Current Trends in Biotechnology and Pharmacy
Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

Suppression of Hotair Gene in Small-Cell Lung Cancer Effect The Differentially Expressed Genes and Inhibit Progression of the Disease

Thamilvaani Manaharan[1], Vijayan Manickam Achari[2]*, Nusrat Rehana Bano Binti Naushad Ahmad[3], and Mogana Darshini Ganggayah[4]

[1]Basic Medical Sciences Division, Faculty of Medicine, University of Cyberjaya, 63000 Cyberjaya, Selangor
[2]Institute of Biological Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[3]Accenture Technology Center Malaysia Sdn. Bhd, Century Square, Ground Floor, North Wing, Block 2320, Jalan Usahawan, 63000 Cyberjaya, Selangor
[4]School of Business, Monash University, Malaysia campus, Jalan Lagoon Selatan, 47500, Bandar Sunway, Selangor
*Corresponding author: vijay.ramana@um.edu.my

Abstract

HOTAIR gene is a long non-coding RNA (lncRNA) that regulates the differentially expressed genes causing intermediate malignant respiratory disorders like small-cell lung carcinoma (SCLC). Thus, HOTAIR gene can be a predictor for chemotherapy response and therapeutic target against chemoresistance in SCLC. In this study, SCLC and HOTAIR suppressed genes were compared with normal lung tissue samples (control) to identify the differentially expressed genes (DEGs). The number of significant DEGs in both SCLC and HOTAIR suppressed vs control analyses were 4906 and 255 respectively with a cutoff value of ($q \leq 0.05$). Among the DEGs, 48 genes were expressed in both SCLC and HOTAIR suppressed conditions with log2 fold change <2. Eight genes (n=8) (ASAH1, CENPE, FABP5, ACOXL, CTSB, RAB11FIP1, VEGFA, JUN) were significantly expressed ($p<0.05$) in both conditions. CENPE, VEGFA and JUN were lowly expressed in SCLC but highly expressed in HOTAIR suppressed condition. Enrichment analysis on the selected genes using Kyoto Encyclopedia of Genes and Genomes (KEGG) tool shows the cell adhesion and cell cycle pathways were significantly activated in both SCLC and HOTAIR suppressed conditions. About 58 reactome pathways were identified and the post translational protein modification pathway was activated in SCLC while the innate immune system pathway in HOTAIR suppressed gene. These potential HOTAIR suppressed DEGs are CENPE, VEGFA and JUN and they could serve as potential biomarkers for cancer drug development. However, further studies are required to develop the protein interaction model of the enriched genes to analyze the correlation and insight of the therapeutic intervention.

Keywords: Small-cell lung cancer, HOTAIR gene, differentially expressed gene, pathway analysis, cellular mechanism

Introduction

Lung cancer is the leading cause of cancer death among men and second leading cause of cancer death among women (1). There is a slow advancement in survival rate of lung cancer due to late-stage diagnosis. Small-cell lung cancer (SCLC) is difficult to investigate clinically owing to a paucity of substantive tumourspecimens. SCLC is grouped with tumors that display neuroendocrine differentiation. It is common for patients of small-cell carcinoma having metastatic disease and most of them may have relapse within first 2 years after undergoing first treatment. The 2-year survival rate is less than 10 percent in metastatic patients and this cancer is commonly located in major airways in the

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

human body (2). Long non-coding RNA (lncRNA) is a non-protein coding RNA with a length of more than 200 nucleotides. It regulates various biochemical and cellular processes in cancer. The lncRNA mechanism requires clarification to be used as a potential biomarker and therapeutic target as tumor progresses (3).

HOTAIR is a long non-coding RNA. It is a splice and poly-adenylated RNA with 2158 nucleotides and exons (4). Deviation of DNA methylation is also a suitable biomarker for cancer diagnosis, prognosis, and prediction in response to chemotherapeutic drug (5). HOTAIR gene has a higher expression level in tumor tissue than the non-tumor tissue in SCLC. This is due to lymphatic invasion and relapse. A previous study stated that (6), different expression levels of HOTAIR can assist in detecting progression stage of cancer. In addition, it helps to predict the survival rate of individuals where different functional SNPs across HOTAIR locus influences cancer risk (1,6). In recent studies, knockdown of HOTAIR increases cell sensitivity to anti-cancer drug, apoptosis and inhibits growth of tumor by decreasing the cell cycle progression. Thus, HOTAIR can be a predictor for chemotherapy response and therapeutic target against chemoresistance in SCLC (2,3).

Omics technology enables the profiling of genomic, transcriptomic, epigenomic and proteomic that can generate various types of biological data. This technology was involved in the studies, namely analysis of gene or RNA expression, epigenetic, protein abundance and even regulatory mechanism (7). This technology makes it possible for diagnosis and prognosis of diseases such as cancers to identify differentially expressed genes (DEGs). This facilitates the identification of potential biomarkers and drug targets for targeted therapeutic medication or treatment providing more insights on cancer or tumorigenesis mechanism, its immune feedback response, and its metastasis condition. Hence, this study aimed to identify HOTAIR suppressed DEGs which could serve as potential biomarkers for cancer drug development.

Materials and Methods

Data pre-processing

In this study, New Tuxedo protocol was applied to analyze the RNA-seq data. It is faster, uses less computing power and produces more accurate results (8). Differentially expressed genes were fed into gene set enrichment analysis (GSEA) tools to determine gene interactive pathways.

Data description

The RNA-seq datasets used in this study are small-cell lung cancer (SCLC) from patients, human normal lung tissue as control and suppressed HOTAIR gene in lung tissue of human. These datasets were downloaded from sequence read archive (SRA), NBCI in FASTQ format. There are 3 biological replicates for each sample to ensure the quality of data (Table 1).

Quality Control

FastQC tool was used to perform quality control on raw sequence data from high throughput sequencing pipeline. This tool checks sequence quality per base, adapters, and GC (Guanine-Cytosine) content per sequence. CutAdapt tool was used to cut the adapters that were attached to the sequence resulting from the Illumina experiments without the help of FastQC. The quality of the sequence was interpreted using boxplot where if the boxplot falls within green threshold, it indicates good quality (9).

Gene mapping and transcriptome assembly

RNA-seq analysis tool comprised of three categories, which are reading alignments, transcriptome assembly or genome annotation and quantification of gene and transcript (10). In this analysis, the New Tuxedo protocol that uses TopHat and BowTie was used for mapping, aligning

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

reads, estimating transcript abundance and analysing differential expression. Functional profiling was performed using high throughput functional enrichment tools such as GSEA or DAVID to determine pathways and functions of enriched genes (11).

Identification of significant differentially expressed genes (DEGs)

RNA-seq has various methods to identify novel genes or transcripts used in genetic engineering, identification of mutations and differential gene expression (DGE). Each RNA-seq experiment requires reference genome mapping, where in this analysis the reference genome hg38.p11 was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.37/). Quantification of expression abundance was done on the mapped reads for gene or transcript, but the read depth varies for different samples and direct comparison of samples could not be accomplished. Normalization was done using the Cufflink sub-tool named CuffDiff, which is embedded with methods such as Reads Per Kilobase per Million mapped reads (RPKM) and Fragments Per Kilobase per Million mapped fragments (FPKM). False Discovery Rate (FDR) in Cufflinks was calculated using Benjamini-Hochberg approach with a cutoff value of 0.05 and fold-change cutoff at 2. Figure 1 shows the pipeline of RNA-seq differential expression gene analysis.

Gene Set Enrichment Analysis (GSEA)

To look at the biological significance of differentially expressed genes, gene set enrichment analysis and pathway analysis were performed to identify all the significant gene ontology, which includes biological

Table 1: RNA-seq datasets summary

Type of Datasets	Experiment ID(GSM) / Individual ID(V)	SRA ID
Small cell lung cancer (SCLC)	GSM1464365	SRX669103
	GSM1464366	SRX669104
	GSM1464367	SRX669105
Normal lung tissue	V133	ERX288531
	V81	ERX288597
	V130	ERX288620
HOTAIR suppressed tissue	GSM3375812	SRX4644228
	GSM3375813	SRX4644229
	GSM3375814	SRX4644230

Source: Sequence Read Archive (2019). Retrieved from https://www.ncbi.nlm.nih.gov/sra)

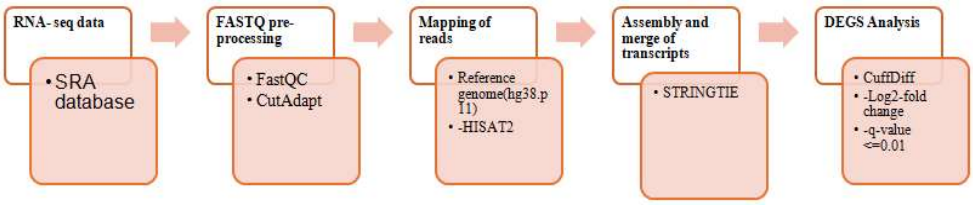

Figure 1: RNA-seq differential expression gene (DEGs) analysis pipeline

Suppression of Hotair Gene

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

process, cellular components, molecular function and the KEGG and REACTOME pathway (12). Normalized enrichment score (NES) was calculated for gene set based on the degree of over-expressed rank gene list (13). The significant genes were selected based on the cutoff value of q <= 0.01. The selected significant genes were then input into GSEA tools and cross-referenced with the significant gene and annotated genes with 1000 permutations. Databases used in this study are Hallmark gene sets, REACTOME, and Kyoto Encyclopedia of Genes and Genomes (KEGG). Hallmark gene sets reduced the redundancy in gene sets cross referencing results. Pathway with adjusted p-value ($p<0.05$) was generated.

Results and Discussion

RNA-sequencing is popular mainly in cancer research. This is due to its ability to find biomarkers for therapeutic drug target design specialized for each patient (11). RNA-sequencing is more widely used as it can detect novel genes or transcripts and measure level of gene expression by abundance estimation of log2-fold change (14,15). In this study, the transcriptome reads were mapped with the reference genome from NCBI (hg38.p11), which was built using HISAT2 and assembled using STRINGTIE tool. Figure 2 shows the alignment rate of each sample after being mapped to reference genome. Alignment rate more than 60 percent indicates that it has high reproducibility (9) and our data shown in Figure 2 exceeded the threshold of good alignment rate.

Initially, the differential expression gene (DEG) analysis was done on lung cancer sample against normal lung tissue to determine significant genes, which playan important role in the diagnosis ofSCLC. Secondly, the HOTAIR suppressed lung sample was analyzed against the normal lung tissue (control) to observe the significant genes, which play an important role in suppressing the important marker for progression of SCLC.

Our results show that SCLC vs normal produces 4906 significant DEG after applying $q < 0.05$ (Figure 3A). Whereas HOTAIR suppressed sample vs normal have 255 significant DEGs after applying $q < 0.05$ (Figure 3B). Similar findings

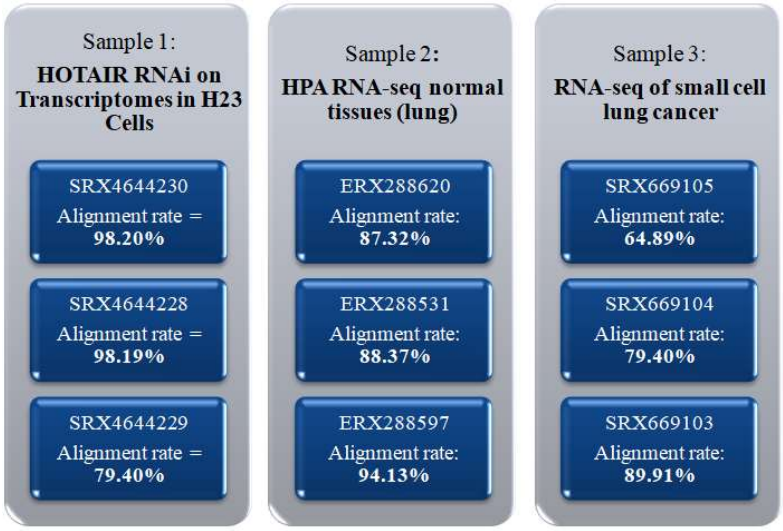

Figure 2: Alignment summary using HISAT2 for each sample. Reference genome used was hg38.p11

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

reported (16) where lesser DEGs identified in HOTAIR suppressed sample. This suppression reduces the SCLC cell proliferation in the patient. Among the 255 significant genes identified in HOTAIR suppressed sample vs normal, 80 genes were upregulated (log2FC >= 2), and 78 genes were downregulated (log2FC <= -2). On the other hand, 101 genes were upregulated and 549 were downregulated in SCLC vs normal.

The heatmap of SCLC vs normal shows that the top 50 significant genes have the highest expression level of 3 (Figure 4). In Figure 5, heatmap of HOTAIR suppressed vs normal shows the top 50 highest expression level in only 2.5, which is lesser than SCLC vs normal. The identified DEGs were subjected to KEGG and REACTOME pathway enrichment and Gene Ontology (GO) analysis with False Discovery Rate (FDR) q <= 0.05. We retrieved the top 25 GO enriched pathways (Table 2); 17 belong to biological process domains, 3 from cellular component and 5 from molecular function. This indicates that in the condition of SCLC vs normal, the most affected domain is the biological process of the human body (14). In addition, SCLC causes more disruption in the biological process of the patient as tumor cells may change the functionality of the genes.

The computed enrichment analysis yielded 43 enriched KEGG pathways in SCLC vs normal sample and the top 10 pathways were depicted in Figure 6. KEGG Oxidative Phosphorylation pathway has the highest number of enriched genes. Oxidative phosphorylation pathway is providing the necessity of the tumor cells to carry out proliferation in the form of nutrient or energy (7). The genes which are involved in this pathway are ATP5PO and ATP6V1F.

On the other hand, the computed enrichment analysis yielded 50 enriched REACTOME pathways from this sample. Ranked by NES value for most enriched pathway, the top 10 pathways are selected as shown in Figure 7. REACTOME Post-Translational Protein Modification has the highest number of enriched genes, which is 2. Post-translational protein modification pathway involves biochemical modification that occurs in one or more amino acids after protein being translated (5). The genes involved in this pathway are SUMF2 and YKT6.

The identified DEGs in HOTAIR suppressed vs normal were subjected for enrichment analysis and observedthar 139 GO terms in correlation to the significant genes in this sample. After selecting the top 25 GO enriched pathways, it was observed

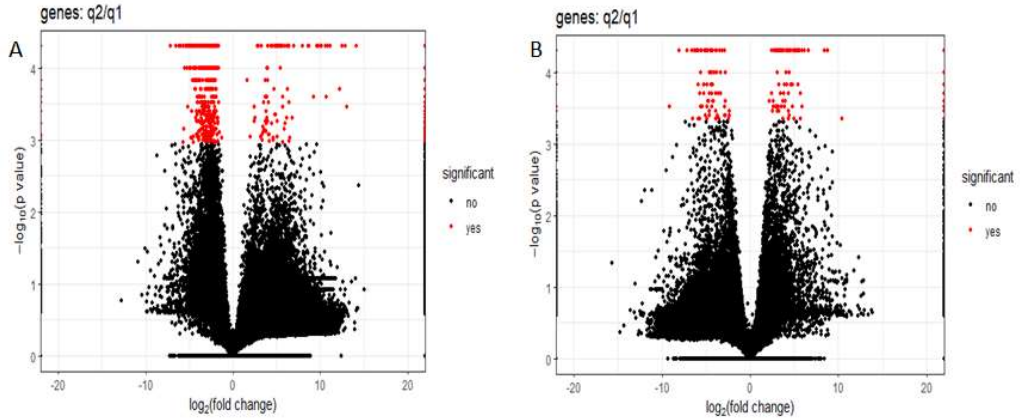

Figure 3A: Volcano plot of SCLC against normal; **3B**: Volcano plot of HOTAIR suppressed against normal

Suppression of Hotair Gene

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

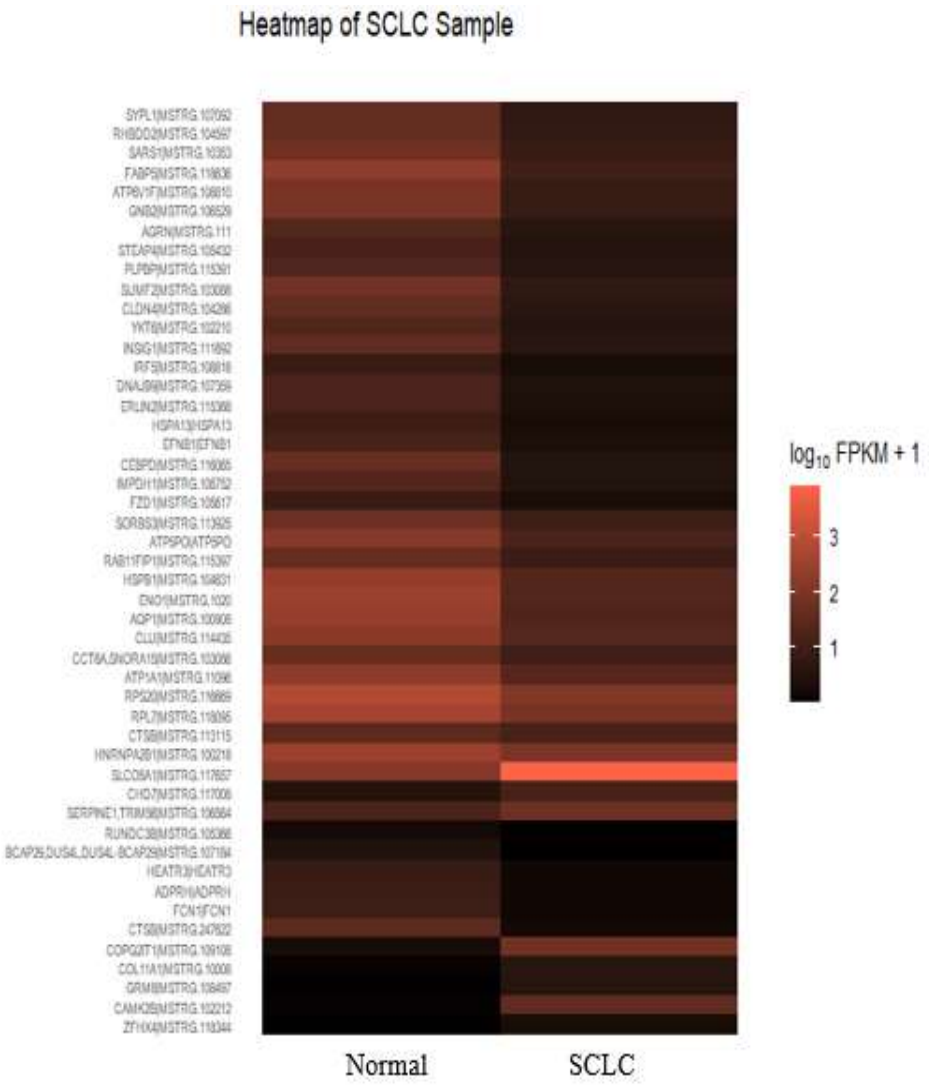

Figure 4: Heatmap of top 50 DEGSs in SCLC vs normal sample (control). Hierarchical clustering of each gene shown in y-axis. Each row represents a gene name, and expression level based on log10FPKM in color key form. Black shows lowest level of FPKM <= 1 and lightest shade of red shows highest level of FPKM >=3 as this sample has very high expression level thus the FPKM is high

that 15 are from biological process domain, 7 are from cellular component and 3 from molecular function (Table 3). The most enriched GO pathway in this sample is GO cell cycle. There were 29 enriched genes in this pathway, which are NCAPG2, NEK2, JUN, FOXM1, MUC1, MAPK13, STIL, TOP2A, CCND2, KIF11, CENPE, IQGAP3,

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

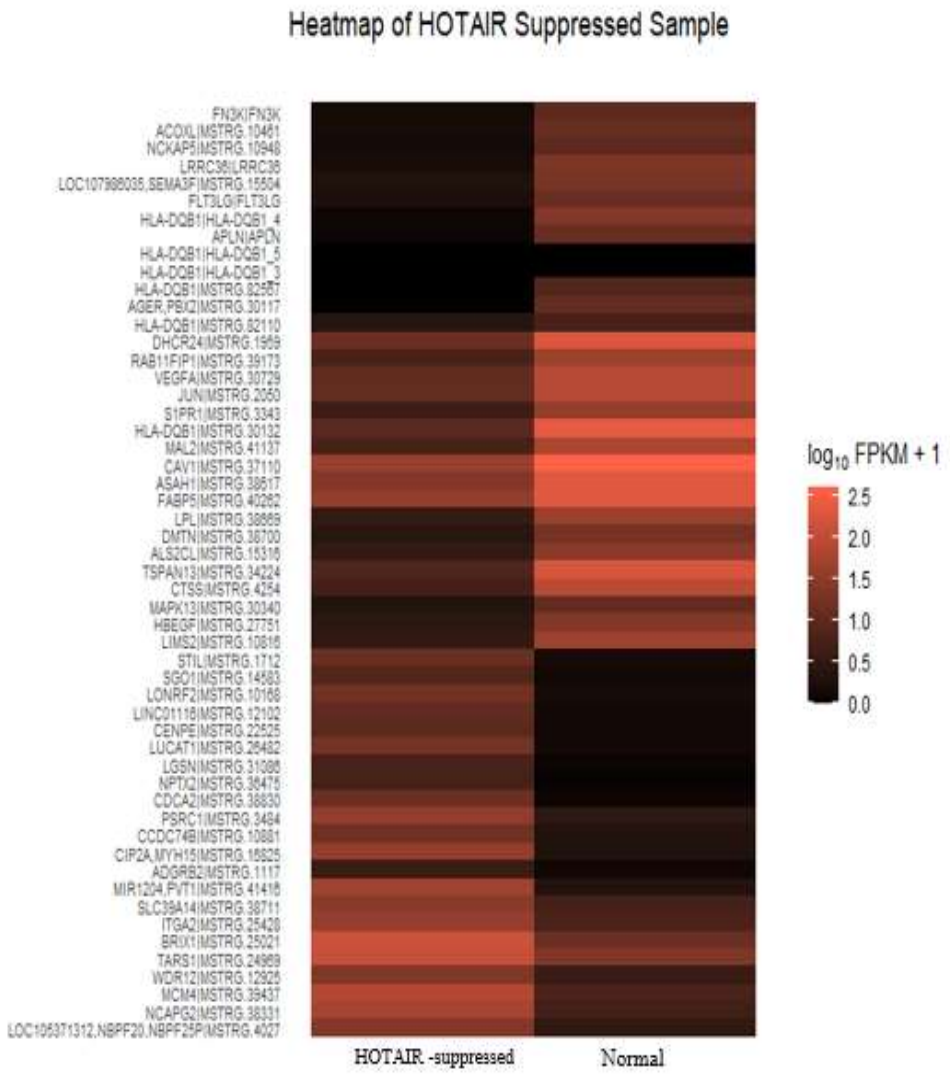

Figure 5: Heatmap of HOTAIR suppressed vs normal sample(control). Hierarchical clustering of each gene shown in y-axis. Each row represents a gene name, and expression level based on log_{10}FPKM in color key form. Black shows lowest level of FPKM <= 1 and lightest shade of red shows highest level of FPKM >=2.5 as this sample has very high expression level thus the FPKM is high

CDCA3, RAD51, TIMELESS, E2F7, SUSD2, FLT3LG, HELLS, CDCA2, PSRC1, SGO1, MCM4, KIF18B, POLE2, MKI67, DHCR24, PPP1R15A and FGFR2. These genes expression level is elevated during knockdown of HOTAIR in which the aid in suppressing development of cancer cell in small cell lung cancer (4).These genes are

Suppression of Hotair Gene

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

Table 2: Summary of 25 most enriched GO pathways in SCLC vs normal sample

GO Pathway	GO Domain	Enriched Gene Name
GO_CATALYTIC_COMPLEX	Cellular component	ENO1, HNRNPA2B1, HSPB1, RHBDD2
GO_NUCLEAR_TRANSPORT	Biological process	HEATR3, HNRNPA2B1
GO_PASSIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	Molecular function	AQP1, ATP5PO, CLDN4
GO_CELLULAR_RESPONSE_TO_OXYGEN_LEVELS	Biological process	AQP1, ENO1
GO_CATION_CHANNEL_ACTIVITY	Molecular function	AQP1, ATP5PO
GO_POSITIVE_REGULATION_OF_SECRETION	Biological process	AQP1, FCN1
GO_POSITIVE_REGULATION_OF_CELL_POPULATION_PROLIFERATION	Biological process	AQP1, EFNB1
GO_ATP_METABOLIC_PROCESS	Biological process	ENO1, ATP5PO
GO_ATP_BIOSYNTHETIC_PROCESS	Biological process	ENO1, ATP5PO
GO_REGULATION_OF_BODY_FLUID_LEVELS	Biological process	AQP1, HSPB1, CLDN4
GO_CELL_ADHESION_MOLECULE_BINDING	Molecular function	ENO1, YKT6
GO_RESPONSE_TO_OXYGEN_LEVELS	Biological process	AQP1, ENO1
GO_MEMBRANE_PROTEIN_COMPLEX	Cellular component	ATP1A1, YKT6, RHBDD2, INSIG1, ATP5PO, ATP6V1F, CLDN4
GO_RESPONSE_TO_DRUG	Biological process	AQP1, FZD1, ATP1A1, IRF5, AGRN, CLDN4
GO_MAGNESIUM_ION_BINDING	Molecular function	ENO1, ADPRH
GO_REGULATION_OF_VASCULATURE_DEVELOPMENT	Biological process	AQP1, SARS1, HSPB1
GO_CELL_CELL_ADHESION	Biological process	EFNB1, HSPB1, CLDN4
GO_GENERATION_OF_PRECURSOR_METABOLITES_AND_ENERGY	Biological process	ENO1, STEAP4, ATP5PO
GO_CADHERIN_BINDING	Molecular function	ENO1, YKT6
GO_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND	Biological process	CLU, AQP1, ATP1A1, INSIG1, IRF5, AGRN, ATP6V1F, CLDN4
GO_REGULATION_OF_ANATOMICAL_STRUCTURE_MORPHOGENESIS	Biological process	AQP1, SARS1, FZD1, CAMK2B, HSPB1, CLDN4
GO_GOLGI_MEMBRANE	Cellular component	STEAP4, YKT6, RHBDD2
GO_RESPONSE_TO_OXIDATIVE_STRESS	Biological process	AQP1, FZD1, HSPB1
GO_CELLULAR_RESPONSE_TO_CHEMICAL_STRESS	Biological process	AQP1, FZD1, HSPB1
GO_REGULATION_OF_PEPTIDASE_ACTIVITY	Biological process	AQP1, CLDN4

The ranking of most enriched pathway determined by Negative normalized Enrichment Score (NES) value. The higher the value of NES, the more enriched the pathway is in the sample. This table is arranged in the order of the most enriched pathway from the top.

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

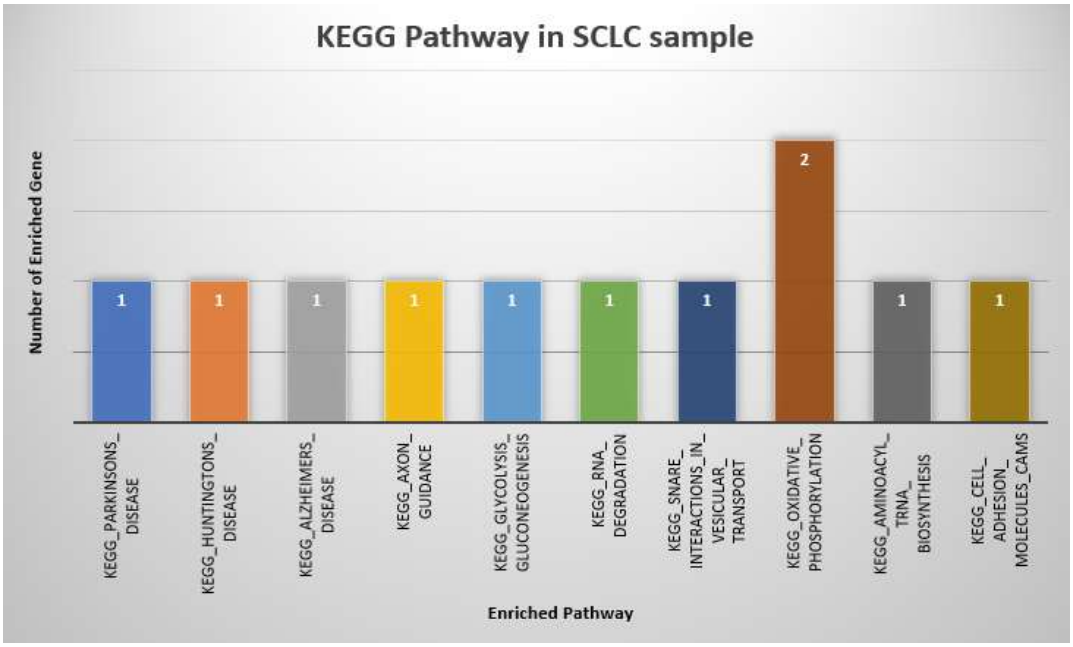

Figure 6: Enriched KEGG pathway in small cell lung cancer sample against normal sample. This bar chart shows the top 10 KEGG pathway out of 43 pathways

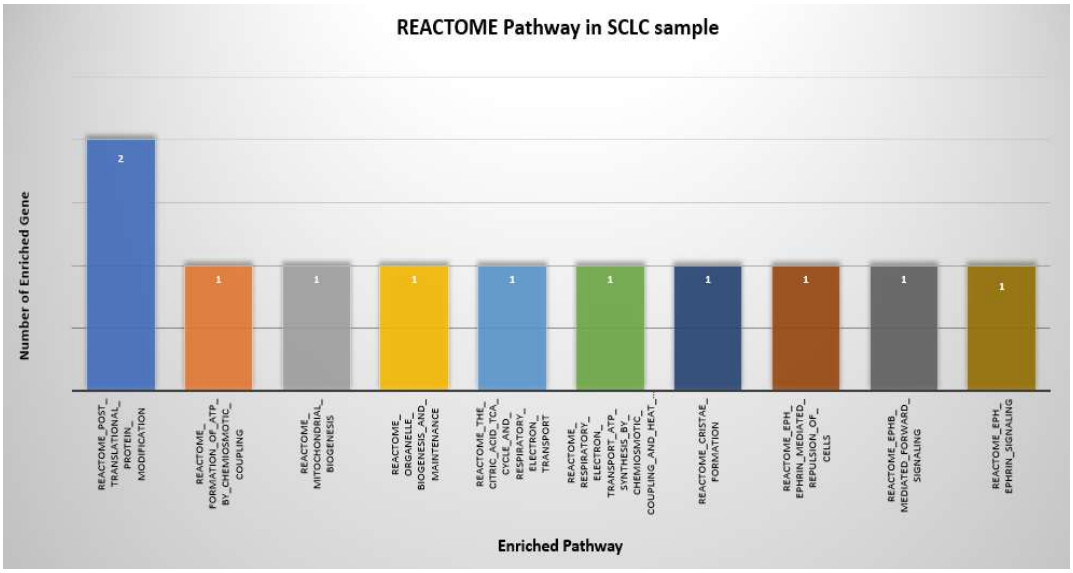

Figure 7: Enriched REACTOME pathway in small cell lung cancer sample vs normal sample.This bar chart shows only the top 10 REACTOME pathways out of 50 pathways

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

Table 3: Summary of 25 most enriched GO pathways in HOTAIR suppressed vs normal sample

GO Pathway	GO Domain	Enriched Gene Name
GO_MITOTIC_CELL_CYCLE	Biological process	RAD51, E2F7. NEK2, FOXM1, PSRC1, SGO1, MUC1, MCM4, KIF18B, STIL, TOP2A, CCND2, POLE2, MKI67. KIF11, CENPE, IQGAP3
GO_ORGANELLE_FISSION	Biological process	KIF18B, TOP2A, RAD51, MKI67, KIF11, CENPE, NEK2, PSRC1, SGO1
GO_CELL_CYCLE	Biological process	NCAPG2, NEK2, JUN, FOXM1, MUC1, MAPK13, STIL, TOP2A, CCND2, KIF11, CENPE, IQGAP3, CDCA3, RAD51, TIMELESS, E2F7, SUSD2, FLT3LG, HELLS, CDCA2, PSRC1, SGO1, MCM4, KIF18B, POLE2, MKI67, DHCR24, PPP1R15A, FGFR2
GO_CHROMOSOME_ORGANIZATION	Biological process	VEGFA, RAD51, NCAPG2, NEK2, HELLS, PSRC1, SGO1, MUC1, MCM4, KIF18B, TOP2A, POLE2, MKI67, CENPE
GO_CELL_CYCLE_PROCESS	Biological process	RAD51, TIMELESS, E2F7, SUSD2, NEK2, FOXM1, PSRC1, SGO1, MUC1, MCM4, KIF18B, STIL, TOP2A, CCND2, POLE2, MKI67, KIF11, CENPE, DHCR24, IQGAP3, PPP1R15A
GO_CHROMOSOME_SEGREGATION	Biological process	KIF18B, TOP2A, MKI67, CENPE, NEK2, CDCA2, PSRC1, SGO1
GO_MITOTIC_NUCLEAR_DIVISION	Biological process	KIF18B, MKI67, KIF11, CENPE, NEK2, PSRC1, SGO1
GO_MICROTUBULE_BASED_PROCESS	Biological process	KIF18B, STIL, NCKAP5, KIF11, CENPE, NEK2, WDR35, TRIM46, PSRC1, SGO1
GO_RNA_BINDING	Molecular function	TOP2A, TARS1, MKI67, KPNA2, BRIX1, PDCD11, NQO1, JUN, RPS18

(*Contd.*)

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

Table 3: Summary of 25 most enriched GO pathways in HOTAIR suppressed vs normal sample (*Contd.*)

GO Pathway	GO Domain	Enriched Gene Name
GO_CONDENSED_CHROMOSOME	Cellular component	TOP2A, RAD51, MKI67, CENPE, NCAPG2, NEK2, SGO1
GO_MICROTUBULE_CYTOSKELETON	Cellular component	CTSC, RAD51, NEK2, WDR35, PSRC1, SGO1, KIF18B, STIL, NCKAP5, TOP2A, KIF11, CENPE, ADGRB2
GO_CELL_DIVISION	Biological process	VEGFA, TIMELESS, E2F7, SUSD2, NCAPG2, CAT, NEK2, HELLS, CDCA2, PSRC1, SGO1, KIF18B, TOP2A, CCND2, KIF11, CENPE, FGFR2, CDCA3
GO_SPINDLE	Cellular component	KIF18B, KIF11, CENPE, NEK2, PSRC1, SGO1
GO_DNA_METABOLIC_PROCESS	Biological process	MCM4, ACVRL1, TOP2A, RAD51, TIMELESS, POLE2, FOS, KPNA2, NEK2, HELLS, TIGAR, FOXM1
GO_ATPASE_ACTIVITY_COUPLED	Molecular function	MCM4, KIF18B, TOP2A, RAD51, KIF11, HELLS
GO_MICROTUBULE_CYTOSKELETON_ORGANIZATION	Biological process	KIF18B, STIL, NCKAP5, KIF11, CENPE, NEK2, TRIM46, PSRC1, SGO1
GO_SISTER_CHROMATID_SEGREGATION	Biological process	KIF18B, TOP2A, CENPE, NEK2, PSRC1, SGO1
GO_NUCLEAR_CHROMOSOME_SEGREGATION	Biological process	KIF18B, TOP2A, CENPE, NEK2, PSRC1, SGO1
GO_DNA_REPAIR	Biological process	MCM4, RAD51, POLE2, TIMELESS, TIGAR, FOXM1
GO_CHROMOSOMAL_REGION	Cellular component	MCM4, RAD51, CENPE, NEK2, HELLS, SGO1
GO_NUCLEAR_BODY	Cellular component	KIF18B, RAD51, POLE2, E2F7, MKI67, NCAPG2, SGO1
GO_MICROTUBULE_ORGANIZING_CENTER_ORGANIZATION	Cellular component	STIL, KIF11, NEK2, SGO1

(*Contd.*)

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

Table 3: Summary of 25 most enriched GO pathways in HOTAIR suppressed vs normal sample (*Contd.*)

GO Pathway	GO Domain	Enriched Gene Name
GO_CHROMOSOME	Cellular component	ARPC5, RAD51, TIMELESS, E2F7, FOS, NCAPG2, NEK2, JUN, HELLS, CDCA2, FOXM1, SGO1, MUC1, CAPN2, MCM4, TOP2A, CCND2, POLE2, MKI67, CENPE
GO_REGULATION_OF_MICROTUBULE_BASED_PROCESS	Biological process	STIL, KIF11, NEK2, TRIM46, PSRC1
GO_ATPASE_ACTIVITY	Molecular function	MCM4, KIF18B, TOP2A, RAD51, KIF11, CENPE, ABCA3, HELLS

The ranking of most enriched pathway determined by Negative normalized Enrichment Score (NES) value. The higher the value of NES, the more enriched the pathway is in the sample. This table is arranged in the order of the most enriched pathway from the top 25 GO pathways

also involved in the progression of cell cycle that comprises of the 4 phases and replication of genome and segregation of chromosome.

The computed enrichment analysis yielded 2 enriched KEGG pathways;Cell Adhesion Molecule (CAMS) and Viral Myocarditis have the same number of enriched genes. Cell Adhesion Molecule (CAMS) pathway involved in the binding of cell surface with other cell or extracellular matrix as it is a subset of the cell adhesion protein. This pathway aids the regulation of a cell or process. The genes which are involved in this pathway are CNTN1, HLA-DQB1 and HLA-DRA. Our data supported where HLA-DQB1 genes are slightly expressed in HOTAIR suppressed condition and highly expressed in normal condition (Figure 5). Next, enriched KEGG pathway is the Viral Myocarditis pathway, which is involved in inflammation of heart muscle due to viral infection. It may be caused by direct cytopathic effect of virus, a pathologic immune response to persistent virus or autoimmunity triggered by viral infection. The genes involved in this pathway are CNTN1, HLA-DQB1 and HLA-DRA and similarly to Cell Adhesion Molecule (CAMS) pathway genes, in which they are under-expressed in HOTAIR suppressed condition and highly expressed in the normal condition.

From REACTOME pathway library, the computed enrichment results yielded 9 enriched REACTOME pathway in HOTAIR suppressed vs normal sample (Figure 8). REACTOME Innate Immune System pathway has the highest number of enriched genes which is 18. Innate Immune System pathway encompasses in the non-specified part of immunity, which is part of individual biological make-up. The genes involved in this pathway are CTSC, CST3, ARPC5, CTSD, CTSB, CTSS, FABP5, FOS, ASAH1, CAT, CREG1, JUN, LYZ, MUC1, MAPK13 SLC11A1, CD81 and JUP. This pathway refers to the non-specified defense mechanism when an antigen occurs. It includes physical barrier and immune system. The main purpose of knockdown of HOTAIR gene is to suppress SCLC and this pathway is the most enriched as the genes involve in the active anti-tumor immune response (1).

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

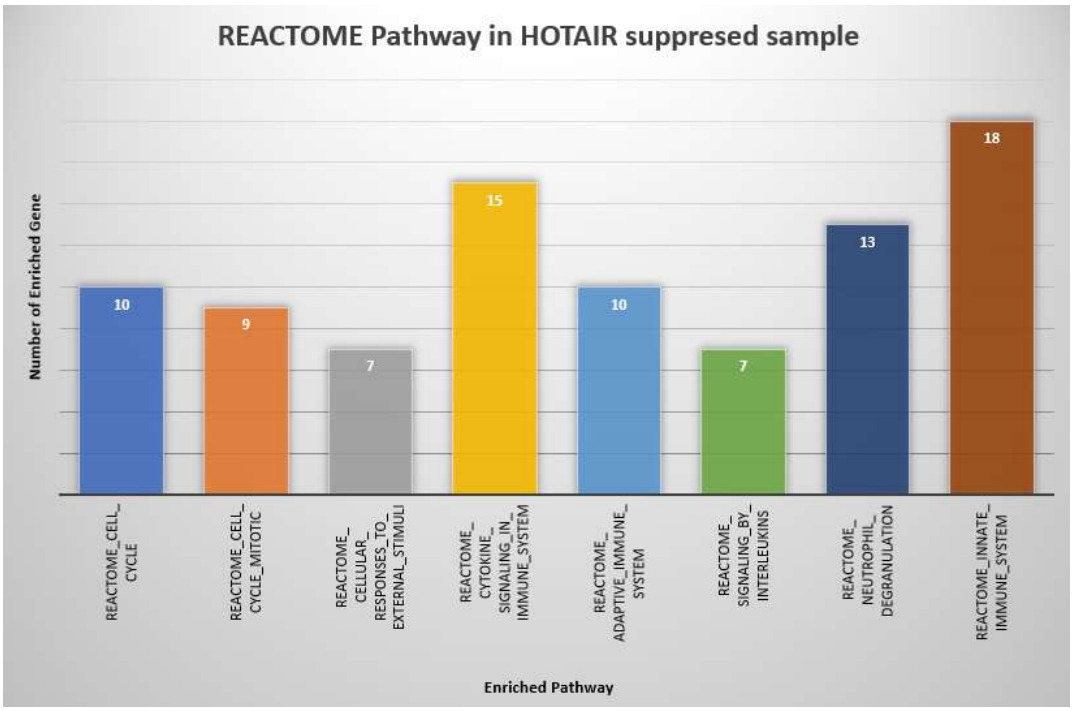

Figure 8: Enriched REACTOME pathway in HOTAIR suppressed sample vs normal sample

Conclusion

SCLC vs normal condition produces 4096 significant differentially expressed genes (DEGs) while HOTAIR suppressed vs normal condition produces 255 significant DEGs. There are apparent differences in total number of DEGs in both conditions because the purpose of knockdown of HOTAIR gene is to suppress small-cell lung cancer progression. Thus, the number of DEGs in a cancer cell is higher when compared to normal cells whereas DEGs of suppressed cancer gene (HOTAIR) is lower than in normal cells.The biological pathway enriched by GSEA shows that SCLC vs normal affects post translational protein modification pathway and for HOTAIR suppressed vs normal effect Innate immune system pathway. In KEGG pathway, the cell adhesion molecule (CAMS) is highly affected because CAMS are tumor suppressant and there is major difference of expression level in both conditions. Seven genes have the tendencyto become potential diagnostic biomarkers for lung cancer. The genes are PLPBP, IRF5, DNAJB9, EFNB1, CEBPD, ENO1 and CLU. In SCLC vs normal sample and HOTAIR suppressed vs normal, there are 8 highly significant genes, which play role in cancer development whereas only 3 genes have significant expression level in HOTAIR suppressed sample. These genes are CENPE, VEGFR and JUN; which serve as possible prognostic biomarkers (17). These genes as potential biomarkers may enhance the effectiveness of drug development for treatment benefits of lung cancer.

References

1. Fang, S., Gao, H., Tong, Y., Yang, J., Tang, R., Niu, Y., Li, M., & Guo, L. (2016). Long noncoding RNA-HOTAIR affects chemoresistance by regulating HOXA1 methylation in small cell lung cancer cells. *Laboratory Investigation*, *96*(1), 60-68. https://doi.org/10.1038/labinvest.2015.123

Vol. 18 (Supplementry Issue 3s) 136 - 149, July-Sept 2024, ISSN 0973-8916 (Print), 2230-7303 (Online)
10.5530/ctbp.2024.3s.10

2. Liu, M., Zhang, H., Li, Y., Wang, R., Li, Y., Zhang, H., Ren, D., Liu, H., Kang, C., & Chen, J. (2018). HOTAIR, a long noncoding RNA, is a marker of abnormal cell cycle regulation in lung cancer. *Cancer science*, *109*(9), 2717-2733. https://doi.org/10.1111/cas.13745

3. Jiang, C., Li, X., Zhao, H., & Liu, H. (2016). Long non-coding RNAs: potential new biomarkers for predicting tumor invasion and metastasis. *Molecular Cancer*, *15*(1), 62. https://doi.org/10.1186/s12943-016-0545-z

4. Loewen, G., Jayawickramarajah, J., Zhuo, Y., & Shan, B. (2014). Functions of lncRNA HOTAIR in lung cancer. *Journal of Hematology & Oncology*, *7*(1), 90. https://doi.org/10.1186/s13045-014-0090-4

5. Roche, J. (2018). The Epithelial-to-Mesenchymal Transition in Cancer. *Cancers*, *10*(2), 52. https://doi.org/10.3390/cancers10020052

6. Hajjari, M., & Salavaty, A. (2015). HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer biology & medicine*, *12*(1), 1-9. https://doi.org/10.7497/j.issn.2095-3941.2015.0006

7. Petrek, H., & Yu, A.-M. (2019). MicroRNAs in non-small cell lung cancer: Gene regulation, impact on cancer cellular processes, and therapeutic potential. *Pharmacology research & perspectives*, *7*(6), e00528-e00528. https://doi.org/10.1002/prp2.528

8. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650-1667. https://doi.org/10.1038/nprot.2016.095

9. Mirsafian, H., Ripen, A.M., Manaharan, T., Mohamad, S.B., & Merican, A.F. (2016). Toward a Reference Gene Catalog of Human Primary Monocytes. *Omics A Journal of Integrative Biology, 20*, 627-634.

10. Tripathi, N., P, S., & Nair, A. (2017). Comparative Study of Transcriptomic profiling and Functional enrichment in Ovarian Cancer Cell lines. *Canadian Journal of Biotechnology*, *1*, 65-65. https://doi.org/10.24870/cjb.2017-a52

11. Bedre, R., Irigoyen, S., Petrillo, E., & Mandadi, K. K. (2019). New Era in Plant Alternative Splicing Analysis Enabled by Advances in High-Throughput Sequencing (HTS) Technologies [10.3389/fpls.2019.00740]. *Frontiers in Plant Science*, *10*, 740. https://www.frontiersin.org/article/10.3389/fpls.2019.00740

12. Hikida, T., Yao, S., Macpherson, T., Fukakusa, A., Morita, M., Kimura, H., Hirai, K., Ando, T., Toyoshiba, H., & Sawa, A. (2020). Nucleus accumbens pathways control cell-specific gene expression in the medial prefrontal cortex. *Scientific Reports*, *10*(1), 1838. https://doi.org/10.1038/s41598-020-58711-2

13. Park, J., & Chun, K.-H. (2020). Identification of novel functions of the ROCK2-specific inhibitor KD025 by bioinformatics analysis. *Gene*, *737*, 144474. https://doi.org/https://doi.org/10.1016/j.gene.2020.144474

14. Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor protocols*, *2015* (11), 951-969. https://doi.org/10.1101/pdb.top084970

15. Marini, F., Linke, J., & Binder, H. (2020). ideal: an R/Bioconductor package for Interactive Differential Expression Analysis. *bioRxiv*, 2020.2001.2010.901652. https://doi.org/10.1101/2020.01.10.901652

16. Martin, J., Wang, Z (2011). Next-generation transcriptome assembly. *Nature Review Genetics,* 12, 671–682. https://doi.org/10.1038/nrg3068

17. Haidar, M. N., Islam, M. B., Chowdhury, U. N., Rahman, M. R., Huq, F., Quinn, J. M., & Moni, M. A. (2020). Network-based computational approach to identify genetic links between cardiomyopathy and its risk factors. *IET Systems Biology*, *14*(2), 75-84.