

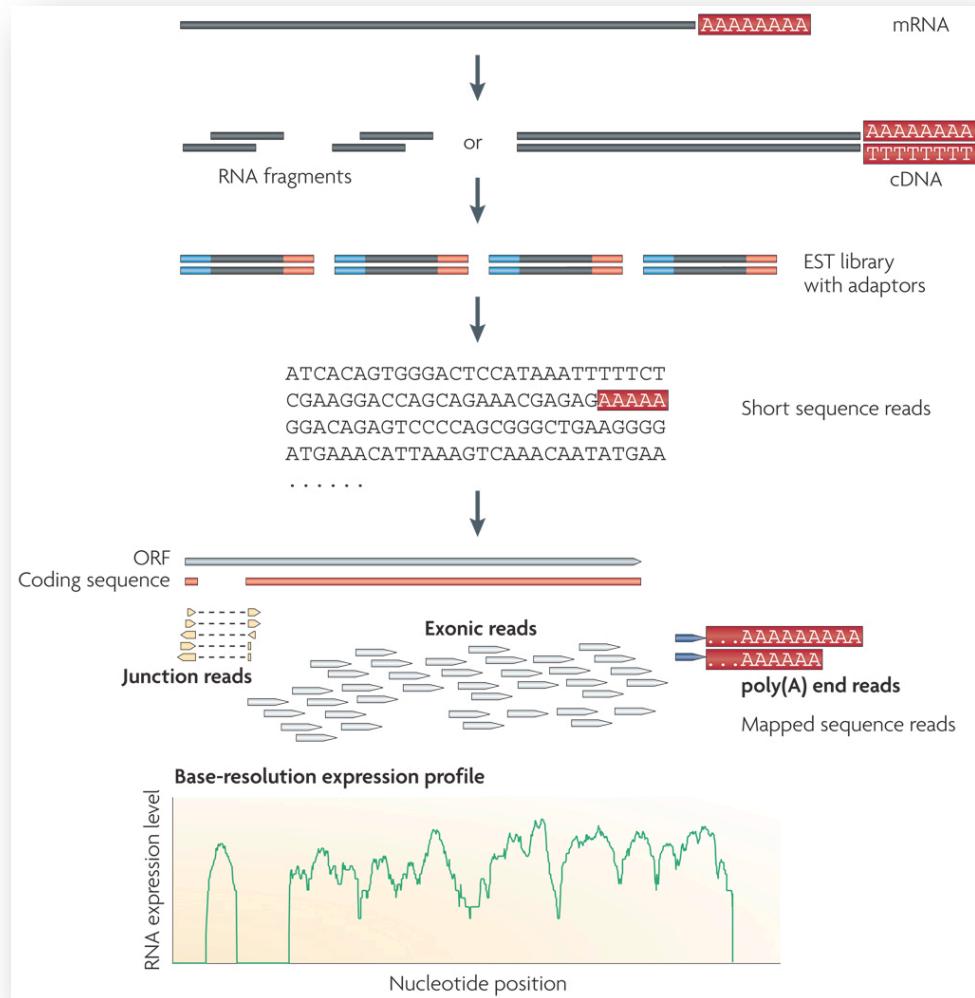
Transcriptome Analysis of the Model Protozoan, *Tetrahymena thermophila*, Using Deep RNA Sequencing

Jie Xiong^{1,5}, Xingyi Lu^{1,5}, Zhemin Zhou², Yue Chang^{1,5}, Dongxia Yuan¹, Miao Tian^{1,5}, Zhigang Zhou⁴, Lei Wang², Chengjie Fu¹, Eduardo Orias³, Wei Miao^{1*}

1 Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, People's Republic of China, **2** Tianjin Economic-Technological Development Area School of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China, **3** Department of Molecular, Cellular and Developmental Biology, University of California Santa Barbara, Santa Barbara, California, United States of America, **4** Feed Research Institute of Chinese Academy of Agricultural Sciences, Beijing, People's Republic of China, **5** Graduate School of Chinese Academy of Sciences, Beijing, People's Republic of China

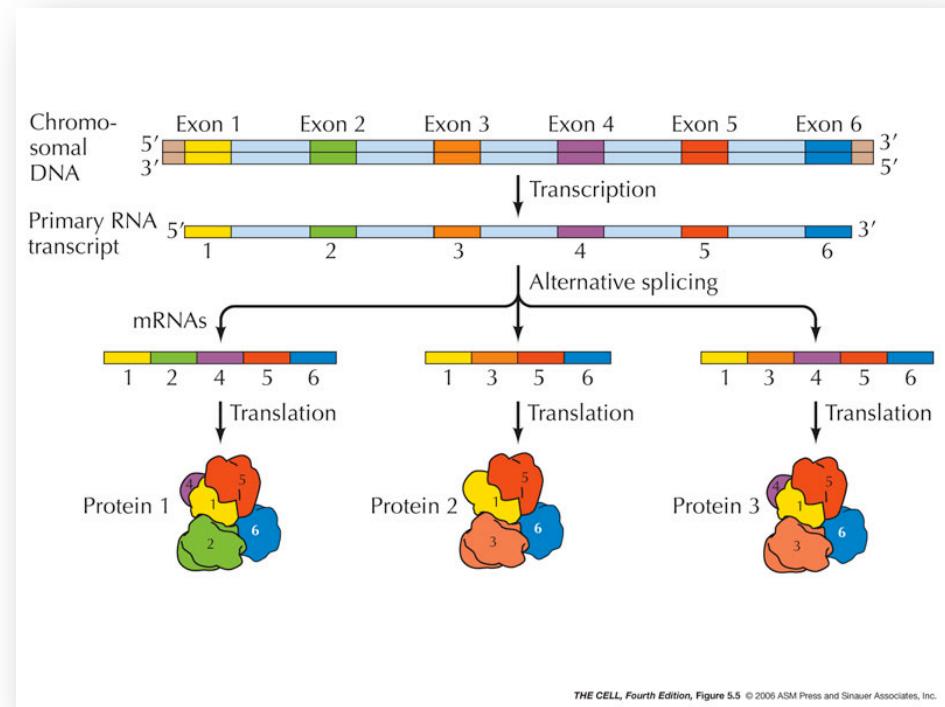
Transcriptome of *Tetrahymena thermophila*

- First transcriptome
- RNA-seq
- Three major life stages
 - Growth
 - Starvation
 - Conjugation



Transcriptome of *Tetrahymena thermophila*

- Covered 96% of the 24,725 predicted genes
- 1 000 new transcribed regions
- Alternative splicing
 - 1,500 events
 - 5.2% genes
 - Stage specific



Transcriptome of *Tetrahymena thermophila*

- Completely confirm 26.8% predicted genes
- Correct previous studies with microarray

Why reproducing this paper?

- Learn the methods
 - Mapping reads to a reference genome
 - Differential expression
 - Gene Ontology enrichment
 - Alternative splicing
- Better understand the data

Sample	Reads length (bp)	Reads generated	Base generated	Unique mapped reads
G-m	53	61,625,410	3,266,146,730	35,553,047
S-3 (mating type VI)	75	31,956,872	2,396,765,400	20,044,802
S-3 (mating type V)	100	17,878,254	1,787,825,400	11,166,805
S-15	100	16,509,042	1,650,904,200	11,402,254
C-2	75	32,658,040	2,449,353,000	23,060,144
C-8	100	27,384,596	2,738,459,600	22,022,362
Total	/	188,012,214	14,289,454,330	123,249,414

The samples G-m, S-3, S-15, C-2 and C-8 refer to growth to mid-log cell density ($\sim 3.5 \times 10^5$ cells/ml), 3 hours of starvation, 15 hours of starvation, 2 hours into conjugation and 8 hours into conjugation, respectively.
doi:10.1371/journal.pone.0030630.t001

Table 1. RNA-seq mapping statistics.

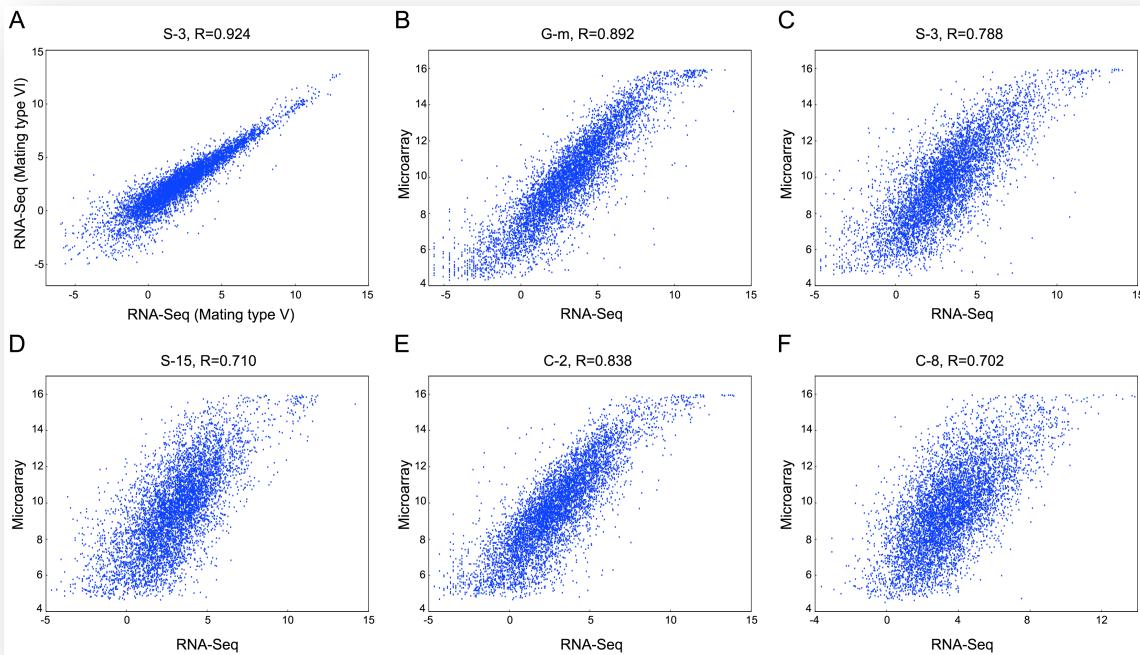


Figure 1. Correlation between RNA-seq and microarray data sets.

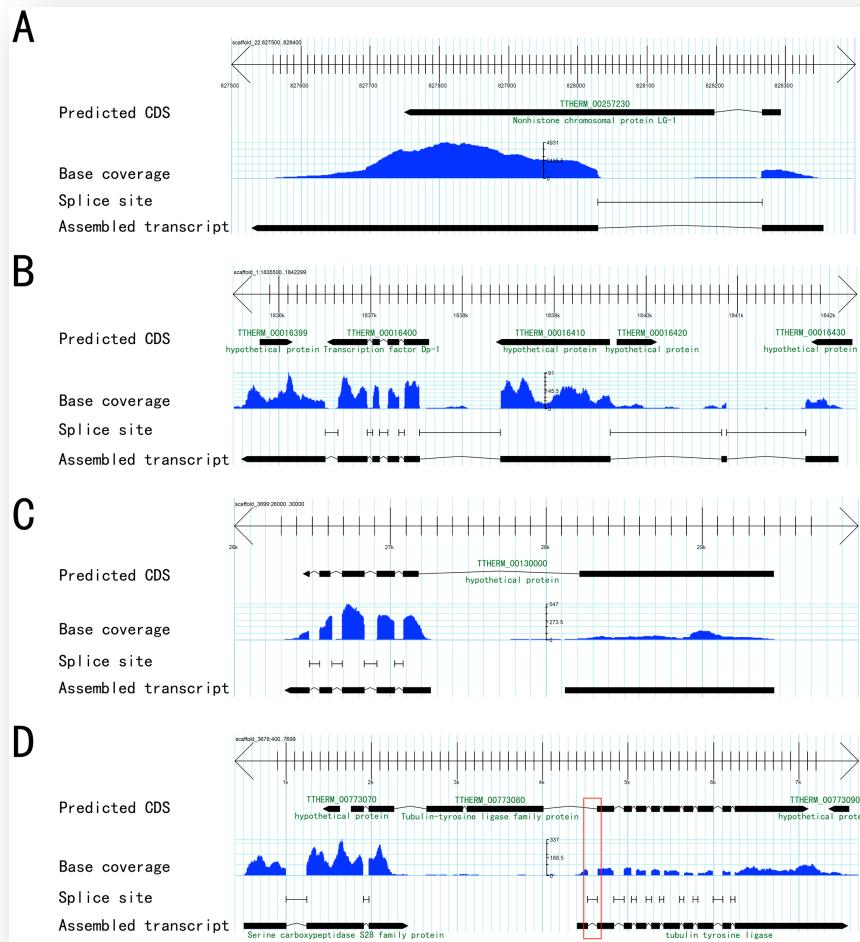


Figure 2. Gbrowse snapshot showing examples of mis-predictions in the current *T. thermophila* genome annotation that were corrected by RNA-seq.

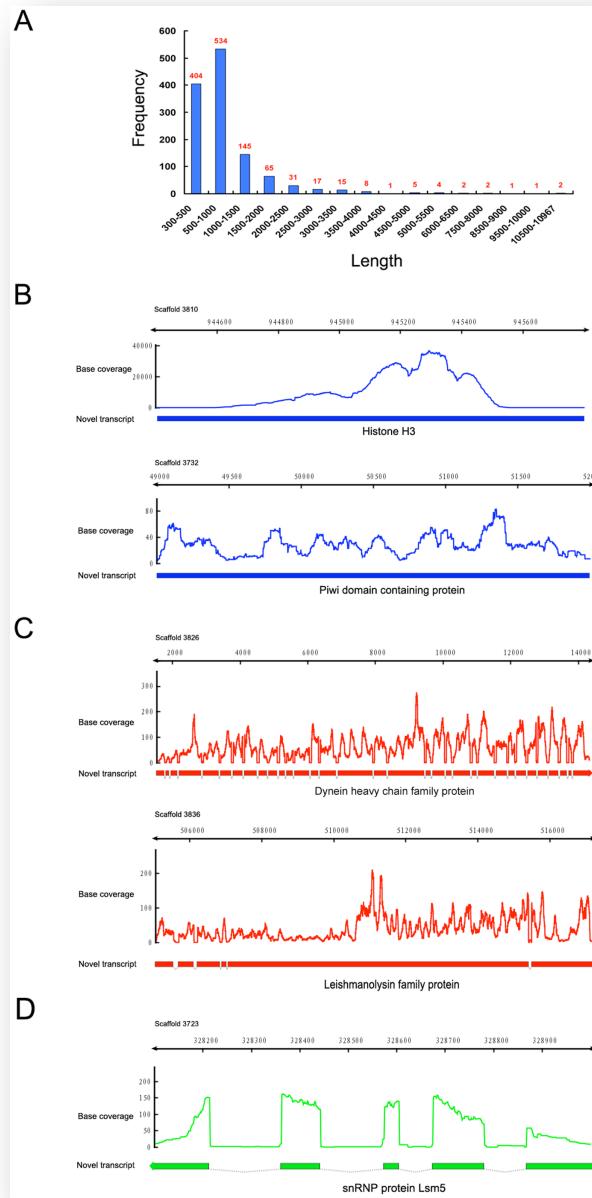


Figure 3. RNA-seq identified novel transcribed regions in *T. thermophila*.

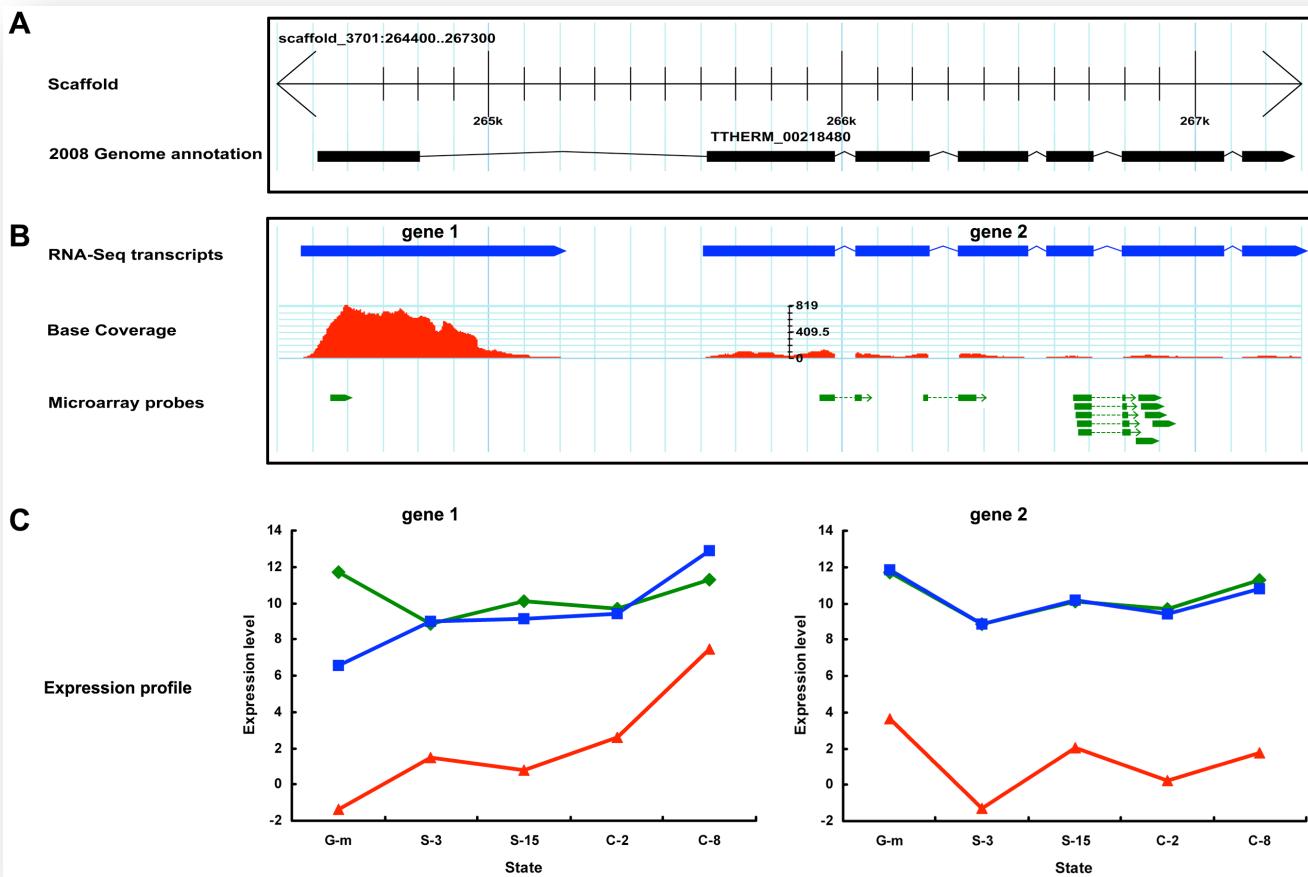


Figure 4. Reassignment of microarray probes and renormalization of microarray expression values using RNA-seq corrected gene models.

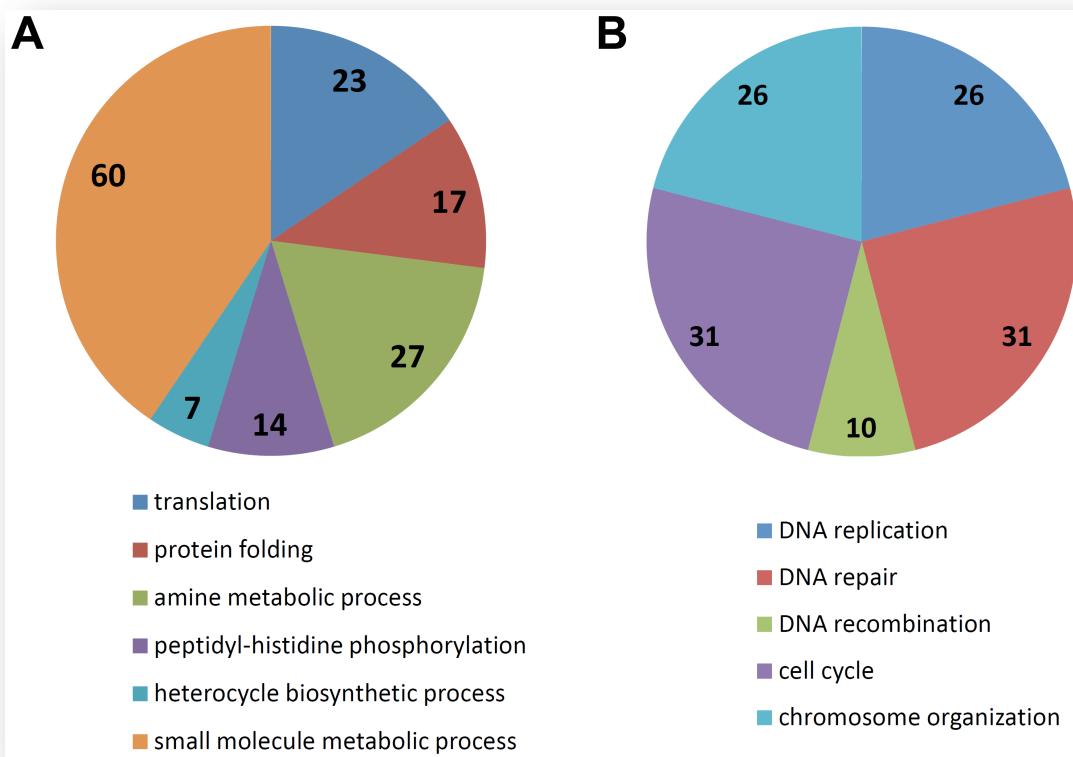


Figure 5. Representative overrepresented functions among genes which are up-regulated during growth and conjugation.

Organism	Evolutionary Group	Total gene number^b	Total Intron number^c	Average Introns per genes	AS/Total genes %	AS Method & Reference
<i>Homo sapiens</i>	metazoa	20,834	258,085	12.39	95%	RNA-seq [32]
<i>Mus musculus</i>	metazoa	23,200	209,066	9.01	45%	EST [56]
<i>Bos taurus</i>	metazoa	22,787	190,317	8.35	21%	Bioinformatics [57]
<i>Danio rerio</i>	metazoa	26,564	240,336	9.05	17%	EST [58]
<i>Caenorhabditis elegans</i>	metazoa	20,183	109,160	5.41	25%	RNA- seq & Microarray [59]
<i>Drosophila melanogaster</i>	metazoa	13,776	45,796	3.32	54.20%	RNA-seq [60]
<i>Oryza sativa</i>	plant	28,453	129,751	4.56	33%	RNA-seq [20]
<i>Arabidopsis thaliana</i>	plant	27,361	120,155	4.39	36.40%	RNA-seq [19]
<i>Chlamydomonas reinhardtii</i>	plant (unicellular)	14,416	104,665	7.26	3%	EST [61]
<i>Aspergillus oryzae</i>	fungi ^a	12,691	29,473	2.32	8.10%	RNA-seq [62]
<i>Tuber melanosporum</i>	fungi	7,496	21,525	2.87	15.4	RNA-seq [63]
<i>Cryptococcus neoformans</i>	fungi (unicellular)	6,273	35,122	5.60	4.20%	EST [64]
<i>Saccharomyces cerevisiae</i>	fungi (unicellular)	5,882	358	0.06	few cases	Not described [40]
<i>Phytophthora sojae</i>	Oomycetes (unicellular) ^a	16,988	34,207	2.01	0.70%	EST [65]
<i>Dictyostelium discoideum</i>	Amebozoa (unicellular)	13,289	16,869	1.27	0.10%	Not described [66]
<i>Plasmodium falciparum</i>	Apicomplexa (unicellular)	5,331	8,826	1.66	4.50%	RNA-seq [67]
<i>Paramecium tetraurelia</i>	Ciliate (unicellular)	40,043	90,574	2.26	0.90%	EST [42]
<i>Tetrahymena thermophila</i>	Ciliate (unicellular)	24,725	89,302	3.61	5.20%	RNA-seq, this work

^aWith hyphal stage in the life cycle.
^bGene number retrieved from the KEGG genome statistics.
^cThe total intron numbers were determined using the GTF or GFF annotation file of each organism.

doi:10.1371/journal.pone.0030630.t002

Table 2. Incidence of AS genes in diverse eukaryotes.

A. Number of introns in the gene	B. Number of genes	C. Number of IR genes	D. Number of IR introns excised			E. Frequency of the most commonly retained intron ^b	
			None	1	2 to (n-1) ^a	Growth & Starvation	Conjugation
2	781	22	1	21	NA	0.93	0.89
3	499	15	0	15	0	0.91	0.93
4	388	22	0	19	3	0.92	0.9
5	322	19	0	18	1	0.92	0.89
6	233	9	0	9	0	0.91	0.86
7	175	14	0	10	4	0.87	0.93
8	116	9	0	7	2	0.92	0.9
>8	184	16	0	15	1	0.85	0.79
Total	2698	126	1	114	11	0.90 ^c	0.89 ^c

^aOne less than the maximum number of encoded introns.
^bCalculated as the number of reads indicating retention of the most commonly retained intron divided by the number of reads indicating retention of *any* intron.
^cWeighted averages. Paired t-test shows no significant difference (paired t-test is not statistically significant, P value = 0.251 > 0.05).
NA: not applicable.
doi:10.1371/journal.pone.0030630.t003

Table 3. Distribution and specificity of intron retention type of AS events among genes encoding at least two introns.