Section 2. General Biology

https://doi.org/10.29013/ELBLS-23-1-10-18

Jessica Xiong,
High School: Adlai E. Stevenson High School IL, USA
Wei Wang,
Dr. Instructor, Beijing University

HNSCC: DIFFERENTIAL GENE EXPRESSION IN PRIMARY VERSUS RECURRENT TUMORS

Abstract. Head and Neck Squamous Cell Carcinoma, also known as HNSCC, is the sixth most common cancer worldwide. Between now and 2030, new cases are anticipated to increase by 30%, totalling approximately 1.08 million new cases annually. Generally, all tumors that originate in the mucosal epithelium lining of the oral cavity, pharynx, larynx, and sinonasal tract are considered part of HNSCC. Most HNSCCs in the oral cavity and larynx develop due to abusive alcohol and tobacco consumption, whereas development of HNSCCs in the pharynx seems connected to human papillomavirus (HPV). Due to the consumption of carcinogen products, such as the areca nut, which local people tend to chew, HNSCC is most prevalent in South Asia and Australia. It is also very prevalent in the United States and Europe due to higher infection rates of HPV. Additionally, HNSCC is known for its genetic instability, needing multiple genetic transformations to occur, which is what this study will focus on.

Keywords: HNSCC, HPV.

1. Background

HNSCC is a typically localized cancer. Compared to other cancers, it spreads to distant parts of the body more slowly. Development of HNSCC is connected to alcohol abuse, tobacco consumption, and prior positive HPV infections. Originating in the oral cavity (which includes lips, buccal mucosa, hard palate, anterior tongue, floor of mouth and retromolar trigone), the nasopharynx, the oropharynx (which includes palatine tonsils, lingual tonsils, base of tongue, soft palate, uvula and posterior pharyngeal wall), the hypopharynx (which includes the bottom part of the throat, extending from the hyoid bone to the cricoid cartilage), and the larynx, HNSCC usually metastasizes to the lungs or nearby lymph nodes.

Progression of invasive HNSCC usually follows a certain pattern: "epithelial cell hyperplasia, followed by dysplasia (mild, moderate and severe), carcinoma in situ and, ultimately, invasive carcinoma." Since HNSCC is very heterogeneous, cell of origin usually depends on "anatomical location and aetiological agent (carcinogen versus virus)"; however, the most common origin is adult stem cells or progenitor cells, which, after oncogenic transformation, turn into cancer stem cells (CSCs) that have self-renewal and pluripotency properties (1). Although HNSCC CSCs constitute only 1–3% of

cells in primary tumors, there have been a number of molecular biomarkers with prognostic significance. Of these CD44, CD133, and ALDH1 are the most validated. CD44 is a "cell surface receptor for hyaluronic acid and matrix metalloproteinases (MMPs) and is involved in intercellular interactions and cell migration. HNSCC cells with high levels of CD44 are capable of self-renewal, and CD44 levels in HNSCC tumours are associated with metastasis and a poor prognosis. Similarly, increased levels of the membrane-spanning protein CD133 are associated with HNSCC invasiveness and metastasis. ALDH1 is an intracellular enzyme that converts retinol into retinoic acid, plays a part in cellular detoxification and is a marker for both normal stem cells and CSCs. High levels of ALDH1 expression or activity are associated with self-renewal, invasion and metastasis and may have prognostic significance in HNSCC" [2].

To pinpoint the specific cell of origin, it is necessary to look at the development of second primary tumors (SPTs). In HNSCC, SPTs appear at an extremely high rate after the diagnosis of the primary tumor, and they are frequently lethal. The development of SPTs reflects CSCs arising from independent oncogenic transformations by looking at the field cancerization, which "involves the formation of multiple patches of premalignant disease with a higher-than-expected rate of multiple local second primary tumors." This suggests that carcinogens damage large anatomical fields.

Symptoms of HNSCC include persistent sore throat, pain, weakness, or numbness near the head and neck, enlarged lymph nodes, and odd patches or openings in the throat and mouth [3].

Survival rates for HNSCC have improved over the past three decades; the 5-year survival increased from 55% from 1992–1996 to 66% during 2002–2006. If treated with Surgery, Chemotherapy, Radiation Therapy, Immunotherapy, or Targeted Therapy, survival rates are now 56–62% across all five stages. However, after treatment, 15–50% patients develop recurrent HNSCC, which is both difficult to treat and a major cause of morbidity. Recurrent HNSCC is difficult to treat because of the loss of effectiveness due to prior treatments and the infiltrative nature of recurrent diseases in the head and neck area. A study done on the most effective treatment for recurrent HNSCC suggests that aggressive treatments, such as surgery and CCRT, reduces deaths of recurrent HNSCC patients most efficiently; however, there is yet a study to be done on the mutated genes responsible for recurrence.

2. Methods

The tools and databases used in this study are publicly available. NCBI's GEO database was used to search for datasets relevant to the objective studied; those with samples separated through arrays were analyzed with GEO2R, while those with samples separated through high throughput sequencing were analyzed with DESeq2. Source Batch Search was used to annotate gene symbols. Results were copied onto Google Sheets, where samples were filtered into upregulated and downregulated according to p-value and fold change. Then, a venn diagram was drawn to find common genes in multiple studies for more accurate results. These genes were then compiled, searched on GenCard Bank, and separated according to their functions.

2.1 Sample Download and Extraction

NCBI's GEO database, a public international archive storing genomics data submitted by the research community (https://www.ncbi.nlm.nih.gov/geo/), was used. Studies with results relevant to this paper's objective were extracted, and samples were downloaded and separated into two types: array and high throughput sequencing. There were two tumor versus normal samples selected for more accurate results, and one primary versus recurrent sample.

2.2 Array Analysis

GEO2R (https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html) was used for the array datasets. GEO2R is an interactive web tool available to GEO

datasets with array samples, which allows users to compare two or more of these samples to identify genes that are differentially expressed across set experimental conditions. The results are processed and presented by significance as a table of ordered genes, and graphic plots are available to help visualize differentially expressed genes and assess data set quality.

2.3 High Throughput Sequencing Analysis

DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) was used for high throughput sequencing analysis. The R program was installed, and the DESeq2 package was downloaded. Then, using code chunks, results were organized into a table.

2.4 Filtering

The resulting tables from both DESeq2 and GEO2R were put separately into google sheets and filtered according to p-value<0.01 and FC<=0.5 (downregulated), FC>=2 (upregulated).

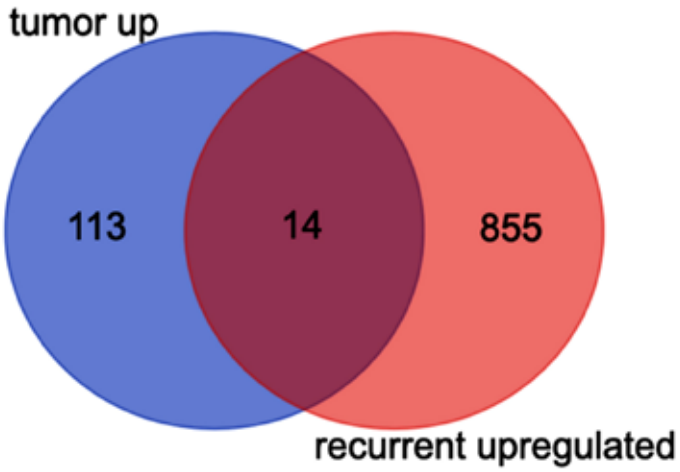
2.5 Commonality Grouping

Resulting gene names were compiled into an online venn diagram tool (https://bioinformatics.psb.ugent.be/webtools/Venn/). The two tumor versus normal upregulated results were inserted, and the common genes were located; this process was repeated for the downregulated genes and the primary versus recurrent genes. The common genes found were separated into upregulated in recurrent tumors and downregulated in recurrent tumors.

Input files:

List names	number of elements	number of unique elements
recurrent upregulated	869	869
tumor up	127	127
Overall number of unique elements		**982**

Image result:



Save Image As PNG Save Image As SVG

Text results:

Save text result

Names	total	elements
recurrent upregulated tumor up	14	MET GJA1 ITGA3 HOXD10 TTYH3 CA2 TENM2 CXCL1 CDSN LAMB3 ITGA6 FMNL2 MMP10 SLC2A1

2.6 Gene Function Research

GeneCards: The Human Gene Database (https://www.genecards.org), an online knowledgebase that automatically integrates gene-centric data from ~150 web sources, was used for the research of functions and locations of genes.

3. Results

The basic information of the genes were compiled into the table below. In total, there are 9 genes that regulate physiological processes, 5 genes that regulate tumor-related functions, 5 genes that regulate inflammation, 4 genes that regulate ion-related functions, 3 genes that regulate cell surface adhesion, 2 genes that regulate immune cells, 2 genes that regulate signaling, 2 that regulate antigens, and 5 whose functions are unrelated to any of the others.

Table 1.

	Full Name	**Main Function**	**Protein/gene family**	**Detailed description**	**Up or Down regulated**
1	***2***	***3***	***4***	***5***	***6***
MET	Mesenchymal Epithelial Transition	Physiological Processes	receptor tyrosine kinase protein family	Regulates proliferation, scattering, morphogenesis; reduces lung fibrosis	Up
GJA1	Gap Junction Protein Alpha 1	Physiological Processes	connexin gene family, encodes protein that's component of gap junctions in the heart	Involved in synchronized heart contraction, embryonic development, bladder capacity, and hearing	Up
ITGA3	Integrin Subunit Alpha 3	Cell surface adhesion	integrin alpha chain protein family	n/a	Up
HOXD10	Homeobox D10	Physiological Processes	Abd-B homeobox protein family	Involved in cell differentiation and limb development; part of developmental regulatory system: provides cells with specific positional identities on anterior-posterior axis	Up
ITYH3	Tweety Family Member 3	Ion Channels	tweety family of proteins	Encoded protein is calcium (2+)- activated large conductance chloride (–) channel; responsible for ion channel transport and transport of inorganic cations/anions and amino acids/oligopeptides	Up
CA2	Carbonic anhydrase 2	Bone reabsorption	isozymes of carbonic anhydrase	Essential for bone resorption and osteoclasts differentiation; regulates fluid secretion into anterior chamber of eye; contributes to intracellular pH regulation in duodenal upper villous epithelium during proton-coupled peptide absorption	Up

1	*2*	*3*	*4*	*5*	*6*
TENM2	Teneurin Transmembrane Protein 2	Physiological Processes, cell surface adhesion, ion channels	tenascin	Enables cell adhesion molecule and signaling receptor binding activity; involved in calcium-mediated signaling using intracellular calcium source; heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules; retrograde trans-synaptic signaling by trans-synaptic protein complex; involved in neural development by regulating proper connectivity within nervous system	Up
CXCL1	C–X-C Motif Chemokine Ligand 1	Inflammation	CXC subfamily of chemokines	Encoded protein is a secreted growth factor that signals through G-protein coupled receptor and CXC receptor 2; plays role in inflammation and as chemoattractant for neutrophils	Up
CDSN	Corneodesmosin	Epidermal	protein found in corneodesmosomes	Epidermal barrier integrity	Up
LAMB3	Laminin Subunit Beta 3	Physiological processes	basement membrane proteins	Mediates attachment, migration, organization of cells into tissues during embryonic development by interacting e other extracellular matrix components	Up
ITGA6	Integrin Subunit Alpha 6)	Cell surface adhesion	integrin alpha chain protein family	Present in oocytes, involved in sperm-egg fusion; plays structural role in hemidesmosome	Up
FMNL2	Formin Like 2	Physiological processes	formin-related protein	Regulates cell morphology and cytoskeleton organization; required in cortical actin filament dynamics	Up
MMP10	Matrix Metallopeptidase 10	Physiological processes	peptidase M10 family of matrix metalloproteinases (MMPs)	Breaks down extracellular matrix in normal physiological processes (embryonic development, reproduction, tissue remodeling, and disease processes like arthritis and metastasis)	Up
SLC2A1	Solute Carrier Family 2 Member 1	Glucose transport	Solute carrier family	Encodes major glucose transporter in mammalian blood-brain barrier; protein mainly found in cell membrane and cell surface, also functions as receptor for HTLV virus I and II	Up
B3GALT5	Beta-1,3-Galactosyltransferase 5	Antigens	membrane-bound glycoproteins	Encoded protein may synthesize type1 Lewis antigens, which are elevated in gastrointestinal and pancreatic cancers	down

1	*2*	*3*	*4*	*5*	*6*
ADH7	Alcohol Dehydrogenase 7	Metabolize substrates	class IV alcohol dehydrogenase 7 mu or sigma subunit	Most active as retinol dehydrogenase, thus may participate in synthesis of retinoic acid (hormone used for cellular differentiation); catalyzes NAD-dependent oxidation of all-trans-retinol, alcohol, and omega-hydroxy fatty acids	down
HPGD	15-Hydroxyprostaglandin Dehydrogenase	Metabolism of prostaglandins, inflammation	short-chain non-metalloenzyme alcohol dehydrogenase protein family	Catalyzes NAD-dependent oxidation of hydroxylated polyunsaturated fatty acids; decreases levels of pro-proliferative prostaglandins such as prostaglandin E2 (whose activity increased in cancer because increase in expression of cyclooxygenase 2); inactivates resolvins E1, D1, D2, which play roles in inflammation	down
SCGB1A1	secretoglobin family 1A member 1	Physiological processes, inflammation	secretoglobin family of small secreted proteins	Anti-inflammation, inhibition of phospholipase A2, sequestering of hydrophobic ligands	down
NUCB2	Nucleobindin-2	Ions, tumor related	calcium binding protein	Calcium level homeostasis, eating regulation in hypothalamus, release of tumor necrosis factor from vascular endothelial cells; non receptor guanine nucleotide exchange factor, binds to and activates guanine nucleotide binding protein (G-protein) alpha subunit GNAI3	down
KRT4	Keratin, type I cytoskeletal 4	Epithelial	keratin gene family	Specifically expressed in differentiated layers of mucosal and esophageal epithelia	down
CXCL12	C–X-C Motif Chemokine Ligand 12	Physiological processes, tumor related, immune cells, ion channels, inflammation	stromal cell-derived alpha chemokine member of intercrone family	Protein functions as ligand for G-protein coupled receptor, chemokine (C–X-C motif) receptor 4; CXCR4 activated to induce rapid and transient rise in level of intracellular calcium ions and chemotaxis. Plays roles in embryogenesis, immune surveillance, inflammation response, tissue homeostasis, tumor growth/metastasis; chemoattractant active on T-lymphocytes and monocytes but not neutrophils, stimulates migration; several critical functions in embryonic development, bone marrow and heart ventricular septum formation, B-cells	down

1	*2*	*3*	*4*	*5*	*6*
DIO2	Iodothyronine Deiodinase 2	Tumor related	iodothyronine deiodinase family	Protein is selenoprotein w non-standard amino acid Sec, which encoded by the UGA codon that signals translation termination	down
CCL21	C–C Motif Chemokine Ligand 21	Immune cells, ions, inflammation	CC cytokines genes	Immunoregulatory and inflammatory processes; encoded protein inhibits hemopoiesis and stimulates chemotaxis; chemotactic in vitro for thymocytes and activated T-cells, not for B cells macrophages or neutrophils; cytokine also plays role in mediating homing of lymphocytes to secondary lymphoid organs	down
GNA14	G Protein Subunit Alpha 14	Signaling	guanine nucleotide binding/ G protein family	Modulators/transducers in various transmembrane signaling systems	down
GCNT3	Glucosaminyl (N-Acetyl) Transferase 3, Mucin Type	Antigens	N-acetylglucosaminyltransferase family	Introduce the blood group I antigen during embryonic development	down
BOC	BOC Cell Adhesion Associated, Oncogene Regulated	Signaling	immunoglobulin/ fibronectin type III repeat family	Cell-surface receptor com-led that mediates cell-cell interactions between muscle precursor cells, promotes myogenic differentiation	down
PLAC8	Placenta Associated 8	Tumor related	cornifelin family	Might enable chromatin binding activity, positive regulation of cold-induced thermogenesis, positive regulation of transcription by RNA polymerase II, acts upstream/within several processes (brown fat cell differentiation, defense response to bacterium, response to cold)	down
GULP1	GULP PTB Domain Containing Engulfment Adaptor 1	Tumor related	nucleocytoplasmic shuttling protein	Protein encoded is adapter protein necessary for engulfment of apoptotic cells by phagocytes; modulates cellular glycosphingolipid and cholesterol transport; may play role in internalization and endosomal trafficking of various LRP1 ligands such as PSAP	down

4. Discussion

According to the resulting table, the most prevalent function of the differentially expressed genes is the regulation of physiological processes. However, as physiological processes is a generally broad topic, it is necessary to take the more detailed description into account. Most notably, there are 5 genes that are used in embryonic development: GJA1, HOXD10, LAMB3, MMP10, CXCL12. The mutation of genes that are responsible for embryonic development causes the increased potential for developing cancer, as the inability of embryonic cells to develop proper structures that regulate important functions may encourage cancer development. It is connected to hereditary cancer and the tendency for certain groups of demographics to develop cancer.

Secondarily, the tumor-related genes. Again, tumor-related is a broad topic, and detailed descriptions are needed; however, the functions of these genes can be easily connected to the reasons behind recurrence and tumor development. As expected, all genes in this section are downregulated – the disappearance of these genes will increase the likelihood of cancer recurring. Thus, this paper will not be discussing their functions in detail.

Third, inflammation. In the inflammation section, there is only one gene that is upregulated: CXCL1. The rest, HPGD, SCGB1A1, CXCL12, and CCL21, are all downregulated. Inflammation is the body's response to tissue damage, which can be caused by physical injury, infection, exposure to toxins, or other types of trauma. Inflammation causes the repairing of damaged tissue and cellular proliferation. If the cause persists or certain control mechanisms fail, inflammation can become chronic. Once this occurs, tissue repair and cell proliferation will often create an environment in which cancers have a tendency to develop [4]. This information supports the notion that the differentially expressed genes in this table are connected to recurrence. If those that usually regulate the inflammatory response shut-down, mutate, and do not occur frequently in the tumor sites, the inflammatory responses will become chronic and cancer will develop again, even if it is removed.

Fourth, ion regulation. Studies show that increases in intracellular calcium may inhibit apoptosis, depending on concentration level, location, and timing [5]. In the results above, all 4 genes, TTYH3, TENM2, NUCB2, and CXCL12, are connected to the calcium ion. Two are upregulated, two are downregulated, respectively. TTYH3 encodes for a calcium channel; if upregulated, it therefore increases the calcium ion channels in the cell membrane, which may cause an increase in the intracellular calcium. Using an intracellular calcium source, TENM2 is involved in calcium-mediated signaling. The upregulation of this gene will therefore cause an influx of calcium into the cell. NUCB2 is responsible for calcium level homeostasis. Downregulation of it will cause a destruction of balance; an increase in intracellular calcium will not be returned to normal. CXCL12 activates CXCR4, which induces a rapid increase of calcium levels inside the cell. All these causes added together increase the possibility of the intracellular level of calcium reaching the point of inhibition of apoptosis. Due to this inhibition, the likelihood of tumor development increases; recurrency can occur.

Fifth, cell surface adhesion. All cell surface adhesion genes are upregulated in recurrent tumors. It is said that cell surface proteins are capable of restricting cell growth through contact inhibition; alterations of these molecules are common in cancer [6]. CAM-DR (Cell adhesion mediated drug resistance) is a significant limitation to the success of cancer therapies, most notably chemotherapy. The explanation to why this occurs can be explained by the FN model, which shows the cellular arrest in the G1 phase of cellular division, which significantly reduces the efficacy of drugs [7]. Thus, the upregulation of cellular adhesion molecules may inhibit the initial development of cancer to a certain extent; however, once cancer develops, it negatively impacts the efficiency of treatment, which explains why it is prevalent in recurrent tumor cells.

5. Conclusion

To summarize, most genetic mutations cause differential gene expressions in genes regulating embryonic development, tumor-regulation, inflammation, intracellular calcium level regulation, and cell surface adhesion molecules. Mutation in these functions are proved to be responsible for cancer development or secondary cancer development. However, the upregulation and downregulation of these genes appears unconnected to HNSCC specifically; instead, they seem more connected to cancers in general. These results are still relevant to the objective of the paper– the understanding of the mechanisms in cancer recurrence is helpful in HNSCC recurrence identification, and potential future treatments. Additionally, since recurrence is connected to survival rates, by looking at the genes in the tumor samples doctors will be able to predict the patients' survival rates after treatment.

Acknowledgments

I would like to thank Dr. Pingzhang Wang for introducing to me the tools used in this study, answering my questions when I was confused, and guiding me to my understanding of this topic.

References:

1. Chang Wu, Yuan T. H. Wu and Wu. Locoregionally recurrent head and neck squamous cell carcinoma: incidence, survival, prognostic factors, and treatment outcomes, NIH 2017.
2. Johnson E., Burtness, C. Leemans, Lui, E. Bauman, and R. Grandis, Head and neck squamous cell carcinoma, NIH 2020.
3. Head and neck squamous cell carcinoma, Medicine Plus 2015.
4. Singh, Baby, Rajguru, Patil, Thakkannavar, Pujari, Inflammation and Cancer, NIH 2019.
5. Fnu, Weber, Alterations of Ion Homeostasis in Cancer Metastasis: Implications for Treatment, NIH 2021.
6. Moh, Shen. The roles of cell adhesion molecules in tumor suppression and cell migration, NIH 2009.
7. Huang, Wang, Tang, Qin, Shen, He, Ju. CAM-DR: Mechanisms, Roles and Clinical Application in Tumors, Frontiers 2021.