OMICS RESEARCH SYMPOSIUM

BIOINFORMATICS APPROACH TO UNDERSTAND THE TRANSCRIPTIONAL CHANGES IN SMALL CELL LUNG CANCER (SCLC) PATIENTS WITH BRAIN METASTASES

PINE.BIO
OMICS DATA SENSE AND MORE

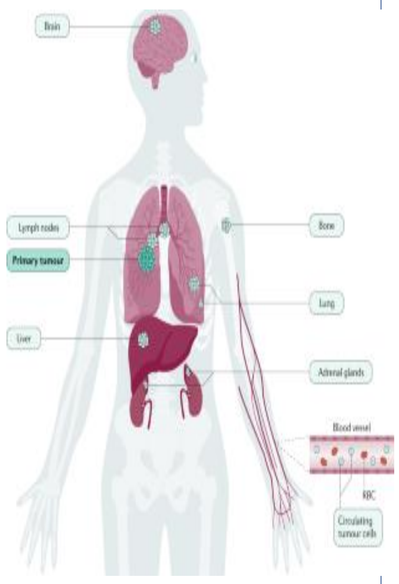
Kanubha Sharma[1], Harpreet Kaur[2], Elia Brodsky[2], Mohit Mazumder[2]

Introduction

Lung cancer is one of the major cause of mortality worldwide, making it a prominent cause of cancer related deaths (Tomazis*et al.*, 2020). It is believed that lung cancer can be cured if detected at an early stage with the help of various screening tools (Thakur *et al.*, 2020). But the early detection of lung cancer is very crucial as the tools available in medical industry are not advanced enough to detect the genes and the pathways responsible for the cause of cancer and even if it is diagnosed at early stage, it cannot be treated completely (Xu *et al.*,2021). The early detection is a crucial task as only 15% of the patients are detected at an early stage while others, that is 75% of the patients at diagnosed with an advanced or last stage (Jacobsen *et al.*,2017).

Lung cancer can be grouped into two main subtypes namely: NSCLC (Non-Small Cell Lung Cancer) and SCLC (Small Cell Lung Cancer). NSCLC is most prominent with its prevalence of 85%-90% of population suffering from lung cancer (Herbst*et al.*, 2018). Whereas SCLC is present in 10%-15% of population suffering from lung cancer (Molina *et al.*, 2008). NSCLC (Non-Small Cell Lung Cancer) can be further grouped into adenocarcinoma and squamous cell carcinoma (Torre *et al.*,2012). But SCLC, when compared to NSCLC, has a fast rate of forming new cells, and also a high malignancy rate, resulting in a huge risk of metastases (Kalemkerian*et al*, 2013). SCLC has a very low (almost rare) survival rate and patients usually lose their lives due to this disease.

Due to high metastatic rate of this disease, the cancerous (or tumour) cells spread from the primary site to the portions of the body including brain, liver, bones etc. Mostly this spread is diagnosed in (or towards) the brain. Approximately, at the time of primary diagnoses, about 10 percent of the SCLC patients are detected to have brain metastases and about 50 percent develops during the course of cancer from stage to stage (Quan*et al*, 2004). Hence, it becomes necessary to identify the cause of this spread and to curb it at its root cause only.



Common sites of metastasis in SCLC

Methods

Dataset:

- ✓ The data for this project was obtained from the Gene Expression Omnibus (GEO) database after a thorough search, which is a part of NCBI (National Center for Biotechnology Information).
- ✓ The dataset of series: GSE161968 was selected as the ideal dataset for the study and analysis of the project and was associated with BioProject: PRJNA680181 (**Zhu H *et al*, 2021**).
- ✓ This dataset consisted of microarray data (mRNA) of SCLC patients with Brain metastases and without Brain metastases. The samples of these patients were extracted from the peripheral blood (tissue). There were 12 samples in total, with 6 samples as BM positive and 6 samples as BM negative

Exploratory Analysis:

- ✓ Exploratory analysis enabled the detection of the variance present among all the samples (patients with and without brain metastases).
- ✓ PCA was performed on all the 12 samples with the help of T-bioinfo server platform that provided with different pipelines. The pipeline first performed Quantile normalization of the data (as the data obtained from dataset was already normalized) and then it performed the PCA.
- ✓ As an output result, file that contained information about the total variance in data with respect to PC components: **PC1** (representing maximum variance) and **PC2** (representing second maximum variance).

Data processing

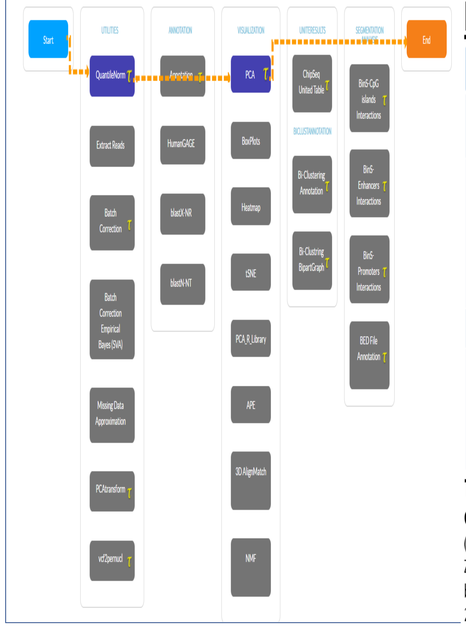
- ✓ The dataset consisted of microarray data. As a result of which genes (present as gene symbols) in samples had multiple probe ids. In order to perform comparative and statistical analysis on the data, the probe ids were matched to gene ids (to avoid and remove duplicate genes (if present)). This process was performed on DAVID (Database for Annotation, Visualization and Integrated Discovery) database.
- ✓ The probe ids were extracted from the supplementary file, which was obtained from the dataset, and were uploaded in DAVID database to obtain gene symbols.
- ✓ From the DAVID database, the gene symbols were obtained that matched with the probe ids. The data of the samples of individual probe ids corresponding to gene symbols were mapped to each other with the help of python code

Statistical analysis:

- ✓ To identify differentially expressed genes between samples with brain metastasis vs. samples without brain metastasis, different parametric tests were performed in excel.
- ✓ T-test was performed to evaluate p-value which signifies significant genes present in sample. After that, log2Fold change and P-adj (Q value) were calculated in order to obtain significant differentially expressed genes that were present in samples.
- ✓ The output data was filtered with p-value <0.05 (threshold) and log2Fold change: >-1.2 or <1.2 (range selected based on the number of genes, varies from sample to sample).
- ✓ From this data, the up-regulated and down-regulated genes were identified based on logfold change values.

Gene Enrichment Analysis:

- ✓ To understand and analyse the biological implications of identified significant genes obtained (both up-regulated and down-regulated) as a result of statistical analysis, we performedenrichment analyses in different pathways.
- ✓ Towards this, the genes were uploaded (in form of list) on ENRICHR database (reference). This database analyses and identifies the role of genes in different pathways, provides information about the enrichment of the genes in different Geno ontology (GO) terms, i.e. biological processes,molecular functions and cellular components



DATASET

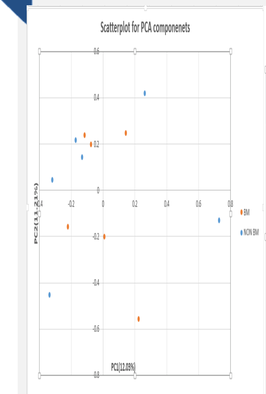
SERIES	GSE161968
EXPERIMENT TYPE	Expression profiling by array array
CHARACTERISTICS	subject status (SCLC patient with patient with BM and (or) without without BM) tissue: Peripheral blood
TOTAL SAMPLES	12
GROUPS	SCLC patients with BM -6 SCLC patients without BM-6

Table 1 – GEO Dataset taken for the identification of biomarkers to detect Brain metastases due to SCLC

(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161968)

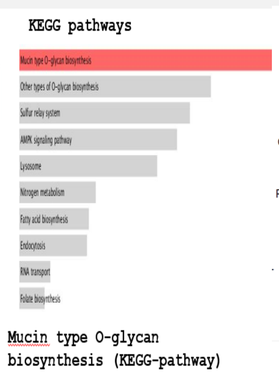
Zhu H, Li J, Jing W, Jia W, Zhai X, Yu J. Downregulation of lncRNA XR_429159. 1 could promote brain metastasis in patients with limited-stage small-cell lung cancer. Frontiers in Oncology. 2021;11:611

Results

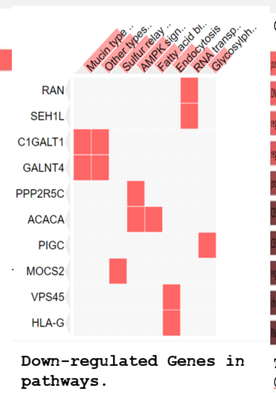


STATISTICAL ANALYSIS

TOTAL GENES	913
GENES WITH P-VALUE LESS THAN 0.05 (DEGs)	555
GENES WITH P-adj VALUE LESS THAN 0.05 (significantly expressed genes)	262
GENES WITH log2Fold change >1.2 or <-1.2	165
DOWN-REGULATED GENES	158
UP-REGULATED GENES	7

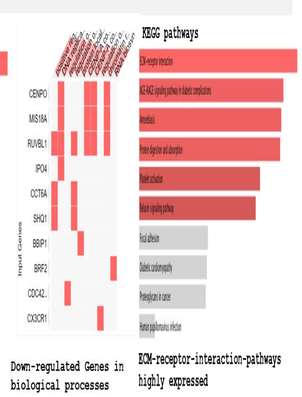


Mucin type O-glycan biosynthesis (KEGG-pathway)



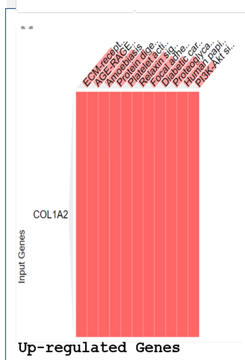
Down-regulated Genes in pathways.

Telomerase RNA localization to Cajal body (biological processes) is highly expressed

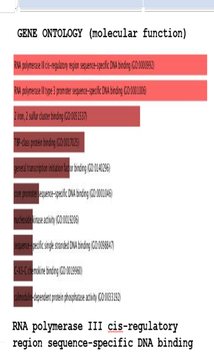


Down-regulated Genes in biological processes

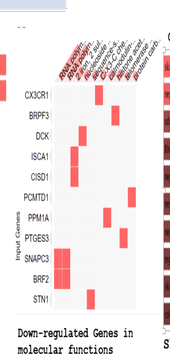
ECM-receptor-interaction-pathways highly expressed



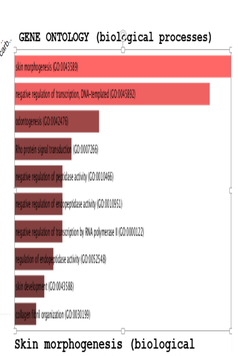
Up-regulated Genes involved in pathways



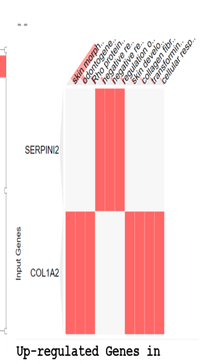
RNA polymerase III cis-regulatory region sequence-specific DNA binding (molecular function) is highly expressed



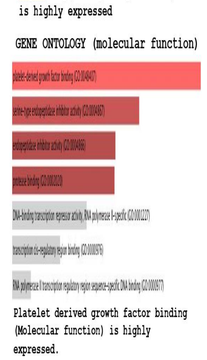
Down-regulated Genes in molecular functions



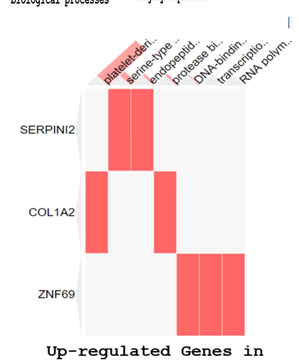
Skin morphogenesis (biological process) is highly expressed.



Up-regulated Genes in biological processes



Platelet derived growth factor binding (Molecular function) is highly expressed.



Up-regulated Genes in molecular functions

Conclusions

The results obtained after analysing data depicted that the genes: **COL1A2 , C1GALT1, GALNT4, .RUVBL1, CCT6A, SHQ1, SNAPC3, BRF2** might act as effective biomarkers.

Main pathways (KEGG) that were involved include:

- ✓ **Mucin type O-glycan biosynthesis pathway (**highly supressed due to down-regulation of genes)
- ✓ **ECM-receptor-interaction-pathways** (highly activated due to up-regulation of genes)

Molecular functions that resulted in metastatic conditions were:

- ✓ **RNA polymerase III cis-regulatory region sequence-specific DNA binding function** (highly supressed due to down-regulation of genes)
- ✓ **Platelet derived growth factor binding (PDGF)** (highly activated due to up-regulation of genes)

Biological processes involved:

- ✓ **Telomerase RNA localization to Cajal body** (highly supressed due to down-regulation of genes)
- ✓ **skin-morphogenesis** (highly activated due to up-regulation of genes)

References

- Toumazis, I., Bastani, M., Han, S. S., & Plevritis, S. K. (2020). Risk-Based lung cancer screening: A systematic review. *Lung Cancer*.
- Thakur, S. K., Singh, D. P., & Choudhary, J. (2020). Lung cancer identification: a review on detection and classification. *Cancer and Metastasis Reviews*, *39*, 989-998.
- Xu, K., Zhang, C., Du, T., Gabriel, A. N. A., Wang, X., Li, X., ... & Zhang, Y. (2021). Progress of exosomes in the diagnosis and treatment of lung cancer. *Biomedicine & Pharmacotherapy*, *134*, 111111
- Jacobsen, M. M., Silverstein, S. C., Quinn, M., Waterston, L. B., Thomas, C. A., Benneyan, J. C., & Han, P. K. (2017). Timeliness of access to lung cancer diagnosis and treatment: a scoping literature review. *Lung cancer*, *112*, 156-164.
- Herbst, R. S., Morgensztern, D., & Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature*, *553*(7689), 446-454
- Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., & Adjei, A. A. (2008, May). Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. In *Mayo clinic proceedings* (Vol. 83, No. 5, pp. 584-594). Elsevier
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, *65*(2), 87-108.
- Kalemkerian GP, Akerley W, Bogner P, Borghaei H, Chow LQ, Downey RJ, Gandhi L, Ganti AK, Govindan R, Grecula JC, Hayman J. Small cell lung cancer. Journal of the National Comprehensive Cancer Network. 2013 Jan 1;11(1):78-98.
- Quan AL, Videtic GM, Suh JH. Brain metastases in small cell lung cancer. ONCOLOGY-WILLISTON PARK THEN HUNTINGTON THE MELVILLE NEW YORK-. 2004 Jul 1;18:961-72.
- Zakowski MF. Pathology of small cell carcinoma of the lung. InSeminars in oncology 2003 Feb 1 (Vol. 30, No. 1, pp. 3-8). WB Saunders.
- Brenner B, Tang LH, Klimstra DS, Kelsen DP. Small-cell carcinomas of the gastrointestinal tract: a review. Journal of Clinical Oncology. 2004 Jul 1;22(13):2730-9.
- Guinee Jr DG, Fishback NF, Koss MN, Abbondanzo SL, Travis WD. The spectrum of immunohistochemical staining of small-cell lung carcinoma in specimens from transbronchial and open-lung biopsies. American journal of clinical pathology. 1994 Oct 1;102(4):406-14.
- Castillo JJ, Vincent M, Justice E. Diagnosis and management of hyponatremia in cancer patients. The oncologist. 2012 Jun;17(6):756

Acknowledgements: I would like to extend my gratitude to Dr.Harpreet Kaur and the entire team of pine biotech for providing me with their expert guidance and support for helping me to complete research project

Video link- [(156) Bioinformatics approach to understand transcriptional changes in SCLC patients with brain metastases - YouTube]