

RNA – RNA level expression of genes

1. **GENE ID** : ensemble ID (repeated for each sample ID) - **text**
2. **SAMPLE** : individual code (repeated for different genes) - **text**
3. **CANCER** : code for tumor type (repeated for many sample IDs) - **text**
4. FPKM : transcript per million (normalized count of RNA molecules for a certain gene) - **double**

Threshold on FPKM + cancer type → Gene ID

Converter – Conversion of gene ID

1. NCBI Gene ID : Unique gene ID - **int**
2. **Ensembl gene ID** : Unique gene ID - **text**
3. Approved symbol : Unique gene ID- **text**

Ensembl → Hugo/Approved symbols (one-to-one)

Onco genes – Onco-annotation for all genes

1. **Hugo Symbol** : Hugo symbol/ Approved symbol - **text**
2. Entrez gene ID : Entrez gene ID - **int**
3. OncoKB Annotated : Can be targeted for treatment - **text**
4. Is Oncogene : its level of expression changes tumor growth (+) - **text**
5. Is Tumor Suppressor gene : its level of expression changes tumor growth (-) - **text**

Hugo Symbol → Hugo symbol filtered according to yes/no (is oncogene)

Onco drug – FDA approved molecules

1. **Gene** : Hugo symbol/ Approved symbol - **text**
2. Alterations : Type of alteration of gene - **text**
3. Cancer Types : Tumor type (extensive version) - **text**
4. Drugs : Name of the drug - **text**

Hugo symbol filtered → Drugs

Queries:

Da RNA calcoarsi view con media per gene

- From drug to sample ID

(Onco drug → Converter → RNA)

1. Scegli una drug → Lista di geni
2. Lista di geni in Hugo → Lista di geni in ensembl
3. Lista di geni in ensembl → Lista di samples id (if fpkm > threshold)

- Drugs associated to cancer type

(RNA → Converter → Onco genes → Onco drug)

1. Scegli un cancer type → Lista di geni in esembl (if fpkm > threshold)
2. Lista di geni in ensembl → Lista di geni in Hugo
3. Lista di geni in Hugo → Lista di oncogeni in Hugo
4. Lista di oncogeni Hugo → Lista di drugs utilizzabili per quel cancer type

- Find tumor suppressor genes

(Onco genes → Converter → RNA)

1. Seleziona solo suppressor genes → Lista di Hugo geni
2. Lista di Hugo geni → Lista di ensembl geni
3. Lista di ensembl geni → Sample id e gene ID dei geni che passano il threshold

- Find alterations for cancer type

(RNA → Converter → Onco genes → Onco drugs)

1. Seleziona un tipo di tumore → Lista di geni in ensembl (if fpkm > threshold)
2. Lista di geni in Ensembl → Lista di geni in Hugo
3. Lista di geni in Hugo → Lista di geni in Hugo (if oncogene == yes)
4. Lista di geni filtrata → Lista di geni filtrata + tipo di alterazione

- Find non annotated oncogenes

(RNA → Converter → Onco genes → Onco drugs)

1. Geni in ensembl (if fpkm > threshold) → Lista di geni filtrati in ensemble
2. Geni in ensemble → Lista di geni in Hugo
3. Lista di geni che sono oncogenes → Lista di geni che non sono annotati in oncokb

- Find ensembl gene id subjected to a specific type of alteration

(Onco drugs → Onco genes → Converter)

- Find those genes that do not pass threshold even if they are annotated in oncokb

(Onco drugs → Onco genes → Converter → RNA)

- Find the percentage of genes that are overexpressed given a cancer type and pass the oncogene filter

(Onco drugs → Onco genes → Converter → RNA)

- For each sample ID return overexpressed gene and possible type of alterations

(RNA → Converter → Onco genes → Onco drugs)

Improvements:

- Come stabilire threshold? Media tra tutti gli FPKM di uno stesso gene in uno stesso cancer type?
- Additional table for conversion of extended cancer type to code
- Pre-processing: binary instead of yes and no