

APPENDICES TO THE GUIDELINES FOR ANNOTATION

HAGIT SHATKAY¹, W. JOHN WILBUR², AND ANDREY RZHETSKY³

1. School of Computing, Queen's University, Kingston, ON, Canada (shatkay@cs.queensu.ca)
2. National Center for Biotechnology Information NLM, NIH, Bethesda, MD, USA (wilbur@ncbi.nlm.nih.gov)
3. Department of Biomedical Informatics, Columbia University, New York, NY, USA (ar345@columbia.edu)

Appendix A

Fragmentation examples. Where to break sentences. Sentences that remain un-fragmented.

Example 1

The Wnt/beta-catenin pathway normally regulates expression of a range of genes involved in promoting proliferation and differentiation. ****ISP3E0**

Unfragmented

Example 2

We show that treatment with ICG-001 induces apoptosis in colon carcinoma cells, ****ISP3E3+** but not in normal colonic epithelial cells. ****2SN3E0**

Fragmented at a conjunction term ("but"); the fragments demonstrate here a change in polarity from positive to negative ("but not") and a change from positive direction (induces) to a neutral trend. The indication of evidence "We show" remains as part of only the first fragment. The second fragments thus has no evidence and receives E0.

Example 3

ICG-001 (molecular weight 548) (Fig. 1A) had an IC50 of 3 μ M (Fig. 1B) against TOPFLASH ****ISP3E3**

Unfragmented.

Example 4

Given that the majority of colorectal cancers involve activation of the beta-catenin signaling pathway, and given that multiple mutations lead to this activation, ****ISP3E0+**

there is a clear need for drugs that attenuate the nuclear functions of beta-catenin (15). ****2GP3E2**

Fragmented at conjunction terms ("there is") where there is a change in:

- a. Focus, from scientific to generic; The subject changes from the specific scientific discussion of the activation of the beta-catenin signaling pathway and its causes, to the state-of-the-world where there is "a clear need for drugs",*
- b. Direction, from positive ("activation") to neutral*
- c. Evidence, from 0 (no evidence) to 2 (a supporting reference "(15)").*

Example 5

ICG-001 selectively blocked the beta-catenin-CBP interaction ****ISP3E3-** without interfering with the beta-catenin-p300 interaction. ****2SN3E0-**

Fragmented at a conjunction term ("without"). Here there is a change of polarity between the two fragments, from positive to negative. Note that the direction in both fragments is "-" since both "blocked" and "interfering" indicate a decline. The past tense of the first fragment "selectively blocked" indicates that this is an experimental report by the authors, and therefore obtains evidence level 3.

Example 6

We demonstrate that ICG-001 binds specifically to CBP ****1SP3E3**
but not the related transcriptional coactivator p300, ****2SN3E0**
thereby disrupting the interaction of CBP with beta-catenin. ****3SP3E0-**

This sentence is fragmented because part is positive and part negative. The E3 is only placed on the first fragment because of the words "We demonstrate" in the first fragment.

Example 7

Furthermore, we show that the increased somal $[Ca^{2+}]_i$ ****1SP3E3+**
and decreased cell survival following proximal transections ****2SP3E0-**
are not due to less frequent or slower plasmalemmal sealing or Ca^{2+} entry through plasmalemmal Na^+ and Ca^{2+} channels. ****3SN3E0-**

Here fragmentation is required to represent the different trends as well as the positive and negative statements.

Example 8

Rather, the increased somal $[Ca^{2+}]_i$ ****1SP3E0+**
and lethality of proximal neurite injuries ****2SP3E0**
may be due to the decreased path length ****3SP1E0-**
/increased diameter for Ca^{2+} entering the transection site to reach the soma. ****4SP3E0+**

Again fragmentation is natural to represent the different trends. Uncertainty is high in the third fragment leading to P1 and no evidence is suggested yielding E0 in all fragments.

Example 9

A ryanodine block of Ca^{2+} release from internal stores before transection has no effect on cell survival; ****1NP3E0**
however, a ryanodine- or thapsigargin-induced buildup of somal $[Ca^{2+}]_i$ before transection ****2SP3E0+**
markedly reduces cell survival, ****3SP3E0-**
suggesting a minor involvement of Ca^{2+} -induced release from internal stores. ****4SP1E0**

No uncertainty is expressed until the 4th fragment so code P3. 4th fragment starts with "suggesting" indicating some uncertainty, hence certainty is P1. Fragmentation is necessary here to allow identification of negation between the first two fragments, changes in trend between every two fragments, and change in certainty between the 3rd and the 4th fragment.

Example 10

Finally, we show that cell survival following proximal injuries can be enhanced by increasing intracellular Ca²⁺ buffering capacity with BAPTA ****1SP3E3+** to prevent the increase in somal [Ca²⁺]_i. ****2SN3E0+**

Fragmentation is necessary here to represent positive and negative statements and is helpful in showing the part of the text describing trends.

Example 11

The peak of TUNEL labeling of neurons coincides with the peak at 4 wk p.i. of up-regulated gene expression of the pro-apoptotic Fas and caspase-1, ****1SP3E0+** downregulation of mRNA for the anti-apoptotic bcl-x, ****2SP3E0-** and upregulation of mRNA for the proinflammatory IL1a and -b, IL6, and TNFa. ****3SP3E0+**

Trend keeps changing... There's no evidence for an experiment by the authors, despite the positive and confident language. The present tense of the report suggests it can be a review of someone else's work (but is not referenced). Therefore the evidence code is E0.

Example 12

Neuronal apoptosis in rabies precedes evidence of immune activation and is considered a direct, virus-induced event (35, 36). ****1SP3E2**

No fragmentation

Example 13

Expression of the pro-apoptotic cytokine TNFb correlates temporally with apoptosis in Cblm and prefrontal Cx: ****1SP3E0**

in Cblm, TNFb mRNA levels and apoptosis are high at wk 6 ****2SP3E0+**

and recede by wk 12; in prefrontal Cx, TNFb mRNA levels and apoptosis diminish earlier (wk 6). ****3SP3E0-**

Trend is changing... There's no evidence for an experiment by the authors, despite the positive and confident language. The present tense of the report suggests it can be a review of someone else's work (but is not referenced). Therefore the evidence code is E0.

Example 14

Another potential factor in promotion of apoptosis, inducible NO synthase (37), ****1SP2E2+**

is limited in distribution to perivascular infiltrates at the peak of inflammation ****2SP3E0**

and is unlikely to contribute to widespread neuronal loss. ****3SP1E0-**

Fragmentation is motivated here by a change of strength of a statement (from "potential factor" P2, to a full confidence statement starting with "is" P3, to "unlikely" P1), and change in evidence support and in trend.

Example 15

Susceptibility of neuronal populations to apoptosis after infection may depend on maturational differences in developmental programs influencing apoptosis (e.g., proliferative capacity, ref. 38; bcl-2/bax ratios, ref. 33; glutamate_N-methyl-D-aspartate receptors, ref. 39; adrenal steroids, ref. 39; protein kinase C, ref. 40; and cytokines, ref. 41), ****ISP1E2**

as presence of virus alone cannot account for the patterns of cell death observed. ****2SN3E0**

Fragmentation is motivated by a change of strength of evidence

Appendix B

Examples of many well-behaved annotations – mostly science focus, some others

Example 1

In the amphibian *Xenopus*, fertilization of the egg results in the establishment of a parallel array of microtubules with the plus end pointing away from the sperm entry point. ****ISP3E0**

The assertion is stated as fact, with high confidence, but no evidence, or even a suggestion of evidence, is provided.

Example 2

Three components of the Wnt/beta-catenin pathway-Dishevelled, glycogen synthase kinase 3 (GSK-3)-binding protein (GBP), and beta-catenin-accumulate on the side of the egg and early embryo that receive these small vesicles. ****ISP3E0**

The assertion is stated as fact, with high confidence, but no evidence, or even a suggestion of evidence, is provided. Note that while the assertion is complex, and discusses three components - all of which are listed, there is still only one assertion here with a single tag.

(The assertion is that three components on the side of the egg/embryo, receive small vesicles)

Example 3

We determined TP53 gene mutation in two cases and the genome-wide allelotype, AXIN1, and CTNNB1/beta-catenin gene mutation in one case. ****ISP3E3**

The words "We determined" in this case provide a direct description of findings resulting from the authors' own experimental work. This is the highest level of evidence - namely the report of experimental work.

Example 4

None of the NBD rats had classic Borna disease or meningoencephalitis. ****1MN3E3**

A negative assertion "None", stated as fact, with high confidence. The reference to "...the NBD rats...", in the context they are given suggests that these are the authors' are describing the setting for their own experiment. These are not the results of the experiments, but just the methodology about to be used. Therefore the focus is "M" while the evidence level is 3 (authors' own experiment).

Example 5

Apparent differences in gene expression for CNTF between NBD and NL groups did not achieve statistical significance. ****ISN2E3**

A negative assertion "...did not achieve", stated with confidence. Basing the negative finding on lack of statistical significance suggests a level of confidence that is slightly lower than a categorical-negative.

The reference to "Apparent differences...", in the context they are given suggests that these are observations made by the authors, reported as experimental results, and therefore obtain evidence level 3.

Example 6

Astrocytosis and microgliosis were evident in all brain regions by 3 wk p.i. (Fig. 7). ****ISP3E3**

The past tense for "were" suggests an observation and the reference to a figure supports this as an observation being reported, thus the E3 evidence code. No uncertainty is expressed so the P3 designation.

Example 7

Neonatally infected rats are reported not to have inflammation (6–10) ****ISN2E2**

Some uncertainty is expressed in the phrase "are reported not to have " (as opposed to just saying "do not have"), thus the P2. There is evidence in the form of references and this gives the E2 code.

Example 8

Inflammatory infiltrates are found transiently in Cx; however, their significance is unclear, as some behavioral disturbances antedate their appearance ****ISP3E1**

and infiltrates are not present in DG and Cblm, the two structures most severely damaged in this system. ****2SN3E0**

*Fragmentation occurs due to change in polarity between the first and second fragment
The occurrence of the phrase "significance is unclear" in Fragment 1 does not imply that the assertion itself is uncertain, thus the P3 rating.*

Also, the words "are found" in fragment 1, suggests that there is experimental evidence, but we are not told where this evidence may be found, thus the E1 for evidence in fragment 1.

Note that the word "found" conveys more commitment to having observed the phenomenon than the words "are not present" in the following fragment. Therefore evidence level is 0 in the second fragment.

Example 9

It could be argued that other molecular mechanisms may increase levels of the associated gene product. ****ISP1E1+**

"It could be argued" suggests the existence of evidence, but not where to find such evidence, and this leads to the E1 code. The statement seems also to be surrounded by considerable uncertainty, thus the P1.

Example 10

At the present time, then, the available data would support the notion that b-catenin mutations are only rarely seen in sporadic colon cancer. ****ISP2E1-**

"available data would support" suggests the existence of data, but we are not told where, thus E1. Some uncertainty seems expressed due to the phrases "at the present time" and "available data" which suggests lack of complete knowledge, leading to P2. The trend is negative since the phrase "rarely seen" corresponds to a low level of observation.

Example 11

Immunoblotting with anti-p300, anti-beta-catenin, or anti-TCF4 antibodies did not detect any bands on the gel (data not shown). ****ISMN3E1**

No fragmentation. Describes both methodology "immunoblotting" and the related scientific finding. Hence, both Science (S) and Methodology (M) focus. "Data not shown" suggests that evidence exists elsewhere. Hence evidence code is E1.

Example 12

ICG-001 has no effect on AP1 and CRE reporter constructs. ****ISN3E0**

No fragmentation.

Example 13

Since She2p forms a stable dimer in solution, ****ISP3E0**

we sought to establish whether or not She2p dimerization is required for RNA binding.

****2SMP0E0**

Example 14

TROSPA antisera interfered with OspA binding to native tick gut proteins; ****ISP3E3-**

however, partial inhibition of this attachment indicated that other pathways might also exist (Figure 4B). ****2SP1E3-**

Fragmentation due to change in certainty.

Example 15

Several lines of evidence demonstrate that *I. scapularis* TROSPA is a specific ligand for *B. burgdorferi* OspA. ****ISP3E1**

Appendix C

Examples of Generic statements

Example 1

Future structural and functional studies will be necessary to understand precisely how She2p binds *ASH1* mRNA and how interactions with She3p influence the formation of a functional localization complex. ****1SGP0E0**

While this is a statement with scientific contents it does not assert a completely scientific fact. Rather it makes an assertion about the necessity for further studies about a topic and finding further information. Therefore the focus is both S and G (Generic). Since it asks "how" questions, the level of certainty is 0, and as no evidence is provided, the level of evidence is also 0.

Example 2

Thus, in this case, the chemistry of the product is similar to that of the signal molecules, ****1GP3E0** but there is no complementary relationship to the signal sequences. ****2GN3E0**

While this statement describes a scientific observation, it does not refer to any specific biological or medical entity; rather, it uses generic terms ("product", "signal molecules" and "signal sequences") making the whole statement generic.

The fragmentation occurs when there is a change of polarity from positive to negative.

Example 3

Although neuropathologic studies of autism are limited, ****1GP3E1-**

reports of Purkinje and granule cell loss in Cblm (16) also suggest overlap with this neonatal infection paradigm. ****2SP2E2**

The first fragment makes a claim only about the number of studies, not their scientific content. Therefore the focus is generic rather than scientific. As there is reference to "studies" but no explicit citation, that evidence is at level 1. The direction is negative since the number is said to be "limited".

The second fragment does discuss an explicit phenomenon and provides citation, hence the fragment is viewed as scientific, with evidence level 2. Because of the term "suggest" the confidence level is not as-high-as-can-be, but just "2".

Example 4

In the present study, we evaluate the contribution to tumorigenesis and the timing of b-catenin mutations in 202 sporadic colorectal tumors, including 48 small (<1 cm) adenomas, 82 large adenomas, and 72 invasive cancers. ****1GP0E3**

The statement here is merely that evaluation is about to be conducted (of certain scientific entities), and therefore pertains to the organization of the paper. There is neither prediction of the outcome nor provided results, making it a generic statement with no scientific assertion and no confidence. However, the numbers and types of cancer/tumor cases, indicate the performance of an experiment and suggest that the evaluation has a strong experimental component, hence the evidence level E3. Since the sentence says that the contribution to tumorigenesis is to be evaluated (and does not say that there is or that there isn't such a contribution with any level of certainty) the certainty level is 0.

Example 5

The term pervasive developmental disorders was first introduced in DSM–III (American Psychiatric Association, 1980), with Asperger disorder only separated from other pervasive developmental disorders in DSM–IV. ****IGP3E2**

A general statement about the history of the terminology, not based on scientific experiment or on scientific knowledge. A reference is provided - therefore evidence code is E2.

Example 6

The bp changes, predicted amino acid change, size (for the small adenomas), and anatomical site of the respective tumors, as well as the age of the individual, are indicated in Table 1. ****IGP3E3**

The sentence simply tells what can be found in a certain table. It is not a scientific statement. It is only a statement about how the paper is structured. Evidence to support the statement is found in Table 1, thus the E3 evidence code.

Example 7

Representative chromatograms of the mutations are shown in Fig. 1. ****IGP3E2**

The sentence simply tells what can be found in a certain figure. It is not a scientific statement. It is only a statement about how the paper is structured.

Example 8

We thank M. Solbrig for valuable comments and M. Chatard, L. O'Rourke, and B. Bauder for technical assistance. ****IGP3E0**

Nothing scientific is stated and no methodology is described.

Example 9

To deal with them, the world needs to reformulate the biomolecular paradigm that has been exploited in the last two centuries. ****IGP3E0**

This statement expresses an opinion and is thus not a scientific statement.

Example 10

This article suggests that medicine recycling may be a possibility (especially if manufacturers are mandated to blister-package and bar-code individual tablets and capsules). ****IGP3E3**

This is not a scientific statement, but a statement of content of the article. Since the article itself is the evidence here we use the E3 code.

Example 11

This work was supported by the CNRS, the Human Frontier Science Program, the “Association pour la Recherche contre le Cancer,” and the “Ligue Nationale contre le Cancer.” ****IGP3E0**

Example 12

The anti-ORC1 antibody was a gift from Dr. J. Blow (Rowles et al., 1999). ****IGP3E2**

A reference is provided - therefore evidence code is E2.

Example 13

The structural features of substrate recognition by calpains are not yet fully understood. ****IGN3E0**

Example 14

As you read on, we hope you experience some of the excitement that inspired us to put this review issue together. ****IGP3E0**

Example 15

We plan to continue organizing collaborative meetings with different themes every year. ****IGP3E0**

Appendix D

Examples of Methodology statements

Example 1

DNA sequence was collected and analyzed on an ABI Prism 377 automated DNA sequencer.

****IMP3E3**

The statement describes what was done, using the past tense verb "was collected". This indicates that the authors are reporting their own work, and makes it an evidence level of 3.

The statement does not discuss results or conclusions - just the methodology used. Therefore the tag M (Methodology) is used.

Example 2

We considered as mutations only those bp changes that were verified by sequencing in both directions and by repeat amplification and sequencing. ****IMP3E3**

A statement describing what was done, using the past tense verb "We considered...were verified". This indicates that the authors are reporting their own work, and makes it an evidence level of 3. The statement does not discuss results or conclusions - just the methodology used.

Example 3

Exon 3 was amplified in two PCR reactions using intron-based nested primers. ****IMP3E3**

Example 4

The outside primers for the first PCR reaction were 59 TTCAATGGGTCATATCACAGATTC-39 and 59-CTAACTTTTTAGTTCTCAAAGTGC-39; the inside tailed primers for the second reaction were 59-UP-TCTACTAATGCTAATACTGTTTCG-39 and 59RP-ATTCTGACTTTCAGTAAGGCAATG-39 (UP and RP refer to universal primer and reverse primer, respectively). ****IMP3E3**

The statement has the M (Methodology) tag as it describes a part of the protocol used by the authors (as indicated by the use of the past-tense "were").

Example 5

This PCR was performed as follows: (a) 1 cycle at 94°C for 2 min; (b) 35 cycles of 94°C for 30 s, 60°C for 30 s, and 65°C for 12 min; and (c) 1 cycle at 72°C for 10 min (10). ****IMP3E3**

Example 6

Anesthetized rats (methoxyflurane) were perfused with 4% buffered paraformaldehyde. ****IMP3E3**

A method is described and since it is described in the past tense "were" the E3 code is appropriate as an indicator of something the investigator did.

Example 7

Results obtained from DHPLC and gel electrophoresis were completely consistent. ****1MSP3E3**

Here we have both methodology and experimental results described in such a way they cannot be readily separated. Thus the MS focus.

Example 8

Thus, DHPLC analysis of Bat26 site may be a favorable method of detecting MSI-H status in gastric cancer, and be of clinical importance. ****1MSP1E1**

Again we have method and result intermingled and use MS. Also the word "Thus" suggests some evidence which has been mentioned in the previous discussion, so E1 is used.

Example 9

Whereas individual treatment of U266 myeloma cells with 10 $\mu\text{mol/L}$ HA14-1 or 100 nmol/L flavopiridol had little effect, ****1MSP3E3-**

exposure of cells to flavopiridol (6 hours) followed by HA14-1 (18 hours) resulted in a striking increase in mitochondrial dysfunction ****2MSP3E3+**

(cytochrome c and Smac/DIABLO release; ****3SP3E0**

loss of mitochondrial membrane potential), ****4SP3E0-**

activation of the caspase cascade, ****5SP3E0+**

apoptosis, and diminished clonogenic survival. ****6SP3E0-**

This is a complex sentence which is conveniently treated as six fragments.

The first two contain some methodology and indicate the results are findings of the current experiment (indication is through the use of the past tense "had little effect" and "resulted in a striking increase..."), and are given the E3 code.

Example 10

After 6 months of tamoxifen therapy, the total cholesterol decreased -9.7% and the LDL cholesterol -16.7% from levels after the chemotherapy, ****1MSP3E3-**

while the cholesterol concentrations remained at increased levels in the control group ($P=0.001$ and $P<0.0001$, respectively). ****2SP3E3+**

Fragmentation is necessary here to represent the difference in trends. Also note that methodology is stated in the first fragment only and that in the first fragment a decrease is signaled by the word "decreased" as well as by the "-" sign for two different entities.

Example 11

To generate p18-deficient mice, we used the gene trap method (Zambrowicz et al., 1998).

****1MP3E3**

Example 12

The cDNAs encoding the FAT and C-terminal domains of ATM were obtained by PCR with their specific primers. ****1MP3E3**

Example 13

The wt *rpoC* was amplified by PCR with a high-fidelity *Pwo* enzyme (Roche) using β' L 5'-gggattaacctggctatgaaagattattaaagtt-3' and β' R 5'-ttcggggccaagcttctcggtatcagaaccgcc-3' primers and cloned into the Nco I and Hind III sites of the pBAD-A plasmid (Invitrogen). ****IMP3E3**

Example 14

To test this hypothesis and to address its physiological significance, ****IGP3E0**
we developed a genetic screen for dominant-negative mutations in the G region of the β' subunit. ****2MP3E3**

Example 15

To study the effects of growth factor withdrawal on cells lacking the intrinsic apoptotic pathway, ****IGP3E0**
immortalized interleukin-3- (IL-3) dependent cell lines were generated from the bone marrow of $Bax^{-/-}Bak^{-/-}$ mice. ****2MP3E3**

Evidence and focus changes in the middle of sentence—hence fragmentation.

Appendix E

Examples for the various types of evidence

Example 1

The Wnt/beta-catenin pathway normally regulates expression of a range of genes involved in promoting proliferation and differentiation. ****ISP3E0**

Stated as a well-accepted fact, but with no indication for any evidence. Hence, E0.

Example 2

We demonstrate that ICG-001 binds specifically to CBP ****ISP2E3**

but not the related transcriptional coactivator p300, ****2SN2E0**

thereby disrupting the interaction of CBP with beta-catenin. ****3SP2E0-**

First fragment: "We demonstrate" indicates that experiments performed by the authors support the statement, E3.

The sentence is fragmented, due to change in polarity to negative ("but not"), and then again to positive polarity - with a negative trend ("thereby disrupting").

The last two fragments do not carry any indication of evidence, therefore the code becomes E0.

Example 3

ICG-001 (molecular weight 548) (Fig. 1A) had an IC50 of 3 ~M (Fig. 1B) against TOPFLASH ****ISP2E3**

The past tense "had" and the references to the results shown in the figures suggest that these are reports from experiments conducted by the authors. Hence, E3.

Example 4

At the present time, then, the available data would support the notion that b-catenin mutations are only rarely seen in sporadic colon cancer. ****ISP2E1-**

The terms: "the available data" indicate that evidence is available, but there is no explicit reference to data produced by the authors (experiments) or to a paper presenting data produced by anyone else.

Therefore the evidence is very indirect, and is indicated as E1.

Example 5

Mechanisms underlying morphogenesis have been studied extensively in the unicellular yeast systems (Lew, 2003). ****IGP3E2**

The explicit reference to Lew, 2003, provides evidence of type 2 (as defined in the guidelines) to the claim. Hence E2.

Example 6

Whereas individual treatment of U266 myeloma cells with 10 μmol/L HA14-1 or 100 nmol/L flavopiridol had little effect, ****1MSP3E3-**

exposure of cells to flavopiridol (6 hours) followed by HA14-1 (18 hours) resulted in a striking increase in mitochondrial dysfunction ****2MSP3E3+**

(cytochrome c and Smac/DIABLO release; ****3SP3E0**

loss of mitochondrial membrane potential), ****4SP3E0-**

activation of the caspase cascade, ****5SP3E0+**

apoptosis, and diminished clonogenic survival. ****6SP3E0-**

This is a complex sentence which is conveniently treated as six fragments.

The first two contain some methodology and indicate the results are findings of the current experiment (indication is through the use of the past tense "had little effect" and "resulted in a striking increase..."), and are given the E3 code. The other fragments do not contain an indication of evidence.

Example 7

Astrocytosis and microgliosis were evident in all brain regions by 3 wk p.i. (Fig. 7). ****1ISP3E3**

The past tense for "were" suggests an observation and the reference to a figure supports this as an observation being reported, thus the E3 evidence code. No uncertainty is expressed so the P3 designation.

Example 8

Neonatally infected rats are reported not to have inflammation (6–10) ****1SN2E2**

Some uncertainty is expressed thus the P2. There is evidence in the form of references and this gives the E2 code.

Example 9

Inflammatory infiltrates are found transiently in Cx; however, their significance is unclear, as some behavioral disturbances antedate their appearance ****1SP3E1**

and infiltrates are not present in DG and Cblm, the two structures most severely damaged in this system. ****2SN3E0**

Fragmentation occurs due to change in polarity between the first and second fragment

The occurrence of the phrase "significance is unclear" in Fragment 1 does not imply that the assertion itself is uncertain, thus the P3 rating.

Also, the words "are found" in fragment 1, suggests that there is experimental evidence, but we are not told where this evidence may be found, thus the E1 for evidence in fragment 1.

Note that the word "found" conveys more commitment to having observed the phenomenon than the words "are not present" in the following fragment. Therefore evidence level is 0 in the second fragment.

Example 10

It could be argued that other molecular mechanisms may increase levels of the associated gene product. ****ISP1E1+**

"It could be argued" suggests the existence of evidence, but not where to find such evidence, and this leads to the E1 code. The statement seems also to be surrounded by considerable uncertainty, thus the P1.

Example 11

This work was supported by the CNRS, the Human Frontier Science Program, the "Association pour la Recherche contre le Cancer," and the "Ligue Nationale contre le Cancer." ****IGP3E0**

This is a peculiar case of a generic positive statement—there is no reference to supporting data (such as a document or legal contract that could have been cited), so the evidence code is E0.

Example 12

Immunoblotting with anti-p300, anti-beta-catenin, or anti-TCF4 antibodies did not detect any bands on the gel (data not shown). ****IMSN3E1**

It is explicitly stated that the experimental data supporting the statement exist, (as evidenced by the statement "data not shown"), but not explained in detail within the paper.

Example 13

Susceptibility of neuronal populations to apoptosis after infection may depend on maturational differences in developmental programs influencing apoptosis (e.g., proliferative capacity, ref. 38; bcl-2/bax ratios, ref. 33; glutamate/N-methyl-D-aspartate receptors, ref. 39; adrenal steroids, ref. 39; protein kinase C, ref. 40; and cytokines, ref. 41), ****ISP1E2**

as presence of virus alone cannot account for the patterns of cell death observed. ****2SN3E0**

Fragment 1: The supporting evidence is provided in the form of references to other explicitly defined papers, so the code is E2.

Fragment 2: It is unclear how the conclusion is supported, so the default code is no evidence—E0.

Example 14

TROSPA antisera interfered with OspA binding to native tick gut proteins; ****ISP3E3-**

however, partial inhibition of this attachment indicated that other pathways might also exist (Figure 4B). ****2SPIE3**

In this case the results were directly observed by authors and even their speculation is supported by an experimental figure - the strongest evidence E3.

Example 15

Since She2p forms a stable dimer in solution, ****ISP3E0**

we sought to establish whether or not She2p dimerization is required for RNA binding.

****2SMP0E0**

No clear indication of any experimental support to the statements.

Appendix F

Examples interesting border-line cases in polarity, direction, methodology - and other basket cases ☺

Example 1

At the present time, then, the available data would support the notion that b-catenin mutations are only rarely seen in sporadic colon cancer. ****ISP2E1-**

The expressions "at the present time" and "would support" suggest an intermediate level of certainty in the statement.

The phrase "only rarely seen" suggests a positive result, but with a very low level of the observed phenomenon (presence of b-catenin mutation in colon cancer). Therefore the negative trend.

The phrase "the available data would support..." suggests that evidence does exist but it is not explicitly given. Therefore, the evidence level is 1.

Example 2

In Cblm, Purkinje cells stained as early as 2 wk; ****ISP3E3**

however, only a small proportion of astrocytes and Bergmann glia were infected, ****2SN2E3**

and neither microglia nor cerebellar granule cells showed signs of infection. ****3SN3E3**

The first fragment indicates an experimental result (through term "stained" and time given), and states affirmatively and with certainty. Hence the tags are P3 (positive & high certainty) and E3 (experimental evidence).

The second fragment starts with the words "however, only" which suggests a result that was contrary to the expected - hence negative. Since there was still a small proportion of infection observed, the negative statement is not categorical, and therefore has certainty level 2. The evidence is experimental as before.

The third fragment also reports negative results, this time categorical (neither, nor) and therefore the certainty level is 3.

Example 3

In fact, as demonstrated using several SOD assays including pulse radiolysis, 2-ME does not inhibit SOD ****ISN3E3-**

but rather interferes with the SOD assay originally used. ****2SP3E3-**

The first fragment indicates an experimental result (through term "using" and "assays"). It is a negative result ("does not") stated with certainty. The direction assigned is "-" as the result is concerned with inhibition. Note that the terms "does not" and "inhibit" do not cancel each other out.

The "does not" is captured by the negative polarity, while the "inhibit" is captured by the negative direction of the result and the tag "-".

The second fragment is stated in the affirmative (interferes), but the word "interfere" has a negative direction indicating decline/reduction, and therefore the direction remains negative and is indicated by "-".

Example 4

Given that the majority of colorectal cancers involve activation of the beta-catenin signaling pathway, and given that multiple mutations lead to this activation, ****ISP3E0+**

there is a clear need for drugs that attenuate the nuclear functions of beta-catenin (15). ****2GP3E2**

The first fragment makes a scientific statement that is positive (P) and certain (level 3), but provides no evidence (simply states it as a "Given"). Hence the evidence level is E0. Since it talks about "activation", the statement has a positive direction, indicated by "+".

The second fragment makes a general statement about the need for a drug in the world outside the scope of the paper, and is therefore tagged as "G". Note that a reference is given, which implies an evidence level 2 (E2).

Example 5

At 4 wk, NBD rats had prolonged locomotor inhibition upon introduction to the novel environment (0–30 min) ****ISP3E3-**

This is a positive and certainly stated fact, based on experimental evidence (the expression "...rats had prolonged..." and the time (0-30 min), makes it clear that this is an experiment-result report. However, the report is about "inhibition", and is therefore tagged with a negative direction "-".

Example 6

Our blood studies of 61 children with autism do not show evidence of BDV infection (unpublished results). ****ISN2E1**

"unpublished results" leads to the E1 designation. The result is stated in the negative resulting in an N2 code. One should code this as N2 as there is some uncertainty expressed.

Example 7

Pilot studies indicate a more adult pattern of CNS disease when rats are inoculated after 12 h of postnatal life or with lower-titer stocks. ****ISP3E1**

"Pilot studies" are mentioned but no reference is given, so the evidence code is E1. No uncertainty is expressed so P3 is appropriate.

Example 8

We conclude that mutation of b-catenin can be an early, perhaps initiating, event in colorectal tumorigenesis. ****ISP1E1**

"can be" suggests considerable uncertainty leading to the P1 code. "We conclude" suggests that evidence has been given in the preceding text to support the assertion and thus the E1 code.

Example 9

Loss of serine or threonine phosphorylation sites from exon 3 of b-catenin has been identified in approximately half of colorectal tumors which lack adenomatous polyposis coli (APC) mutations, ****ISP2E1-**

The pluperfect tense "has been identified" indicates some experiments by someone in the past and provide the E1 code. The words "has been identified" also suggest some level of uncertainty leading to the P2 designation. The words "lack . . . mutations" give rise to the - code.

Example 10

The frequency of COX-2 and PPARgamma detection was significantly increased ****ISP3E3+** and decreased as lesions progressed to carcinoma, respectively. ****2SP3E3-**

Fragmentation is necessary here to capture the opposing trends, but note that the statement must be understood as a whole in order to capture its meaning.

Example 11

The requirement for a heterologous surface region for She3p binding could explain ****1SP1E0** why She3p interacts efficiently with She2p: RNA ****2SP3E0** and not with She2p alone (Long et al., 2000) ****3SN3E2** and why She3p stabilizes She2p-RNA interaction (Böhl et al., 2000). ****4SP3E2**

The first two fragments indicate no evidence. There is a change in certainty between the two. The second two fragments provide evidence by references, which is evidence code E2.

Example 12

Confocal immunofluorescence studies showed TROSPA labeling on the surface of *Drosophila* cells, ****1SP3E3**

indicating that the protein was predominantly associated with the plasma membrane of transfected cells (Supplemental Figure S1B). ****2SPIE3**

The evidence supporting the statements is provided in a figure—evidence code E3.

Example 13

To test whether macroautophagy plays a role in maintaining growth factor-independent cell survival, ****1GP1E0**

shRNA against *ATG5* were introduced into the IL-3 dependent cells. ****2MP3E3**

The words "were introduced" indicate an experimental support—evidence code E3.

Example 14

We found that treatment of Cdk inhibitor (olomoucine) decreased the phosphorylation levels of CRMP-2 in COS7 cells (data not shown) ****1SP3E1-**

and hippocampal neurons (Supplemental Figure S1B) and that GSK-3 β phosphorylated CRMP-2 that was prephosphorylated by Cdk5 in vitro (Figure 1C). ****2SP3E3**

The first fragment says "We found" which would suggest evidence code 3, but "data not shown" changes this into evidence code E1, since the evidence is not given here but is only known to exist. The second fragment has evidence code E3, as the figures show the data and provide the evidence.

Example 15

It has come to our attention that panels D, E, and G in Figure 6 of Nolan et al. (Cell 119, 719–732, November 2004) are incorrectly labeled. ****1GP3E12**

Evidence code is E12; Someone (unspecified) brought it to their attention, hence E1, and the item under dispute is in the referenced paper which is evidence code E2.