



Bioinformatics approaches for cross-species liver cancer analysis based on single-cell RNA sequencing data in rats mice and human insights causative genes and pathways

Sudheer Menon

Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India

DOI: <https://doi.org/10.33545/26646501.2021.v3.i1.a.19>

Abstract

In this research, we first recognized 2322 differentially expressed genes via way of means of evaluating gene expression profiles for 2 control, adenoma, and carcinoma samples by the use of an F-test. These genes have been sooner or later mapped to the rat chromosomes through the use of a unique visualization tool, the Chromosome Plot. Using the equal plot, we similarly mapped the enormous genes to orthologous chromosomal places in people and mice. Several genes expressed in rat 1q which is probably amplified in rat liver maximum cancers map to the human chromosomes 10, 11, and 19 and to the mice chromosomes 7, 17, and 19, which have been implicated in studies of human and mice liver cancers. Using Comparative Genomics Single cell RNA sequencing data Analysis (CGMA), we recognized areas of capacity aberrations in people. Lastly, a pathway evaluation became performed to expect altered human pathways primarily based totally on statistical evaluation and extrapolation from the rat records. All of the recognized pathways were regarded to be vital with-inside the etiology of the human liver most cancers, inclusive of molecular cycle control, molecular boom and differentiation, apoptosis, transcriptional regulation, and protein metabolism.

Background: The final touch of the sequencing of the human, mouse, and rat genomes and understanding of cross-species gene homologies permit research of differential gene expression in animal fashions. These styles of research have the capacity to substantially beautify our information of sicknesses inclusive of liver most cancers in people. Genes co-expressed throughout a couple of species are maximum in all likelihood to have conserved features. We have used numerous bioinformatics methods to take a look at single cell RNA sequencing data expression profiles from liver neoplasms that stand up in albumin-SV40 transgenic rats to explain genes, chromosome aberrations, and pathways that are probably related to human liver most cancers.

Aim and objectives: The main aim of this research study was to evaluate the bioinformatics approaches of liver cancer with single cell RNA sequencing Data. This study will analyze the role of RNA changes that would help in liver cancer analysis. Another aim was to evaluate and differentiate the pattern on liver cancer in rats, mice and humans.

Keywords: bioinformatics approaches, liver cancer analysis, single-cell RNA

Introduction

For decades, classical toxicology has used chance checks primarily based totally on animal research for regulatory decisions. The underlying assumption is that vital organic features are frequently conserved throughout species. In continuation of this paradigm, the attempt in toxico-genomics is positioned on analyzing rodents and different surrogates the use of superior genomics technologies, inclusive of Single-cell RNA sequencing data. Single cell RNA sequencing data research permits simultaneous dimensions of the expression of big numbers of genes. Given the final touch of the RNA series of the human, mouse, and rat genomes [1-3], genes recognized in single cell RNA sequencing data research may be simply in comparison throughout species regarding the gene orthologs [4, 5]. This assumes that genes co-expressed throughout a couple of species are in all likelihood to have conserved features [6-8]. Thus, single cell RNA sequencing data evaluation gives the opportunity of furthering our information of cross-species commonalities and variations that might result in the extra powerful use of animal fashions to recognize the reason and development of sicknesses on the human mechanistic stage. Hepatocellular carcinoma (HCC) is the main reason for loss of life global and, like

maximum cancers, is a genetic ailment because of the buildup of genetic and epigenetic molecular changes. The development of hepatic neoplasia is characterized via way of means of growing genetic instability, inclusive of duplication and deletion of components of chromosomes and growing proliferative boom benefit of the affected cells. Molecular cytogenetic techniques, inclusive of Comparative genomic hybridization (CGH) and Spectral karyotyping (SKY) [9-11], have allowed the assessment of chromosomal aberrations in HCC. More recently, Crawley [12] has established the cap potential of comparative genomic single cell RNA sequencing data evaluation (CGMA) to explain the alteration of particular genes collectively with the genetic modifications on the chromosome stage primarily based totally on single cell RNA sequencing data records. Thus, single cell RNA sequencing data evaluation presents an unparalleled possibility to similarly the information of the etiology and development of liver most cancers. Bioinformatics techniques and equipment are crucial to investigate and interpret records from single cell RNA sequencing datas. The essential and pressing mission is to accomplice altered styles of gene expression with the ailment. Interpreting single cell RNA

sequencing data records with-inside the context of signaling and regulatory pathways is a specifically powerful bioinformatics technique to convert records into organic which means and to generate hypotheses for similar studies. Using pathways, ailment mechanisms may be interpreted as disturbances of the elaborate interconnections amongst genes, molecules, and cells. Most stated pathway evaluation of single cell RNA sequencing data records has tested the position of differentially expressed genes in pathways decided on with a priori understanding. Alternatively, enormous pathways may be recognized primarily based totally on statistical evaluation, doubtlessly main to discoveries and a extra entire interpretation of single cell RNA sequencing data records with-inside the context of organic approaches on the mechanistic stage. The number one mechanism for the evaluation of HCC is the management of carcinogenic agents. Several version structures were evolved to recognize the pathogenesis of liver most cancers [13–15]. Additionally, the improvement of transgenic fashions lets in the evaluation of the genetic foundation for the induction and development of HCC [16–19]. The albumin-Simian virus forty (SV40) T antigen transgenic rat carries the mouse albumin-promoter/enhancer related to the coding place of the SV40 big T antigen (SV forty tag). SV40 T antigen inactivates each p53 and Rb, ensuing with-inside the spontaneous improvement of hepatic neoplasms (adenoma and carcinoma) inside 6–nine months. Thus, the Albumin-SV40 T antigen transgenic rat may be used to take a look at liver most cancers improvement and maintenance [20–22]. In this manuscript, we describe a bioinformatics manner in which single cell RNA sequencing data statistics from the SV40 transgenic rat changed into tested for the utility to the examine of HCC in humans. We first used a unique visualization device to analyze liver most cancers through mapping the chromosomal vicinity of differentially expressed genes from the rat version to the chromosomal areas of human orthologs. Then, CGMA evaluation changed into used to narrate gene expression bias styles to cytogenetic aberration profiles on humans. Lastly, a statistical technique changed into used to, pick out numerous pathways worried in human HCC primarily based totally at the rat single cell RNA sequencing data. The pathway evaluation well-known shows that the predicted involvement in apoptosis, molecular cycle, boom and differentiation, genetic stability, and methionine metabolism is vital for most cancers development, upkeep, and progression. The effects imply that the gene expression profiles of the transgenic rat version can be beneficial with-inside the examination of human liver most cancers.

Methods

Single Cell RNA Sequencing Data

The information of the single cell RNA sequencing data experimental technique is mentioned elsewhere [21]. Briefly, RNA samples had been remoted from the rat liver tissues of six samples, controls, adenomas, and carcinomas. The laser seizes micro-dissected samples that had been amplified earlier than single cell RNA sequencing data hybridization. An NCI cRNA array (IncyteGem2) changed into used that incorporates 10238 probes representing 9984 specific genes. Gene expression profiles had been produced for all six samples with dye flip, which ended in a complete of 12 arrays.

The log₂ ratio-primarily based totally suggest worldwide normalization changed into first carried out and the normalized

ratios of the swapped dye labels had been then averaged. An overall of 9150 genes remained for additional evaluation after getting rid of none-hybridized,enes because of low intensity. Significantly differentially expressed genes had been decided by the use of an F-take a look at with $P < 0.05$.

Data evaluation the use of Array Track

Most analyses had been performed with the use of the in-residence software program, Array Track <http://edkb.fda.gov/webstart/arraytrack/>. ArrayTrack is a bioinformatics software program, in which statistics management, evaluation, and interpretation are completely integrated [37]. Array Track includes 3 components: (1) Single cell RNA sequencing dataDB for storing single cell RNA sequencing data statistics; (2) LIBRARY for statistics interpretation that incorporates many kinds of purposeful records approximately genes, proteins, and pathways; and (three) TOOL that offers the capability for single cell RNA sequencing data statistics evaluation. LIBRARY incorporates many sub-libraries and statistics in those sub-libraries are extracted from one-of-a-kind organic databases with-inside the public domain (e.g., NCBI bioinformatics resources) [38]. In this project, records for orthology evaluation, chromosome-primarily based totally evaluation, and pathway evaluation had been retrieved from the LIBRARY. More specifically,

Gene Orthology Analysis

The human and mice orthologs to rats were obtained from Library of the Orthologene in ArrayTrack. ArrayTrack allows the fast matching of a large variety of genres in the course of humans, mice, and rats for gene orthology evaluation. Chromosome-primarily based totally evaluation – The cytogenetic places of genes had been exported without delay from the Gene Library of ArrayTrack. A novel visualization device, the Chromosome Plot, changed into evolved to examine the impact of a gene expression sample on liver most cancers via figuring out the altered cytogenetic areas of every chromosome. Figure 1 suggests a bar chart depiction with the y-axis giving cytogenetic vicinity alongside every rat chromosome represented through 20 vertical bars extending alongside the x-axis. This sort of plot has been used. It depicts rat genes of their cytogenetic places on every chromosome the use of color-coding expression records as purple for up-regulation, inexperienced for down, and gray for unaffected genes (e.g., Figure 1). Thus, the plot presents for a selected species a compact visible show of cytogenetic blocks and/or chromosomes altered. Alternatively, the genes also can be mapped to the chromosomal vicinity of any other species, and color-coded in step with the chromosome of the test species. (e.g., Figures 2 for rat mapping to mouse and Figure 3 for rat mapping to human).

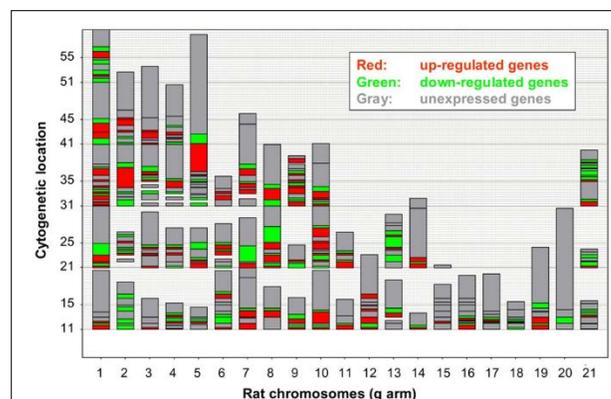


Fig 1

Expressed genes in Single-cell RNA sequencing mapped to the rat cytogenetic region and chromosome. The genes had been acquired from an ANOVA evaluation amongst controls, adenomas, and carcinomas samples of the transgenic rat. The cytogenetic region of genesis at the y-axis for every of the 20 rat chromosomes which can be displayed as separate bars alongside the x-axis. Red and inexperienced regions are the big genes that can be up or down-regulated, respectively, and gray represents the one's genes now no longer differentially expressed.

$$p_i = \frac{m!(N-m)!n_i!(N-n_i)!}{N!m_i!(m-m_i)!(n_i-m_i)!(N-n_i-m+m_i)!}$$

Where N is a complete range of genes at the chip (i.e., 9150), m is the range of differentials expressed genes recognized the use of the F-take a look at (i.e., 2322), n_i is the range of genes out of N that belong to pathway I at the same time as m_i is the range of genes out of M differential expressed genes that are related pathway i. The -sided Fisher's Exact Test p-cost much less than 0.05, shows that the possibility of vast genes on this pathway isn't always predicted through hazard alone.

Comparative Genomic Single cell RNA sequencing data Analysis (CGMA)

CGMA identifies cytogenetic areas containing unidirectional gene expression biases. The biased areas in all likelihood imply chromosomal profits and losses [12]. Of the entire 9255 genes, GenBank accession numbers (Refseq in NCBI) for human orthologs to rodent genes were acquired for 2995 genes out of 3484 genes with Homologene ID using the Orthologene Library and Gene Library in ArrayTrack; Genes and ESTs that may be confined to rodents were excluded.. -tailed z-statistic changed into then computed to check whether or not chromosomal areas exhibited gene-expression biases [12]. CGMA changed into finished for every of adenoma samples and every of the 2 carcinoma samples the use of a web model of the software program at <http://www.vai.org/vari/supplemental/kyle.furge/development/clam-0.2.2.pl> Output human CGMA effects contained 2728 genes from an entry of 2925 Refseq genes. A Z-information of 1. ninety-six corresponds to 95% confidence (that the expression bias with-inside the chromosome changed into now no longer because of hazard) and 2. fifty-eight corresponds to 99% confidence.

Results

An overall of 2322 differentially expressed genes had been diagnosed throughout 3 agencies of samples (i.e., ordinary, adenoma, and carcinoma) primarily based totally on an F-test. The differentially expressed genes had been first mapped to the rat chromosome. As depicted in Figure 1, the differentially expressed genes mainly happened in numerous chromosomes, indicating that those chromosomes had been altered in rats with neoplasm in comparison to ordinary rats. Specifically, a huge range of up-regulated gene mapping with-inside the rat chromosome 1q is constant with preceding findings of excessive amplification in rat liver most cancers [22].

To look at the cross-species extrapolation primarily based totally on the outcomes from the transgenic rat, the differentially expressed genes had been first mapped to the orthologous chromosomal region of the mouse chromosomes. As depicted in Figure 2 the bulk of the differentially expressed genes from the rat 1q which can be acknowledged to be vital for rat liver most cancers improvement seem particularly at the mouse chromosomes 7, 17, and 19 (displayed because the orange band in Figure 2). A contrast of rats to human beings suggests that the differentially expressed genes from rat sat 1q seem mainly on human chromosomes 10, 11, and 19 (Figure 3). The outcomes endorse that the mouse chromosomes (7, 17, and 19) and human chromosomes (10, 11, and 19) are probably vital in the liver most

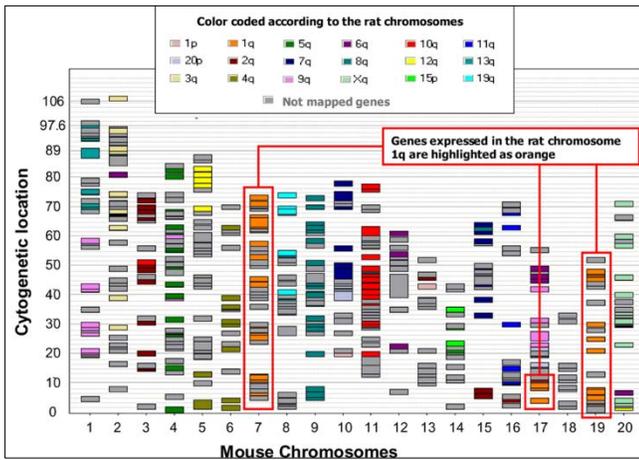


Fig 3

Genes substantially differentially expressed in the rat (proven in Figure 1) mapped to the orthologous genes on chromosomes of mice. Different colorations denote the corresponding rat chromosome range of the orthologous genes

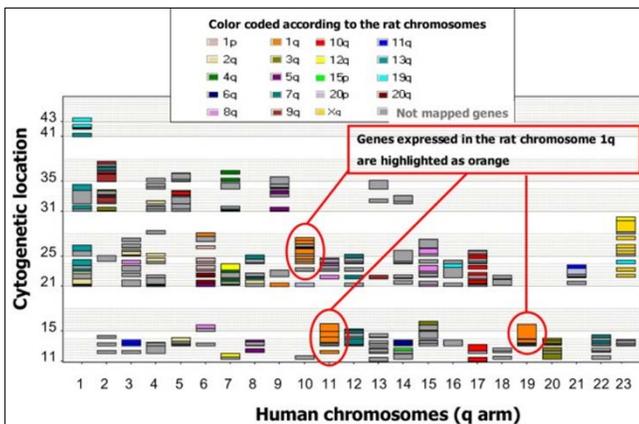


Fig 3

In figure 3 Genes that are significantly differentially expressed in rat (shown in Figure 1) mapped to the orthologous genes on chromosomes of human. Different colors denote the number of the orthologous genes corresponding rat chromosome.

Pathway evaluation – The pathway statistics had been received from the Pathway Library in ArrayTrack. The Pathway Library incorporates pathways from each the Kyoto Encyclopedia of Genes and Genomes (KEGG) [35] and PathArt (Jubilant Biosys Ltd, Columbia, MD 21045) that may be searched one by one or in the mixture in ArrayTrack. The Fisher Exact Test [39] changed into used to estimate the statistical importance of pathway i:

cancers for those species. The findings are supported with the aid of using numerous reports [12, 23, 24].

Table 1 lists the cytogenetic vicinity of the differentially expressed genes from the rat 1q and the vicinity of the orthologous gene in humans and mice. There are seven agencies of substantially expressed genes (known as gene blocks); genes in every organization are consecutive to every different and throughout species. The genes with-inside the identical blocks can be coordinately expressed to carry out comparable transcriptional applications or physiological strategies

throughout species in liver most cancers improvement and maintenance. For example, the human gene blocks 11p15.5, 10q24-26, 11q13.5-15.5, and 19q13.2 have corresponding blocks on rat 1q, and corresponding blocks on mouse chromosomes 7 and 19. These blocks are related to numerous most cancers-associated strategies and functions, inclusive of apoptosis, M phase, molecular communication, and nuclear department as visible in a statistical evaluation primarily based totally on Gene Ontology (outcomes now no longer proven).

Table 1: Seven blocks of the large genes from the rat 1q conserved throughout rat, mouse, and human. The log₂-converted expression of common fold extrude (common over 4 tumor samples) for every gene is given in column with a route up or down indicated through the sign, wherein genes with common fold extrude extra than an arbitrary +1.87 and much less than -1.87 are highlighted.

Gene	Average Fold Change Log ₂	Cytogenetic Location (cM)			Description
		Rat	Mice	Human	
Tctex1	2.07	1q11	173.2		t-complex testis expressed 1
Slc22a1	-0.34	1q11-q12	177.23	6q26	solute carrier family 22, member 1
Ig2fr	2.45	1q11	177.34	6q26	insulin-like growth factor 2 receptor
Psmbl	1.43	1q12	178.32	6q27	proteasome (prosome, macropain) subunit, beta type 1
Calm3	3.34	1q22	74.1	19q13, 2q13	calmodulin 3 carcinoembryonic antigen-related cell adhesion
Ceacam1	-0.41	1q21-q22	75.4	19q13.1	molecule 1
Grik5	1.93	1q21	76.2	19q13.3	glutamate receptor, ionotropic, kainate 5
Zfp36	1.04	1q21	711.1	19q13.2	zinc finger protein 36
Usf2	0.98	1q21	710.2	19q13	transcription factor USF2
Sth2	-1.63	1q21.3-122.1	7 A1		sulfotransferase, hydroxysteroid preferring 2
Rps16	2.56	1q21	7 A3	19q13.1	ribosomal protein S16
Pafah1b3	2.65	1q21	7 A3	19q13.4	ribosomal protein S19
Ech1	1.63	1q21	7 B1	19q13.1	platelet-activating factor acetylhydrolase, isoform 1b, alpha1 subunit
Atf5	0.92	1q21	7 B	19q13.2	enoyl coenzyme A hydratase 1
Myd116	-0.67	1q22	7 B3	19q13.3	activating transcription factor 5
Nkg7	-.87	1q22	7 B3		myeloid differentiation primary response gene 116
Nucb	-.87	1q22	7 B3	19q13.4	natural killer cell group 7 sequence
Pold1	0.97	1q22	7 B3		Nucleobindin
Emp3	2.06	1q22	7 23.1	19q13.3	RNA polymerase delta, catalytic subunit
Tead2		1q22	7 24.4	19q13.2	epithelial membrane protein 3
Bax	0.95	1q22	7 23.1	19q13.4	TEA domain family member 2
Fzd4	2.17	1q31.2	7 23.1	19q13.2	Bcl2-associated X protein
Nox4	1.64	1q32	7 43.8	11q14.3	frizzled homolog 4 (Drosophila)
Rps3	0.54	1q32	7 D3	11q14.1	NADPH oxidase 4
Skd3	-2.42	1q32	7 E1	11q14.2	ribosomal protein S3
Thrsp	1.28	1q32	7 E1	11q13.2	suppressor of K ⁺ transport defect 3
Ccnd1	1.26	1q32	7 E1	11q13.3	thyroid hormone responsive protein
Ppp1ca	1.93	1q42	7 72.4	11q13.4	cyclin D1
Cfl1	2.45	1q43	7 E3-F2	11q13.1	11q13 protein phosphatase 1, catalytic subunit, alpha isoform
Fen1	4.76	1q43	19 A	11q13.2	cofilin 1
Gstp2	1.43	1q43	19 A	11q12	Flap structure-specific endonuclease 1
Ctp1a	-0.42	1q42	19 0.0		glutathione S-transferase, pi 2
Plcb3	1.34	1q43	19 2.0	11q13.3	carnitine palmitoyltransferase 1, liver
Chk	1.07	1q43	19 3.0	11q13	phospholipase C, beta 3
Emk(Mark2)	1.62	1q33	19 3.0	11q13.3	choline kinase/MAP/microtubule affinity-regulating kinase 2 membrane-spanning 4-domains, subfamily A, member
Ms4a2	-0.84	1	19 8.4	11p15.1	2
ILk	1.45	1q33	7 E1	11p15.4	integrin-linked kinase
Arhg	1.65	1q34	7 E2	11p15.3	Ras homolog gene family, member G
Mprl17	0.93	1q41	7 E3	11p15.2	mitochondrial ribosomal protein L17
Arnti	3.57	1q41	7 52.1	11p15	aryl hydrocarbon receptor nuclear translocator-like
Cd81	-0.59	1q41	7 69.2	11p15.1	CD 81 antigen
Tssc3	2.01	1q42	7 68.7	11p15.4	pleckstrin homology-like domain, family A, member 2
Tnni2	0.96	1q36	7 70.1	11p15.5	troponin 1, type 2
Cdkn1c(p57)	4.23	1q36	7 79.45	11p15.6	cyclin-dependent kinase inhibitor 1C, p57
Coro1a	1.11		7 62.2		coronin, actin binding protein 1A

Mapk3	1.82	1q36	7 61.3	11p16.1	protein kinase, mitogen activated 3 (extracellular-signal-regulated kinase 1, ERK1)
Maz	1.98	1q36-q37	7 F3	11p16.4	MYC-associated zinc finger protein (purine-binding transcription factor)
Pp4c	1.24	1q41-q42	7 F3	11p16.2	protein phosphatase 4 (formerly X), catalytic subunit

In addition, to affirm the validity of cross-species extrapolation, we investigated chromosomal aberration in human beings primarily based totally on the differentially expressed genes from the rat version of the usage of CGMA. Table 2 summarizes the Z facts for most cancers patterns from the CGMA evaluation. Chromosomes showing unidirectional bias with a minimum 95% self-assurance have the desk cells with fantastic cost denoting up-law or cells with poor cost denoting down-law. Of forty-six chromosomal areas (23 p and 23 q arms), 15 reveal unidirectional bias in gene expression. Of the 15 affected chromosomal areas,

14 display up-law, and the maximum of those are related to adenoma. The CGMA outcomes had been in addition in comparison with Karyotype outcomes with-inside the Cancer Genome Anatomy Project (CGAP) with-inside the Mitelman Database [25]. Of 15 affected chromosomal areas diagnosed from rat gene expression information, 10 areas also are stated in CGAP. This is proven with-inside the ultimate column of Table 2 that lists each the range of citations and the range of sufferers in CGAP.

Table 2: Summary of the Z information for human chromosomes for every check pattern from the CGMA analysis. A Z-information of 1. Ninety-six corresponds to $\alpha = 0.05$, or 95% self-belief that the expression bias withinside the chromosome became now no longer because of chance. A Z-statistic of 2. Fifty-eight corresponds to 99% self-belief. Chromosomes displaying unidirectional bias with at the least 95% self-belief have tremendous signal denoting up-law or poor signal for denoting down-law.

Chromosome Region	Adenoma sample 1	Adenoma sample 2	Carcinoma sample 1	Carcinoma sample 2	Karyotype results in CGAP # Citations (# Patients)
1.p	2.43	2.67	-1.1	1.56	1(1)
7.p	0.56	2.56	-0.56	1.45	8(34)
9.p	1.56	1.45	-2.34	-2.23	5(32)
11.p	1.57	2.74	0.45	0.67	4(10)
11.q	0.74	2.25	0.67	0.45	4(10)
12.p	1.45	2.45	0.0	-1.45	2(4)
12.q	2.34	2.64	1.95	1.98	2(4)
14.-	1.65	2.43	2.75	2.45	
16.-	2.7	3.34	0.34	0.0	1(1)
17.q	0.67	2.23	0.86	0.0	1(6)
19.p	2.64	3.56	2.37	1.65	
19.q	2.34	2.74	2.24	1.34	
20.q	3.6	0.34	1.78	1.34	2(20)
21.-	2.4	3.23	0.45	1.75	
22.-	1.53	2.24	0.23	1.34	

We additionally investigated which pathways in human beings had been substantially affected primarily based totally on the differentially expressed genes diagnosed with-inside the transgenic rat version. Pathway evaluation is an especially powerful manner to take a look at how the findings with-inside the rat version relate to human beings with-inside the context of organic functions. Table 3 summarizes the outcomes of the pathways evaluation. Fifteen pathways had been substantially altered in a Fisher's Exact Test with $p < 0.05$. They predominately contain the molecular cycle, molecular increase, and

differentiation. Most diagnosed pathways are shown with the aid of using a huge literature to be related to many cancers types [26, 27]. Examples are 1) the p53 pathway worried in reaction to RNA damage, 2) the Rb pathway entails with-inside the manipulate of a molecular cycle, and three) the reworking increase factor-beta (TGF-beta) pathway is worried in increased inhibition. In addition, altered methionine metabolism pathway and law of P27 at some stage in molecular cycle development are acknowledged to be important for most cancers development [28].

Table 3: Pathway analysis of 2322 significant genes using a Fisher's Exact Test identified the listed pathways ($p < 0.05$) that might be related to human liver cancer.

Pathways	The wide variety of sizable genes observed in a pathway divided via way of means of the genes from the chip observed in the same pathway	P value
Apoptosis	18/34	0.00064
TGF Beta Induced apoptosis		
Cell cycle	11/22	0.0078
P53 signaling pathway	6/7	0.0034
Regulation of P27 during cell cycle progression	9/12	
Influence of Ras and Rho Protein on G1 to S transition	6/10	0.0032
ECM and Integrin mediated signaling pathway	10/12	0.015

Rb signaling pathway	13/18	0.00012
Cyclin mediated Pathway	14/20	0.000064
G1-S checkpoint	15/23	0.000072
Growth and differentiation		
Insulin signaling pathway	7/13	0.034
TGF beta signaling pathway		
Amino Acid/Metabolic pathway	4/5	0.021
Methionine Metabolism		
Other Signaling pathways	8/13	0.014
PKC-mediated pathway	13/35	0.0046
Heregulin signaling pathway	9/20	0.028
AR Mediated Signaling Pathway	5/7	0.036
Apigenin Signaling Pathway	6/9	0.0084

Discussion

This observation investigates the results of the usage of single cell RNA sequencing data outcomes from the albumin-SV40 transgenic rat for the observation of human liver most cancers. We validated the significance of bioinformatics to interpret single cell RNA sequencing data information for cross-species comparisons. Specifically, in-residence bioinformatics equipment is of significance for the evaluation, the Chromosome Plot, and ArrayTrack. The Chromosome Plot now no longer simplest offers a visible presentation of the gene expression sample at the extent of gene order throughout chromosomes (e.g., Figure 1), however additionally may be used to map chromosome and cytogenetic region of differentially expressed genes from one species to another (e.g., Figures 2 and 3). ArrayTrack software program that integrates information from public repositories become used to perceive the cross-species orthologous genes, their chromosomal locations, and, maximum importantly, the pathways that can be associated with liver most cancers. In addition, CGMA evaluation becomes done to analyze the range of a couple of chromosome aberration styles primarily based totally on gene expression information, that is in comparison with the outcomes offered in CGAP.

Implications of orthologs and chromosome-primarily based totally evaluation

The lately finished sequencing of the rat genome offers a foundation for destiny studies to clarify how variations and commonalities have an effect on the capacity of rat fashions to expect human ailment. The rat genome challenge stated that the majority of human genes acknowledged as being related to the ailment have orthologous genes with-inside the rat genome and that human, mouse, and rat genomes are about 90% orthologous [1]. We additionally analyzed orthologous genes among human beings, rats, and mice many of the 9150 genes at the chip the usage of ArrayTrack. The chip becomes located in Orthologene Library in ArrayTrack to include 3414 human beings, 3365 mice, and 1950 rat genes, with the relaxation of the genes being both EST tags or Riken genes (approximately 1500). The outcomes confirmed that 92% of human genes are orthologous to both rats or mice. Although a massive variety of genes have been diagnosed to be differentially expressed from the rat model, a number of those genes may also end result from most cancers as opposed to causally related. In addition, the feature of a selected gene and its involvement in most cancers won't be conserved throughout the species. Thus, as essential as structural and purposeful homology of precise genesis, the conservation of feature of blocks of genes is probable to be extra essential in

cross-species comparison. We determined seven awesome blocks of extensively differentially expressed genes inside exclusive cytogenetic areas of the rat 1q with homologous chromosomal segments in human beings and mice (Table 1). However, human beings, mice, and rats have very exclusive chromosomal arrangements. The genes in those blocks seem consecutively in contiguous cytogenetic areas, no matter species and chromosomal location. This locating isn't sudden thinking about the near evolutionary distance among the species in which 278 orthologous segments are stated to be shared among humans and rats, and 280 segments are stated to be shared among humans and mice [1]. It is proposed that those seven blocks of genes can be of importance for liver most cancers improvement, maintenance, and development throughout human beings, rats, and mice rats. For example, genes with-inside the blocks can be coordinately expressed to percentage transcription applications or to reply to the genomic instability determined in liver most cancers. Several genes in Table 1 display massive fold modifications and are implicated in most cancers improvement and maintenance. For example, Rps16, Rps19, and Rps3 code for ribosomal proteins and their altered expression has been related to liver and different tumors [29, 30]. Insulin-like increase factor2 receptor (Igf2r) is mutated in lots of human HCC tumors and the gene's haploid insufficiency has been advised as an early occasion in human hepato-carcinogenesis [31]. Cyclin D1 (Cnd1) and cyclin-based kinase inhibitor 1C (Cdkn1c = p57) are important for the molecular cycle, which includes G1 development and G1/S transition. Cyclin D1 is amplified in 10 to 20% of HCCs.

The implication of CGMA evaluation

Chromosomal aberrations are not unusual to place in cancers, especially in superior stages. CGH has been hired to decide gross RNA profits and losses at chromosomal and sub-chromosomal levels [10]. CGH, however, is time-consuming, and lacks the decision and sensitivity to locate modifications on the gene degree; for example, CGH is not able to locate reproduction variety modifications inside slim areas of chromosomes (alternation of <1 Mb). It fails to discover putative tumor-suppressor genes or oncogenes [32]. These boundaries are probably conquer through the usage of CGMA [12]. CGMA identifies cytogenetic areas containing unidirectional gene expression biases. Such area-based expression extrude can be the end result of allelic imbalances usually determined in the liver and different cancers. Evidence suggests that RNA reproduction variety alterations (deletion, low, mid, and higher-degree amplification) with a mean 2-fold extrude in RNA reproduction variety corresponds to a 1.5-fold extrude in mRNA degree [33].

Therefore, CGMA primarily based totally on single cell RNA sequencing data facts measuring mRNA degree may be associated with the RNA degree. Using CGMA, we diagnosed 15 out of 46 (23 p and 23 q arms) human chromosomal areas that might be concerned in liver most cancers improvement, maintenance, and development. These chromosomal aberrations are steady with the CGAP record for 10 out of 15 chromosome areas through karyotypes (Table 2). Although the CGAP database (Table 2) cites no proof of involvement of chromosome 19 in the human liver in most cancers, we determined that genes in each chromosome 19q and 19p are extensively down-regulated for 3 out of 4 tumor samples. In addition, there's additionally a block of genes in 19q that corresponds to rat 1q (Table 1) at the same time as a massive variety of differentially expressed genes additionally happens in 19p the usage of the Chromosome Plot (Figure isn't shown). Analysis of each human 19p and q shows the viable relevance of the chromosome in the human liver most cancers. The genes extensively altered in rat single cell RNA sequencing data similar to human 19p13 are JunB, Rab8a (Mel), Tnfsf9, and Dnmt1. Further research is needed to verify their relevance to human HCC. Comparing with the findings through Crawley *et al* [12], we anticipated chromosomal profits for five (12q, 16, 17q, 19p, and 20q) out of 8 of these stated through Crawley who done CGMA with human HCC gene expression arrays. Both our evaluation and that of Crawley's advise the significance of 19p in human liver most cancers, an area of aberration now no longer formerly located with CGH evaluation. These outcomes suggest human 19p as an area of aberration now no longer formerly located with CGH evaluation.

Implications of pathway evaluation

Pathways are an excellent car for the translation of the organic capabilities of genes. An essential intention of current biology is to discover the interplay and regulatory networks amongst organic molecules. A logical technique is to research the gene expression modifications with-inside the context of recognized organic pathways [34, 35]. Several human pathways have been determined to be extensively altered the usage of the Fisher's Exact Test through evaluating the variety of genes with altered expression in a pathway to the variety of genes at the chip with-inside the identical pathway. We inferred which pathways are concerned in the human liver most cancers from the differentially expressed genes with-inside the transgenic rat liver most cancers model. It is essential to factor out that the statistically extensive pathways diagnosed on this procedure have been entirely primarily based totally at the evaluation of single cell RNA sequencing data facts collectively with the Orthologene and Pathway Libraries in ArrayTrack, and for that reason required no a priori understanding concerning most cancers genes and the pathways that they control. The outcomes of the pathway evaluation given in Table 3 encompass the ones concerned in apoptosis, molecular cycle, molecular increase, and differentiation, and others which are extensive in liver most cancers. Most of the altered pathways are worried in molecular cycle law. In most cancers, disruption of molecular cycle law is completed through coordinating the hobby of cyclin-based kinases, checkpoint controls, and RNA restores pathways, which, whilst perturbed cause out-of-control molecular increase [27]. Not surprisingly, our research guide that P53 and Rb signaling pathways in addition to cyclin-mediated pathways together with

G1-S checkpoints are altered. Both p53 and Rb are tumor-suppressor genes, and their merchandise is transcription elements that reply to plenty of pressure alerts and are frequently related to the development of neoplastic diseases. The transgenic version means that each P53 and Rb signaling pathway may be disrupted in the human liver most cancers. Without P53 and Rb, molecular cycle arrest and/or programmed molecular death (apoptosis) are inhibited, main to the buildup of mutations and genetic instability. Since P53 is deactivated on this transgenic version, we additionally located that the pathway that impacts Ras and Rho proteins at some stage in G1 to S transition is altered. Ras is a proto-oncogene this is worried in more than one sign transduction pathway transmitting pro-proliferative alerts to the nucleus, even as Rho proteins are individuals of the prolonged Ras own circle of relatives that modulate gene expression, molecular cycle development, and molecular proliferation and survival.

Conclusion

Categories of altered pathways are related to increasing and differentiation. Genes with-inside the ECM (extracellular matrix) and Integrin-mediated signaling pathway were suggested to be over-expressed in human HCC, even though the mechanism isn't absolutely understood [36]. In addition, an extra of TGF-beta is a concept to crush the molecular in ways. Initially, it promotes the blood vessels to overgrow. Second, extra TGF-beta suppresses T cells and different additives of the immune device that could commonly assault aberrant cells. The human-applicable liver most cancers pathways primarily based totally on SV forty transgenic rat liver version are showed through reviews on human liver most cancers models. Therefore, the pathway evaluation of the usage of Fisher's Exact Test is novel and efficient. We provided numerous bioinformatics procedures to extrapolate single cell RNA sequencing data statistics regarding rat liver most cancers to the human. Single cell RNA sequencing data has been broadly used in lots of fields of clinical and organic research. The modern task of bioinformatics of single cell RNA sequencing data is now not to pick out a listing of differentially expressed genes however to expand powerful bioinformatics strategies and equipment for statistics interpretation and understanding discovery. In this examination, we first advanced a Chromosome Plot that gives a compact visible precis of gene expression statistics at the extent of the chromosomal region for the identity of altered chromosomal regions. This device allows cross-species comparison. The statistics to be had in ArrayTrack on gene ontology, gene orthologs, and gene pathways turned into then used to interpret the single cell RNA sequencing data statistics. Finally, the CGMA bioinformatics device turned into used to deduce HCC chromosomal aberrations with-inside the human-primarily based totally on single cell RNA sequencing data statistics from rats. The crucial lesson of this examination is a way to restrict the statistics of the usage of bioinformatics assets and the statistical way to give an unbiased (or statistical) view to interpret single cell RNA sequencing data consequences regarding genes, pathways, chromosomes, and functions. Based on thorough bioinformatics evaluation, we observed that the albumin-SV40 transgenic rat is a beneficial animal version for the prediction of human liver most cancers. The examination demonstrates that the hepatic gene expression profiles from the albumin-SV40 transgenic rat version found out genes, pathways,

and chromosome changes regular with experimental and scientific studies in human liver most cancers. The bioinformatics equipment offered on this paper is crucial for cross-species extrapolation and mapping of single cell RNA sequencing data records, its evaluation, throughout species, and interpretation.

References

- Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*,2004;428(6982):493-521.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al.*: Initial sequencing and analysis of the human genome. *Nature*,2001;409(6822):860-921.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature*,2002;420(6915):520-562.
- Sudheer Menon. "Preparation and computational analysis of Bisulphite sequencing in Germfree Mice" *International Journal for Science and Advance Research In Technology*,2020;6(9):557-565.
- Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. "Computational identification of promoter regions in prokaryotes and Eukaryotes" *EPRA International Journal of Agriculture and Rural Economic Research (ARER)*,2021;9(7)21-28.
- Sudheer Menon. "Bioinformatics approaches to understand gene looping in human genome" *EPRA International Journal of Research & Development (IJRD)*,2021;6(7):170-173.
- Sudheer Menon. "Insilico analysis of terpenoids in *Saccharomyces Cerevisiae*" *international Journal of Engineering Applied Sciences and Technology* ISSN No. 2455-2143,2021;6(1):43-52.
- Tatusov RL, Koonin EV, Lipman DJ: A genomic perspective on protein families. *Science*,1997;278(5338):631-637.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, *et al.* The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*,2003;4(1):41.
- Zhou XJ, Gibson G: Cross-species comparison of genome-wide expression patterns. *Genome Biol*,2004;5(7):232.
- Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J *et al.* Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* 2003, 4(3):223-238.
- Lee YK, El-Nezami H, Haskard CA, Gratz S, Puong KY, Salminen S *et al.* Kinetics of adsorption and desorption of aflatoxin B1 by viable and nonviable bacteria. *J Food Prot*,2003;66(3):426-430.
- Sudheer Menon. "Computational analysis of Histone modification and TFBS that mediates gene looping" *Bioinformatics, Pharmaceutical, and Chemical Sciences (RJBPCS)*,2021;7(3)53-70.
- Sudheer Menon, Shanmughavel piramanayakam, Gopal Prasad Agarwal. "FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes" *EPRA International Journal of Multidisciplinary research*,2021;7(7)620-623.
- Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. Computational Identification of promoter regions in fungal genomes, *International Journal of Advance Research, Ideas and Innovations in Technology*,2021;7(4):908-914.
- Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. Bioinformatics methods for identifying hirschsprung disease genes, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*,2021;9(7):2974-2978.
- Kallioniemi A, Kallioniemi OP, Piper J, Tanner M, Stokke T, Chen L. Detection and mapping of amplified RNA sequences in breast cancer by comparative genomic hybridization. *Proc Natl Acad Sci U S A*,1994;91(6):2156-2160.
- Kallioniemi OP, Kallioniemi A, Piper J, Isola J, Waldman FM, Gray JW, Pinkel D. Optimizing comparative genomic hybridization for analysis of RNA sequence copy number changes in solid tumors. *Genes Chromosomes Cancer*,1994;10(4):231-243.
- Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, *et al.*: Multicolor spectral karyotyping of human chromosomes. *Science*,1996;273(5274):494-497.
- Sudheer Menon, Bioinformatics approaches to understand the role of African genetic diversity in disease, *International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)*,2021;4(8):1707-1713.
- Sudheer Menon. Comparison of High-Throughput Next generation sequencing data processing pipelines, *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS)*,2021;3(8):125-136.
- Sudheer Menon. Evolutionary analysis of SARS-CoV-2 genome and protein insights the origin of the virus, Wuhan, *International Journal of Creative Research Thoughts (IJCRT)*,2021;9(8):b696-b704.
- Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. A step-by-step work flow of Single Cell RNA sequencing data analysis, *International Journal for Scientific Research and Development (IJSRD)*,2021;9(6):1-13.
- Crawley JJ, Furge KA: Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression single cell RNA sequencing data data. *Genome Biol*,2002;3(12):0075.
- Dragan YP, Sargent L, Xu YD, Xu YH, Pitot HC. The initiation-promotion-progression model of rat hepatocarcinogenesis. *Proc Soc Exp Biol Med*,1993;202(1):16-24.
- Pitot HC. Altered hepatic foci: their role in murine hepatocarcinogenesis. *Annu Rev Pharmacol Toxicol*,1990;30:465-500.
- Fausto N. Mouse liver tumorigenesis: models, mechanisms, and relevance to human disease. *Semin Liver Dis*,1999;19(3):243-252.
- Adams JM, Cory S: Transgenic models of tumor development. *Science*,1991;254(5035):1161-1167.
- Hanahan D. Dissecting multistep tumorigenesis in transgenic mice. *Annu Rev Genet*,1988;22:479-519.

30. Merlino G. Transgenic mice as models for tumorigenesis. *Cancer Invest*,1994;12(2):203-213.
31. Sudheer Menon. Computational characterization of Transcription End sites in Human Genome, *International Journal of All Research Education and Scientific Methods (IJRESM)*,2021:9(8):1043-1048.
32. Sudheer Sivasankaran Menon, Shanmughavel Piramanayakam. Insilico prediction of gyr A and gyr B in *Escherichia coli* insights the DNA-Protein interaction in prokaryotes, *International Journal of Multidisciplinary Research and Growth Evaluation, (IJMRD)*,2021:2(4):709-714.
33. Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. Bioinformatics tools and methods to analyze single cell RNA sequencing data, *International Journal of Innovative Science and Research Technology, (IJISRT)*,2021:6(8):282-288.
34. Sudheer Menon. Computational genome analysis for identifying Biliary Atresia genes, *International Journal of Biotechnology and Microbiology, (IJBM)*,2021:3(2):29-33.
35. Grisham JW: Interspecies comparison of liver carcinogenesis: implications for cancer risk assessment. *Carcinogenesis*,1997;18(1):59-81.
36. Dragan YP, Sargent LM, Babcock K, Kinunen N, Pitot HC: Alterations in specific gene expression and focal neoplastic growth during spontaneous hepatocarcinogenesis in albumin-SV40 T antigen transgenic rats. *Mol Carcinog*,2004;40(3):150-159.
37. Yim SH, Ward JM, Dragan Y, Yamada A, Scacheri PC, Kimura S, *et al.* Single cell RNA sequencing data analysis using amplified mRNA from laser capture microdissection of microscopic hepatocellular precancerous lesions and frozen hepatocellular carcinomas reveals unique and consistent gene expression profiles. *Toxicol Pathol*,2003;31(3):295-303.
38. Sargent LM, Dragan YP, Sattler G, Xu YH, Wiley J, Pitot HC: Specific chromosomal changes in albumin simian virus 40 T antigen transgenic rat liver neoplasms. *Cancer Res*,1997;57(16):3451-3456.
39. Sudheer Menon. Recent Insilco advancements in genome analysis and characteristics of SARS-Cov2. *International Journal of Biology Research, (IJBR)*,2021:6(3):50-54.
40. Sudheer Menon. Bioinformatics methods for identifying Human disease genes, *International Journal of Biology Sciences, (IJBR)*,2021:3(2):1-5.
41. Sudheer Menon. SARS-CoV-2 Genome structure and protein interaction map, insights to drug discovery, *International Journal of Recent Scientific Research, (IJRSR)*,2021:12(8):42659-42665.
42. Sudheer Menon (2021) Insilico Insights to Mutational and Evolutionary aspects of SARS-Cov2, *International Journal of Multidisciplinary Research and Development, (IJMRD)*,2021:8(8):167-172.
43. Mohammad Shahid Masroor, Mohammad Salim, Shagufta Parween, Mayuri Singh. Recent trends in the study of Roseoloviruses causing diseases, complications and cancer in human. *Int. J Adv. Biochem. Res.* 2020;4(2):08-10. DOI: 10.33545/26174693.2020.v4.i2a.48
44. Forozan F, Mahlamaki EH, Monni O, Chen Y, Veldman R, Jiang Y, *et al.* Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary RNA single cell RNA sequencing data data. *Cancer Res*,2000;60(16):4519-4525.
45. Saffroy R, Pham P, Lemoine A, Debuire B: Biologie moleculaire et carcinome hepatocellulaire: donnees actuelles et developpements futurs. *Ann Biol Clin (Paris)*, 2004;62(6):649-656.
46. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med*,2004;10(8):789-799.
47. Nesbit MA, Hodges MD, Campbell L, de Meulemeester TM, Alders M, Rodrigues NR, *et al.*: Genomic organization and chromosomal localization of a member of the MAP kinase phosphatase gene family to human chromosome 11p15.5 and a pseudogene to 10q11.2. *Genomics*,1997;42(2):284-294.