***In silico* analysis of the PTCome distribution (Kernel plots)**

This method consists on *in silico* methodology to visualize and analyse the qualitative distribution of the collection of PTC (PTCome) found on genes of interest in association to disease (PTEN is used as example), as annotated in the literature and in gene mutation and gene variant databases. The qualitative analysis of PTC distribution along the domains of the protein under study may provide predictive genotype/phenotype information and potential attributes of pathogenicity to the distinct protein regions.

1. **Obtaining the potential PTCome** (PTC that can be generated by single‐nucleotide substitutions from a protein-coding nucleotide sequence)
   1. To obtain the potential PTCome, extract from Gene database (NCBI, https://www.ncbi.nlm.nih.gov/gene/) the cDNA sequence of the desired gene. Copy it in a text document.
   2. Import the text document in the PTCMAKER program to obtain a text file with potential PTC residues.
   3. Import the obtained text file in RStudio by File > Import Dataset > From text (readr) or using the *read\_csv* function.
   4. Run the following code in R (<https://github.com/leiretorices/Kernel-plots>, section 1) to obtain the list of potential PTC residues.
2. **Germline-associated PTCome** (PTC generated by single‐nucleotide substitution mutations found in the germline)
   1. To obtain the germline-associated PTCome, browse for the desired PTP in the HGMD (<https://www.hgmd.cf.ac.uk>; https://digitalinsights.qiagen.com/products-overview/clinical-insights-portfolio/human-gene-mutation-database/) database. Click on Missense/Nonsense, which displays a dataframe with missense and nonsense mutations together. Germline-associated PTCome can also be obtained from ClinVar (NCBI, <https://www.ncbi.nlm.nih.gov/clinvar>) database.
   2. Import the downloaded file in RStudio by File > Import Dataset > From Excel, or using the *read\_excel* function. Additional PTC described in the literature or in other databases should be added manually if the number of different PTC in HGMD is low.
   3. Run the following code in R (<https://github.com/leiretorices/Kernel-plots>, section 2) to obtain unique PTC mutated residues.
3. **Cancer-associated PTCome** (PTC generated by single‐nucleotide substitution somatic mutations found in tumors), of utility for PTP associated with tumor suppression of oncogenesis.
   1. To obtain the cancer-associated PTCome, browse for the desired PTP in the COSMIC database (Wellcome Trust Sanger Institute; https://cancer.sanger.ac.uk/cosmic). Go to the mutation distribution section and click on nonsense substitution.
   2. Download/Export the desired dataset in a comma delimited (.csv) format. Import the downloaded file in RStudio by File > Import Dataset > From text (readr) or using the *read\_csv* function, specifying comma as a delimiter. Reliable cancer-associated PTCome can be obtained only for PTP whose function is related with cancer and display a significant number of PTC mutations distributed along their sequence.
   3. Run the code in R (<https://github.com/leiretorices/Kernel-plots>, section 4) to obtain unique PTC mutated residues.
4. **PTCome qualitative representation: Kernel density plot**
   1. Insert in lines 40 and 41 of the code (<https://github.com/leiretorices/Kernel-plots>, section 3) the amino acid length and the name of your protein of interest, respectively (PTEN protein name and amino acid length are given as examples).
   2. Run the code in (<https://github.com/leiretorices/Kernel-plots>, section 3) to obtain a Kernel density plot representation. The vectors (potential PTCome, germline-associated PTCome, cancer PTCome) are rescaled to 100 to facilitate comparisons between proteins. Mirror vectors are generated to avoid bias at boundaries and a density is calculated of sum of the parental vector and the mirror vectors. The final plot contains curves representing the potential PTCome (in black) and the germline-associated PTCome (in blue).