repair. *Am J Physiol Lung Cell Mol Physiol* 282, L1253-65 (2002)

Abbreviations: CF: cystic fibrosis; CFTR: cystic fibrosis transmembrane conductance regulator; FBS: fetal bovine serum; TU: transducing unit.

Key Words: intron, function, micro-RNA, regulation, transcription, gene signalling

Send correspondence to: Eric J. Sorscher, M.D., 1918 University Blvd. (MCLM 796), Birmingham, AL 35294-0005, Tel: 205-934-9640, Fax: 205-934-7593, E-mail: sorscher@uab.edu

http://www.bioscience.org/current/vol11.htm