

Single-cell exome sequencing identifies mutations in *KCP*, *LOC440040*, and *LOC440563* as drivers in renal cell carcinoma stem cells

Cell Research (2017) 27:590-593. doi:10.1038/cr.2016.150; published online 16 December 2016

Dear Editor,

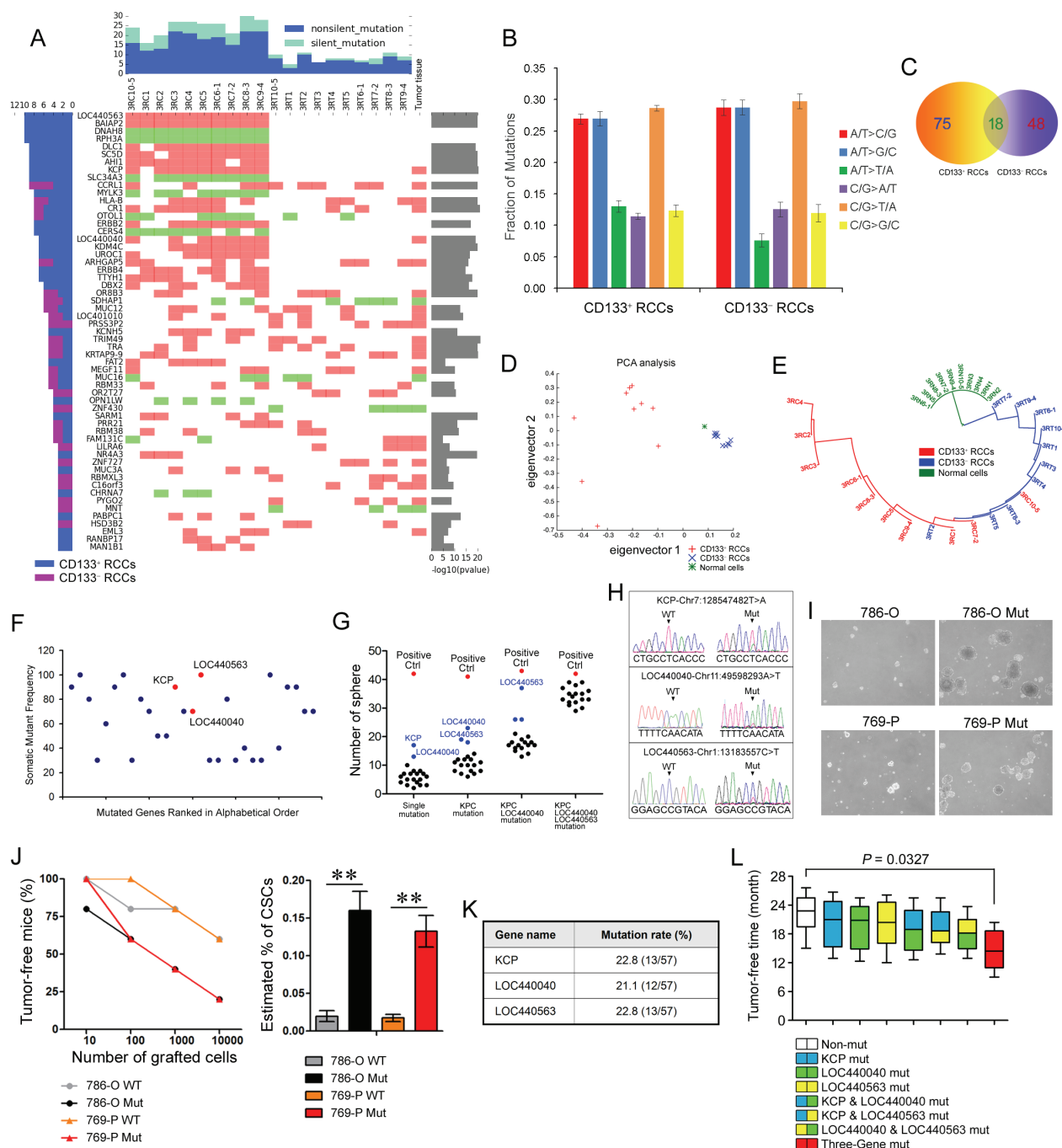
Renal cell carcinoma (RCC) accounts for ~3% of adult malignancies and 90%-95% of kidney neoplasms [1]. Metastatic disease is usually resistant to radiation and chemotherapy. Immunotherapy shows limited response rates of 15% to 20% [2]. Cancer stem-like cells (CSCs) are key players in RCC initiation, development and resistance to therapy [3]. Previous studies have shown that RCC is a genetically distinct adult carcinoma with a relatively low mutation rate [4]. Sequencing analyses have revealed that some common somatic mutations are shared among RCC patients [5]. Recently, single-cell exome sequencing has been used to evaluate somatic mutations in many tumor types [6]. However, a systematic effort applying this new technique to identify key driver genes in the CSCs of RCC has not been made.

CD133 has been identified as a common CSC marker for many solid tumors, including kidney cancer [7]. We investigated whether CD133⁺ RCC cells isolated from patients showed stem-like characteristics. We sorted CD133⁺ and CD133⁻ RCC cells from the specimens of a 57-year-old male patient diagnosed with stage T3aN1M0 RCC (Supplementary information, Figure S1A and Supplementary information, Table S1A), and only CD133⁺ cells had sphere-forming capability (Supplementary information, Figure S1A). We next grafted isolated CD31⁻CD45⁻, CD31⁻CD45⁻CD133⁻, or CD31⁻CD45⁻CD133⁺ RCC cells subcutaneously into NOD/SCID mice. Nearly all mice that received 10 000 CD31⁻CD45⁻CD133⁺ cells developed visible tumors, and among mice that received only 10 cells, ~25% developed visible tumors (Supplementary information, Figure S1B). However, more than 50% of the mice transplanted with the other two cell types (10 000 cells/mouse) remained tumor-free (Supplementary information, Figure S1B). In addition, to directly evaluate the frequency of stem-like RCC cells, we applied an extreme limiting dilution assay. We estimated the percentage of CSCs in the CD31⁻CD45⁻CD133⁺ cell

sample was more than seven fold higher than in the other two cell types (Supplementary information, Figure S1C), indicating that CD133⁺ cells isolated from RCC have stem-like properties.

We next sorted 10 CD133⁺ RCC cells, 10 CD133⁻ RCC cells, and 10 normal renal cells from the same patient for single-cell whole-exome sequencing (WES). The RCC and adjacent normal tissues of the same patient were also subjected to WES (Supplementary information, Figure S1D). For each cell and tissue sample, we acquired WES data with over 140× coverage after whole-genome amplification (WGA) using the MAL-BAC technique.

To eliminate the false-positive variations arising from WGA, we only selected single-nucleotide variants (SNVs) present in CD133⁺ or CD133⁻ RCC cells but absent in normal cells. Similarly, for whole-tissue WES analysis, SNVs specifically present in the RCC tissue were chosen. WES of the cancer tissue revealed 160 somatic SNVs and single-cell sequencing of the 20 tumor cells identified 297 somatic SNVs (Supplementary information, Figure S1E). Commonly mutated genes in RCC, such as *VHL*, *BAP1*, *TRA* and *CHD4*, were found to carry variations in both RCC single cells and tissue samples, demonstrating the reliability of our WES analysis (Supplementary information, Table S1B). Among the SNVs in single cells, 141 were located in coding regions. These SNVs were more enriched in CD133⁺ cells than in CD133⁻ cells (Figure 1A and 1C). More importantly, we discovered several coding region mutations that are unique to CD133⁺ RCC cells. Among them, three missense mutations, c.A241T>p.R81W in *Kielin/chordin-like protein (KCP)*, c.G316A>p.G106S in *LOC440563* and c.A406T>p.N136Y in *LOC440040*, have not been reported in RCC. While other mutations in *KCP* can be found in the RCC TCGA database (c.G2590A>p.A864T, c.C1250G>p.A417G, and c.A2680G>p.R894G), no mutations in *LOC440563* and *LOC440040* have been linked to RCC and other cancer before (Figure 1A, Supplemen-



tary information, Table S1C). In addition, we found that *KCP* and *LOC440040* mutations could be detected in both single cells and in the original cancer tissue (Figure 1A and Supplementary information, Table S1B). Furthermore, we identified that *C/G>T/A*, *A/T>C/G* and *A/T>G/C* transitions were the most common mutations in RCC (Figure 1B). Notably, the *A/T>T/A* transitions were significantly more frequent in $CD133^+$ cells than in $CD133^-$ cells (Figure 1B, $P = 0.31 \times 10^{-9}$).

To further confirm that the isolated single cells are indeed tumor cells, we performed principle component analysis of all somatic mutation data obtained from the 30 cells (Figure 1D). The normal renal and $CD133^-$ RCC cells formed two independent groups, whereas the $CD133^+$ RCC cells were scattered and were different from these two groups (Figure 1D). We constructed a neighbor-joining tree for the 30 cells; in the tree, the 10 normal cells were grouped together, whereas both sets of

10 CD133⁺ RCC cells and 10 CD133⁻ RCC cells were nearly completely separated between the groups (Figure 1E). The evolution distance between normal cells and CD133⁺ RCC is larger than that between normal cells and CD133⁻ RCC, suggesting that CD133⁺ RCC cells more likely originated from cancer cells than normal cells.

We then selected mutated genes that are shared by more than three cells in each group for further analysis. Twenty-nine mutated genes were detected in more than 3 CD133⁺ RCC cells (Figure 1F and Supplementary information, Table S1C), while only 11 mutated genes were shared by over 3 CD133⁻ RCC cells (data not shown). Among the 29 mutated genes detected in CD133⁺ RCC cells, 18 genes, including *KCP* (not *LOC440563* or *LOC440040*), were listed in the TCGA RCC database as mutated genes, and the frequency of the 18 mutated genes in the 416 RCC cases in TCGA was less than 2% (Supplementary information, Table S1D).

To determine the tumor-propagating potential of each mutation, we introduced heterozygous mutations in the 20 candidate genes, which are related to the function of chromatin remodeling, transcription regulation and self-renewal, in RCC 786-O cells using the CRISPR-Cas9 technique (one mutation/gene), and assay for single cell spherogenicity. In all, 3 out of 20 gene mutations are nonsense mutations and the remaining are missense mutations (Supplementary information, Table S1C). All 20 mutations introduced by the CRISPR-Cas9 system were verified by DNA sequencing (Figure 1H and Supplementary information, Figure S2). Two of the 20 mutations in 786-O cells significantly enhanced the spherogenesis capabilities by more than 10% (Figure 1G,

left). The mutation with the highest cancer spherogenesis was in the *KPC* gene. Because multiple mutations may be required for the maintenance and development of CD133⁺ RCC cells, we mutated each of the remaining 19 genes in combination with *KPC*. The *LOC440040* mutation was most effective, which enhanced spherogenesis capabilities of the *KPC* mutation (Figure 1G, middle). We also additionally mutated *LOC440563* and found *KPC-LOC440040-LOC440563* triple mutation could increase the sphere-formation abilities by more than 30% (Figure 1G, right). To validate that *KCP* is indeed a prominent driver gene among the 20 candidates, we mutated the *LOC440040* and one of the other 19 genes and found that *LOC440040* plus *KCP* double mutations had significantly higher spherogenesis than other combination. Consistently, combining the *KPC* mutation could increase the sphere-forming ability of cells with the *LOC440040-LOC440563* double mutations (Supplementary information, Figure S1F). Moreover, 786-O and 769-P RCC cells with *KCP/LOC440040/LOC440563* triple mutations had significant enhancement in cancer spherogenesis capabilities compared with wild-type cells (Figure 1I). Results of xenograft experiments confirmed that more mice developed tumors and had higher enrichment of CSCs in the group that received triple-mutated 786-O or 769-P cells compared with mice that received wild-type cells (Figure 1J), indicating that mutations in the three genes played a key role in the CD133⁺ RCC tumor-propagating features. Taken together, these results suggest that mutations in *KCP*, *LOC440040* and *LOC440563* may facilitate RCC cells to acquire CSC properties.

Next, we assessed the mutation rates of *KCP*, *LOC440040*,

Figure 1 Identification of driver genes in renal cell carcinoma stem cells via single-cell exome sequencing. **(A)** Detection of somatic mutations in CD133⁺CD133⁻ RCC cells and in cancer tissue. The main plot shows information for genes with mutations for 20 cells and original cancer tissue. The red color represents non-silent mutations and green color represents silent mutations. **(B)** Somatic mutation graph. Two substitutions (A/T>G/C and C/G>T/A) are clearly frequent. **(C)** Venn plots show the somatic mutations in CD133⁺ and CD133⁻ RCC cells. **(D)** Principle component analysis (PCA) of the mutations in the CD133⁺ RCC cells (red), CD133⁻ RCC cells (green) and normal cells (blue). Eigenvector is defined as the Covariance Matrix. **(E)** A neighbor-joining tree was constructed using the somatic mutation data set. The normal cells are labeled in green, CD133⁻ RCC cells are labeled in blue, and CD133⁺ RCC cells are labeled in red. **(F)** The average mutation frequency of 29 genes with variations in at least 3 CD133⁺ RCC cells. The mutation frequency indicates the percentage of CD133⁺ RCC cells with the mutated gene. **(G)** Data points indicate the average number of spheres of RCC cells with distinct mutations in serum-free conditions. Each of the 20 mutations was tested alone (first column, 'single mutation'), in combination with a *KCP* mutation (second column) or in combination with *KCP* and *LOC440040* mutations (third column). Other mutations were also tested in combination with *KCP*, *LOC440040*, and *LOC440563* mutations (fourth column). Mutation combinations that enhanced the *in vitro* spherogenicity (blue) were selected for *in vivo* validation. CD133⁺ cells spheres served as the positive control (red). **(H)** Representative Sanger-sequencing data of *KCP*, *LOC440040*, and *LOC440563* in wild-type (WT) and mutated (Mut) renal cancer cells are listed below. **(I)** Representative oncospheres in mutated (Mut) and vehicle renal cancer cells. **(J)** The 18-week tumor-free rate of NOD/SCID mice after subcutaneous injection at the indicated dilutions of 786-O WT, 786-O Mut, 769-P WT, and 769-P Mut cells in xenografted mice using extreme limiting dilution analysis ($n = 6$ grafted tumors per dilution; right panel). **(K)** The CD31⁺CD45⁻CD133⁺ cells from 57 RCC patients were individually sorted and pooled together for the indicated targeted sequencing. The mutation rates of *KCP*, *LOC440040*, and *LOC440563* are indicated. **(L)** The average tumor-free time of 57 renal cancer patients with or without *KCP*, *LOC440040*, and/or *LOC440563* mutation(s) after primary tumor resection.

and *LOC440563* in a cohort of 57 RCC patients. CD133⁺ RCC cells in each patient were sorted and subjected to targeted Sanger sequencing. More than 20% of the patients harbored at least one mutation (Figure 1K) and 5.26% (3/57) harbored all three mutations (Supplementary information, Table S1A). The patients with triple mutations showed a significantly shorter tumor-free time after primary tumor resection (Figure 1L). The outside region mutations in RCC remain to be investigated in the future because of the current limitations of targeted Sanger sequencing.

In summary, using single-cell exome sequencing, we discovered that CD133⁺ RCC cells have CSC properties and likely originate from cancer cells instead of from normal renal cells. *KCP*, *LOC440040*, and *LOC440563* mutations are novel renal cancer stem cell drivers. *LOC440563* encodes an RNA-binding protein, which belongs to the subfamily of heterogeneous nuclear ribonucleoproteins (hnRNPs). It influences pre-mRNA processing, metabolism and transport [8, 9]. *LOC440563* is significantly mutated in colon cancer; however, the mutations are different from what we identified in RCC CSCs [10]. Considering the contribution of the hnRNP family to DNA damage repair [11], it is possible that the *LOC440563* mutation may drive CSC stemness via impairing the DNA repair process. *LOC440040* is a GRM5 (glutamate receptor, metabotropic 5) pseudogene. GRM5 belongs to the mGluR family, and dysregulated glutamatergic signaling is involved in many cancer types, such as glioma and melanoma [12]. However, the relationship between *LOC440040* and cancer is currently unknown. It would be interesting to determine whether mutant *LOC440040* affects glutamatergic signaling in RCC CSC populations. *KCP* encodes a secreted protein, which plays complex roles in the regulation of the TGF- β pathway. Abnormal level of *KCP* may cause myriad renal pathology and cancer [13-15]. It would be interesting to study whether the *KCP* mutation identified in this work promotes CSC self-renewal via impacting TGF- β signaling. Finally our study indicates *KCP*, *LOC440040*, *LOC440563* mutations, which are present in at least 20% of patients in our survey, constitute robust and dangerous drivers promoting reprogramming of RCC cancer cells into CSCs. They thus should serve as important prognostic factors and therapeutic targets for RCC.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81472413 and 81472679), 863 Program

(SS2014AA020607 and 2015AA020405), and the National Key Research and Development Program (2016YFC0902603).

Chong Li^{1,2,3,*}, Song Wu^{4,*}, Zhao Yang^{5,*},
Xiaolong Zhang^{4,*}, Qi Zheng⁶, Ling Lin⁶,
Zexiong Niu⁶, Ruiqiang Li⁶, Zhiming Cai⁴, Lei Li¹

¹Department of Urology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shanxi 710061, China; ²Chinese Academy of Sciences Protein Science core facility center, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China; ³Beijing Jianlan Institute of Medicine, Beijing 100190, China; ⁴The Affiliated Luohu Hospital of Shenzhen University, Shenzhen Luohu Hospital Group, Shenzhen, Guangdong 518000, China; ⁵CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; ⁶Novogene Bioinformatics Institute, Beijing 100083, China
*These four authors contributed equally to this work.

Correspondence: Lei Li

E-mail: lilydr@hotmail.com

References

- Rini BI, Campbell SC, Escudier B. *Lancet* 2009; **373**:1119-1132.
- Fyfe GA, Fisher RI, Rosenberg SA, et al. *J Clin Oncol* 1996; **14**:2410-2411.
- Peired AJ, Sisti A, Romagnani P. *Stem Cells Int* 2016; **2016**:8342625.
- Hung RJ, Moore L, Boffetta P, et al. *Cancer Epidemiol Biomarkers Prev* 2007; **16**:1287-1290.
- Farley MN, Schmidt LS, Mester JL, et al. *Mol Cancer Res* 2013; **11**:1061-1071.
- Xu X, Hou Y, Yin X, et al. *Cell* 2012; **148**:886-895.
- Florek M, Haase M, Marzesco AM, et al. *Cell Tissue Res* 2005; **319**:15-26.
- Caporali R, Bugatti S, Bruschi E, et al. *Autoimmunity* 2005; **38**:25-32.
- Liu N, Dai Q, Zheng G, et al. *Nature* 2015; **518**:560-564.
- Futreal PA, Coin L, Marshall M, et al. *Nat Rev Cancer* 2004; **4**:177-183.
- Kai M. *Int J Mol Sci* 2016; **17**:310.
- Wu CH, Apweiler R, Bairoch A, et al. *Nucleic Acids Res* 2006; **34**:D187-191.
- Lin J, Patel SR, Cheng X, et al. *Nat Med* 2005; **11**:387-393.
- Soofi A, Zhang P, Dressler GR. *J Am Soc Nephrol* 2013; **24**:897-905.
- Watabe T, Miyazono K. *Cell Res* 2009; **19**:103-115.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2016