

# A genome-wide loss-of-function screening method for minimizing false-negatives caused by functional redundancy

*Cell Research* (2016) 26:1067-1070. doi:10.1038/cr.2016.97; published online 26 August 2016

## Dear Editor,

Loss-of-function screens are powerful tools for identifying gene contribution in a given biological context. Over the past decade, the RNA interference technology has become a dominant approach in the loss-of-function screen-based gene discovery [1]. However, about 3 quarters of genes in mammalian genomes belong to gene families and have functional redundant homologs. While this functional redundancy protects cells and organisms from deleterious mutations [2], it also masks phenotypic outcomes, resulting in false-negatives in loss-of-function screens that target individual genes. However, all of reported loss-of-function screens have been designed to target individual genes and thus suffer from limitations of the false-negatives.

The Wnt- $\beta$ -catenin pathway plays pivotal roles in embryogenesis as well as in adult tissue homeostasis. Aberrant Wnt- $\beta$ -catenin signaling has been linked to a wide range of pathologies in humans, including cancer. Wnt3A, a prototypic canonical Wnt, initiates its signaling by binding to a Frizzled (Fz) family receptor and a low density lipoprotein-related protein (LRP) 5/6 coreceptor, causing, via the intracellular signaling protein dishevelled (Dvl), the stabilization and accumulation of  $\beta$ -catenin in both cytoplasm and nucleus by inhibiting the degradation function of destruction complex composed of GSK3, APC, Axin and  $\beta$ -TrCP. The nuclear  $\beta$ -catenin binds to the TCF family of transcription factors and regulates gene transcription [3]. Like other biological system, functional redundancy has been observed in the Wnt signaling pathway [4-6]. For instance, there are ten Fz (Fz1-10) and two LRP (LRP5/6), three Dvl (Dvl1-3), two GSK3 (GSK3 $\alpha/\beta$ ), two Axin (Axin1/2), and two  $\beta$ -TrCP ( $\beta$ -TrCP1/2).

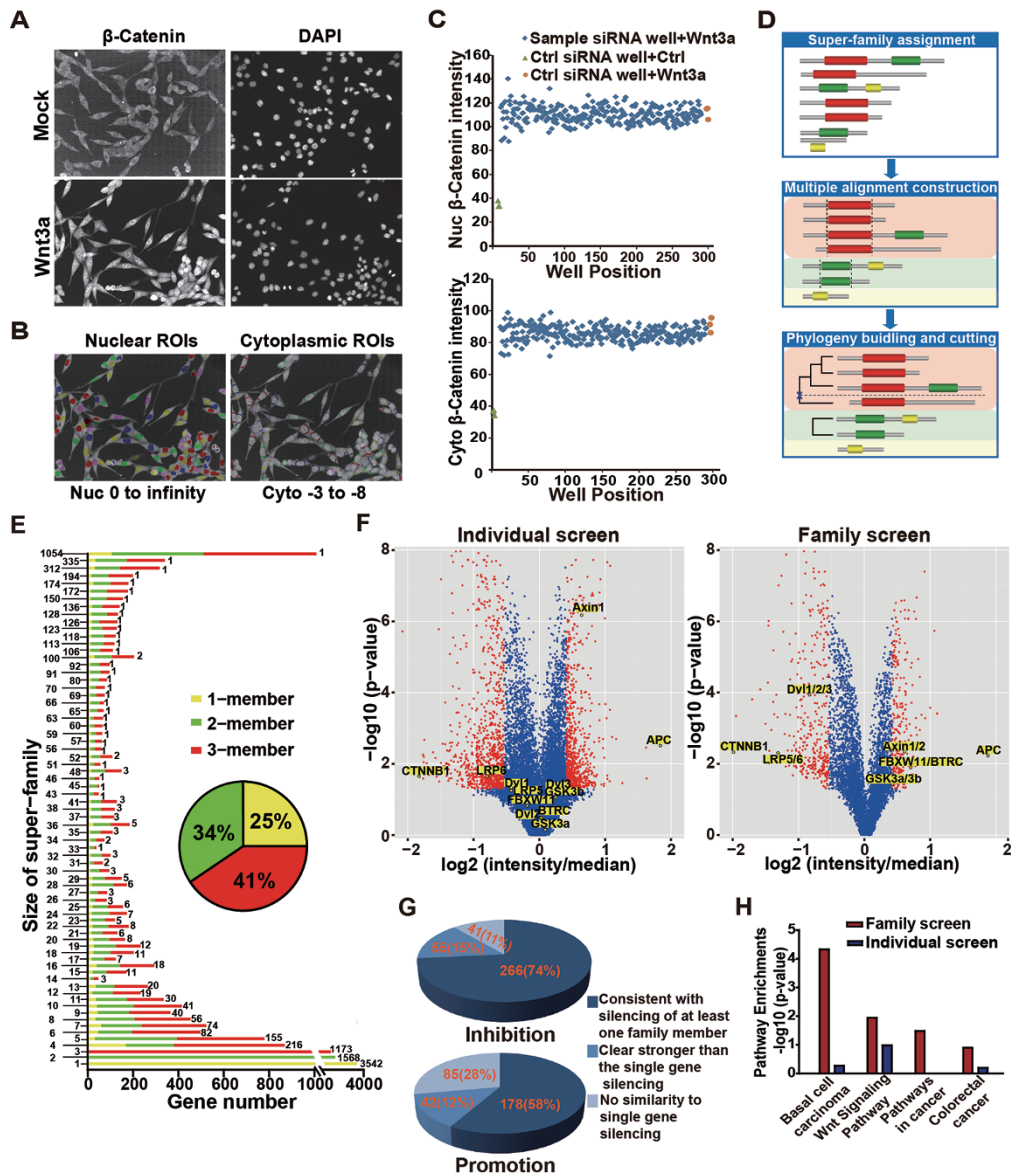
Here, we report a gene family screen approach that can circumvent the false-negative issue resulted from gene functional redundancy. Using a genome-wide siRNA screen for regulators important for Wnt3A-induced

$\beta$ -catenin accumulation as an example, we demonstrate that a gene family-based loss-of-function screen can effectively minimize the functional redundancy problem that plagues the individual gene screens.

We used the Opera high-content imaging system to perform genome-wide siRNA screening for regulators of  $\beta$ -catenin content and subcellular localization in response to Wnt3a treatment in mouse L cells. Figure 1A shows an example of  $\beta$ -catenin and DAPI staining in cells transfected with the control siRNA. For  $\beta$ -catenin content quantification, the regions of interest (ROIs) of the nucleus and cytoplasm were delineated by the Acapella software (Perkin Elmer) based on the composite images of the DAPI and  $\beta$ -catenin staining (Figure 1B). The average pixel intensities of the  $\beta$ -catenin staining in the nuclear and cytoplasmic ROIs are taken as relative nuclear and cytoplasmic  $\beta$ -catenin contents, respectively, whereas the sum of the average nuclear and cytoplasmic intensities is taken as the total  $\beta$ -catenin content of the cell.

To validate our approach, L cells were transfected with siRNAs for a number of known Wnt signaling components, including APC, LRP6 and  $\beta$ -catenin. Expected changes in both nuclear and cytoplasmic  $\beta$ -catenin contents were observed using the aforementioned detection and quantification approach (Supplementary information, Figure S1A). To assess the suitability of our approach for high-throughput screening, the Z factors [7] for the nucleus and cytoplasmic  $\beta$ -catenin contents in a 384-well plate were determined (Figure 1C). They are 0.61 and 0.63, respectively, indicating that our assay system has a reliable reproducibility and uniformity and is thus well suited for high-throughput screening.

We next performed a high-content screen using the Dharmacon Mouse Genome siRNA Library, which contains siRNA smartpools targeting 19 059 genes, using L cells stimulated with Wnt3A in triplicates. The screen data were normalized and analyzed using a BioConductor bioinformatics package OperaMate [8] and are sum-



**Figure 1** Genome-wide individual gene-based and gene family-based high-content siRNA screen. **(A, B)** Detection and quantification of  $\beta$ -catenin contents. L cells were treated with mock or Wnt3A (2 h) and stained with anti- $\beta$ -catenin antibody and DAPI. The images were acquired using an Opera LX system **(A)**. The Acapella software was used to generate the composite images **(B)** from the  $\beta$ -catenin and DAPI staining images **(A)**. The nuclear ROI is defined as the DAPI staining area, while the cytoplasmic ROI is defined as the region 3 to 8 pixels outside the nuclear ROI by the software. **(C)** Scatter plot of  $\beta$ -catenin contents detected in a 384-well plate. L cells were treated with mock or Wnt3A (2 h) and  $\beta$ -catenin contents were detected and quantified as aforementioned. **(D)** Schematic representation of the generation of the gene family library. First, proteins were assigned into super-families according to Pfam annotations. Second, multiple sequence alignment was carried out. Finally, based on the phylogenetic tree, proteins of each super-family were classified into groups with a maximum of 3 members. **(E)** Overview of the composition of the gene family library. Vertical axis shows the size of the super-families, whereas the numbers at the right side of the bars indicate how many gene superfamilies are with that size. The horizontal axis shows the number of genes assigned to families with 1 (yellow), 2 (green) and 3 (red) members. The proportions of all of the genes assigned to families with 1, 2, or 3 members are shown in the pie chart. **(F)** Volcano plots of individual and family siRNA screen results. Log<sub>2</sub> transformed normalized intensity is plotted against  $-\log_{10}$  transformed  $P$  for each sample. Hits meeting both quantitative (t-score < 0.1) and statistical ( $P$  < 0.05) criteria are colored red. Some Key components of Wnt3A signaling are highlighted in yellow. **(G)** Sub-division of gene family screen results based on the comparison with the individual gene screen results. In Group 1, at least one individual member shows a consistent effect (i.e., with an inhibition t-score < 0.1 or promotion t-score < 0.2) as the family screen. In Group 2, the family has at least one member that shows a weak effect (i.e., with an inhibition t-score < 0.2 or promotion t-score < 0.3) as the family screen. In Group 3, the family has no individual member that shows any effect. **(H)** Pathway analysis. Pathway analysis for individual gene screen and gene family screen hits was carried out by the DAVID functional annotation software using KEGG and Biocarta reference pathway database. The bars show the enrichment degree by  $-\log_{10}$  transformed  $P$  (the EASE scores in the DAVID reports).

marized in Supplementary information, Table S1A. The putative positive hit candidates, which fulfill the hit-calling criteria of a  $t$ -score  $< 0.1$  with a multiple student's  $t$ -test  $P < 0.05$ , contain many previously known Wnt signaling components (Supplementary information, Table S1B) and are listed in Supplementary information, Table S1C. However, a number of other well-characterized Wnt signaling components, including Dvl,  $\beta$ -TrCP, and GSK3, are not in the list. We compared our individual gene screen with three previously published screens [9–11], and found  $\beta$ -TrCP and GSK3 were also missing in those screens (results of the comparison are listed in Supplementary information, Table S1D). Given the known function redundancy in Wnt signaling pathway [4–6], it is reasonable to postulate that the failure to identify these well characterized Wnt signaling components might be due to the presence of multiple functionally redundant homologs of these components. Indeed, when all three Dvl isoforms were silenced simultaneously, significant inhibition of Wnt3A-induced  $\beta$ -catenin accumulation was observed (Supplementary information, Figure S1B). This is in clear contrast to the lack of effect of depleting each individual Dvl isoform. Similar results were also observed for silencing  $\beta$ -TrCP1/2 (Supplementary information, Figure S1C). These results together support the idea that functional redundancy is a real issue that can cause many of the false-negatives in a loss-of-function screen.

To circumvent the problem, we generated an siRNA library to target, instead of each individual gene, gene families that consist of functionally redundant homologs. The foremost difficulty is to obtain a gene family database for an entire genome as there is no such database available. We thus developed a bioinformatic method to group genes into functionally related gene families based on the sequence similarity of their coded proteins reasoning that functionally related proteins in general share highest degree of amino acid sequence similarity (Figure 1D). The protein sequence for each gene was first retrieved from GenBank [12]. Pfam [13] was used to annotate the protein sequences with an expectation cutoff value of  $1 \times 10^{-4}$ . If a protein had multiple Pfam annotations, only the one with the most significant expectation value was used. Proteins with the same Pfam annotations were assigned into families. The Dharmacon mouse genome siRNA library we used is made of siRNA smartpools. Since most of the known functionally redundant gene families in Wnt signaling consist of less than three genes, and pooling of three siRNA smartpools was shown to be effective in silencing the Dvl family (Supplementary information, Figure S1B), as well as pooling of more than three siRNA smartpools might compromise

gene silencing efficiency, we decided to limit the family pool to no more than three genes. For those families with more than three genes, the amino acid sequences of the members were aligned by ClustalW [14], and phylogenetic trees were constructed based on the pairwise Kimura protein distance and the UPGMA (unweighted pair group method with arithmetic mean) algorithm implemented in Bioperl [15]. Two proteins with the closest distance to each other in the tree were grouped together, and the third protein was added if it was the closest one to the group as well as it was not assigned into other groups. In total, the 19 059 genes in the Dharmacon mouse genome siRNA library were grouped into 2 580 3-member families and 3 270 2-member families with 4 779 genes left as individuals (1-member families). The proportions of genes contained in the 3-member, 2-member and 1-member families were 41%, 34%, and 25%, respectively (Figure 1E and Supplementary information, Table S2A). The gene family siRNA library was physically constructed by pooling the siRNAs using the Beckman Coulter liquid handling robotic system under a sterile condition using a customerized robot-controlling software.

We then performed a screen with the custom gene family siRNA library. Those siRNAs that are not grouped into any of the families were not included, as we have done the individual gene screen. The screen data were analyzed the same way as the data from the individual gene screen were analyzed and are listed in Supplementary information, Table S2B. We used the same cutoff criteria as for the individual gene screen to select putative positive gene family hits, which are listed in Supplementary information, Table S2C. The outcomes of the individual and family screens were compared side-by-side using the Volcano plot, in which the putative positive hits are labeled with red (Figure 1F). The Dvl1/2/3,  $\beta$ -TrCP1/2 and GSK3 $\alpha/\beta$  families were all identified as hits in this gene family screen, validating our experimental design and approach.

Taking a different approach to compare the results of the individual and family screens, we subdivided the putative positive hits from the gene family screen into three groups based on the comparison of the effects of gene family silencing with those of silencing of each individual genes in the family (Figure 1G, Supplementary information, Table S2D and S2E). The first group are the families in which at least one individual member shows the consistent effect with that of the family in the screen; in the second group, one or more members in a family shows only a weaker effect than that of the family; and in the third group, only the families, but not their individual members, show any effect in the screens. Most of the family hits belong to the first group (Figure 1G), suggest-

ing good consistency between individual gene and gene family screens. On the other hand, Group 2 and particularly Group 3 are of greater interest, as these groups may contain regulators for  $\beta$ -catenin contents, which have been missed in individual gene loss-of-function screens. The top 10 inhibition and promotion hits in Group 3 were then validated by western blot analysis. 60% of them showed results consistent with the screening data (Supplementary information, Figure S1D).

We also carried out pathway analysis for the positive hits based on functional annotations of the genes and gene families using DAVID functional annotation software. There are more gene family screen hits related to the Wnt signaling, cancer and colorectal cancer pathways than those from the individual gene screen (Figure 1H and Supplementary information, Table S2F). This result provides further support for the effectiveness of our approach in identifying novel gene functions and novel crosstalks between different signaling pathways.

In summary, we used the Opera high-content imaging system to carry out siRNA-based loss-of-function screens for potential regulators of  $\beta$ -catenin contents in cells treated with Wnt3A. Our novel gene family-based screen strategy circumvents the false-negative issue resulting from gene functional redundancy. Comparison of the results from the individual gene screen with those of the gene family screen clearly demonstrates the advantage of our new screen approach. This new gene family-based screen strategy should be applicable to all of the loss-of-function screens, including the CRISPR/Cas9-based screens. Moreover, in contrast to previous cell-based loss-of-function screens of the Wnt signaling pathway, which used the Wnt reporter gene assays [9, 10], we directly examined  $\beta$ -catenin content and localization using a high-content imaging system. Thus, our study also provides valuable resources for the Wnt research community. Nevertheless, the current gene family screen approach has a key limitation: its effectiveness may be compromised for the gene families that have more than three family members. However, this limitation may be overcome with new technologies that increase transfection efficiency or with the use of viral transduction.

## Acknowledgments

The work was supported by the National Natural Science Foundation

of China (31230044 and 31530094 to LL) and the strategic priority research program of Chinese Academy of Sciences (CAS; XDB19000000 to LL), NIH grant (GM112182 to DW) and the National Basic Research Program of China (973 Program; 2011CB910204 to YL). The work was also supported by the CAS/SAFEA International Partnership Program for Creative Research Teams.

Li Mao<sup>1,\*</sup>, Chenglin Liu<sup>2,3,\*</sup>, Zhen Wang<sup>2</sup>, Xiaofeng Niu<sup>1</sup>, Liang Xue<sup>1</sup>, Zhilei Zhou<sup>1</sup>, Zhenying Cai<sup>1</sup>, Meng Yu<sup>1,4</sup>, Yixue Li<sup>2,3,5</sup>, Dianqing Wu<sup>6</sup>, Lin Li<sup>1</sup>

<sup>1</sup>State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Innovation Center for Cell Signaling Network, Institute of Biochemistry and Cell Biology, <sup>2</sup>Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; <sup>3</sup>School of Life Sciences and Biotechnology, Shanghai Jiaotong University, Shanghai, China; <sup>4</sup>School of Life Science and Technology, Shanghai Tech University, Shanghai 200031, China; <sup>5</sup>Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China; <sup>6</sup>Vascular Biology and Therapeutic Program and Department of Pharmacology, Yale School of Medicine, New Haven, CT 06520, USA

\*These two authors contributed equally to this work.

Correspondence: Lin Li<sup>a</sup>, Dianqing Wu<sup>b</sup>, Yixue Li<sup>c</sup>

<sup>a</sup>E-mail: lli@sibs.ac.cn

<sup>b</sup>E-mail: dianqing.wu@yale.edu

<sup>c</sup>E-mail: yxli@sibs.ac.cn

## References

- 1 Gao S, Yang C, Jiang S, *et al.* *Protein Cell* 2014; **5**:805-815.
- 2 Zhang, J. *Adv Exp Med Biol* 2012; **751**:279-300.
- 3 Clevers H, Nusse R. *Cell* 2012; **149**:1192-1205.
- 4 Guardavaccaro D, Kudo Y, Boulaire J, *et al.* *Dev Cell* 2003; **4**:799-812.
- 5 Doble BW, Patel S, Wood GA, *et al.* *Dev Cell* 2007; **12**:957-971.
- 6 Etheridge SL, Ray S, Li S, *et al.* *PLoS Genet* 2008; **4**:e1000259.
- 7 Zhang JH, Chung TD, Oldenburg KR. *J Biomol Screen* 1999; **4**:67-73.
- 8 Gentleman RC, Carey VJ, Bates DM, *et al.* *Genome Biol* 2004; **5**:R80.
- 9 Tang W, Dodge M, Gundapaneni D, *et al.* *Proc Natl Acad Sci USA* 2008; **105**:9697-9702.
- 10 Conrad W, Major MB, Cleary MA, *et al.* *F1000Res* 2013; **2**:134.
- 11 Madan B, Walker MP, Young R, *et al.* *Proc Natl Acad Sci USA* 2016; **113**:E2945-E2954.
- 12 Coordinators NR. *Nucleic Acids Res* 2014; **42**:D7-D17.
- 13 Finn RD, Bateman A, Clements J, *et al.* *Nucleic Acids Res* 2014; **42**:D222-D230.
- 14 Larkin MA, Blackshields G, Brown NP, *et al.* *Bioinformatics* 2007; **23**:2947-2948.
- 15 Stajich JE, Block D, Boulez K, *et al.* *Genome Res* 2002; **12**:1611-1618.

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