An *in silico* analysis of differentially expressed genes and their prognostic utility in cholangiocarcinoma

Waleed Ali[1], Andre Kajdacsy-Balla, M.D.[2]
1.Albert Einstein College of Medicine
2.University of Illinois at Chicago, Dept. of Pathology

Introduction

Cholangiocarcinoma (CCA) encompasses a heterogeneous set of rare cancers, with an incidence as ranging from as low to .5 to around 6/100,000 in most areas of the world.[1] In tandem with its comparatively rare occurrence rate, it has among the most histologically varied presentation, presenting problems regarding its accurate diagnosis due to a lack of tumor biomarkers among other criteria. Most prominent in Asian countries such as China and Thailand, linked to parasitic infestation, rates of CCA have been increasing in the West.[1,2] With generalized symptoms such as jaundice, weight loss, and abnormal LFT values, it can be difficult for clinicians to accurately diagnose. When it is detected, it is often at later stages of progression with most cases of CCA going undetected early on, at which point it has developed chemoresistance and demonstrated metastasis to distal lymph nodes.[3] At the point of identification of CCA for most patients, surgical intervention is not as effective in comparison to other cancers, with a 5 year survival following resection being less than 20%. CCA can be classified into its intrahepatic form affecting the segmental ductules to the bile ducts, perihilar form affecting the common hepatic ducts, or distal form targeting the common bile duct (iCCA, pCCA, or dCCA respectively). Through the use of previously collected mRNA, CCA gene expression and online tools, genes of interest can be identified and lay the groundwork for further wet-lab analysis.[1,4]

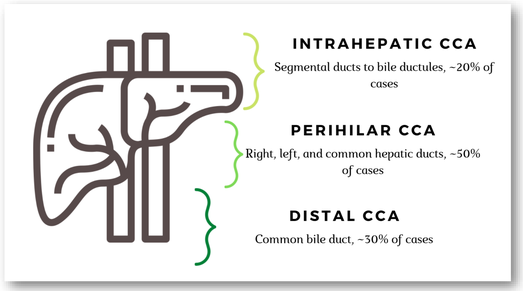

Fig 1: Schematic of various CCA subtypes

AMMF
THE CHOLANGIOCARCINOMA CHARITY

Results

Table 1: List of differentially expressed genes in CCA in cases from the TCGA differentiated in outcome status, along with p-value, and mutation frequency as determined through the CbioPortal (n=600-700 depending in gene), as well as if there is an OncoKB annotation for it. All p-values <.01, with the rest of the genes without a mut frequency being listed on the right.

(Pseudo)Gene	P-Value	Frequency of Mut	OncoKB
ZNF471	7.74E-04	0.30%	No
COL4A3	1.84E-03	0.30%	No
PLCG2	2.00E-03	0.30%	Yes
GABRB3	2.17E-03	0.30%	No
PIWIL4	3.19E-03	0.10%	No
SIX1	5.47E-03	0.10%	Yes
RBM42	6.03E-03	0.10%	No
PROS1	7.04E-03	0.10%	No
PLOD2	8.62E-03	0.10%	No
PRRT1	9.00E-03	0.10%	No
TNFRSF14	9.02E-03	0.10%	Yes

Additional genes (no mut freq. observed)
LINC01587, CTD-2033C11.1, RP11-775C24.3, LMNTD2, RP11-573D15.2, DYM, EIF5A, EIF4EBP3, CTD-2376I4.2, XXbac-B476C20.9, AL358852.1, BRWD1-AS2, SOX15, RP11-212P7.2, AC004057.1, PMS2CL, AL023806.1, RP11-428K3.1, RP11-1348G14.1, AP000640.10, UFSP1, RP11-264M12.2, FAM185A, RP6-109B7.2, RP11-131L23.2, WAC-AS1, KRT17, SIX1, CECR6, RPSAP52, CHMP4A, GSTT2B, RP11-162A12.2, MPDU1, RP11-274H2.5, NDUFV2-AS1, SGSH, CTD-3247F14.2, SH2D6, PI4KAP1, MAN1B1-AS1, AC004156.3, SLC25A27, ZNF20, RP11-729L2.2, CTD-2270P14.5, EIF2S2P4, TMEM242, RPL23AP97, RP5-899E9.1, PRPF38A, AC017116.11, CFHR3, ENPP7, SHOC2, BET1L, ZFP41, ARMC10, NOP14-AS1, LINC00484

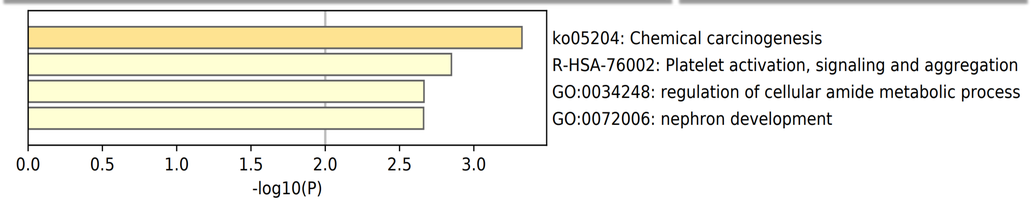

Fig 2: Map of enriched pathways found in the above gene list, taken from Metascape & corroborated through CPDB

Table 2: Correlated Gene network to the list of CCA differentially expressed genes

Pathway Name (Genes)	q-value
Translation Factors (EIF5A, EIF4EBP3, EIF4EBP2, EIF4EBP1)	**.000463**
Glucocorticoid receptor regulatory network (KRT14, KRT17, KRT5)	**.030**
intrinsic prothrombin activation pathway (PROS1, COL4A3)	**.0303**
Keratinization (KRT5, KRT14, KRT17)	.0507
Gamma carboxylation, hypusine formation and arylsulfatase activation (EIF5A, PROS1)	.0507
EGFR1 (PLCG2, KRT5, PIAS3, EIF4EBP1, KRT17)	.0507
Validated transcriptional targets of deltaNp63 isoforms (KRT14, KRT5)	.0507
Platelet Aggregation Inhibitor Pathway, Pharmacodynamics (COL4A3)	.0507
Interferon type I signaling pathways (PIAS3, EIF4EBP1)	.0544
RNA transport - Homo sapiens (human) (EIF4EBP3, EIF4BP1, EIF4EBP2)	.0556

Materials & Methods

The focus of this project was to create a battery of genes which would provide unique insight into CCA, in particular, to its survival, and set the groundwork for future projects. To do so, a collections of different tools and techniques were used (in addition to further R manipulation).[5]

Differentially expressed genes: A list of significantly (p<.01) differentially expressed genes in CCA cases delineated based on survival status was generated through the Gene Expression profiling Interactive analysis tool[6]

Examination of genomic abnormalities: Through the CbioPortal, the generated survival delineated gene set was examined for any copy number alterations (CNA), mutations, as well as any other genomic abnormality[7]

Gene enrichment/pathway analysis: Those differentially expressed genes were examined through CPDB and Metascape, utilizing pathway/ontology databases such as KEGG, Reactome, Biocarta, and more to examine for any enriched[8,9] involvement in biological functions

Creation of related gene network: Through GeneMania, a battery of genes related to the aforementioned survival-based gene list was generated based on finding genes which significantly (p<.01) shared protein domains, co-expression values, and both observed and hypothesized genetic interactions[10]

Conclusions

Within this study, we have observed a multitude of differentially expressed genes correlated to CCA survival, with involvement in a wide range of biological processes ranging from nephron development to platelet signaling. Many of these genes show no OncoKB descriptor, suggesting possible launching points for further studies/ dearth in our knowledge base. Additionally, analysis highlighted a battery of genes based on co-expression, genetic interactions, and shared protein domains that may also have an effect on the disease course of CCA.

Acknowledgement & References

1. Banales, J.M., Marin, J.J.G., Lamarca, A. *et al.* Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol* **17,** 557–588 (2020). https://doi.org/10.1038/s41575-020-0310-z
2. Bridgewater JA, Goodman KA, Kalyan A, Mulcahy MF (2016). "Biliary Tract Cancer: Epidemiology, Radiotherapy, and Molecular Profiling". *American Society of Clinical Oncology Educational Book. American Society of Clinical Oncology. Annual Meeting.* **35** (36): e194-203.
3. Chong DQ, Zhu AX. The landscape of targeted therapies for cholangiocarcinoma: current status and emerging targets. *Oncotarget.* 2016;7(29):46750-46767. doi:10.18632/oncotarget.8775
4. Rizvi, S., Khan, S. A., Hallemeier, C. L., Kelley, R. K. & Gores, G. J. Cholangiocarcinoma - evolving concepts and therapeutic strategies. *Nat. Rev. Clin. Oncol.* **15,** 95–111 (2018).
5. R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/
6. Tang, Z. et al. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res, 10.1093/nar/gkx247.
7. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012 May;2(5):401-4. doi: 10.1158/2159-8290.CD-12-0095. Erratum in: Cancer Discov. 2012 Oct;2(10):960. PMID: 22588877; PMCID: PMC3956037.
8. Konstantin Pentchev, Keiichiro Ono, Ralf Herwig, Trey Ideker, Atanas Kamburov, Evidence mining and novelty assessment of protein–protein interactions with the ConsensusPathDB plugin for Cytoscape, *Bioinformatics*, Volume 26, Issue 21, 1 November 2010, Pages 2796–2797, https://doi.org/10.1093/bioinformatics/btq522
9. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019 Apr 3;10(1):1523. doi: 10.1038/s41467-019-09234-6. PMID: 30944313; PMCID: PMC6447622.
10.Franz, M., Rodriguez, H., Lopes, C., Zuberi, K., Montojo, J., Bader, G. D., & Morris, Q. (2018). GeneMANIA update 2018. *Nucleic acids research, 46*(W1), W60–W64. https://doi.org/10.1093/nar/gky311

We acknowledge Jessica Juarez and Daniel Jacobs (Einstein) for their advice and proofreading