Gene set enrichment analysis of hepatocellular carcinoma in Hawai`i

Scott T. Nishioka[1]; Miles M. Sato[1], MS, MPH; Linda L. Wong[2], MD; Maarit Tiirikainen[2], PhD; Sandi A. Kwee[1], MD, PhD

[1]Office of Research and Development, Queen's Medical Center, Honolulu, HI;
[2]University of Hawai`i Cancer Center, University of Hawai`i, Honolulu, HI

Background

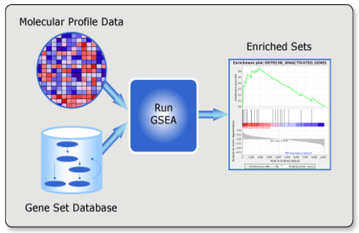
Liver cancer, which mainly manifests as hepatocellular carcinoma (HCC), currently represents the second leading cause of cancer-related deaths worldwide [1]. With a mortality to diagnosis ratio of 0.95, the prognosis for HCC remains very grim, thus illustrating the global-scale burden that characterizes HCC [1]. While the greatest rates have been known to occur in developing regions of Eastern Asia and Sub-Saharan Africa, regions historically known for low rates of HCC, including Oceania, Europe, and the United States, have recently observed a rapid increase in HCC occurrence. In the U.S. alone, HCC diagnoses have tripled since 1975. With the 5-year survival rate at only 12%, HCC has become the country's fastest rising cause of cancer-related mortality [2]. Pacific regions, specifically the state of Hawai`i, are geographically and demographically unique from both East Asia and the U.S., yet share the high and increasing rates of HCC incidence. From 2009-2013, Hawai`i experienced an age adjusted incidence rate of 11.0 per 100,000 cases, which is higher than the U.S. rate of 7.6 [3]. Therefore, HCC poses a notable and urgent issue for the state of Hawai`i as well as the United States. The poor prognosis and high occurrence of HCC in a multitude of regions around the globe convey the need for more investigation into HCC and its mechanisms of carcinogenesis.

Aim

Use GSEA to evaluate differences in the gene expression data between clinically and histologically distinct hepatocellular carcinoma tumors in an effort to sub-classify HCC patients for more personalized targets for treatment.

Materials and Methods

- Written informed consent was obtained from 40 patients with HCC.
- Tumor and liver tissue samples were placed in preservative (RNALater, ThermoFisher) upon surgical resection and stored at -80 C.
- RNA extraction was performed using AllPrep DNA/RNA (Qiagen) with RNA quality confirmed by Bioanalyzer using RNA 6000 nano chips (Agilent).
- RNA was processed using WG-DASL and hybridized onto Human HT-12 v4 Expression BeadChips (Illumina) covering 24,000 transcripts of genes, gene candidates, and splice variants.
- After array scanning, expression levels were quantified using GenomeStudio software (Illumina), resulting in a data array corresponding to 20792 genes x 40 patients. This data was pre-processed by log2 transform, background subtracted, quantile normalized, and centered in preparation for loading into GSEA software platform (Broad Institute, Massachusetts Institute of Technology).
- GSEA was used to test the hypothesis that a priori defined gene sets corresponding to specific molecular pathways differed between samples belonging to two distinct clinical classes (ie. phenotypes) of patients. Phenotypes examined were age (<> 63.5 years), gender (M/F), liver disease severity (FIB4 <> 2.87), elevated AFP (<> 400 ng/mL), Hepatitis B virus infection (+/-).
- Gene sets were obtained from the Molecular Signature Database MSigDB (Broad Institute), an online, curated collection of gene signatures assembled from medical literature.
- GSEA was performed using phenotype permutation at 1000 iterations to compute enrichment scores, with a False Discovery Rate (FDR) q-value <0.25 as criteria for significant enrichment.



(Gene Set Enrichment Analysis, Broad Institute)

Results

N	40
Median Age, years	63.5
Gender M/F	30/10
HBV infected +/- (%)	9/40 (22.5%)
HCV infected +/- (%)	16/40 (40%)
EtOH Abuse +/- (%)	15/40 (37.5%)

Table 1. Various characteristics and risk factors of sample population.

Gene Set Collections:
The following gene set collections were employed for GSEA:
H: Hallmark (50 gene sets): Gene sets accepted as major contributors in well-defined biological processes.
C2 CGP: Chemical and Genetic Perturbations (3402 gene sets): Gene sets curated from published biomedical literature on gene expression signatures represented in various genetic and chemical perturbations.
C2 CP: Canonical Pathways (1329 gene sets): Gene sets representing known biological pathways from well established pathway databases such as KEGG and Reactome.
C5: Gene Ontology (GO) (5917 gene sets): Gene sets derived from GO collective's efforts to represent genes in combination with biological products. Includes three sub-collections: Molecular Functions (MF), Cellular Components (CC), and Biological Processes (BP).
C6: Oncogenic Signatures (189 gene sets): Signature gene sets often mutated or dis-regulated, leading to tumorigenesis.
In addition to these curated collections, we also applied investigator-generated gene set collection consisting of only liver-related signatures (i.e. gene sets) derived from homo sapiens (All Human Liver) as well as only signatures for HCC molecular sub-classification published in Pubmed-Index journals (All Hoshida).

Age, Gender, FIB-4, and HBV Infection
No significant enrichment between the aforementioned phenotypes.

Alpha-fetoprotein
H: 7 significantly enriched gene sets in elevated AFP class. Noteworthy gene sets included DNA Repair (q = 0.185), Unfolded Protein Response (q = 0.026), and MYC targets V1/V2 (q = 0.051, 0.083)
C2 CGP: 334 significantly enriched gene sets in AFP class. See Figure 2.
C2 CP: 50 significantly enriched gene sets in elevated AFP class. See Figure 2.
C5 GO CC: 58 significantly enriched gene sets in elevated AFP class.
C5 GO MF: 1 significantly enriched gene set in elevated AFP class.
C6: 19 significantly enriched gene sets in elevated AFP class.
All Human Liver: 12 significantly enriched gene sets in elevated AFP class. Noteworthy gene sets include Liver Cancer Survival Down (q = 0.01) and Liver Cancer Subclass Proliferation Up (q = 0.087).
All Hoshida: 6 significantly enriched gene sets in elevated AFP class. Noteworthy gene sets include Liver Cancer Survival Down (q = 0.035) and Hoshida Liver Cancer Subclass S2 (0.084).

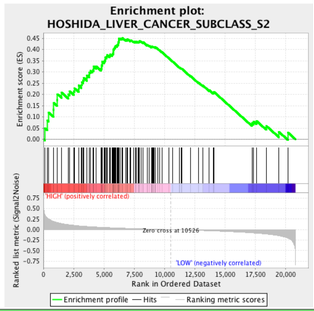
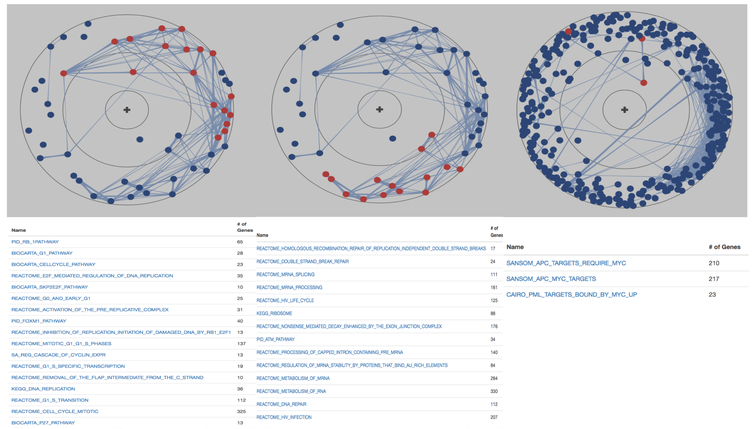


Figure 1. An example enrichment plot for the gene set "Liver Cancer Subclass S2" displaying a running sum of the enrichment score. The enrichment score (ES) represents the degree to which a gene set is overrepresented at the top or bottom of an expression-ranked list of genes.

Results (Cont.)



Name	# of Genes
PID_RB_1PATHWAY	65
BIOCARTA_G1_PATHWAY	28
BIOCARTA_CELLCYCLE_PATHWAY	23
REACTOME_E2F_MEDIATED_REGULATION_OF_DNA_REPLICATION	35
BIOCARTA_SKP2E2F_PATHWAY	10
REACTOME_G0_AND_EARLY_G1	25
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	31
PID_FOXM1_PATHWAY	40
REACTOME_INHIBITION_OF_REPLICATION_INITIATION_OF_DAMAGED_DNA_BY_RB1_E2F1	13
REACTOME_MITOTIC_G1_G1_S_PHASES	137
SA_REG_CASCADE_OF_CYCLIN_EXPR	13
REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	19
REACTOME_REMOVAL_OF_THE_FLAP_INTERMEDIATE_FROM_THE_C_STRAND	10
KEGG_DNA_REPLICATION	36
REACTOME_G1_S_TRANSITION	112
REACTOME_CELL_CYCLE_MITOTIC	329
BIOCARTA_P27_PATHWAY	13
REACTOME_HOMOLOGOUS_RECOMBINATION_REPAIR_OF_REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS	17
REACTOME_DOUBLE_STRAND_BREAK_REPAIR	24
REACTOME_MRNA_SPLICING	111
REACTOME_MRNA_PROCESSING	161
REACTOME_HIV_LIFE_CYCLE	125
KEGG_RIBOSOME	88
REACTOME_NONSENSE_MEDIATED_DECAY_ENHANCED_BY_THE_EXON_JUNCTION_COMPLEX	176
PID_ATM_PATHWAY	34
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	140
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS	84
REACTOME_METABOLISM_OF_MRNA	284
REACTOME_METABOLISM_OF_RNA	330
REACTOME_DNA_REPAIR	112
REACTOME_HIV_INFECTION	207

Name	# of Genes
SANSOM_APC_TARGETS_REQUIRE_MYC	210
SANSOM_APC_MYC_TARGETS	217
CAIRO_PML_TARGETS_BOUND_BY_MYC_UP	23

Figure 2. Constellation Map visualization of GSEA results suggest association of elevated AFP with DNA damage [left], cell cycle progression [middle], and MYC oncogene pathway activation [right].

Conclusions

- These results indicated that HCC tumors in the elevated AFP class were enriched for genes corresponding to molecular signatures and pathways associated with poor survival, increased cell proliferation and cell damage response, and overall biologic aggressiveness.
- S2 molecular subclass of HCC found to display elevated AFP in combination with greater tumor size, enrichment of MYC oncogene, and AKT activation [4].
- Proliferation subclass of HCC found to display elevated AFP as well as chromosomal instability and insulin-like growth factor (IGF) enrichment [5].

References

1. Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012." *International Journal of Cancer* 136, no. 5 (March 1, 2015): E359-386. doi:10.1002/ijc.29210.
2. Mittal, Sahil, and Hashem B. El-Serag. "Epidemiology of HCC: Consider the Population." *Journal of Clinical Gastroenterology* 47, no. 0 (July 2013): S2–6. doi:10.1097/MCG.0b013e3182872f29.
3. Hawai`i Tumor Registry, University of Hawai`i Cancer Center 2009-2013. Available at http://www.uhcancercenter.org/images/pdf/HTR_cancer_booklet.pdf
4. Hoshida, Yujin, Sebastian M.B. Nijman, Masahiro Kobayashi, Jennifer A. Chan, Jean-Philippe Brunet, Derek Y. Chiang, Augusto Villanueva, et al. "Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma." *Cancer Research* 69, no. 18 (September 15, 2009): 7385–92. doi:10.1158/0008-5472.CAN-09-1089.
5. Chiang, Derek Y., Augusto Villanueva, Yujin Hoshida, Judit Peix, Philippa Newell, Beatriz Minguez, Amanda C. LeBlanc, et al. "Focal Gains of VEGFA and Molecular Classification of Hepatocellular Carcinoma." *Cancer Research* 68, no. 16 (August 15, 2008): 6779–88. doi:10.1158/0008-5472.CAN-08-0742.

Acknowledgements

Huge thanks to Lori Tsue and Sherry Chan for all their hard work in facilitating the 2017 Queen's Summer Research Internship, organizing informative lectures and shadowing opportunities, and overseeing our work on this project. I would also like to congratulate and thank all the other summer research interns for their hard work in completing their research alongside me this summer.