A Metabolic Signature of Colon Cancer Initiating Cells

Kai-Yuan Chen, Xiaojing Liu, Pengcheng Bu, Chieh-Sheng Lin, Nikolai Rakhilin, Jason W. Locasale, Xiling Shen

Abstract— **Colon cancer initiating cells (CCICs) are more tumorigenic and metastatic than the majority of colorectal cancer (CRC) cells. CCICs have also been associated with stem cell-like properties. However, there is a lack of system-level understanding of what mechanisms distinguish CCICs from common CRC cells. We compared the transcriptomes of CD133+ CCICs and CD133- CRC cells from multiple sources, which identified a distinct metabolic signature for CD133^{high} CCICs. High-resolution unbiased metabolomics was then performed to validate this CCIC metabolic signature. Specifically, levels of enzymes and metabolites involved in glycolysis, the citric acid (TCA) cycle, and cysteine and methionine metabolism are altered in CCICs. Analyses of the alterations further suggest an epigenetic link. This metabolic signature provides mechanistic insights into CCIC phenotypes and may serve as potential biomarkers and therapeutic targets for future CRC treatment.**

I. Introduction

A rare population isolated from primary CRC tumors, CCICs play important roles in CRC tumorigenesis [1][2]. CCICs also possess certain stem cell-like traits, including self-renewal, differentiation, and asymmetric division[3]. CCICs were originally identified by the marker CD133[1][2][4]. Since then, they have also been associated with other markers such as CD44, ALDH1, and Lgr5[3][5].

However, it has remained largely unclear whether CCICs isolated from different CRC tumors indeed share common mechanisms that account for their phenotype, or alternatively they are completely different cells that were categorized simply by their tumorigenic capacity.

To address this question, we first analyzed 5 GEO microarray datasets that measured the transcriptomes of CD133+ versus CD133- CRC cells[6]. The transcriptome analysis suggested that CD133+ cells consistently regulate certain metabolic enzymes differentially from CD133- cells. Unbiased metabolomics by high-resolution mass spectrometry further corroborated the metabolic signature of CD133+ CCICs, which involve the glycolysis, TCA cycle, and cysteine/methionine metabolism pathways.

Resrach supported by NIH NIGMS R01GM95990 and NSF 1137269.

K-Y. Chen, Author is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (phone: 607-793-0198; e-mail: kc632@ cornell.edu).

X.J. Liu, Author is with the Division of Nutrition Sciences. Cornell University, Ithaca, NY 14853 USA (e-mail: xl296@cornell.edu).

P. Bu, Author is with the Department of Biomedical Engineering, Cornell University, Ithaca, NY 14850 USA (e-mail: pb345@cornell.edu)

C-S. Lin, Author is with Department of Biomedical Engineering, Cornell University, Ithaca, NY 14853 USA. Cornell University, Ithaca, NY 14850 USA (e-mail: cl979@cornell.edu).

N. Rakhilin, Author is with the School of Electrical and Computer Engineering Cornell University, Ithaca, NY 14850 USA. Cornell University, Ithaca, NY 14853 USA (e-mail: nr233@cornell.edu).

J.W. Locasale. Author is with the Division of Nutrition Sciences. Cornell University, Ithaca, NY 14853 USA (e-mail: locasale@cornell.edu).

X. Shen. Author is with the School of Electrical and Computer Engineering and the Department of Biomedical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: xs66@cornell.edu).

II. Material and Methods

A. Microarray Data From GEO

Five sets of microarray data from GEO were analyzed. These data sets include 28 FACS (Fluorescence-Activated Cell Sorting) sorted CD133+ vs. CD133- pairs from 3 CRC patient tumors and 4 CRC cell lines (Table I).

B. Statistical Analysis

An R-package, *Bioconductor*[7], was used to extract pre-analyzed GEO data and for post-processing. Genes that were significantly (p-value <0.05) up-regulated or down-regulated in CD133+ vs. CD133- cells were identified by differential analysis using t-test and fold-change analysis. Gene frequencies and Venn diagrams were further generated by the R-packages *limma*[8] and *vennDiagram*[9] to integrate analytical results from the 5 GEO datasets.

C. Metabolomics Data

We isolated CD133+ and CD133- populations from patient-derived CRC lines we have previously described. 6 samples were collected and FACS sorted using CD133 antibodies, and their metabolites levels were detected and measured by a high-resolution qExactive liquid chromatography–mass spectrometer (LC-MS).

TABLE I. GEO Microarray Data on CRCs.

Microarray Data List			
GEO accession number	***Bio marker***	***Cell Type***	***Sample Number***
GSE11757[12]	CD133	CACO-2	3 x CD133+ 3 x CD133-
GSE23295	CD133	SW620	2 x CD133+ 2 x CD133-
GSE24747	CD133	CACO-2	3 x CD133+ 3 x CD133-
GSE34053[13]	CD133	Patient specimen	3 x CD133+ 3 x CD133-
GSE38049	CD133	HCT116	3 x CD133+ 3 x CD133-

D. Pathway Analysis

We performed pathway analyses on the CCIC-regulated genes identified from the GEO datasets using Gene Set Enrichment Analysis (GSEA)[10] and the Kyoto Encyclopedia of Genes Genomes (KEGG)[11]. The metabolomics and transcriptomics data were then integrated by using the KEGG Pathway Online Module.

III. RESULTS

A. Differential Transcriptional Profiling in CD133+ Cells

Transcriptional profiling of the five CRC sets showed that CD133+ and CD133- cells have on average 3178 significant-differential (P-value < 0.05) genes, of which 1628 genes are upregulated and 1550 genes are downregulated in CD133+ cells (Figure 1, Table II). To compare the gene expression profiling of the 5 sets of microarray data from 4 different platforms, we converted the probe IDs of each platform into unified Entrez format. A total of 5521 genes were identified as upregulated in CD133+ cells and 5527 genes as downregulated in CD133+ cells. Based on the frequencies of these identified genes across the 5 data sets, 946 genes were upregulated in more than one CRCs set and 718 genes were downregulated in more than one CRCs set (darker regions in Figure 2). However, only one gene (CD133) was consistently upregulated in all 5 datasets, and no gene was consistently downregulated in all 5 datasets.

Our differential analysis suggests that the transcriptional landscapes are consistently altered in CD133+ CCICs vs. CD133- CRC cells, and many genes are involved. However, CD133+ CCICs from different CRC tumor origins involve diverse genes, reflecting their various genetic backgrounds.

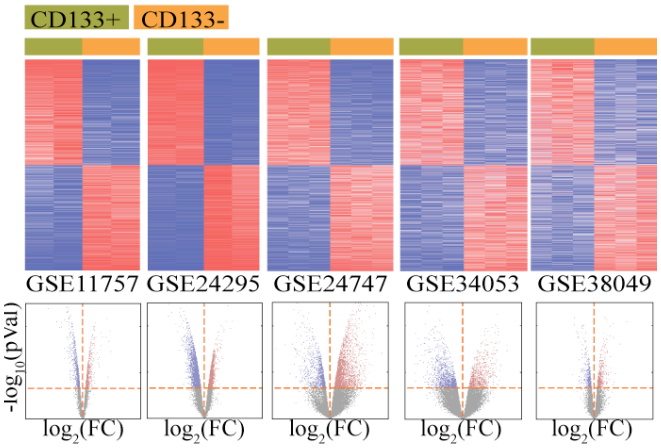

Figure 1. Differentially regulated genes in the 5 GEO datasets. Upper panels: Heatmaps of microarray datasets. Each row is one gene, and each colum is a sample. Red and blue colors represent high and low expression levels respectively. Lower panels: Volcano plots of differentially regulated genes calculated by p-values and fold changes, Each dot represents a gene. The red dots represent significantly upregulated genes, and the blue dots represent significantly downregulated genes.

TABLE II. SUMMARY OF DIFFERENTIAL ANALYSIS

GEO accession number	*Total Significant-Differential Genes*	*Significant-Differential Upregulated Genes*	*Significant-Differential Downregulated Genes*
GSE11757	1676	861	815
GSE23295	6935	3089	3846
GSE24747	4011	2723	1288
GSE34053	2230	954	1276
GSE38049	1038	512	526
AVERAGE	3178	1628	1550

B. Pathway Enrichment Analysis

To find out what pathways these genes are involved with, we performed GSEA and KEGG pathway analyses on the identified list of differential genes. Pathway analyses showed that the curated gene sets of METABOLIC PATHWAYS, PATHWAY IN CANCER, and TRANSCRIPTIONAL MISREGULATION IN CANCER are most highly enriched. KEGG identified 282 enriched pathways, among which metabolic pathways scored highest gene hit rates – 50 curated metabolic pathways involve 614 differentially regulated metabolic genes. The metabolic alterations in CD133+ CCICs are visualized in the global metabolic map shown in Figure 3, wherein red lines represent upregulated enzymatic reactions and blue lines represent downregulated reactions. Darker color indicates a higher frequency of the metabolic gene across the 5 datasets. The global analysis suggests that metabolic pathways are highly reprogramed in CD133+ CCICs.

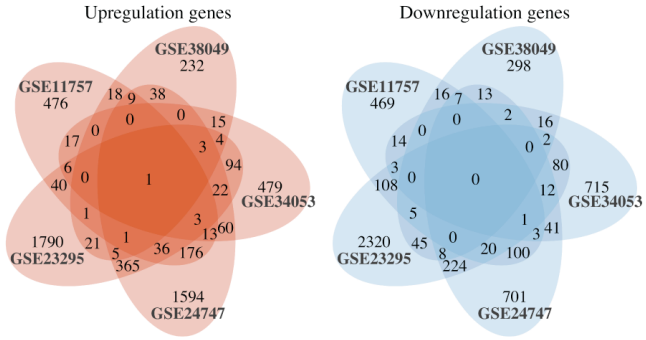

Figure 2. Venn Diagrams of differentially regulated genes that are statistically significant. Each circle represents a GEO dataset. The number in each intersecting region represents the number of overlapping genes.

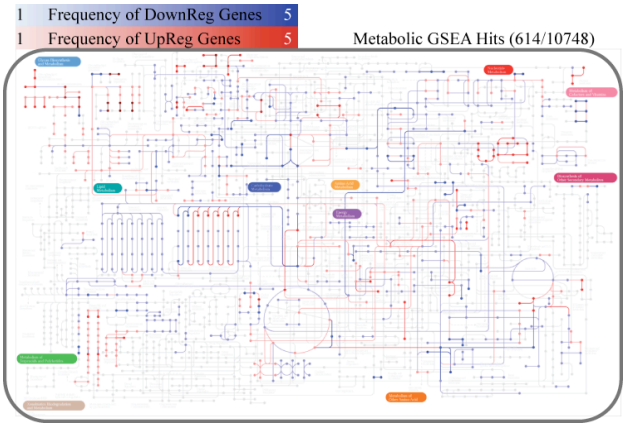

Figure 3. A global metabolic map showing metabolic alterations in CD133+ CCICs. Nodes represent metabolites, and lines represent enzymatic reactions. Red color refers to significantly upregulated enzymetic genes, and blue color refers to significantly downregulated enzymetic genes. Color saturation reprents the frequency of the gene across GEO datasets, from 5 (darkest) to 1 (lightest).

C. Metabolomic Profiling

The gene and pathway enrichment analyses suggest that metabolic reprogramming may be a signature of CD133+

CCICs. To validate this hypothesis, we FACS sorted pure CD133+ and CD133- populations from patient-derived CRC lines we have previously published [3] and performed unbiased, high-resolution metabolomics to measure metabolite levels in triplicate samples using LC-MS. Differential analysis identified 54 metabolites that were differentially expressed in a statistically significant way (P-value < 0.05), among which 28 metabolites were upregulated in CD133+ CCICs and 26 metabolites were downregulated in CD133+ CCICs. The list of differential expressed metabolites is summarized in Table III.

TABLE III. LIST OF SIGNIFICANTLY-DIFFERENTIAL METABOLITES

54 differentially expressed metabolites (p-value < 0.05)	
Upregulated Metabolites (28 metabolites)	Downregulated Metabolites (26 metabolites)
fructose-16-bisphosphate	serine
inosine	adenine
Phosphorylcholine	betaine aldehyde
UTP	2-keto-isovalerate
CDP-ethanolamine	sarcosine
malate	2-hydroxygluterate
ATP	Atrolactic acid
dGTP	Phenyllactic acid
S-methyl-5-thioadenosine	hydroxyphenylpyruvate
orotate	aspartate
uridine	Pyroglutamic acid
Maleic acid	histidine
octulose-monophosphate (O8P-O1P)	folate
NAD+	2-Isopropylmalic acid
glutathione disulfide	glycolate
4-aminobutyrate	phenylpyruvate
dimethylglycine	Hydroxyisocaproic acid
S-adenosyl-L-methionine	glucono-D-lactone
UDP-N-acetyl-glucosamine	Pyridoxamine
citrulline	Ascorbic acid
aconitate	N-acetyl-glutamine
ADP	acetoacetate
dGDP	sn-glycerol-3-phosphate
citrate-isocitrate	Pyrophosphate
betaine	6-phospho-D-glucono-15-lactone
3-phosphoglycerate	2,3-dihydroxybenzoic acid
alanine	
a-ketoglutarate	

D. Integrated Transcriptomic and Metabolomic Network Analysis

Above transcriptomic and metabolomic analyses suggest that metabolic pathways are altered in CD133+ CCICs across different CRC backgrounds. To investigate the connections between CCIC transcriptome and metabolome, we performed integrated KEGG pathway analysis on the list of differential genes and metabolites. The analysis identified carbohydrate metabolism (glycolysis, the TCA cycle) and cysteine and methionine metabolism as consistently altered in CCICs, with highly altered enzyme and metabolite levels (Figures 4, 5).

In most cells, glycolysis converts glucose into pyruvate, which then enters the TCA cycle. Compared to CD133- CRC cells, major glycolysis enzymes PGK1, BPGM, and ENO1 are significantly downregulated in CD133+ CCICs, indicating suppressed upstream activities (Figure 4). Moreover, ACSS1 (Acyl-CoA Synthetase Short-chain family member 1), a mitochondrial acetyl-CoA synthetase enzyme that catalyzes acetate to acetyl-CoA, is significantly upregulated in CD133+ CCICs, which potentially convert more upstream acetate into acetyl-CoA. On the other hand, downstream reactions to catalyze acetyl-CoA into malonyl-CoA and acetoacetyl-CoA are suppressed, hence forcing the extra acetyl-CoA to enter the TCA cycle. Consistently, the upstream metabolites in the TCA-cycle, i.e., citrate, aconitate, isocitrate, and α-ketoglutarate, are highly accumulated in CD133+ CCICs. At the end point of the TCA cycle, significantly upregulated PCK1 and PCK2 form a positive feedback loop to further facilitate the conversion of oxaloacetate into acetyl-CoA to fuel the TCA cycle. It has been shown that reprogramming of energy metabolism is a hallmark of pluripotent stem cells reprogram [14][15]. Stem cells influence epigenetic regulation such as histone acetylation and histone/DNA demethylation by controlling intermediate metabolite substrates Acetyl-CoA and α-ketoglutarate [16][17]. Therefore, altered energy metabolism and elevated Acetyl-CoA and α-ketoglutarate levels may explain some of the stem cell-like phenotypes observed in CD133+ CCICs.

Another consistently altered metabolic pathway is cysteine and methionine metabolism (Figure 5). In this pathway, enzyme CBS (Cystathionine-β-Synthase) combines homocycteine and serine to generate cystationine and subsequently CTH catalyzes cystationine into cysteine. Both CBS and CTH are significantly upregulated in CD133+ CCICs. Interestingly, cysteine has been reported to regulate neural stem cells through the CBS/H_2S pathway[18]. In methionine metabolism, DNMT1/3L, SRM, AMD1, and MTAP, downstream enzymes that catalyze S-AdenosylMethionine (SAM), are significantly downregulated in CD133+ CCICs. The metabolomics data confirm that the downregulated enzymatic reactions lead to accumulation of SAM in CD133+ CCICs. Among the downregulated enzymes, DNMT1/3L is an important DNA methyltransferase that catalyzes the transfer of methyl groups from SAM to DNA. Therefore, altered methionine metabolism might impact DNA methylation in CD133+ CCICs.

IV. DISCUSSION

CD133+ CCICs from different CRC tumors are likely to have diverse mechanisms. However, by perform system-level transcriptomic and metabolomic analyses on various CRC sources, we identified a distinct metabolic signature of CD133+ CCICs that involve glycolysis, the TCA cycle, and cysteine/methionine metabolism. The metabolite substrates involved in epigenetic regulations are highly altered, suggesting a potential epigenetic link. RNA-seq, metabolic flux analysis (MFA), and functional assays are currently being performed to further establish such links. The identified metabolic signature provides insights into reported stem cell-like properties of CD133+ CCICs. The involved metabolic enzymes and metabolites may provide biomarkers for CRC diagnosis and prognosis. New CRC treatments may also target them to suppress CCICs in the tumor population to reduce the risk of relapse and metastasis.

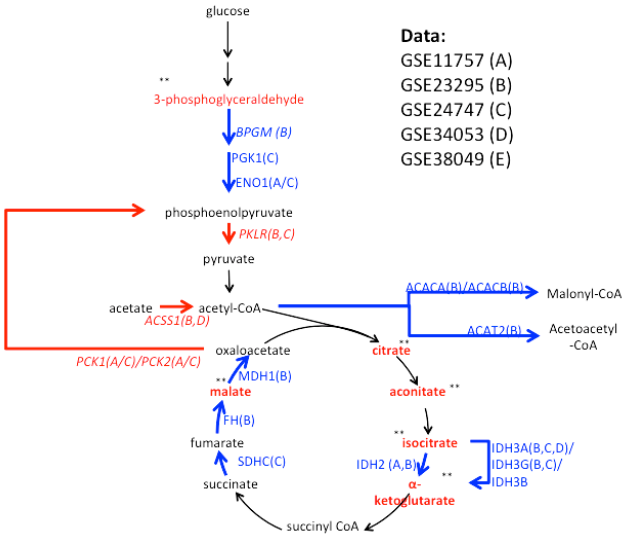

Figure 4. Altered carbohydrate metabolism in CCIC. Arrows represent metabolic flows regulated by marked metabolic enzymes. Metabolites are shown as nodes. Upregulated genes and metabolites are shown in red, downregulated genes and metabolites are shown in blue.

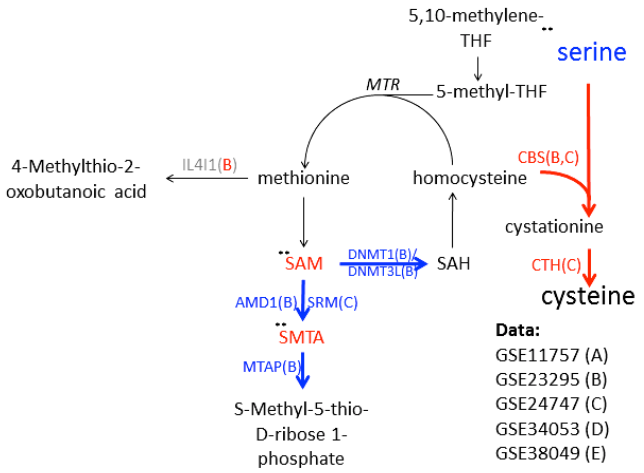

Figure 5. Altered cysteine and methionine metabolism in CCIC. The annotation is consistent with that of Figure4.

ACKNOWLEDGMENT

We thank the other members of the Shen lab and the Locasale lab for useful discussion and generous support.

REFERENCES

[1] L. Ricci-Vitiani, D. G. Lombardi, E. Pilozzi, M. Biffoni, M. Todaro, C. Peschle, and R. De Maria, "Identification and expansion of human colon-cancer-initiating cells.," *Nature*, vol. 445, pp. 111–115, 2007.

[2] C. A. O'Brien, A. Pollett, S. Gallinger, and J. E. Dick, "A human colon cancer cell capable of initiating tumour growth in immunodeficient mice.," *Nature*, vol. 445, pp. 106–110, 2007.

[3] P. Bu, K.-Y. Chen, J. H. Chen, L. Wang, J. Walters, Y. J. Shin, J. P. Goerger, J. Sun, M. Witherspoon, N. Rakhilin, J. Li, H. Yang, J. Milsom, S. Lee, W. Zipfel, M. M. Jin, Z. H. Gümüş, S. M. Lipkin, and X. Shen, "A microRNA miR-34a-regulated bimodal switch targets notch in colon cancer stem cells.," *Cell Stem Cell*, vol. 12, no. 5, pp. 602–15, May 2013.

[4] F. Ren, W.-Q. Sheng, and X. Du, "CD133: a cancer stem cells marker, is used in colorectal cancers.," *World J. Gastroenterol.*, vol. 19, pp. 2603–11, 2013.

[5] J. P. Medema, "Cancer stem cells: the challenges ahead.," *Nat. Cell Biol.*, vol. 15, pp. 338–44, 2013.

[6] T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva, "NCBI GEO: archive for functional genomics data sets--10 years on.," *Nucleic Acids Res.*, vol. 39, pp. D1005–D1010, 2011.

[7] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang, "Bioconductor: open software development for computational biology and bioinformatics.," *Genome Biol.*, vol. 5, p. R80, 2004.

[8] J. W. Davis, "Bioinformatics and Computational Biology Solutions Using R and Bioconductor," *Journal of the American Statistical Association*, vol. 102. pp. 388–389, 2007.

[9] H. Chen and P. C. Boutros, "VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R.," *BMC Bioinformatics*, vol. 12, p. 35, 2011.

[10] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 43, pp. 15545–50, Oct. 2005.

[11] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, "The KEGG resource for deciphering the genome.," *Nucleic Acids Res.*, vol. 32, no. Database issue, pp. D277–80, Jan. 2004.

[12] M. Jaksch, J. Múnera, R. Bajpai, A. Terskikh, and R. G. Oshima, "Cell cycle-dependent variation of a CD133 epitope in human embryonic stem cell, colon cancer, and melanoma cell lines.," *Cancer Res.*, vol. 68, pp. 7882–7886, 2008.

[13] C. Chao, J. R. Carmical, K. L. Ives, T. G. Wood, J. F. Aronson, G. A. Gomez, C. D. Djukom, and M. R. Hellmich, "CD133+ colon cancer cells are more interactive with the tumor microenvironment than CD133− cells," *Laboratory Investigation*, vol. 92. pp. 420–436, 2012.

[14] J. Zhang, E. Nuebel, G. Q. Daley, C. M. Koehler, and M. a Teitell, "Metabolic Regulation in Pluripotent Stem Cells during Reprogramming and Self-Renewal.," *Cell Stem Cell*, vol. 11, no. 5, pp. 589–95, Nov. 2012.

[15] N. Shyh-Chang, Y. Zheng, J. W. Locasale, and L. C. Cantley, "Human pluripotent stem cells decouple respiration from energy production.," *EMBO J.*, vol. 30, no. 24, pp. 4851–2, Dec. 2011.

[16] C. Lu and C. B. Thompson, "Metabolic regulation of epigenetics.," *Cell Metab.*, vol. 16, pp. 9–17, 2012.

[17] C. Johnson, M. O. Warmoes, X. Shen, and J. W. Locasale, "Epigenetics and cancer metabolism," *Cancer Letters*, 2013.

[18] Z. Wang, D. Liu, F. Wang, Q. Zhang, Z. Du, J. Zhan, Q. Yuan, E.-A. Ling, and A. Hao, "L-cysteine promotes the proliferation and differentiation of neural stem cells via the CBS/H(2)S pathway.," *Neuroscience*, 2013.