CancerMine Annotation Guidelines

CancerMine will display a sentence with two entities highlighted: a cancer type and a gene. You must identify whether those two entities are in one of the following relationships:

Driver	The sentence must be strongly stating that the gene, or aberration of the gene drives cancer progression. While the term "driver" is used differently by different researchers, the sentence must explicitly use the words drive/driver/drives or similar to imply a special significance to this gene/aberration. Example: "Imatinib was initially developed for its selective action against the Bcr-Abl fusion protein, a key driver of chronic myeloid leukemia ." (PMID:24119229)
Oncogenic	The sentence must state that the gene, or aberration, is linked with cancer development/progression. It may discuss it being an oncogene, having oncogenic function or other similar phrasing. Example: "These in vivo data confirm our in vitro results and suggest the oncogenic role of Eag in osteosarcoma." (PMID:23450231)
Tumor Suppressive	The sentence must state that the gene/aberration is a tumor suppressor for that type of cancer. This is most commonly phrased explicitly as a tumor suppressor or suppressing but other similar terms are acceptable. Example: "CLIC4 is a tumor suppressor for cutaneous squamous cell cancer." (PMID: 22387366)
None	Select this if none apply Example: "Variants in this gene could lower or raise the sensitivity of this DNA damage detector and impact upon the efficiency of p53 and its tumor suppressor pathway and can predispose women to breast cancers"

Additional Notes

- Only annotate explicit relations of a cancer type and gene. Use None if the relation is only implied.
- Use None if the sentence is only a hypothesis and not an explicit claim.
    - e.g. "We hypothesis that IRF1 is an oncogene in osteosarcoma"
- An annotation of Driver/Oncogenic/Suppressive may also relate to a subprocess of cancer (e.g. EMT or angiogenesis).

- Annotate all sentences regardless of the model system used. Even if it is in mouse, or another model, it is still valuable information.
- CancerMine may also show miRNA as well as genes. Please annotate them also.
- CancerMine will also attempt to spot fusion genes (two gene names separate by a hyphen or slash).
- There is two additional annotation type:
    - "**Entity Error**", where you believe that it has failed to properly extract one of the entities.
        - For instance a sentence that describes "squamous cell carcinoma", but CancerMine only highlights "carcinoma"
    - "**Requires Discussion**": where you believe that there is not an easy answer and that we should discuss the sentence. Please give your best guess for the other annotations and we can revisit it later
- The sentence may describe multiple relations between the two entities. For instance, it may state that a gene is a "driver" and "oncogene". Please select both annotations for it. The sentence may state contradictory results: that a gene has found to be an oncogene as well as a tumor suppressor. Please annotate both.
- Include cases with some speculation: i.e. it is a probable oncogene, likely a tumor suppressor. But not straight up hypothesis: we wonder if X is an oncogene
- CancerMine tries to avoid processing both the acronym and full version if they're both present (e.g. non-small cell lung cancer (NSCLC)). But if they are present, annotate both relations the same
- Annotate the entities that are closest to relation of interest. For example if the sentence describes a relation between CDKN2B and NSCLC, and CDKN2B is mentioned several times in the sentence, only annotate those cases which are reasonably part of the relation.
- mTOR pathway does not mean mTOR so don't tag it as such
- If the gene is only an oncogene under certain conditions (i.e. when expressed with another gene), just tag it as oncogene.
- Associated isn't enough for a positive annotation (i.e. mutations in BCL-2 are associated with NSCLC). In this case, it should be annotated as None
- If a gene is referred to in the context of "<GeneName> oncogene", or "<GeneName> tumor suppressor", tag it as such, even if that's not the main focus of the sentence (i.e. We studied the structure of the TP53 tumor suppressor gene in colon cancer").

Challenging Examples

Below are several challenging sentences with the appropriate annotation explained below.

1. *Two proto-oncogenic chromatin regulators, Bmi-1 and high-mobility-group A2 (Hmga2), were identified in 12 pair cases of HNSCC tumor regions by comparison with their non-cancerous background tissues using cDNA microarray. (PMID: 24145240)*
gene: Hmga2 // cancer: HNSCC

The gene is described as proto-oncogenic, so it should be tagged as an Oncogene.

2. *The oncogenic properties of eEF1A2 have also been studied in different ovarian tumors and established cell lines. (PMID: 25905039)*
gene: eEF1A2 // cancer: ovarian tumors

eEF1A2 has oncogenic properties so should be tagged as an Oncogene

3. *miR-152 is thought to play an important role in oncogenesis of endometrial cancer. (PMID: 23291663)*
gene: miR-152 // cancer: endometrial cancer

It doesn't specifically say that it drives it but that it is involved in oncogenesis. It should just be tagged as an Oncogene.

4. *Furthermore, myeloid progenitors from Gab2 −/− mice are resistant to transformation by Bcr-Abl, indicating that Gab2 is required to sustain the leukemogenesis evoked by this oncogenic fusion protein in a model of chronic myelogenous leukemia (CML) . (PMID: 26045784)*
gene: Bcr-Abl // cancer: chronic myelogenous leukemia

The wording is a little difficult, but it does describe Bcr-Abl as an oncogenic fusion protein, so this should be tagged as Oncogene.

5. *Strikingly, transgenic expression of miR-155 induced pre-B cell leukemias and lymphomas in mice , thus miR-155 seems to be a potent oncogene. (PMID: 19890372)*
gene: miR-155 // cancer: leukemias
gene: miR-155 // cancer: lymphomas

It states it is an oncogene. In this case, you would only want to tag the final "miR-155" as oncogene, as the first one isn't discussed as an oncogene.

6. *Inactivation of TBL1 in human and mouse pancreatic cancer cells reduced cellular growth and aggressiveness and also regulated the activity of the oncogenic PI3 kinase signaling pathway in vitro and in vivo . (PMID: 26070712)*
gene: TBL1 // cancer: pancreatic cancer

The sentence implies that the gene is very important for cancer development. However it doesn't explicitly discuss it using the words driver/driving etc, no discusses oncogenic processes explicitly. Hence this should be tagged as None.

7. *Based on these findings, we concluded that down-regulation of SAE2 expression enhanced tumor suppression and sensitivity of chemotherapy in SCLC, and targeting SAE2 may be a new method for patients with SCLC. (PMID: 26063074)*
gene: SAE2 // cancer: SCLC

As in the previous case, this implies that SAE2 is important in cancer development but doesn't state it explicitly. It should be tagged as None.

8. *Overexpression of the proto-oncogene ERBB2 (HER2/NEU) has been observed in 20-30% of breast cancers involving poor prognosis. (PMID: 15550452)*
gene: ERBB2 // cancer: breast cancers
gene: HER2/NEU // cancer: breast cancers

ERBB2 is discussed as a proto-oncogene and should be tagged as an Oncogene.

9. *It remains unclear whether oncogenic Stat3 signaling pathway is involved in the oncogenesis of bladder cancer. (PMID: 18939995)*
gene: Stat3 // cancer: bladder cancer

This discusses a signaling pathway, not a single gene, and should be tagged as None.