Associations between DPP9 expression, survival and gene expression signature in human hepatocellular carcinoma: Comprehensive *in silico* analyses

Jiali Carrie Huang1,2, Abdullah Al Emran1,2, Justine Moreno Endaya1,2, Geoffrey W McCaughan1,2,3, Mark D Gorrell1,2\*, Hui Emma Zhang1,2\*

1 Centenary Institute, Camperdown, New South Wales, Australia; j.huang@centenary.org.au (J.C.H.), a.alemran@centenary.org.au (A.A.E.), g.mccaughan@centenary.org.au (G.W.M.)

2 The University of Sydney, Faculty of Medicine and Health, Sydney, New South Wales, 2006, Australia; jend8597@uni.sydney.edu.au (J.M.E.)

3 AW Morrow GE & Liver Centre, Royal Prince Alfred Hospital, Camperdown, New South Wales, 2050, Australia;

\* Correspondence: e.zhang@centenary.org.au (H.E.Z.), m.gorrell@centenary.org.au (M.D.G.); Tel.: 61-2-95656156

**Supplementary Table 1.** *DPP8* LoF variants in TCGA. Simple somatic mutation (SSM) affected frequency is calculated as the number of cases affected by a specific mutation in a TCGA disease project divided by the number of cases tested for SSM in that disease project in TCGA. \*Termination codon (Ter).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Position | Reference | Alternate | Protein consequence | Annotation | Case ID | Disease | # SSM affected cases and frequency |
| 65479016delA | | | p.Gln458Lys\*29 | Frameshift | TCGA-AZ-6598 TCGA-AD-5900 | Colon Adenocarcinoma | 2/400 (0.50%) |
| 65479011\_65479012insG | | | p.Gln458Pro\*8 | Frameshift | TCGA-HU-A4GT | Stomach Adenocarcinoma | 1/440 (0.23%) |
| 65474260delC | | | p.Ile512\* | Frameshift | TCGA-12-0778 | Glioblastoma  Multiforme | 1/393 (0.25%) |
| 65467180\_65467181insAGGTAAATTATTAGTCAATT | | | p.Thr543Lys\*13 | Frameshift | TCGA-13-1500 | Ovarian Serous Cystadenocarcinoma | 1/463 (0.23%) |
| 65456248\_65456249insAAATTTAAGCCCTCGGTGACAGGATCCCCTGTTGAGGGCTTAAATTTGAAGGCGCCTTT | | | p.Glu715Lys\*31 | Frameshift | TCGA-4V-A9QI | Thymoma | 1/123 (0.81%) |
| 65454301delAGAGGTATCC | | | p.Gly758Pro\*2 | Frameshift | TCGA-AP-A1DO | Uterine Corpus Endometrial Carcinoma | 1/530 (0.19%) |
| 65500743 | C | A | p.Glu153\* | Stop gained | TCGA-BK-A6W3 | Uterine Corpus Endometrial Carcinoma | 1/530 (0.19%) |
| 65500731 | C | A | p.Glu157\* | Stop gained | TCGA-D1-A17Q | Uterine Corpus Endometrial Carcinoma | 1/530 (0.19%) |
| 65500639 | A | T | p.Tyr187\* | Stop gained | TCGA-VQ-A91D | Stomach Adenocarcinoma | 1/440 (0.23%) |
| 65490237 | C | A | p.Glu276\* | Stop gained | TCGA-AX-A05Z TCGA-AJ-A5DW  TCGA-AA-3510 | Uterine Corpus Endometrial Carcinoma  Colon Adenocarcinoma | 2/530 (0.38%)  1/400 (0.25%) |
| 65480371 | G | A | p.Gln399\* | Stop gained | TCGA-D3-A8GK | Skin Cutaneous Melanoma | 1/469 (0.21%) |
| 65478907 | G | A | p.Arg493\* | Stop gained | TCGA-AG-A02N | Rectum Adenocarcinoma | 1/137 (0.73%) |
| 65467149 | G | C | p.Tyr553\* | Stop gained | TCGA-Q1-A6DW | Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma | 1/289 (0.35%) |
| 65467135\_65467136insTTGTCATCCACCTACCTCGG | | | p.Val558A\*6 | Stop gained | TCGA-24-1431 | Ovarian Serous Cystadenocarcinoma | 1/436 (0.23%) |
| 65467124 | C | A | p.Glu562\* | Stop gained | TCGA-29-1768 | Ovarian Serous Cystadenocarcinoma | 1/436 (0.23%) |
| 65456263 | G | A | p.Arg710\* | Stop gained | TCGA-AZ-4615 | Colon Adenocarcinoma | 1/400 (0.25%) |
| 65454394 | G | A | p.Gln730\* | Stop gained | TCGA-ZP-A9CY | Liver Hepatocellular Carcinoma | 1/364 (0.27%) |

**Supplementary Table 2.** *DPP8* LoF variants in COSMIC.Nonsense mutation is a substitution mutation resulting in a termination codon (\*). CDS= coding sequence; AA= amino acid, SSM= simple somatic mutation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CDS mutation | AA mutation | Legacy mutation ID | Type | Disease | #SSM affected cases |
| c.2262G>A | p.Trp754\* | COSM5946730 | Nonsense | Lymphoid neoplasm | 1 |
| c.2017C>T | p.Gln673\* | COSM6574976 | Nonsense | ER-PR-positive breast carcinoma | 1 |
| c.1659C>G | p.Tyr553\* | COSM4856031 | Nonsense | Cervical squamous cell carcinoma | 1 |
| c.328G>T | p.Glu110\* | COSM1678543 | Nonsense | Colon adenocarcinoma | 2 |
| c.2128C>T | p.Arg710\* | COSM3690497 | Nonsense | Colon adenocarcinoma | 1 |
| c.826G>T | p.Glu276\* | COSM964071 | Nonsense | Endometrioid carcinoma  Colon adenocarcinoma | 3 |
| c.1477C>T | p.Arg493\* | COSM167000 | Nonsense | Colon adenocarcinoma | 1 |
| c.2170C>T | p.Gln724\* | COSM6648997 | Nonsense | Colon adenocarcinoma | 1 |
| c.457G>T | p.Glu153\* | COSM8970383 | Nonsense | Endometrioid carcinoma | 1 |
| c.469G>T | p.Glu157\* | COSM964073 | Nonsense | Endometrioid carcinoma | 1 |
| c.2188C>T | p.Gln730\* | COSM8423812 | Nonsense | Hepatocellular carcinoma | 1 |
| c.2649C>G | p.Tyr883\* | COSM88480 | Nonsense | Ovarian clear cell carcinoma | 1 |
| c.1684G>T | p.Glu562\* | COSM1323886 | Nonsense | Ovarian serous carcinoma | 1 |
| c.1816G>T | p.Glu606\* | COSM3981612 | Nonsense | Ovarian mixed adeno-squamous carcinoma | 1 |
| c.2218C>T | p.Arg740\* | COSM5929565 | Nonsense | Skin basal cell carcinoma | 1 |
| c.1927G>T | p.Gly643\* | COSM7894053 | Nonsense | Malignant melanoma | 1 |
| c.1195C>T | p.Gln399\* | COSM8050061 | Nonsense | Malignant melanoma | 1 |
| c.1228G>T | p.Glu410\* | COSM7945996 | Nonsense | Malignant melanoma | 1 |
| c.2352G>A | p.Trp784\* | COSM135677 | Nonsense | Skin squamous cell carcinoma | 1 |
| c.561T>A | p.Tyr187\* | COSM8209774 | Nonsense | Stomach adenocarcinoma | 1 |

**Supplementary Table 3.** Genome-wide significant loci for severe COVID-19: Intronic *DPP9* variants rs12610495 and rs2109069. hg19\_coordinates: the hg19 chromosome position. Hg38\_coordinates: the hg38 chromosome position. a1: the effect allele (aligned to the + strand). a2: the non-effect allele (aligned to the + strand). afr/amr/eas/eur/sas: the allele frequency for A1 in AFR/AMR/EAS/EUR/SAS population in 1000 Genomes. beta: association between the trait and the SNP expressed per additional copy of the effect allele (odds ratio is given on the log-scale). efo: the experimental factor oncology term for the phenotype or disease. AFR= African; AMR= American; EAS= East Asian; EUR= European; SAS= South Asian.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Genomic location | |  |  | Allele frequencies | | | | |  |  |  |  |  |  |  |  |
| Gene | **rsid** | **hg19\_coordinates** | **hg38\_coordinates** | **a1** | **a2** | **afr** | **amr** | **eas** | **eur** | **sas** | **beta** | **Standard error of beta** | **p value** | **# individuals** | **Dataset ID** | **Trait (phenotype or disease)** | **p**  **value**  **COVID** | **efo** |
| *DPP9* | rs12610495 | chr19:4717672 | chr19:4717660 | A | G | 0.872 | 0.797 | 0.857 | 0.706 | 0.828 | NA | NA | 1.68E-12 | 47644 | GRASP | Fibrotic idiopathic interstitial pneumonias pulmonary fibrosis | 5.20E-06 | NCIT\_C35714 |
| *DPP9* | rs12610495 | chr19:4717672 | chr19:4717660 | A | G | 0.872 | 0.797 | 0.857 | 0.706 | 0.828 | -0.255 | 0.0362 | 2.00E-12 | - | NHGRI-EBI\_GWAS\_Catalog | Interstitial lung disease | 5.20E-06 | EFO\_0004244 |
| *DPP9* | rs2109069 | chr19:4719443 | chr19:4719431 | A | G | 0.196 | 0.219 | 0.14 | 0.321 | 0.186 | NA | NA | 2.42E-11 | 47644 | GRASP | Fibrotic idiopathic interstitial pneumonias pulmonary fibrosis | 2.41E-05 | NCIT\_C35714 |

**Supplementary Table 4.** The LIHC/HCC (n=360) and UCEC (n=540) patient demographics

|  |  |  |
| --- | --- | --- |
| Characteristic | LIHC/HCC | UCEC |
| **Number and percentage or median and range \*** | **Number and percentage or median and range \*** |
| Sex  Male  Female | 244 (67.8%)  116 (32.2%) | /  540 (100%) |
| Age at diagnosis (years) | 61 (16, 85) | 64 (31, 90) |
| Tumour site  Liver  Endometrium  Fundus uteri  Corpus uteri  Isthmus uteri | 360 (100%) | 525 (97.2%)  6 (1.11%)  4 (0.74%)  3 (0.56%) |
| Overall death | 126 (35%) | 91 (16.9%) |
| Follow up time (days) | 587 (1, 3675) | 885 (0, 6859) |
| AJCC stage  I  II  III  IV | 169 (46.9%)  83 (23.1%)  83 (23.1%)  4 (1.11%) | 334 (61.9%)  52 (9.6%)  123 (22.8%)   1. 5.4%) |

\*Missing data excluded in percentage calculation

A picture containing text, screenshot

Description automatically generated

**Supplementary Figure 1.** *DPP9* variants mapping to the DPP9 protein structure. Exonic variants were exported from **(A)** gnomAD, **(B)** TCGA and **(C)** COSMIC databases. The PDB for DPP9 apo structure is 6EOQ. The figures were created using UCSF Chimera (Alpha version 1.15, University of California, USA).

Chart, pie chart

Description automatically generated

**Supplementary Figure 2.** Enriched Reactome pathways of genes that were negatively correlated and in-common between *DPP9* and *DPP8*. **(A)** Venn diagram of the genes negatively correlated with *DPP9* and *DPP8*. “Neg cor” refers to negative correlation. **(B)** Enriched Reactome pathways associated with negatively correlated genes in-common between *DPP9* and *DPP8*.

Diagram

Description automatically generated

**Supplementary Figure 3.** Survival analysis on genes that were positively correlated and in-common amongst *DPP9*, *DPP8*, *DPP4* and *FAP* in HCC. **(A)** Venn diagram showing the numbers of genes that were positively correlated in-common with *DPP9*, *DPP8*, *DPP4* and *FAP*. The 7 genes in-common genes with all four genes are listed to the right-hand side. The genes in the blue section and pink section are *TPRA1* and *SH3BP5* respectively. The two genes on the purple section are *VAMP3* and *TTL*. “Pos cor” refers to positive correlation. **(B)** Kaplan-Meier curve for the 7 genes that were positively correlated in-common amongst *DPP9*, *DPP8*, *DPP4* and *FAP* in HCC patients. The high (red) and low (green) mRNA expression levels of genes in liver tumours were stratified based on median expression value. P values were calculated by Log-rank (Mantel-Cox) test.