A linked data approach to discover HPV oncoproteins and RB1 induced mutation associations for the retinoblastoma research.

Alokkumar Jha[1], Yasar Khan[1], Dietrich Rebholz-Schuhmann[1], Ratnesh Sahay[1]

Insight centre for data analytics, National university of Ireland, Galway, Ireland [1]

Background: *LOSS* or *GAIN* in tumor suppressor gene *RB1* play a significant role as in the case of loss low penetrance where only *39%* of the eye at risk develops in *retinoblastoma*. This research covers the multiple mutation types and its effects and identification of the major type of mutation involved in retinoblastoma because of HPV and RB1.
Methods: First, we focus on exploring gene expression (GE) patterns for RB1 and HPV-associated genes from TCGA. Second, identification of validated and non-validated standard CNV ensured using the COSMIC. Finally, the clinical profiles of filtered mutations have been validated based on ICGC pathological profiling data to infer the prognostic behavior from RB1 and HPV-associated genes. In order to link and retrieve patterns of a gene from TCGA, COSMIC, and ICGC repositories, we performed following steps: transform heterogeneous data repositories and their storage formats into standard Resource Description Framework (RDF) format; to discover associations by finding specific patterns (i.e. correlations) in the GE data sets; scalable querying the large volume and frequently updating datasets covering the GE data from different repositories.
Results: HPV mutations indicated in more than 127 cancer studies shows deletion and *amplifications* are rare mutations.
Retinoblastoma: Expression profile of *RB1* shows mutations such as nonsense, Missense or splice events and in GBM and gliomas the expression values in splice mutations (1500-200), nonsense mutations (200-600). In principal HPV-associated retinoblastoma the higher expression of HPV genes results in splice junctions and lower in nonsense mutations**.**
Other tissues: Pattern of *RB1* where the results coming from more 123 studies show the pattern of mutations similar to the results obtained from HPV-associated genes. Alteration with HPV genes study based on the alteration in Altered in 90% samples of 61 cases where TP53 is holding 90% occurrence majorly as normal mutations and SNRNP70 and BRCA1 is majorly responsible truncating mutation other highly mutated genes are *AP3D1, BRD4, CCHCR1, CPSF4, CREBBP, CUL3, DDX11, EP300, EP400, FRZ1, GNB2L1, GTF2B, KDM5C, NR4A1, PRP1, PLK1, SF1, SRSF1, SRSF7, SMARCB1, SNRNP70, TAF1, TBP, TMF1, TOPBP1*. Whereas *RB1* is associated with 11% cases of deep deletion where the *V654L* is the normal mutation and all others are a highly truncating mutation. Survival graph for HPV and RB1 associated genes median months(m) of survival with alteration in these query genes are 103m whereas in RB1 the median month of survival are 7.63m, however, the disease-free survival in RB1 cases are 4.50m. The *p-value* are 0.76, 0.67, 0.38 respectively. To demonstrate a pattern of survival gene set enrichment have been performed on both gene lists where in the case of RB1 genes with highest interactions are MDM2, CDK4, and TP53. In HPV genes interacting hubs are *TOP1, PARP1, TP53,* and *ODF2*. Higher interacting genes are associated with drugs. RB1 corresponded to *Insulin, p* a non-cancer FDA approved drug whereas HPV genes and especially TOP1 is associated with Lucanthone, Irinotecan, BTBD1 and Topotecan in the case of FDA approved drugs category. Cancer drugs with HPV genes are majorly associated with *TOP1, PARP1,* and *PLK1* namely BTBD1, AZD 2281, AG14361, BI2536, and GW84382X. In ICGC effect of *RB1* is on cancers e.g. melanoma 51.91% Esophageal 45.38% ovarian 31.18% liver 27.96% Pancreatic 22.55%. Associations of RB1 with another cancer mutations are either splice as in TCGA and other associated mutation with RB1 is

LPAR6 majorly these mutations are in exon-region and further understand the other mutations in TCGA ICGC reveals most of these are SNPs at chromosome-13 which defines the locus of RB1 and for HPV. Higher interacting hubs from TCGA *TOP1*, *PAPR1,* and *ODF2*. *TOP1* is associated with *melanoma* two donor hubs of 42.08% and 40.91% liver cancer *Hepatocellular carcinoma (Virus)* with 23.08 % Esophageal (15.05%) and Ovarian (11.36%). Mutations types are SNP and Splice junctions.