TEXT MINING BASED *IN SILICO* DRUG DISCOVERY IN CANCER THERAPY INDUCED ORAL MUCOSITIS

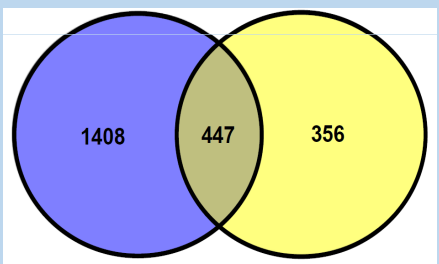
Jon Kirk; Nirav Shah, PhD; Braxton Noll; Farah Mougeot, PhD; Jean-Luc Mougeot, PhD
Carolinas HealthCare System, Cannon Research Center, Oral Medicine Research Laboratory

Abstract

INTRODUCTION: Oral mucositis (OM) is a major dose-limiting side effect of cancer therapy. Available drug-based treatments have limited effectiveness.
OBJECTIVES: Our objectives were to utilize publicly available biological data and computational tools to determine genes associated with OM, and to subsequently identify drugs targeting the relevant molecular pathways.
METHODS: OM-associated genes were determined by text mining with the pubmed2ensembl tool. Gene set enrichment and gene ontology analyses were performed using the GeneCodis program for the identification of OM and wound healing-associated pathways. Protein interaction network analysis was carried out using STRING program. Genes belonging to the identified pathways were queried against the Drug-Gene Interaction Database to identify drug candidates based on their potential ability to impede OM development.
RESULTS: We identified 447 genes common between the concepts 'OM' and 'wound healing' using text mining. Enrichment by gene ontology analysis yielded 89 genes. A final list of 20 genes representing six pathways were found to be targetable by a total of 32 drugs having the potential to be formulated for topical application. Twenty-five of the 32 drugs can directly affect the pathway that has been previously targeted for OM treatment, namely PTGS2 (COX-2) pathway. Among these drugs, only two have been previously investigated for OM treatment using systemic administration.
CONCLUSIONS: Drug target discovery based on *in silico* text mining and molecular pathway analysis can facilitate the identification of existing drugs which can be repurposed for treatment of OM.

Results

Genes in common between wound healing (WH) and oral mucositis (OM) using text mining

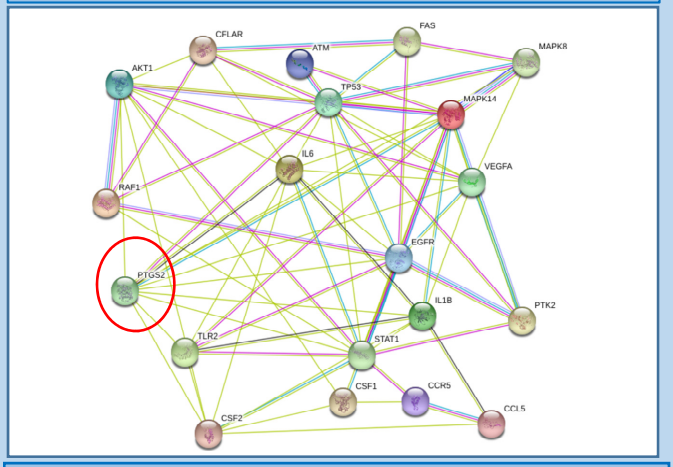


447 genes (Common between WH and OM)

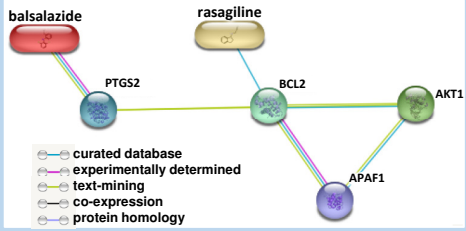
KEGG pathways of enriched gene set

Pathway Name	Genes in Query Set	Total Genes in Genome	FDR Corrected Hypergeometric p-value*	Genes
Cytokine-cytokine receptor interaction	53	259	8.51E-64	IL1B,IL6,KITLG,TNFSF14,IFNA2,IL11,TNFRSF10B,KIT,CXCL9,IL15,IL1R1,IL3,EPOR,IL2,TNFRSF1A,FAS,EPO,IFNG,TGFB3,IL1A,LIF,KDR,IFNA1,IL1RL1,CSF1R,IFNB1,CCL3,IL8,IL6ST,CSF1,IL13,EGF,IL1B,EGFR,FLT1,CCL2,TGFB1,LEP,CSF3,CCR5,CCL5,CD40LG,CCL4,TNFRSF11B,IL4,TNFRSF8,TNFSF10,VEGFA,CSF2,CCL20,CCL27,TNFSF11,CCR3
Toll-like receptor signaling pathway	26	101	8.25E-34	IL1B,IL6,MAPK1,STAT1,TLR3,IFNA2,TLR2,CXCL9,PIK3CA,IBK1,IFNA1,TLR4,IFNB1,CCL3,IL8,CD86,MAPK8,AKT1,NFKB1,CD14,MAPK14,CCL5,RAC1,CCL4,PIK3CG,IRF3
MAPK signaling pathway	26	262	6.16E-23	IL1B,FGF9,MAPK1,FGF10,DUSP1,IL1R1,DDIT3,TNFRSF1A,FAS,TGFB3,IL1A,MAPK8,EGF,AKT1,EGFR,NFKB1,RAF1,MYC,TGFB1,CD14,NR4A1,MAPK14,RAC1,FGF7,TP53,FLNB
Apoptosis	17	86	2.26E-20	IL1B,BCL2L1,TNFRSF10B,PIK3CA,IL1R1,IL3,TNFRSF1A,FAS,IL1AAKT1,NFKB1,BCL2,PIK3CG,CFLAR,TP53,TNFSF10,ATM
T cell receptor signaling pathway	17	107	1.06E-18	MAPK1,CD4,PIK3CA,IL2,IFNG,IL5,IL10,AKT1,NFKB1,RAF1,MAPK14,CD40LG,PTPRC,PIK3CG,IL4,CD28,CSF2
VEGF signaling pathway	12	75	1.2E-13	MAPK1,PIK3CA,KDR,SRC,AKT1,RAF1,MAPK14,RAC1,PTK2,PIK3CG,VEGFA,PTGS2

Enriched functional gene network



Potential drugs targeting genes associated with OM



PTGS2 (COX-2) gene found to be the most targetable gene by various drugs which may be formulated for topical application

Methods

- Used text mining with pubmed2ensembl tool to find an initial set of genes associated with oral mucositis and wound healing processes
- Analyzed gene ontology enrichment with GeneCodis based on biological processes and pathways and selected highly enriched gene sets of interest
- Determined protein-protein interactions by STRING to select most closely associated network of target genes
- Used gene-drug interaction analysis by STITCH on final gene set to produce list of drug candidates with potential for oral mucositis treatment

Conclusion

- 447 genes were found to be common to the text mining results for oral mucositis and wound healing
- The gene ontology analysis resulted in the selection of 89 genes belonging to six key pathways
- The set of genes was then narrowed to include 20 genes belonging to a high confidence network based on the protein interaction analysis
- PTGS-2 (COX-2) and associated genes within the identified high confidence molecular network may be further investigated for their potential use as drug targets in oral mucositis.

References

1. Baran, Joachim, et al. "pubmed2ensembl: a resource for mining the biological literature on genes." *PloS one* 6.9 (2011): e24716.
2. Carmona-Saez, Pedro, et al. "GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists." *Genome biology* 8.1 (2007): R3.
3. Kanehisa, Sato, et al. "KEGG as a reference resource for gene and protein annotation". Nucleic Acids Res 2016; 44 (D1): D457-D462.
4. Szklarczyk, Damian, et al. "STRING v10: protein–protein interaction networks, integrated over the tree of life." *Nucleic acids research* (2014): gku1003.
5. Kuhn, Michael, et al. "STITCH: interaction networks of chemicals and proteins." *Nucleic acids research* 36.suppl_1 (2007): D684-D688