Journal of Visualized Experiments

Discovery of driver genes in cancer stem-like tumorspheres using RNAseq with bioinformatics analysis

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61077R2
Full Title:	Discovery of driver genes in cancer stem-like tumorspheres using RNAseq with bioinformatics analysis
Section/Category:	JoVE Cancer Research
Keywords:	Colorectal cancer, Cancer stem cell, CD133, HT29, LGR5, RNAseq, NetworkAnalyst
Corresponding Author:	Chun-Chia Cheng Chang Gung University Taoyuan City, TAIWAN
Corresponding Author's Institution:	Chang Gung University
Corresponding Author E-Mail:	cccheng.biocompare@gmail.com
Order of Authors:	Chun-Chia Cheng
	Bo-Ray Hsu
	Fang-Hsin Chen
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Taoyuan, Taiwan

TITLE:

Discovery of Driver Genes in Colorectal HT29-derived Cancer Stem-Like Tumorspheres

AUTHORS AND AFFILIATIONS:

Chun-Chia Cheng[1], Bo-Ray Hsu[2], Fang-Hsin Chen[1,2]

[1]Radiation Biology Research Center, Institute for Radiological Research, Chang Gung University, Chang Gung Memorial Hospital at Linkou, Taiwan

[2]Department of Medical Imaging and Radiological Sciences, Chang Gung University, Taiwan

Corresponding Author:

Fang-Hsin Chen (fanghsin0808@gmail.com)

Email Addresses of Co-authors:

Chun-Chia Cheng (cccheng.biocompare@gmail.com)

Bo-Ray Hsu (rayhsu@nhri.edu.tw)

KEYWORDS:

colorectal cancer, cancer stem cell, CD133, HT29, LGR5, RNAseq, NetworkAnalyst

SUMMARY:

Presented here is a protocol to discover the overexpressed driver genes maintaining the established cancer stem-like cells derived from colorectal HT29 cells. RNAseq with available bioinformatics was performed to investigate and screen gene expression networks for elucidating a potential mechanism involved in the survival of targeted tumor cells.

ABSTRACT:

Cancer stem cells play a vital role against clinical therapies, contributing to tumor relapse. There are many oncogenes involved in tumorigenesis and the initiation of cancer stemness properties. Since gene expression in the formation of colorectal cancer-derived tumorspheres is unclear, it takes time to discover the mechanisms working on one gene at a time. This study demonstrates a method to quickly discover the driver genes involved in the survival of the colorectal cancer stem-like cells in vitro. Colorectal HT29 cancer cells that express the *LGR5* when cultured as spheroids and accompany an increase *CD133* stemness markers were selected and used in this study. The protocol presented is used to perform RNAseq with available bioinformatics to quickly uncover the overexpressed driver genes in the formation of colorectal HT29-derived stem-like tumorspheres. The methodology can quickly screen and discover potential driver genes in other disease models.

INTRODUCTION:

Colorectal cancer (CRC) is a leading cause of death with high prevalence and mortality worldwide[1,2]. Due to gene mutations and amplifications, cancer cells grow without proliferative control, which contributes to cell survival[3], anti-apoptosis[4], and cancer stemness[5-7]. Within a tumor tissue, tumor heterogeneity allows tumor cells to adapt and survive during therapeutic

treatments[8]. Cancer stem cells (CSCs), with a higher rate of self-renewal and pluripotency than differential cancer types, are principally responsible for tumor recurrence[9,10] and metastatic CRC[11]. CSCs present more drug resistance[12-14] and anti-apoptosis properties[15,16], thus surviving tumor chemotherapies.

Here, in order to investigate the potential mechanism for stemness in the selected CRC stem cells, RNAseq was performed to screen differentially expressed genes in tumor spheroids. The cancer cells can form spheroids (also called tumorspheres) when grown in low adherence conditions and stimulated by growth factors added to the cultured medium, including EGF, bFGF, HGF, and IL6. Therefore, we selected CRC HT29 tumor cells that resist chemotherapies with an increase in phosphorylated STAT3 when treated with oxaliplatin and irinotecon[17]. In addition, HT29 expressed higher stemness markers when cultured in the described culture conditions. The HT29-derived CSC model expressed higher amounts of leucine-rich repeat-containing G-protein-coupled receptor 5 (LGR5)[18], a specific marker of CRC stem cells[19,20]. Moreover, *CD133*, considered a general biomarker for cancer stem cells, is also highly expressed in the HT29 cell line[21]. This protocol's purpose is to discover groups of driver genes in the established cancer stem-like tumorspheres based on bioinformatics datasets as opposed to investigating individual oncogenes[22]. It investigates potential molecular mechanisms through RNAseq analysis followed by available bioinformatics analyses.

Next generation sequencing is a high-throughput, easily available, and reliable DNA sequencing method based on computational help, used to comprehensively screen driver genes for guiding tumor therapies[23]. The technology is also used for detecting gene expression from reverse transcription of an isolated RNA sample[24]. However, when screening with RNAseq, the most important genes to target with therapy may not have the highest expression differential between experimental and control samples. Therefore, some bioinformatics were developed for classifying and identifying genes based on current datasets such as KEGG[25], GO[26,27], or PANTHER[28], including Ingenuity Pathway Analysis (IPA)[29] and NetworkAnalyst[30]. This protocol shows the integration of RNAseq and NetworkAnalyst to quickly discover a group of genes in the selected HT29-derived spheroids compared to parental HT29 cells. Application of this method to other disease models is also suggested for discovering differences in important genes.

Compared to investigation of individual gene expression, a high-throughput technique provides advantages to find potential driver genes easily for tumor precision medicine. With useful datasets such as KEGG, GO, or PANTHER, specific genes can be identified based on the disease models, signaling pathways, or specific functions, and this allows quickly focusing on specific, important genes, saving time and research costs. A similar application is used in previous studies[14,18,31]. Particularly, a tumor is more complicated because different types of tumors express distinguishing genes and pathways for survival and proliferation. Therefore, this protocol can pick up genes distinguishing different tumor types under different circumstances. There is the potential to find effective strategies against cancers by understanding the mechanism of specific gene expression.

PROTOCOL:

1. Cell culture and tumorsphere formation

1.1. Culture HT29 cells in a 10 cm dish containing Dulbecco's modified eagle medium (DMEM) with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin antibiotic (P/S).

1.2. Grow the cells in an incubator at 37 °C with 5% CO_2 and 95% humidity under aseptic conditions, until they reach 80% confluency.

1.3. Trypsinize HT29 cells with 1 mL of 0.25% trypsin for 5 min at 37 °C and consequently neutralize the trypsin by adding 2 mL of DMEM with 10% FBS and 1% P/S.

1.4. Count the HT29 cells using a hemocytometer.

1.5. Add 2,000 cells/well to a low attached 6 well plates with 2 mL of serum-free DMEM with 1% P/S and supplemented with 0.2% B27, 20 ng/mL of epidermal growth factor (EGF), 20 ng/mL of fibroblast growth factor (bFGF), 20 ng/mL of hepatocyte growth factor (HGF), and 20 ng/mL of interleukin 6 (IL6).

1.6. Grow the cells at 37 °C with 5% CO_2 and 95% humidity under aseptic conditions.

1.7. Add 0.5 mL of cancer stem cell medium every 2 days until the tumorspheres measure >100 µm in diameter for at least 7 days.

1.8. Observe and measure the tumorsphere diameter using an inverted microscope with a digital cell imaging system.

1.9. When the tumorspheres reach >100 µm in diameter, trypsinize the tumorspheres with 0.25% trypsin for 5 min at 37 °C and neutralize the trypsin by adding 2x the volume of growth medium. Count the cells using a hemocytometer.

1.10. Centrifuge at 1,200 rpm for 10 min and remove the supernatant.

1.11. Incubate 5 x 10^4 cells with 2 µL of anti-LGR5-PE and 2 µL of anti-CD133-PE in 100 µL of DMEM individually for 30 min at room temperature, shaking at 200 rpm.

NOTE: CD133 is a general cancer stem cell biomarker that is highly expressed in HT29 cells.

1.12. Add 900 µL of PBS and analyze the LGR5 and CD133 expression using flow cytometry. A change in fluorescence in the FL2-H channel indicates gene expression.

2. RNA isolation

NOTE: Use a commercial kit (see **Table of Materials**) with a rapid column for RNA isolation

following the manufacturer's instructions.

2.1. Add 50 µL of PBS to the harvested cells (2 x 10^5 cells) and resuspend them with pipetting.

2.2. Add 200 µL of lysis buffer containing 2 µL of beta-mercaptoethanol (β-ME). Vortex quickly and let stand for 5 min.

2.3. Centrifuge the solution at 16,000 x *g* for 10 min. Collect the supernatant and mix with 200 µL of 70% ethanol.

2.4. Use the attached column to remove the solvent by centrifugation at 14,000 x *g* for 1 min.

2.5. Wash using wash solution 1 and 2 to completely remove non-RNAs by centrifugation at 14,000 x *g* for 1 min.

2.6. Centrifuge again at 14,000 x *g* for 2 min to remove residual ethanol.

2.7. Add 50 µL of distilled water, centrifuge at 14,000 x *g* for 1 min, and collect the solution.

2.8. Measure the RNA concentration using OD_{260} with a spectrophotometer.
RNA concentration (µg/mL) = (OD_{260}) x (40 µg RNA/mL) and OD_{260}/OD_{280} > 2. RNA samples should have an RNA integrity number (RIN) > 7.

3. RNAseq profiling and bioinformatics analysis

NOTE: RNAseq analysis was performed commercially (see **Table of Materials**) to investigate the differential genes in the HT29-derived tumorspheres compared to parental HT29 cells.

3.1. Use commercial services for RNAseq steps, including library construction, library quality control, and DNA sequencing.

3.2. The data report should contain important information, including the read counts, log2 fold change, and p value. Select the differential genes according to the following parameters: genes with a > 1 log2 fold change with read counts >100 in the HT29 tumorsphere group, and genes <-1 log2 fold change with read count >100 in the HT29 parental group. In this case, a p value < 0.05 was considered acceptable and the data were used (**Table 1**).

NOTE: Here, a gene count >100 was used as the threshold to continue the study of a particular gene and validate its expression.

3.3. Use statistical analysis software (see **Table of Materials**), to show a heatmap and identify overexpressed genes >1 and downregulated genes < 1 in log2 fold change.

3.4. Use R software to draw the Volcano Plot with x: log2 fold change; y: -log10 (p value) to

show the differential genes.

3.7.1 Install R library

```
install.packages(library(calibrate))
```

3.7.2 Read the data in RStudio with the following program:

```
res <-read.csv("/Users/xxx.csv", header=T)
head(res)
with(res, plot(log2FoldChange, -log10(pvalue), pch=19, main="HT29CSC vs HT29", xlim=c(-6,6) ,
col="#C0C0C0"))
with(subset(res, pvalue<.05 & log2FoldChange>1), points(log2FoldChange, -log10(pvalue),
pch=19, col="red"))
with(subset(res, pvalue<.05 & log2FoldChange<(-1)), points(log2FoldChange, -log10(pvalue),
pch=19, col="blue"))
with(subset(res, pvalue>.05), points(log2FoldChange, -log10(pvalue), pch=19, col="#444444"))
abline(h=1.3, lty=2)
abline(v=1, lty=2)
abline(v=(-1), lty=2)
```

3.7.3 Execute the run to obtain the Volcano Plot.

4. Driver gene selection

4.1. Select **Single gene Input** in NetworkAnalyst.

4.2. Copy and paste the selected overexpressed genes from **Table 1** with “human” specified as the organism and ID type **Official Gene Symbol**.

NOTE: Alternatively, use **Ensembl Gene ID** for copy and paste.

4.3. Insert the data by clicking **Upload** and **Proceed** to analyze it using protein-protein interaction (PPI) following genetic PPI.

4.4. Use the STRING interactome database with a confidence score cutoff of 900 to show the seed genes cross-linking the uploaded genes. The seed genes associating with more individual genes were selected as driver genes that may be involved in maintaining formation of HT29-derived tumorspheres.

NOTE: There are three interactome datasets for use: IMEx, STRING, and Rolland. STRING contains higher confidence experimental evidence. With lower uploaded gene numbers, IMEx can be selected to predict and pick up the driver genes in the interactome networks.

4.5. Select **Proceed** in the mapping overview.

4.6. Select **White** in the **Background** and **Force Atlas** in the **Layout** knob.

4.7. Select **PANTHER BP** to analyze the upregulation gene group.

NOTE: This shows that HSPA5 was responsible for anti-apoptosis in the HT29-derived tumorspheres in this study (**Figure 3A**). To narrow down the specific functional field, KEGG, GO, or PANTHER classification can be used alternatively to select the specific driver genes.

REPRESENTATIVE RESULTS:

To establish the model for investigating the mechanism in cancer stem cells, colorectal HT29 cells were used to culture the cancer stem-like tumorspheres in vitro in a low-attachment plate containing B27, EGF, bFGF, HGF, and IL6. The tumorspheres >100 µm in diameter were formed in 7 days (**Figure 1A**). The tumorspheres were trypsinized to single cells and analyzed using flow cytometry to detect LGR5 and CD133 expression. LGR5 increased in the HT29-drived tumorspheres from 1.1% to 11.4% and the cells were detected using flow cytometry (**Figure 1B**). Another stemness marker, CD133, also increased from 61.8% to 81.1% in the cultured HT29-derived tumorspheres compared to parental HT29 cells (**Figure 1C**). Then, the tumorspheres were ready for RNAseq investigation.

RNAseq was used to investigate the gene expression profile in the HT29-derived tumorspheres compared to the parental HT29 cells. The error rate in nucleotide reading was 0.03% for both samples. The total gene mapping rate was 87.87% for HT29 and 87.25% for HT29-derived tumorspheres. The fragments per kilobase of transcript per million (FPKM), normalizing the detective counts for indicating the transcript (mRNA) expression, interval between 0 and 1 was 75.87% for HT29 cells, and 77.16% for HT29-derived tumorspheres. The results were suitable for consequent sequencing. After sequencing reads, the genes with log2 fold change >1 in upregulation and <-1 in downregulation with p value < 0.05 shown by the heatmap (**Figure 2A**,**Table 1**,**Table 2**) were selected. There were 79 upregulated genes and 33 downregulated genes selected. In addition, the Volcano plot using log2 fold change and p value (-log10 p value) was used to distinguish the significant genes between HT29-derived tumorspheres and parental HT29 cells (**Figure 2B**). Based on the preliminary selection, three potentially upregulated genes were identified, including *ACSS2*, *HMGCS1*, and *PCSK9*, in the HT29-derived tumorspheres (**Figure 2A**). Furthermore, in order to identify the driver genes not solely according to the log2 fold change, NetworkAnalyst was used. The upregulated genes with log2 fold change >1 with p value <0.05 were analyzed and resulted in 10 seed genes cross-linking the gene networks, including *HSPA5*, *HSP90AA1*, *BRCA1*, *SFN*, *E2F1*, *CYCS*, *CDC6*, *ALYREF*, *SQSTM1*, and *TOMM40* (**Figure 3A**). To identify the function and significance of the seed genes, the classification interface could be used to perform PANTHER BP to determine the genes involved in anti-apoptosis in HT29-derived tumorspheres. The results indicated that *HSPA5* and *SQSTM1* were associated with negative regulation of apoptosis (**Figure 3A**). Moreover, the selected 10 genes were consequently validated; expression increased in the HT29-derived tumorspheres as confirmed using qPCR (**Figure 3B**).

FIGURE LEGENDS:

Figure 1: Establishment of cancer stem-like tumorsphere in vitro. (**A**) HT29 was used to form tumorspheres, and (**B**) LGR5 was detected in the HT29-derived tumorspheres (Scale bar = 100 μm). (**C**) CD133 was used as another marker to identify the stemness characters in the established tumorspheres.

Figure 2: Heatmap and Volcano plot were used to select the differential genes in the HT29-derived tumorspheres. Log2 fold change >1 and <-1 with read count >100, p value < 0.05 were used to exclude nonsignificant genes and make sure the data were accurate.

Figure 3: NetworkAnalyst was used to identify the driver genes in the HT29-derived tumorspheres. (**A**) The upregulated genes were selected and analyzed consequently, and a classification interface, PANTHER BP, provided the potential functions for the differential driver genes. (**B**) Then, qPCR was used to validate the 10 genes upregulated in HT29-derived tumorspheres compared to parental HT29 cells.

Table 1: The upregulated genes in HT29-serived tumorspheres compared to parental HT29 cells analyzed by RNAseq.

Table 2: The downregulated genes in HT29-serived tumorspheres compared to parental HT29 cells analyzed by RNAseq.

DISCUSSION:

In this study, cultured cancer stem-like tumorspheres were used as a model in analyzing RNAseq data with available bioinformatics. For a disease model, HT29-derived tumorspheres were used. Because the tumorspheres have drug resistance against tumor therapies, the established model can be used to investigate the detailed mechanisms of resistance by investigating differences in gene expression. Moreover, genomic technology using RNAseq with available bioinformatics provides rapid understanding of the study model so the genes potentially involved can be validated with higher confidence. Also, the kind of genes involved in the formation of tumorspheres can be identified.

RNA quality is critical for the RNAseq analysis[32]. Ensure that the sample has RIN >7, because it increases certainty between mapping reads, mapping genes, and RPKM. When analyzing RNAseq data, IPA[29] and NetworkAnalyst[30] were available to identify potential genes and signaling pathways. However, it is essential to rule out the unnecessary genes according to the following parameters: genes showing a >1 log2 fold change with read counts >100 in the experimental group, and genes <-1 log2 fold change with read count > 100 in the control group. Higher read counts are easier for consequent validation using qPCR or Western blots.

Based on the understanding of biological functions and processes, there are many bioinformatics tools allowing rapid investigation of the potential mechanisms of diseases of interest. Combined with a high-throughput tool to screen for gene expression such as RNAseq, potential mechanisms regulating the development of the studied diseases can be proposed and investigated. Here, the method for investigating the formation of CRC stem-like tumorspheres derived from HT29

revealed that *SQSTM1* and *HSPA5* were the target genes upregulated and involved in anti-apoptosis in tumorspheres. Therefore, more experiments can be designed to investigate the detailed mechanism of these genes, which results in more confidence and efficacy when conducting research studies.

Here, only the upregulated genes in the tumorspheres were analyzed because upregulation was considered to be induced by the addition of the growth factors. Otherwise, if the experiment used gene knockdown in the tumor cells, the downregulated genes would be considered as targets that can be selected for investigation using the bioinformatics. Although the methodology is rapid and dependable, subsequent validation is still needed via qPCR and Western blots. It is also suggested that more cell lines be used for gene expression validation, especially for clinical samples.

ACKNOWLEDGEMENTS:

The authors thank the Radiation Biology Core Laboratory of Institute for Radiological Research, Chang Gung Memorial Hospital, for technical support. This study was supported by grants from Cheng Hsin General Hospital (CHGH 106-06), and Mackay Memorial Hospital (MMH-CT-10605 and MMH-106-61). Funding bodies did not have any influence in the design of the study and data collection, analysis and interpretation of data or in writing the manuscript.

DISCLOSURES:

The authors have no relevant financial disclosures.

REFERENCES:

1 Rawla, P., Sunkara, T., Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Przegląd Gastroenterologiczny.* **14** (2), 89-103 (2019).

2 Wong, M. C., Ding, H., Wang, J., Chan, P. S., Huang, J. Prevalence and risk factors of colorectal cancer in Asia. *Intestinal Research.* **17** (3), 317-329 (2019).

3 Liu, Q. et al. Positive expression of basic transcription factor 3 predicts poor survival of colorectal cancer patients: possible mechanisms involved. *Cell Death & Disease.* **10** (7), 509 (2019).

4 Slattery, M. L. et al. Dysregulated genes and miRNAs in the apoptosis pathway in colorectal cancer patients. *Apoptosis.* **23** (3-4), 237-250 (2018).

5 Arteaga, C. L., Engelman, J. A. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell.* **25** (3), 282-303 (2014).

6 Yarden, Y., Pines, G. The ERBB network: at last, cancer therapy meets systems biology. *Nature Reviews Cancer.* **12** (8), 553-563 (2012).

7 Cheng, C. C. et al. YM155 as an inhibitor of cancer stemness simultaneously inhibits autophosphorylation of epidermal growth factor receptor and G9a-mediated stemness in lung cancer cells. *PLoS One.* **12** (8), e0182149 (2017).

8 Prasetyanti, P. R., Medema, J. P. Intra-tumor heterogeneity from a cancer stem cell perspective. *Molecular Cancer.* **16** (1), 41 (2017).

9 Zhao, Y. et al. CD133 expression may be useful as a prognostic indicator in colorectal

cancer, a tool for optimizing therapy and supportive evidence for the cancer stem cell hypothesis: a meta-analysis. *Oncotarget.* **7** (9), 10023-10036 (2016).

10 Choi, J. E. et al. Expression of epithelial-mesenchymal transition and cancer stem cell markers in colorectal adenocarcinoma: Clinicopathological significance. *Oncology Reports.* **38** (3), 1695-1705 (2017).

11 Massard, C., Deutsch, E., Soria, J. C. Tumour stem cell-targeted treatment: elimination or differentiation. *Annals of Oncology.* **17** (11), 1620-1624 (2006).

12 Grillet, F. et al. Circulating tumour cells from patients with colorectal cancer have cancer stem cell hallmarks in ex vivo culture. *Gut.* **66** (10), 1802-1810 (2017).

13 Dallas, N. A. et al. Chemoresistant colorectal cancer cells, the cancer stem cell phenotype, and increased sensitivity to insulin-like growth factor-I receptor inhibition. *Cancer Research.* **69** (5), 1951-1957 (2009).

14 Chang, Y. F. et al. STAT3 induces G9a to exacerbate HER3 expression for the survival of epidermal growth factor receptor-tyrosine kinase inhibitors in lung cancers. *BMC Cancer.* **19** (1), 959 (2019).

15 Catalano, V. et al. Colorectal cancer stem cells and cell death. *Cancers (Basel).* **3** (2), 1929-1946 (2011).

16 Piggott, L. et al. Suppression of apoptosis inhibitor c-FLIP selectively eliminates breast cancer stem cell activity in response to the anti-cancer agent, TRAIL. *Breast Cancer Research.* **13** (5), R88 (2011).

17 Chung, S. Y. et al. Two novel SHP-1 agonists, SC-43 and SC-78, are more potent than regorafenib in suppressing the in vitro stemness of human colorectal cancer cells. *Cell Death Discovery.* **4**, 25 (2018).

18 Cheng, C. C. et al. STAT3 exacerbates survival of cancer stem-like tumorspheres in EGFR-positive colorectal cancers: RNAseq analysis and therapeutic screening. *Journal of Biomedical Science.* **25** (1), 60 (2018).

19 Kleist, B., Xu, L., Li, G., Kersten, C. Expression of the adult intestinal stem cell marker Lgr5 in the metastatic cascade of colorectal cancer. *International Journal of Clinical and Experimental Pathology.* **4** (4), 327-335 (2011).

20 Medema, J. P. Targeting the Colorectal Cancer Stem Cell. *New England Journal of Medicine.* **377** (9), 888-890 (2017).

21 Sahlberg, S. H., Spiegelberg, D., Glimelius, B., Stenerlow, B., Nestor, M. Evaluation of cancer stem cell markers CD133, CD44, CD24: association with AKT isoforms and radiation resistance in colon cancer cells. *PLoS One.* **9** (4), e94621 (2014).

22 Xia, J., Gill, E. E., Hancock, R. E. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nature Protocols.* **10** (6), 823-844 (2015).

23 Gagan, J., Van Allen, E. M. Next-generation sequencing to guide cancer therapy. *Genome Medicine.* **7** (1), 80 (2015).

24 Panichnantakul, P., Bourgey, M., Montpetit, A., Bourque, G., Riazalhosseini, Y. RNA-Seq as a Tool to Study the Tumor Microenvironment. *Methods in Molecular Biology.* **1458**, 311-337 (2016).

25 Kanehisa, M., Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science.* **29** (1), 28-35 (2020).

26 Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene

Ontology Consortium. *Nature Genetics.* **25** (1) (2000).
27 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research.* **47** (D1), D330-D338 (2019).
28 Mi, H., Muruganujan, A., Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research.* **41** (Database issue), D377-386 (2013).
29 Yu, F., Shen, X. Y., Fan, L., Yu, Z. C. Genome-wide analysis of genetic variations assisted by Ingenuity Pathway Analysis to comprehensively investigate potential genetic targets associated with the progression of hepatocellular carcinoma. *European Review for Medical and Pharmacological Sciences.* **18** (15), 2102-2108 (2014).
30 Zhou, G. et al. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research.* **47** (W1), W234-W241 (2019).
31 Cheng, C. C. et al. Epidermal growth factor induces STAT1 expression to exacerbate the IFNr-mediated PD-L1 axis in epidermal growth factor receptor-positive cancers. *Molecular Carcinogenesis.* **57** (11), 1588-1598 (2018).
32 Gallego Romero, I., Pai, A. A., Tung, J., Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology.* **12**, 42 (2014).

Figure 1

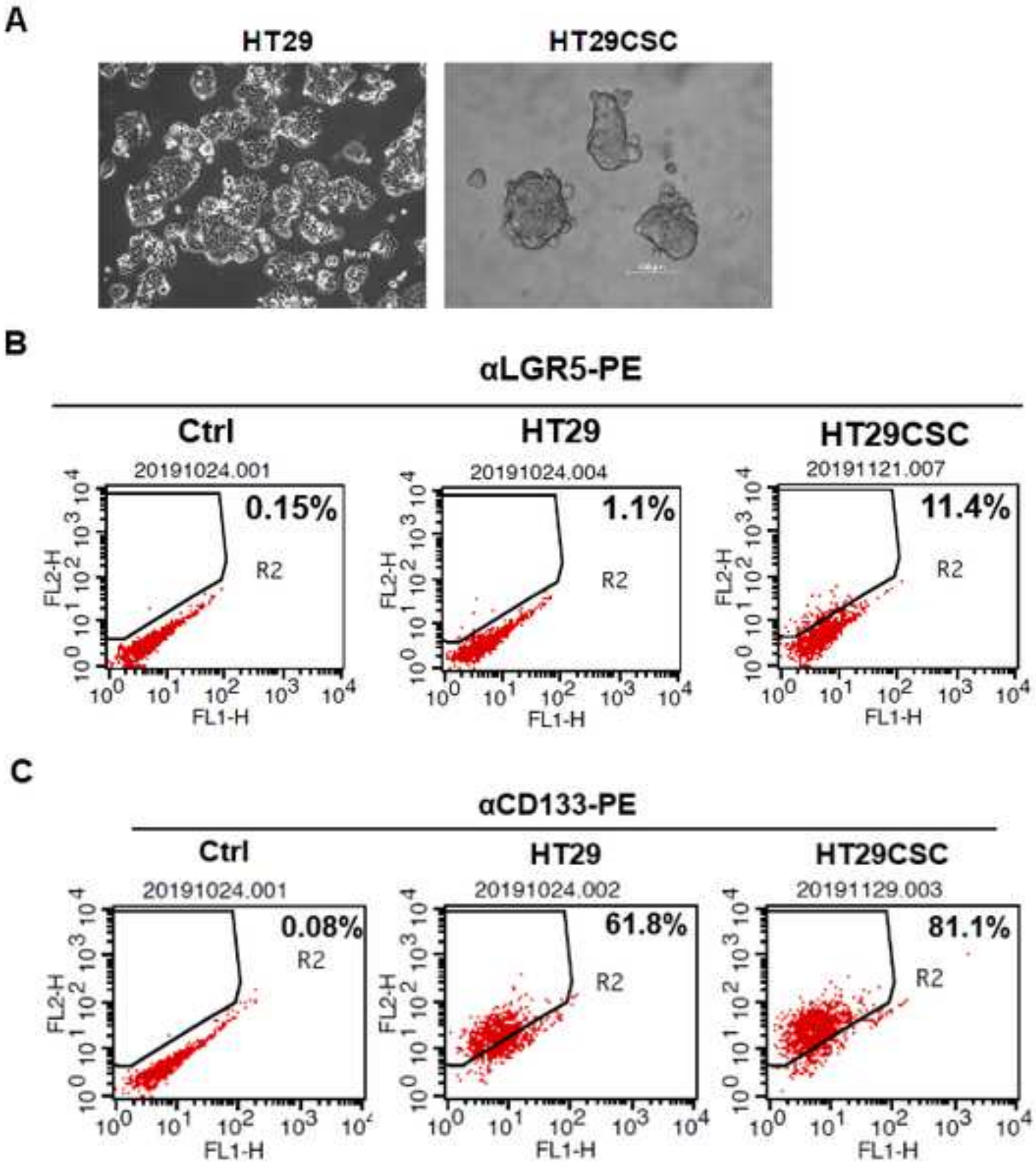


Figure 2

Click here to access/download;Figure;Fig 2.tif

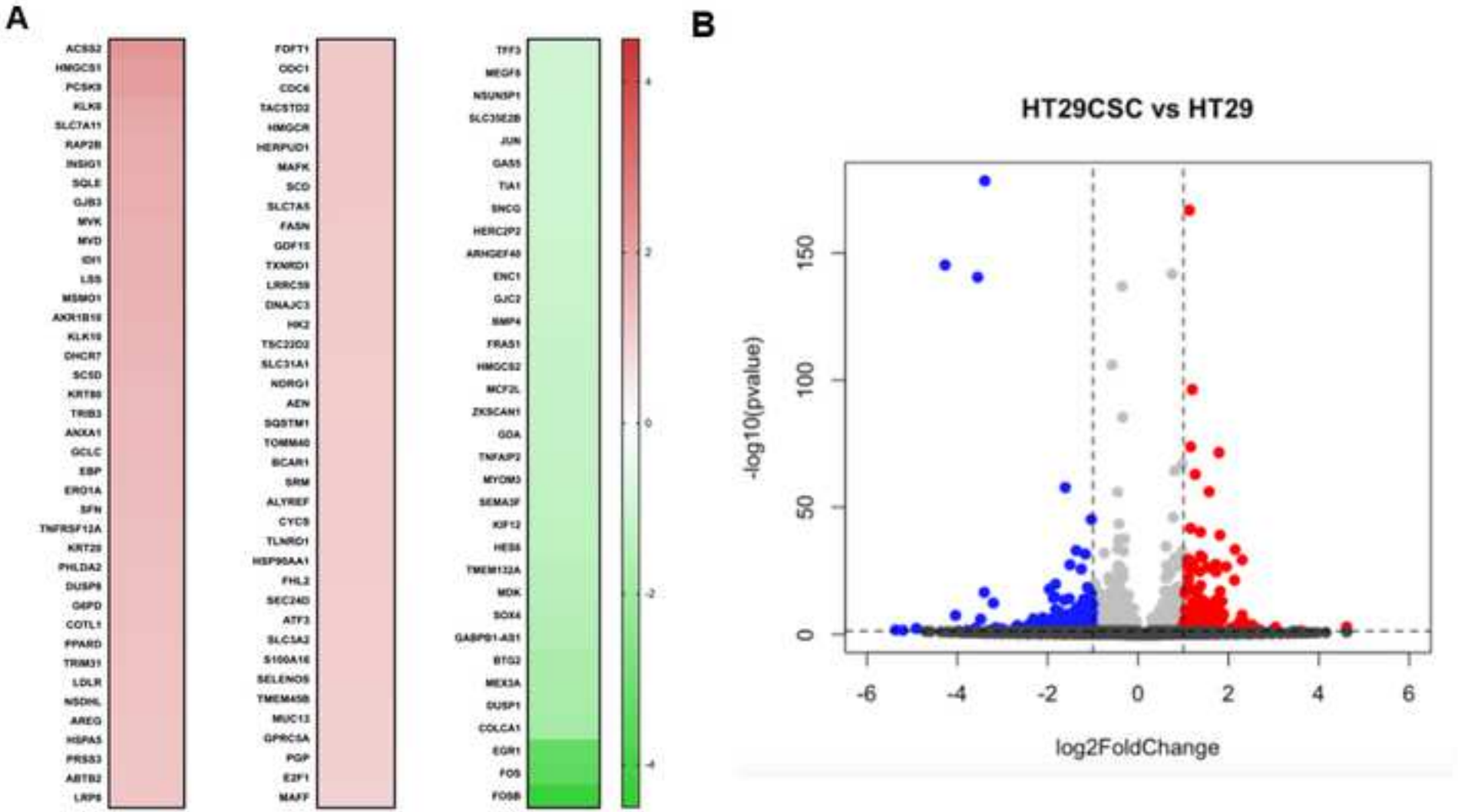
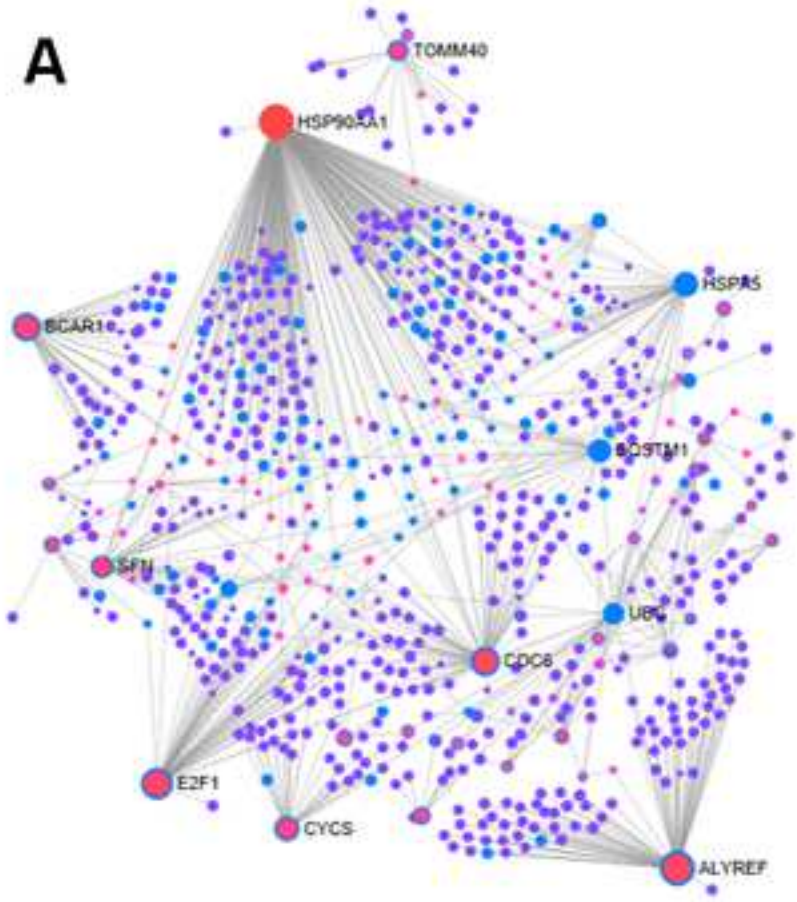


Figure 3



Query: All nodes
Database: PANTHER BP Submit Save

Name	Hits	Pval	AdjP	Col
Protein phosphorylation	106/62	1.32e-30	2.56e-28	
MRNA 3'_end processing	31/57	5.09e-26	4.94e-24	
Negative regulation of apoptos	93/577	3.38e-25	2.19e-23	
Protein folding	47/157	1.13e-24	5.48e-23	
Regulation of cell cycle	44/148	5e-23	1.94e-21	
Cell cycle	92/647	5.79e-21	1.87e-19	
MRNA splicing, via spliceoso	51/236	1.1e-19	3.04e-18	
Viral process	72/448	1.71e-19	4.14e-18	
RNA splicing	55/289	1.86e-18	4.02e-17	
Apoptotic process	88/699	1.27e-16	2.47e-15	
MRNA processing	59/370	5.94e-16	1.05e-14	
DNA replication	36/155	2.87e-15	4.64e-14	
Cholesterol metabolic process	27/89	7.65e-15	1.14e-13	
Steroid metabolic process	28/121	4.06e-12	5.62e-11	
Rhythmic process	26/124	2.49e-10	3.22e-9	

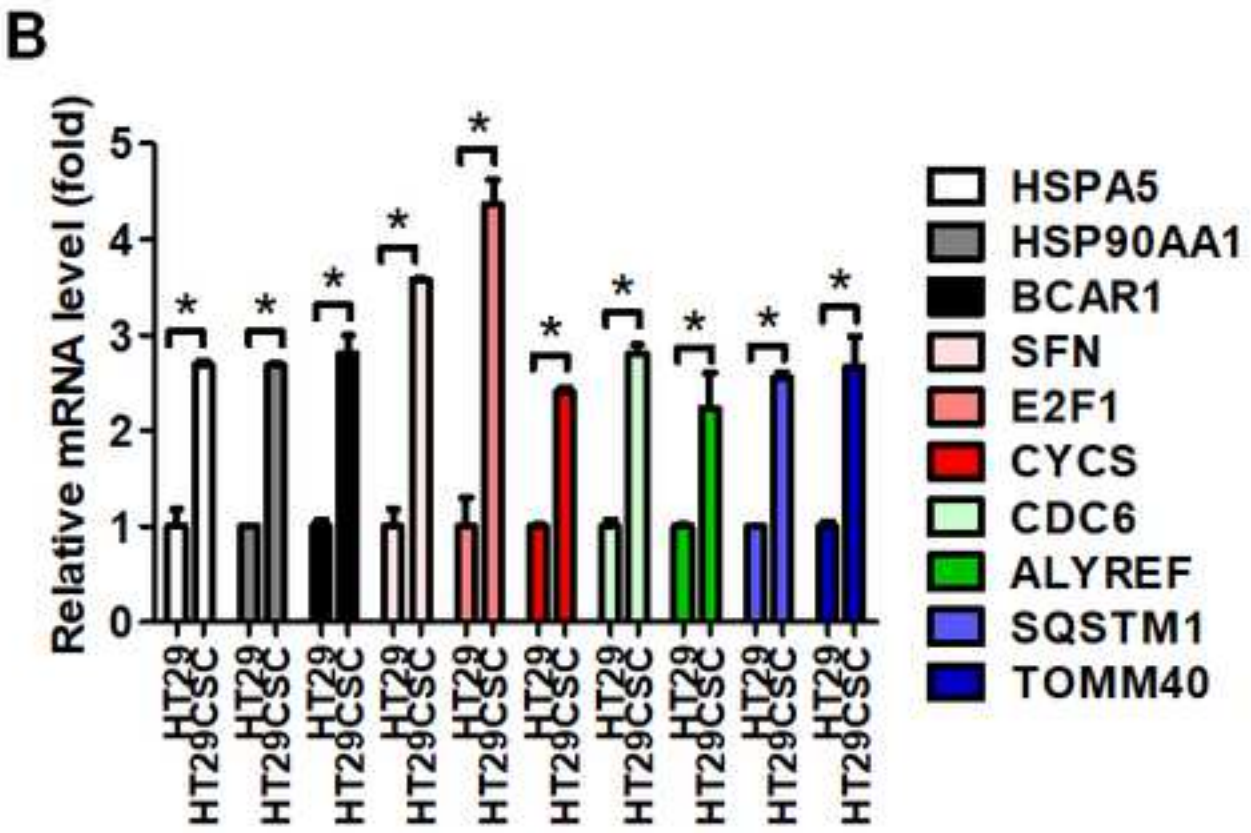


Table 1

Click here to access/download;Table;Table 1. HT29CSCgene.description.up analysis.xlsx

Gene_id	chromosor	Strand	start	end	length	HT29Luc_r	HT29CSC_r
ENSG0000(	20	+	34872146	34927962	3065	471	2086
ENSG0000(	5	-	43289395	43313512	3506	661	2624
ENSG0000(	1	+	55039548	55064852	3891	423	1664
ENSG0000(	19	-	50958631	50969673	1708	687	2382
ENSG0000(	4	-	1.38E+08	1.38E+08	9645	373	1211
ENSG0000(	3	+	1.53E+08	1.53E+08	8351	526	1660
ENSG0000(	7	+	1.55E+08	1.55E+08	3010	1249	3936
ENSG0000(	8	+	1.25E+08	1.25E+08	2961	2399	7445
ENSG0000(	1	+	34781189	34786369	2217	355	1100
ENSG0000(	12	+	1.1E+08	1.1E+08	4109	297	885
ENSG0000(	16	-	88651935	88663161	2352	886	2632
ENSG0000(	10	-	1039908	1049170	2301	478	1412
ENSG0000(	21	-	46188141	46228824	4936	1019	3001
ENSG0000(	4	+	1.65E+08	1.65E+08	2269	344	979
ENSG0000(	7	+	1.35E+08	1.35E+08	1590	1120	3093
ENSG0000(	19	-	51012739	51020175	3111	509	1378
ENSG0000(	11	-	71428193	71452868	2666	2676	7138
ENSG0000(	11	+	1.21E+08	1.21E+08	2456	326	861
ENSG0000(	12	-	52168996	52192000	3894	589	1529
ENSG0000(	20	+	362835	397559	2499	1329	3443
ENSG0000(	9	+	73151757	73170393	2209	405	1044
ENSG0000(	6	-	53497341	53616970	3813	604	1504
ENSG0000(	X	+	48521158	48528716	1775	572	1407
ENSG0000(	14	-	52639916	52695900	5285	760	1853
ENSG0000(	1	+	26863138	26864457	1320	1867	4480
ENSG0000(	16	+	3018445	3022383	1678	1251	2928
ENSG0000(	17	-	40875941	40885227	1737	2693	6297
ENSG0000(	11	-	2928273	2929455	955	2079	4842
ENSG0000(	11	-	1554044	1571920	4480	712	1646
ENSG0000(	X	-	1.55E+08	1.55E+08	2631	1690	3901
ENSG0000(	16	-	84565594	84618077	5330	470	1083
ENSG0000(	6	+	35342558	35428191	3774	408	912
ENSG0000(	6	-	30102897	30113106	2043	1273	2841
ENSG0000(	19	+	11089362	11133816	5337	1157	2579
ENSG0000(	X	+	1.53E+08	1.53E+08	1929	611	1344
ENSG0000(	4	+	74445134	74455009	1240	402	869
ENSG0000(	9	-	1.25E+08	1.25E+08	3908	5393	11617
ENSG0000(	9	+	33750466	33799231	1144	523	1116
ENSG0000(	11	-	34150988	34358008	4902	728	1547
ENSG0000(	1	-	53242364	53328070	7648	468	984
ENSG0000(	8	+	11795573	11839309	2660	2528	5294
ENSG0000(	2	-	10439968	10448504	2653	1569	3273
ENSG0000(	17	+	40287633	40304657	4810	402	837
ENSG0000(	1	-	58575423	58577773	2351	483	1002
ENSG0000(	5	+	75336329	75362104	4585	891	1843

ENSG0000(	16	+	56932048	56944863	2946	666	1376
ENSG0000(	7	+	1530714	1543043	3368	1145	2365
ENSG0000(	10	+	1E+08	1E+08	5362	9587	19715
ENSG0000(	16	-	87830023	87869488	4537	2893	5816
ENSG0000(	17	-	82078333	82098332	8565	7902	15868
ENSG0000(	19	+	18374731	18389176	1200	4429	8890
ENSG0000(	12	+	1.04E+08	1.04E+08	4277	2643	5303
ENSG0000(	17	-	50375059	50397553	2910	978	1962
ENSG0000(	13	+	95677139	95794989	5591	664	1327
ENSG0000(	2	+	74833981	74893359	5772	994	1974
ENSG0000(	3	+	1.5E+08	1.5E+08	11154	706	1399
ENSG0000(	9	+	1.13E+08	1.13E+08	4744	598	1179
ENSG0000(	8	-	1.33E+08	1.33E+08	3983	19151	37748
ENSG0000(	15	+	88621296	88632282	5039	533	1044
ENSG0000(	5	+	1.8E+08	1.8E+08	2986	3385	6615
ENSG0000(	19	+	44890569	44903689	3415	2546	4954
ENSG0000(	16	-	75228187	75268053	4659	2569	4993
ENSG0000(	1	-	11054584	11060024	1266	485	941
ENSG0000(	17	-	81887844	81891586	1088	609	1178
ENSG0000(	7	-	25120091	25125361	4098	2192	4235
ENSG0000(	15	+	81000944	81005788	4845	2496	4822
ENSG0000(	14	-	1.02E+08	1.02E+08	3510	3326	6361
ENSG0000(	2	-	1.05E+08	1.05E+08	4881	529	1008
ENSG0000(	4	-	1.19E+08	1.19E+08	4030	458	863
ENSG0000(	1	+	2.13E+08	2.13E+08	2103	489	919
ENSG0000(	11	+	62856102	62888875	2222	2267	4234
ENSG0000(	1	-	1.54E+08	1.54E+08	1146	1184	2211
ENSG0000(	15	-	1.01E+08	1.01E+08	2444	529	985
ENSG0000(	11	+	1.3E+08	1.3E+08	2234	746	1381
ENSG0000(	3	-	1.25E+08	1.25E+08	2879	1248	2310
ENSG0000(	12	+	12890782	12917937	7362	2208	4068
ENSG0000(	16	-	2211997	2214807	2727	493	904
ENSG0000(	20	-	33675683	33686404	2503	729	1331
ENSG0000(	22	+	38200767	38216511	3284	535	965

HT29Luc_f	HT29CSC_f	HT29Luc_g	HT29CSC_g	HT29CSC_r	HT29Luc_r	log2FoldCh	pval(HT29(
52.63497	238.9589	52.63497	238.9589	251.2988	50.85436	2.305	5.25E-30
50.72119	206.3988	50.72119	206.3988	316.1112	71.36886	2.1471	4.55E-34
12.65916	51.04734	12.65916	51.04734	200.4608	45.67175	2.1339	4.95E-22
62.71271	222.8928	62.71271	222.8928	286.9577	74.1761	1.9518	2.31E-27
8.047651	26.78304	8.047651	26.78304	145.8882	40.2732	1.857	9.52E-14
7.095145	22.95297	7.095145	22.95297	199.9789	56.79277	1.8161	9.96E-18
88.59842	286.2029	88.59842	286.2029	474.1669	134.8558	1.814	9.68E-40
213.7948	680.1223	213.7948	680.1223	896.8933	259.0225	1.7919	2.91E-72
19.92485	63.28703	19.92485	63.28703	132.5161	38.32972	1.7896	5.12E-12
22.39339	68.40094	22.39339	68.40094	106.6153	32.0674	1.7332	1.54E-09
120.6819	367.4936	120.6819	367.4936	317.075	95.66234	1.7288	2.56E-25
45.9817	139.2346	45.9817	139.2346	170.1025	51.61016	1.7207	3.27E-14
52.99437	159.9841	52.99437	159.9841	361.5281	110.0225	1.7163	2.47E-28
29.62543	86.42598	29.62543	86.42598	117.9394	37.14204	1.6669	7.18E-10
134.3589	380.3502	134.3589	380.3502	372.6113	120.9276	1.6235	7.67E-27
33.86685	93.98567	33.86685	93.98567	166.0066	54.95726	1.5949	1.70E-12
286.2673	782.74	286.2673	782.74	859.9093	288.9305	1.5735	7.35E-57
25.20415	68.23591	25.20415	68.23591	103.724	35.19856	1.5592	4.33E-08
17.86436	47.53735	17.86436	47.53735	184.1974	63.59494	1.5343	5.88E-13
100.1377	265.9287	100.1377	265.9287	414.7755	143.4935	1.5313	3.71E-27
41.02648	108.4089	41.02648	108.4089	125.7699	43.72827	1.5241	3.22E-09
42.15484	107.6003	42.15484	107.6003	181.1857	65.21451	1.4742	4.78E-12
70.41885	177.5589	70.41885	177.5589	169.5002	61.75944	1.4566	3.64E-11
61.45773	153.6009	61.45773	153.6009	223.2295	82.05799	1.4438	4.73E-14
159.3252	391.8984	159.3252	391.8984	539.7021	201.5819	1.4208	6.15E-31
161.2352	386.8379	161.2352	386.8379	352.7339	135.0718	1.3849	5.82E-20
296.8243	711.4628	296.8243	711.4628	758.5947	290.766	1.3835	5.64E-41
245.2252	585.4513	245.2252	585.4513	583.312	224.4718	1.3777	1.10E-31
25.13432	59.56243	25.13432	59.56243	198.2923	76.87538	1.367	1.18E-11
150.372	355.8046	150.372	355.8046	469.9504	182.4711	1.3648	1.85E-25
29.2021	68.97632	29.2021	68.97632	130.4682	50.74639	1.3623	4.13E-08
13.90181	31.85381	13.90181	31.85381	109.8679	44.05218	1.3185	9.96E-07
164.0707	375.3437	164.0707	375.3437	342.2531	137.4471	1.3162	6.54E-18
65.46002	149.5717	65.46002	149.5717	310.6901	124.9225	1.3144	2.29E-16
45.3402	102.2344	45.3402	102.2344	161.9106	65.97031	1.2953	4.98E-09
45.92648	101.7682	45.92648	101.7682	104.6878	43.40436	1.2702	3.81E-06
155.4497	343.2486	155.4497	343.2486	1399.491	582.2878	1.2651	1.16E-63
84.646	185.1503	84.646	185.1503	134.4436	56.46885	1.2515	2.36E-07
29.99486	65.3373	29.99486	65.3373	186.3659	78.60292	1.2455	1.37E-09
22.28154	48.02305	22.28154	48.02305	118.5417	50.53045	1.2302	1.76E-06
201.677	432.9309	201.677	432.9309	637.7641	272.9508	1.2244	2.43E-28
122.822	262.6362	122.822	262.6362	394.2958	169.4066	1.2188	5.39E-18
32.41482	69.18285	32.41482	69.18285	100.8327	43.40436	1.2161	1.29E-05
23.14241	49.21352	23.14241	49.21352	120.7102	52.15001	1.2108	1.99E-06
76.20894	161.5879	76.20894	161.5879	222.0248	96.2022	1.2066	1.29E-10

70.97628	150.3186	70.97628	150.3186	165.7656	71.90871	1.2049	2.90E-08
57.12097	120.9418	57.12097	120.9418	284.9097	123.6268	1.2045	3.58E-13
201.4049	424.5603	201.4049	424.5603	2375.051	1035.118	1.1982	4.72E-97
105.1237	216.6367	105.1237	216.6367	700.649	312.3602	1.1655	1.82E-28
547.0959	1126.17	547.0959	1126.17	1911.606	853.1872	1.1639	1.62E-74
723.0538	1487.723	723.0538	1487.723	1070.971	478.2037	1.1632	1.78E-42
148.268	304.9491	148.268	304.9491	638.8483	285.3675	1.1627	5.45E-26
67.75359	139.331	67.75359	139.331	236.3606	105.5957	1.1624	1.44E-10
19.02254	38.96968	19.02254	38.96968	159.8627	71.69277	1.1569	1.52E-07
29.34985	59.74781	29.34985	59.74781	237.8062	107.3232	1.1478	2.06E-10
29.76338	60.45756	29.76338	60.45756	168.5364	76.22755	1.1447	9.42E-08
25.80921	52.16061	25.80921	52.16061	142.0332	64.56668	1.1374	1.11E-06
1682.742	3399.972	1682.742	3399.972	4547.472	2067.753	1.137	1.28E-167
30.94849	62.13956	30.94849	62.13956	125.7699	57.54856	1.1279	5.40E-06
338.0365	677.1582	338.0365	677.1582	796.9039	365.482	1.1246	3.29E-30
196.8398	392.614	196.8398	392.614	596.8045	274.8943	1.1184	8.20E-23
149.7084	298.2628	149.7084	298.2628	601.5028	277.3776	1.1167	6.39E-23
81.66381	162.4175	81.66381	162.4175	113.3615	52.36595	1.1142	1.96E-05
118.4821	234.9287	118.4821	234.9287	141.9127	65.75436	1.1098	1.94E-06
155.0999	307.1707	155.0999	307.1707	510.1871	236.6725	1.1081	2.01E-19
58.03164	114.9219	58.03164	114.9219	580.9026	269.4957	1.108	6.81E-22
299.2485	586.6653	299.2485	586.6653	766.3047	359.1117	1.0935	1.03E-27
36.26871	70.84225	36.26871	70.84225	121.433	57.11668	1.0882	1.54E-05
27.34057	52.80904	27.34057	52.80904	103.9649	49.45074	1.072	8.03E-05
44.56608	85.85516	44.56608	85.85516	110.7112	52.79784	1.0683	4.99E-05
236.6705	453.1049	236.6705	453.1049	510.0667	244.7703	1.0593	5.73E-18
138.7849	265.6652	138.7849	265.6652	266.3574	127.8377	1.0591	4.37E-10
57.74176	110.2112	57.74176	110.2112	118.6622	57.11668	1.0549	3.33E-05
76.04848	144.3114	76.04848	144.3114	166.368	80.5464	1.0465	1.08E-06
147.2059	279.3045	147.2059	279.3045	278.2839	134.7479	1.0463	2.87E-10
109.4724	206.7483	109.4724	206.7483	490.0688	238.4001	1.0396	8.99E-17
30.58054	57.48069	30.58054	57.48069	108.9042	53.22972	1.0328	9.79E-05
32.80807	61.40257	32.80807	61.40257	160.3445	78.71089	1.0265	2.60E-06
29.42645	54.4085	29.42645	54.4085	116.2528	57.76451	1.009	8.28E-05

padj(HT29CS(	significant(	Gene name
2.55E-27	UP	ACSS2
2.99E-31	UP	HMGCS1
1.34E-19	UP	PCSK9
8.86E-25	UP	KLK6
1.46E-11	UP	SLC7A11
2.06E-15	UP	RAP2B
7.75E-37	UP	INSIG1
4.88E-69	UP	SQLE
6.37E-10	UP	GJB3
1.46E-07	UP	MVK
8.43E-23	UP	MVD
5.20E-12	UP	IDI1
1.08E-25	UP	LSS
7.15E-08	UP	MSMO1
2.77E-24	UP	AKR1B10
2.30E-10	UP	KLK10
8.46E-54	UP	DHCR7
3.24E-06	UP	SC5D
8.20E-11	UP	KRT80
1.39E-24	UP	TRIB3
2.95E-07	UP	ANXA1
6.00E-10	UP	GCLC
4.16E-09	UP	EBP
7.38E-12	UP	ERO1A
3.24E-28	UP	SFN
1.45E-17	UP	TNFRSF12A
4.72E-38	UP	KRT20
5.97E-29	UP	PHLDA2
1.41E-09	UP	DUSP8
6.20E-23	UP	G6PD
3.12E-06	UP	COTL1
5.68E-05	UP	PPARD
1.42E-15	UP	TRIM31
4.26E-14	UP	LDLR
4.50E-07	UP	NSDHL
0.00018525	UP	AREG
1.52E-60	UP	HSPA5
1.52E-05	UP	PRSS3
1.30E-07	UP	ABTB2
9.39E-05	UP	LRP8
1.08E-25	UP	FDFT1
1.20E-15	UP	ODC1
0.00054895	UP	CDC6
0.00010511	UP	TACSTD2
1.41E-08	UP	HMGCR

2.26E-06	UP	HERPUD1
5.19E-11	UP	MAFK
1.09E-93	UP	SCD
8.40E-26	UP	SLC7A5
2.99E-71	UP	FASN
1.56E-39	UP	GDF15
1.89E-23	UP	TXNRD1
1.56E-08	UP	LRRC59
1.03E-05	UP	DNAJC3
2.16E-08	UP	HK2
6.62E-06	UP	TSC22D2
6.22E-05	UP	SLC31A1
1.18E-163	UP	NDRG1
0.00025262	UP	AEN
1.64E-27	UP	SQSTM1
2.32E-20	UP	TOMM40
1.87E-20	UP	BCAR1
0.00079299	UP	SRM
0.00010297	UP	ALYREF
4.87E-17	UP	CYCS
1.82E-19	UP	TLNRD1
4.13E-25	UP	HSP90AA1
0.00064338	UP	FHL2
0.0026504	UP	SEC24D
0.0017621	UP	ATF3
1.26E-15	UP	SLC3A2
4.47E-08	UP	S100A16
0.0012666	UP	SELENOS
6.10E-05	UP	TMEM45B
2.99E-08	UP	MUC13
1.74E-14	UP	GPRC5A
0.0031604	UP	PGP
0.00013432	UP	E2F1
0.0027237	UP	MAFF

Table 2

Click here to access/download;Table;Table 2. HT29CSCgene.description.down analysis.xlsx

Gene_id	chromosor	Strand	start	end	length	HT29Luc_r	HT29CSC_r
ENSG0000(	19	+	45467995	45475179	3775	5809	270
ENSG0000(	14	+	75278774	75282230	2158	6258	479
ENSG0000(	5	+	1.38E+08	1.38E+08	3138	8235	704
ENSG0000(	11	-	1.11E+08	1.11E+08	8098	1348	310
ENSG0000(	5	-	1.73E+08	1.73E+08	2019	1137	278
ENSG0000(	1	-	1.56E+08	1.56E+08	6124	1198	300
ENSG0000(	1	+	2.03E+08	2.03E+08	2712	1645	416
ENSG0000(	15	+	50354959	50372202	16182	1271	368
ENSG0000(	6	+	21592768	21598619	5852	5834	1714
ENSG0000(	11	+	46380756	46383837	1339	1453	454
ENSG0000(	11	+	60924463	60937159	3981	2960	934
ENSG0000(	2	-	2.38E+08	2.38E+08	1600	4088	1421
ENSG0000(	9	-	1.14E+08	1.14E+08	2564	1225	427
ENSG0000(	3	+	50155045	50189075	3802	1022	361
ENSG0000(	1	-	24056035	24112175	5804	989	356
ENSG0000(	14	+	1.03E+08	1.03E+08	4683	1030	380
ENSG0000(	9	+	72114595	72257193	5362	1266	475
ENSG0000(	7	+	1E+08	1E+08	9418	3542	1329
ENSG0000(	13	+	1.13E+08	1.13E+08	6580	1644	624
ENSG0000(	1	-	1.2E+08	1.2E+08	2428	1966	766
ENSG0000(	4	+	78057570	78544269	15624	1095	429
ENSG0000(	14	-	53949736	53958761	1917	4859	1939
ENSG0000(	1	+	2.28E+08	2.28E+08	2243	1101	454
ENSG0000(	5	-	74627406	74641424	5657	2955	1219
ENSG0000(	14	+	21070270	21090240	5919	1389	579
ENSG0000(	15	+	22495570	22590815	5746	1093	468
ENSG0000(	10	+	86958618	86963260	794	1141	492
ENSG0000(	2	-	70209444	70248660	4647	968	422
ENSG0000(	1	-	1.74E+08	1.74E+08	1698	2851	1246
ENSG0000(	1	-	58780788	58784327	3540	8354	3658
ENSG0000(	1	-	1659529	1692728	6283	2285	1006
ENSG0000(	7	+	75410322	75416787	4792	995	440
ENSG0000(	19	+	42325609	42378769	11034	1484	661
ENSG0000(	21	-	42311667	42315651	1109	2146	961

HT29Luc_f	HT29CSC_f	HT29Luc_g	HT29CSC_g	HT29CSC_r	HT29Luc_r	log2FoldCh	pval(HT29(
257.3173	12.2599	257.3173	12.2599	32.52669	627.2038	-4.2692	4.46E-146
704.2316	55.2549	704.2316	55.2549	57.70476	675.6828	-3.5496	2.95E-141
295.6138	25.90533	295.6138	25.90533	84.81033	889.1415	-3.3901	3.28E-179
30.57101	7.206708	30.57101	7.206708	37.34546	145.545	-1.9625	1.67E-18
63.43635	15.89929	63.43635	15.89929	33.49044	122.7631	-1.8741	4.55E-15
39.76116	10.20655	39.76116	10.20655	36.14077	129.3493	-1.8396	1.79E-15
92.92975	24.09004	92.92975	24.09004	50.1152	177.6124	-1.8254	1.74E-20
43.50425	12.91187	43.50425	12.91187	44.33267	137.2312	-1.6302	3.21E-14
112.2991	33.82018	112.2991	33.82018	206.4842	629.9031	-1.6091	1.68E-58
202.5668	64.88045	202.5668	64.88045	54.69303	156.8819	-1.5203	9.04E-15
226.2081	73.16759	226.2081	73.16759	112.5183	319.5943	-1.5061	3.97E-28
519.1601	184.9864	519.1601	184.9864	171.1868	441.3856	-1.3665	9.95E-34
88.85434	31.74869	88.85434	31.74869	51.44036	132.2645	-1.3625	3.78E-11
66.97135	24.24938	66.97135	24.24938	43.48939	110.3464	-1.3433	2.21E-09
31.81223	11.73825	31.81223	11.73825	42.88704	106.7834	-1.3161	6.54E-09
71.75319	27.13582	71.75319	27.13582	45.7783	111.2102	-1.2806	6.22E-09
59.69412	22.95868	59.69412	22.95868	57.22288	136.6913	-1.2563	2.05E-10
76.23052	29.31977	76.23052	29.31977	160.1036	382.4334	-1.2562	2.07E-26
95.11519	37.00735	95.11519	37.00735	75.17279	177.5044	-1.2396	7.09E-13
151.5819	60.54077	151.5819	60.54077	92.27942	212.2711	-1.2018	1.57E-14
26.8671	10.78995	26.8671	10.78995	51.6813	118.2283	-1.1939	1.16E-08
478.8668	195.8849	478.8668	195.8849	233.5898	524.6313	-1.1673	2.61E-32
55.29324	23.372	55.29324	23.372	54.69303	118.8761	-1.12	4.49E-08
110.2211	46.60861	110.2211	46.60861	146.852	319.0544	-1.1194	3.27E-19
47.44229	20.27203	47.44229	20.27203	69.75168	149.9718	-1.1044	1.18E-09
48.30191	21.20047	48.30191	21.20047	56.3796	118.0123	-1.0657	1.43E-07
182.31	80.58336	182.31	80.58336	59.27086	123.195	-1.0555	9.34E-08
88.72326	39.64879	88.72326	39.64879	50.83801	104.516	-1.0397	1.14E-06
442.9689	198.4492	442.9689	198.4492	150.1046	307.8254	-1.0361	8.15E-17
265.8308	119.3191	265.8308	119.3191	440.6764	901.9901	-1.0334	5.68E-46
66.18541	29.8696	66.18541	29.8696	121.192	246.7138	-1.0255	1.34E-13
71.52669	32.42299	71.52669	32.42299	53.00646	107.4312	-1.0192	1.16E-06
32.81625	14.98345	32.81625	14.98345	79.63015	160.229	-1.0087	3.75E-09
404.9202	185.8739	404.9202	185.8739	115.7709	231.7059	-1.001	1.79E-12

padj(HT29CS(	significant(	Gene name
2.74E-142	DOWN	FOSB
1.09E-137	DOWN	FOS
6.05E-175	DOWN	EGR1
3.75E-16	DOWN	COLCA1
7.90E-13	DOWN	DUSP1
3.18E-13	DOWN	MEX3A
4.40E-18	DOWN	BTG2
5.14E-12	DOWN	GABPB1-AS1
2.07E-55	DOWN	SOX4
1.53E-12	DOWN	MDK
1.70E-25	DOWN	TMEM132A
6.32E-31	DOWN	HES6
4.27E-09	DOWN	KIF12
2.07E-07	DOWN	SEMA3F
5.77E-07	DOWN	MYOM3
5.54E-07	DOWN	TNFAIP2
2.16E-08	DOWN	GDA
7.32E-24	DOWN	ZKSCAN1
9.83E-11	DOWN	MCF2L
2.58E-12	DOWN	HMGCS2
9.67E-07	DOWN	FRAS1
1.46E-29	DOWN	
3.34E-06	DOWN	GJC2
7.63E-17	DOWN	ENC1
1.13E-07	DOWN	ARHGEF40
9.72E-06	DOWN	HERC2P2
6.59E-06	DOWN	SNCG
6.37E-05	DOWN	TIA1
1.60E-14	DOWN	GAS5
5.50E-43	DOWN	JUN
2.05E-11	DOWN	SLC35E2B
6.44E-05	DOWN	NSUN5P1
3.42E-07	DOWN	MEGF8
2.41E-10	DOWN	TFF3

	Name of Material/Equipment	Company	Catalog Number
A	iRiS Digital Cell Imaging System	Logos Biosystems, Inc	I10999
B	Flow cytometry	BD biosciences	FACSCalibur
C	anti-LGR5-PE	Biolegend	373803
D	anti-CD133-PE	Biolegend	372803
E	EGF	GenScript	Z00333
F	bFGF	GenScript	Z03116
G	HGF	GenScript	Z03229
H	IL6	GenScript	Z03034
I	PureLink RNA extraction kit	Invitrogen	12183025
J	RNAseq performance	Biotools, Taiwan	
K	NetworkAnalyst	Institute of Parasitology, McGill University, Montreal, Quebec, Canada	
L	Prism	GraphPad Software	

Comments/Description
for observing the formation of tumorspheres
for detecting the LGR5 and CD133 in the tumorspheres
LGR5 detection reagent
CD133 detection reagent
for culture of tumorspheres
for culture of tumorspheres
for culture of tumorspheres
for culture of tumorspheres
isolate total RNA for RNAseq analysis
RNAseq analysis is done commerially by Biotools, Ttaiwan
http://www.networkanalyst.ca/
a statistical analysis software

Dear Dr. Editor:

Enclosed please find the revised article entitled "Discovery of Driver Genes in Colorectal HT29-derived Cancer Stem-Like Tumorspheres" by Chun-Chia Cheng, Bo-Ray Hsu, Fang-Hsin Chen. We all have agreed to the submission of this final article, and would like this article considered for publication in *JOVE*. In this study, we describe the methodology for finding the driver genes in the colorectal cancer-derived tumorsphere model through RNAseq and available analytical tools. Here, we demonstrated the methodology, and found HSPA5 and SQSTM1 were the target genes upregulated and involved in anti-apoptosis in tumorspheres. Therefore, we may design more experiments to investigate the detailed mechanism, that results in more confidence and efficacy. We believe that the protocol and finding results will be of interest to the readership of the Journal, including physicians and scientists interested in molecular mechanism of cancer. This manuscript consists of original data and is revised according to the reviewers' comments, and has not been published in any other journal, nor is it under consideration for publication elsewhere. The responses to the comments are followed behind this rebuttal letter. Thank you very much for consideration.

Yours sincerely,

Chun-Chia Cheng, Radiation Biology Research Center, Institute for Radiological Research, Chang Gung University / Chang Gung Memorial Hospital at Linkou, Taiwan.
Telephone: +886 958 766112. E-mail: cccheng.biocompare@gmail.com.

1. **Editorial comments:**

The editor has formatted the manuscript to match the journal's style. Please retain and use the attached version for revision.
Response: Many thanks. I used the attached file for revising.

2. Please address all the specific comments marked in the manuscript.
Response: We appreciate and revise as possible as we can.

3. The manuscript needs thorough copyediting. Please employ professional copyediting services.
Response: We already look for help and the manuscript is revised and checked carefully for the English part.

4. Please ensure that the protocol section is a step by step description of how the actions are performed for your experiment.

Response: Okay. Make sure.
5. Please reword lines 94-95, 100-101, 103, 110, 162-167, 174-175 as it matches with previously published literature
Response: Okay, we have revised the part.

6. Once done, please ensure that the highlight is no more than 2.75 pages including headings and spacings.
Response: Done.

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

The authors describe the generation of colon cancer stem cell enriched spheroids and then use these spheroids to analyze gene expression through RNAseq. While this is of interest for the wide cancer research community, I do not consider great value in the first part of the manuscript, where spheroid formation is described. Spheroid formation is a well established procedure, specially from conventional cell lines, that can even be commercially purchased nowadays. I see, however, the value in describing the bioinformatic process for data analysis. This is where I see the value of the manuscript.

Minor Concerns:

I would add some data about the yield of RNA and sequencing throughout from the spheroids.

Response: We appreciate the comment. We have shown the results in the second paragraph in the Perspective Results.

Reviewer #2:

Major Concerns:

LGR5 antibody has not been validated. Our lab has bought more than 10 of them and we never found neither LGR5 expression nor CSC enrichment. In addition LGR5 is not a consensus for CSC marker like in healthy tissue in which there is no doubt. Why in the list of genes LGR5 do not appear?

Response: We appreciate the reviewer's comment. LGR5 antibody is bought from Biolegend (Cat. 373803) which is labeled with PE for flow cytometry analysis. We actually found the fluorescent signal (anti-LGR5-PE) increased in the cultured HT29-derived spheroids from 1.1% to 11.4% (Figure 1B). Moreover, CD133 also increased from 61.8% to 81.1%. The results indicated the tumorspheres may be more like cancer stem cells that was used for RNAseq analysis consequently. On the other hand, we have tested another colorectal HCT116 cell line (97% CD133-positive) but which had no increase (data not shown).

LGR5 gene expressed very low level in the tumor cells, 1.3 readcount in HT29 and 7.5 readcounts (without significant) in HT29CSC. The readcount is quite low and without significant, therefore, LGR5 data is excluded in the list when we picked up according to the parameters: genes showed a > 1 log2 fold change with read counts > 100 in HT29 tumorsphere group, and genes <-1 log2 fold change with read count > 100 in HT29 parental group.

Reviewer #3:

Manuscript Summary:

In this manuscript the authors describe the protocol for culturing tumorspheres from HT29 colorectal cancer cells. They used RNASeq to study the genes upregulated and downregulated in the LGR5-overexpressing cells and performed bioinformatics analyses on the RNASeq data.

Major Concerns:

1. It is not clear what the control cells are for comparison in RNASeq. Is it HT29 cells before tumorsphere culturing or LGR5-low expressing cells?

Response: We appreciate the reviewer's comment. We have revised in the method (RNAseq profiling and bioinformatics analysis) of the manuscript "compared to the parental HT29 cells".

2. It is not clear if replicates were performed for the RNASeq experiment. Usually at least 3 replicates need to be performed.

Response: We appreciate the reviewer's constructive comment. This manuscript

majorly introduces the methodology how to pick up the driver genes from the RNAseq data. Therefore, we only performed 1 individual sample. But we have validated the results using 3 replicates in qPCR (Figure 3B).

3. A fold change of >1 and <-1 were used as cut-off. Usually for RNASeq experiments a fold change of >1.2 and <-1.2 is used.

Response: We appreciate the reviewer's comment. Here is log2 fold change of >1 and <-1. It means fold change actually >2 or <-2.

4. A p-value of 0.05 was used. For multiple analyses, Bonferroni correction should be considered for p-value.

Response: We appreciate the reviewer's comment. The p value in the RNAseq analysis is done by Biotools, Taiwan, for technical validation. The p value in the qPCR we used non-paired t test with Bonferroni correction using Parism software.

5. As the authors wrote in the manuscript, there was no validation of the genes found by RNASeq by PCR or Western blot

Response: We appreciate the reviewer's comment. We have added and validated the results using qPCR (Figure 3B).

Minor Concerns:

1. There was little discussion about the genes identified. The top 3 genes identified did not feature in the network analysis in Figure 3.

Response: We appreciate the reviewer's comment. The genes were selected using NetworkAnalyst based on the STRING interactome database.