Supplementary file S1: Example of use case applying the MI2CAST concepts for the manual curation of a causal statement from a publication.

The selected paper for this example is the following: “Characterization of E2F8, a novel E2F-like cell-cycle regulated repressor of E2F-activated transcription” [49], that highlights a causal interaction between a transcription factor (E2F8) and a target gene (CCNE1). By following the rules of MI2CAST, it is possible to annotate the following information:

Rule 1: Source and Target entities
UniprotKB:A0AVK6 (E2F8): the source entity is the E2F8 transcription factor annotated with a Uniprot identifier.
Entrez:898 (CCNE1): the target entity is the CCNE1 gene annotated with an Entrez gene identifier.

Rule 2: Effect
MI:2240 (down-regulates): the effect of the causal interaction is a down-regulation, annotated with a Molecular Interactions Controlled Vocabulary (MI) term. E2F8 down-regulates CCNE1.

Rule 3:
3.1: Reference
PMID:16179649: the causal statement is assessed in the article entitled “Characterization of E2F8, a novel E2F-like cell-cycle regulated repressor of E2F-activated transcription”. The PubMed identifier is given.

3.2: Evidence type
ECO:0005648 (luciferase reporter gene assay evidence used in manual assertion): the observation of the causal interaction is done via a luciferase reporter gene assay experiment and has been extracted through a human review. An Evidence and Conclusion Ontology (ECO) identifier is given.

3.2.1: Experimental setup
MI:0506 (over expressed level) and MI:0331 (engineered) for the source entity: The source entity has been over-expressed and engineered during the experiment, annotated with an MI controlled vocabulary term.
MI:0331 (engineered) and SO:0001679 (transcription_regulatory_region) for the target entity: the target entity has being engineered and its transcription regulatory region is used during the experiment, annotated with an MI controlled vocabulary term and a Sequence Ontology (SO) term.

Rule 4:
4.1: Biological activity or mechanism
MI:2247 (transcriptional regulation): the biological mechanism of the causal interaction is a transcriptional regulation, annotated with an MI controlled vocabulary term. E2F8 is a regulator of the transcription of CCNE1.

4.2: Biological type
This is not necessary to be annotated as the correct identifiers have been given: the source entity is a transcription factor (i.e., protein) annotated with a UniProtKB identifier and the target entity is a gene annotated with an Entrez gene identifier.
4.3: Biological modification
There is no information about a specific biological modification of the source nor the target entity in this article, thus no annotation is added for Rule 4.3.

4.4: Taxon
The taxon information is defined by the entities' identifiers: the UniProtKB identifier “A0AVK6” is a human E2F8 protein and the Entrez gene identifier “898” is a human CCNE1 gene.

4.5.2: Cell type or cell line
**BTO:0001938** (Human osteosarcoma cell line): the causal interaction is observed in the human osteosarcoma cell line.

4.5.3: Cellular component
**GO:0005634** (nucleus): the causal interaction occurs in the nucleus and both entities are located in the nucleus, annotated with a gene ontology term.

*It can be noted that this information can be inferred from the biological nature of this causal interaction: as the target entity is a gene and the source entity a transcription factor (TF), the TF exerts its influence on the gene inside the nucleus.*

More examples can be found at [https://github.com/MI2CAST](https://github.com/MI2CAST).
Supplementary file S2: Extensive list of identifiers for “entities”.

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
<th>Ontology/Controlled vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene</td>
<td>A region of a sequence that encodes an RNA transcript.</td>
<td>Ensembl Entrez gene HGNC MOD (Model Organism Database)</td>
</tr>
<tr>
<td>non coding RNA</td>
<td>An RNA sequence that is not translated into a protein.</td>
<td>RNAcentral Wormbase</td>
</tr>
<tr>
<td>mRNA</td>
<td>An RNA sequence that can be translated into a protein.</td>
<td>Ensembl transcript Ensembl gene MOD</td>
</tr>
<tr>
<td>protein</td>
<td>A large biomolecule composed of one or more chains of amino acid residues.</td>
<td>UniProtKB PRO (Protein Ontology)</td>
</tr>
<tr>
<td>chemical</td>
<td>A small molecule.</td>
<td>ChEBI PubChem ChEMBL DrugBank</td>
</tr>
<tr>
<td>family</td>
<td>A group of entities that share a common structure and/or function.</td>
<td>List of components Biological type: family</td>
</tr>
<tr>
<td>transient complex</td>
<td>A group of entities that temporarily interact together to perform a function (all components are necessary for the complex to perform that function).</td>
<td>List of components Biological type: complex</td>
</tr>
<tr>
<td>stable complex</td>
<td>A group of entities that permanently interact together to perform a function (all components are necessary for the complex to perform that function).</td>
<td>ComplexPortal List of components Biological type: complex</td>
</tr>
<tr>
<td>phenotype</td>
<td>An observable phenomenon related to a biological process.</td>
<td>Gene Ontology Biological Process HPO (Human Phenotype Ontology) PATO (Phenotype And Trait Ontology) DO (Disease Ontology)</td>
</tr>
</tbody>
</table>