Cell Reports, Volume *38*

Supplemental information

Systematic illumination of druggable genes in cancer genomes

Junjie Jiang, Jiao Yuan, Zhongyi Hu, Youyou Zhang, Tianli Zhang, Mu Xu, Meixiao Long, Yi Fan, Janos L. Tanyi, Kathleen T. Montone, Omid Tavana, Robert H. Vonderheide, Ho Man Chan, Xiaowen Hu, and Lin Zhang

Supplementary Figure S1

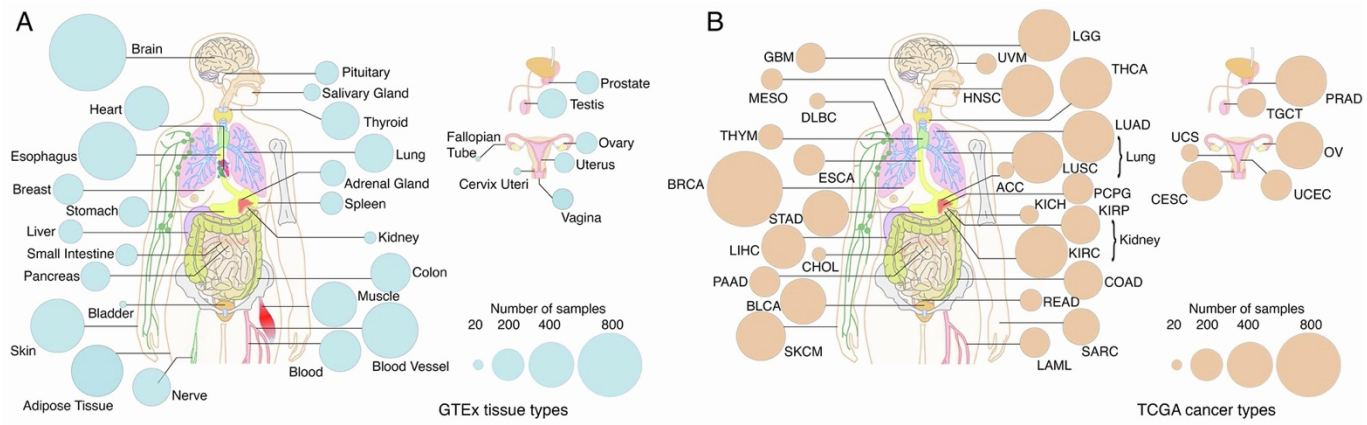

Figure S1. The sample cohorts for expression analysis. Related to Figure 2.
A. and **B.** Summary of the tissue/cancer types and numbers of RNA-seq specimens of the GTEx (healthy normal tissues; A) and TCGA (primary tumor specimens; B) cohorts. The size of each circle corresponds to the number of samples of a given tissue/cancer type.

Supplementary Figure S2

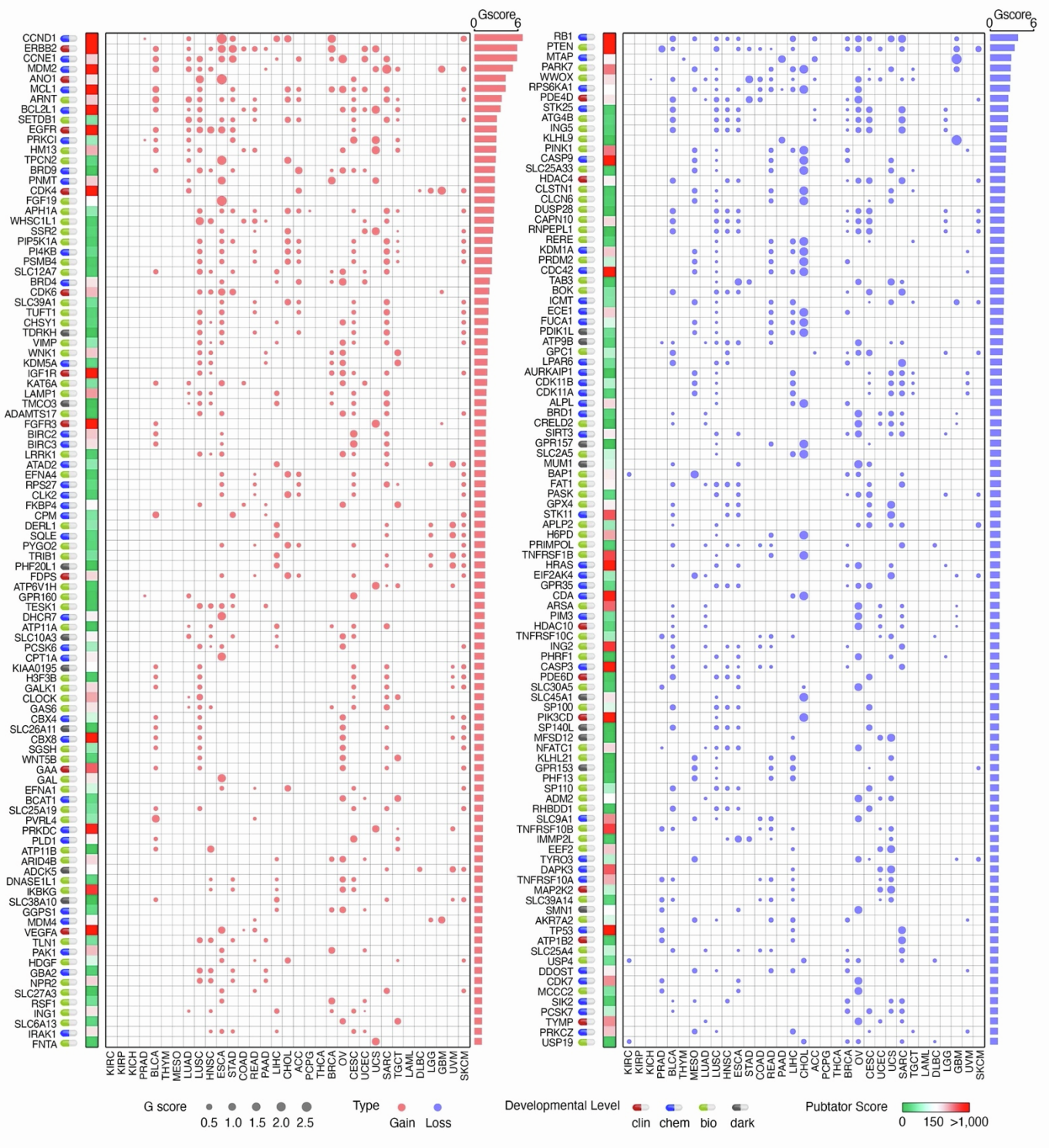

Figure S2. G-scores of top 100 PDGs driven by SCNA. Related to Figure 3.
Details for Figure 3C. Left penal: copy number gain; right penal: copy number loss. Size of bubbles: G-score; red: gain; blue: loss. Heatmap (left) showing the PubTator scores. Green: <150 (understudied genes); red: >150. Target development level of each gene is indicted by color codes. Red: Tclin; blue: Tchem; green: Tbio; grey: Tdark.

Supplementary Figure S3

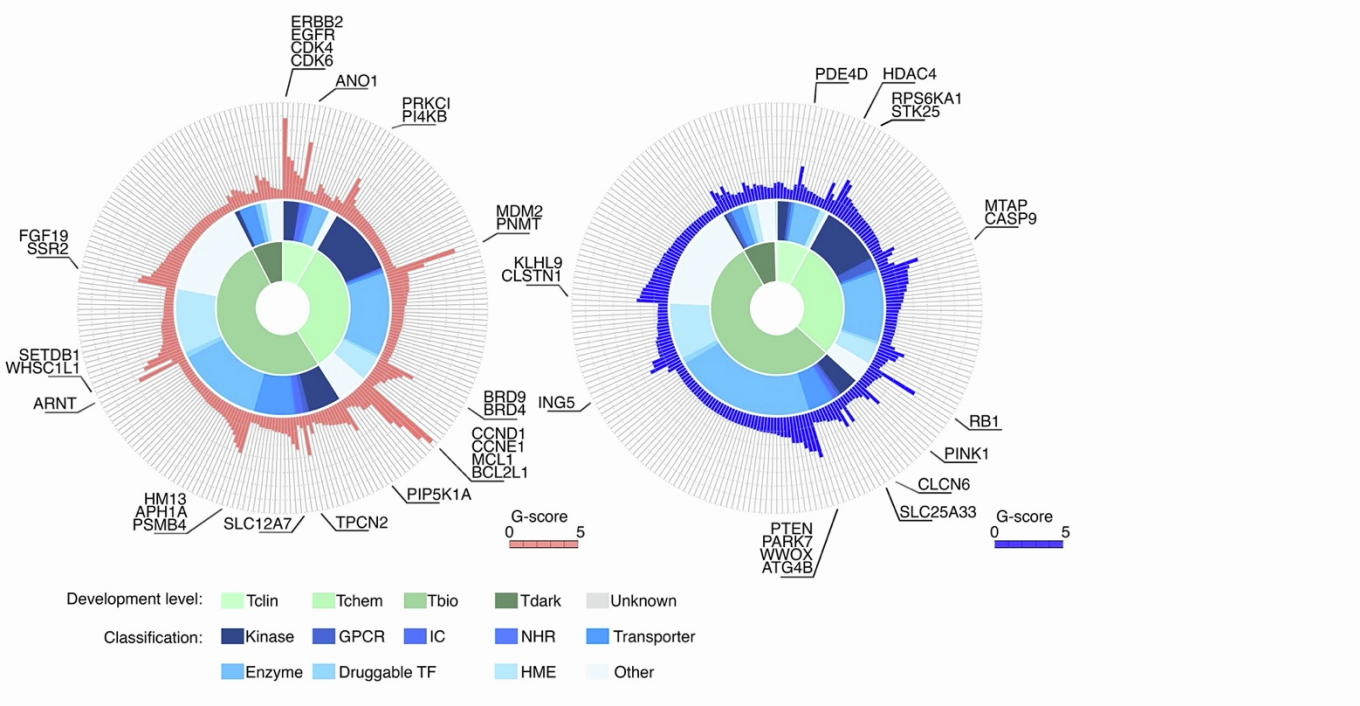

Figure S3. Summary of PDGs driven by SCNA. Related to Figure 3.
Summary of PDGs driven by somatic copy number gain (left) and loss (right), with G-score passed cutoffs at a pan-cancer level. Inner circle: target development level; middle circle: gene classification; outer circle: G-score.

Supplementary Figure S4

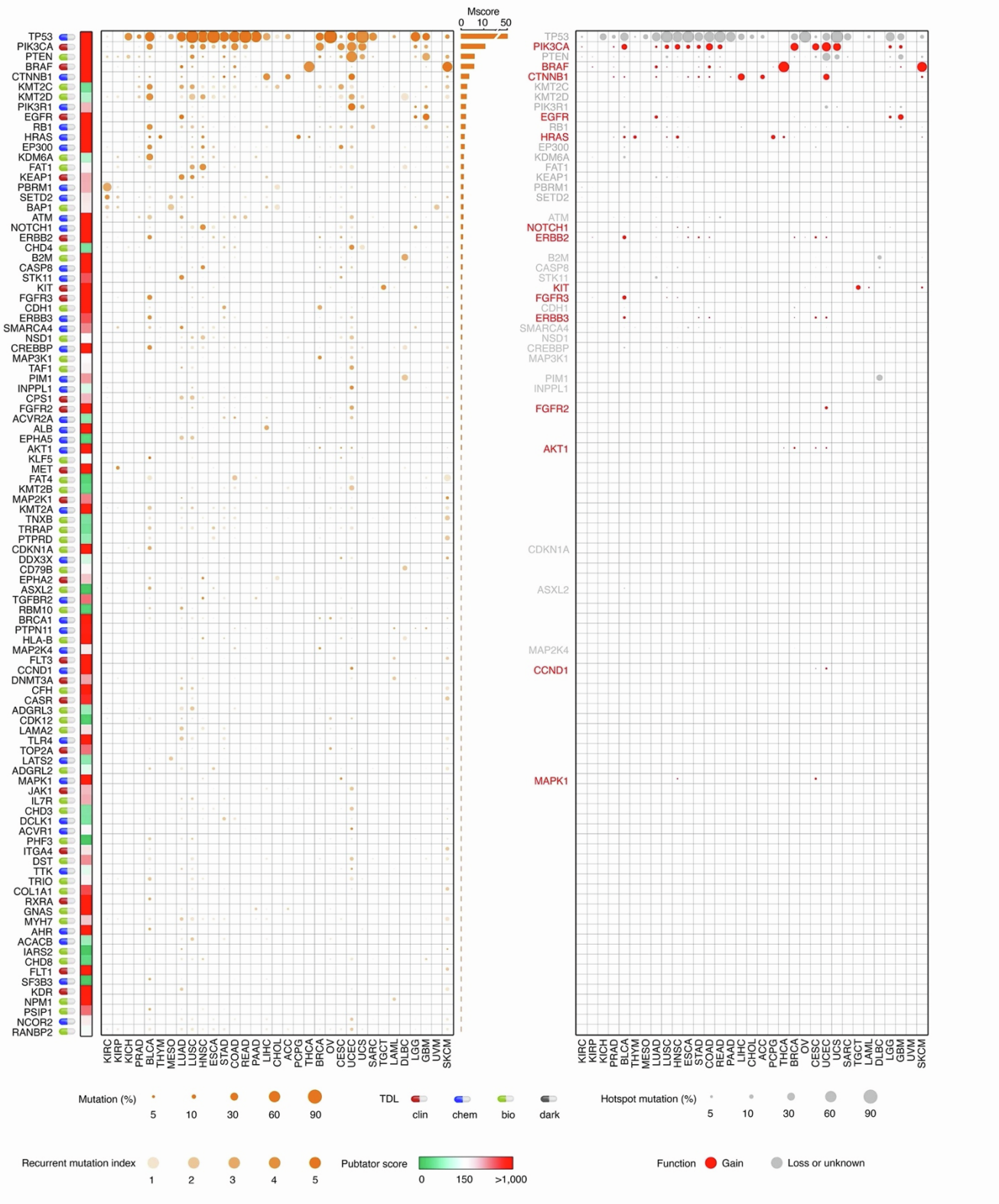

Figure S4. M-scores of top 100 PDGs driven by somatic mutation. Related to Figure 4.
Details for Figure 4C. Left penal: overall mutations; right penal: hotspot mutations. Size of bubbles: mutation frequency; intensity of color: recurrent mutation index. Heatmap (left) shows PubTator scores. Green: <150 (understudied genes); red: >150. Target development level of each gene is indicted by color codes. Red: Tclin; blue: Tchem; green: Tbio; grey: Tdark.

Supplementary Figure S5

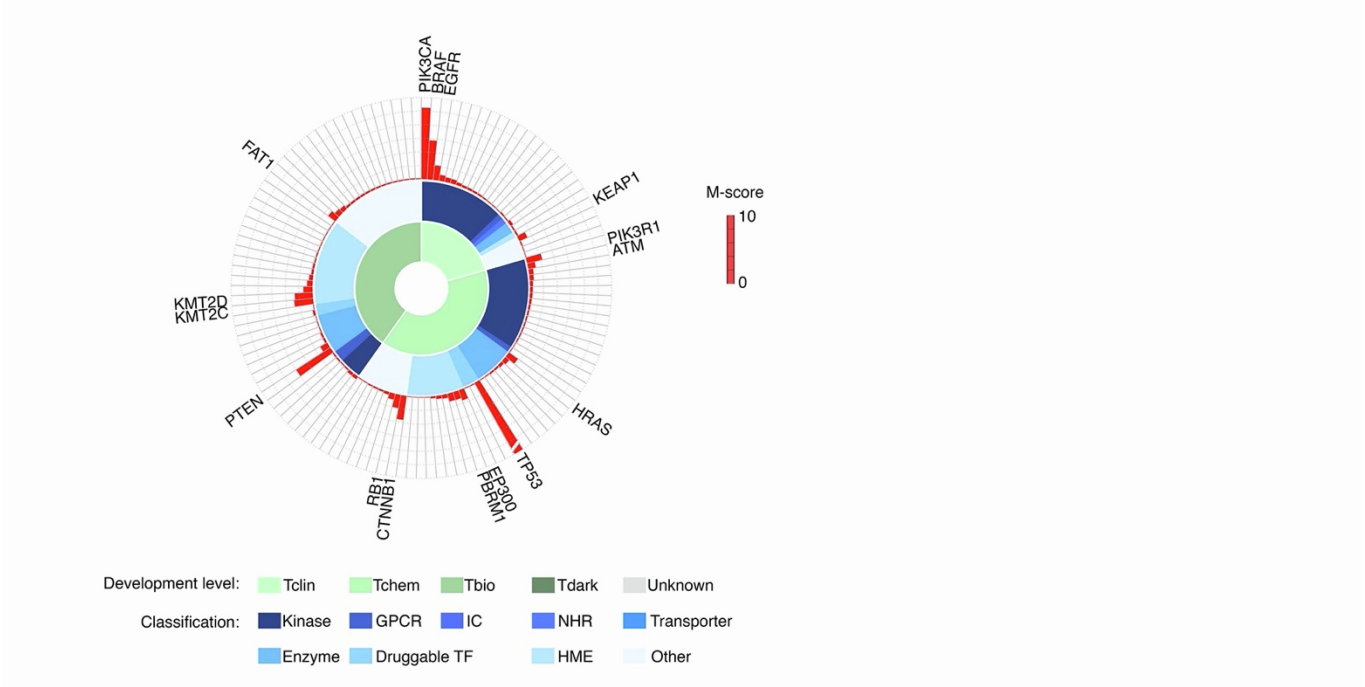

Figure S5. Summary of PDGs driven by somatic mutation. Related to Figure 4.
Summary of PDGs driven by somatic mutation with M-score passed cutoff at a pan-cancer level. Inner circle: target development level; middle circle: gene classification; outer circle: M-score.

Supplementary Figure S6

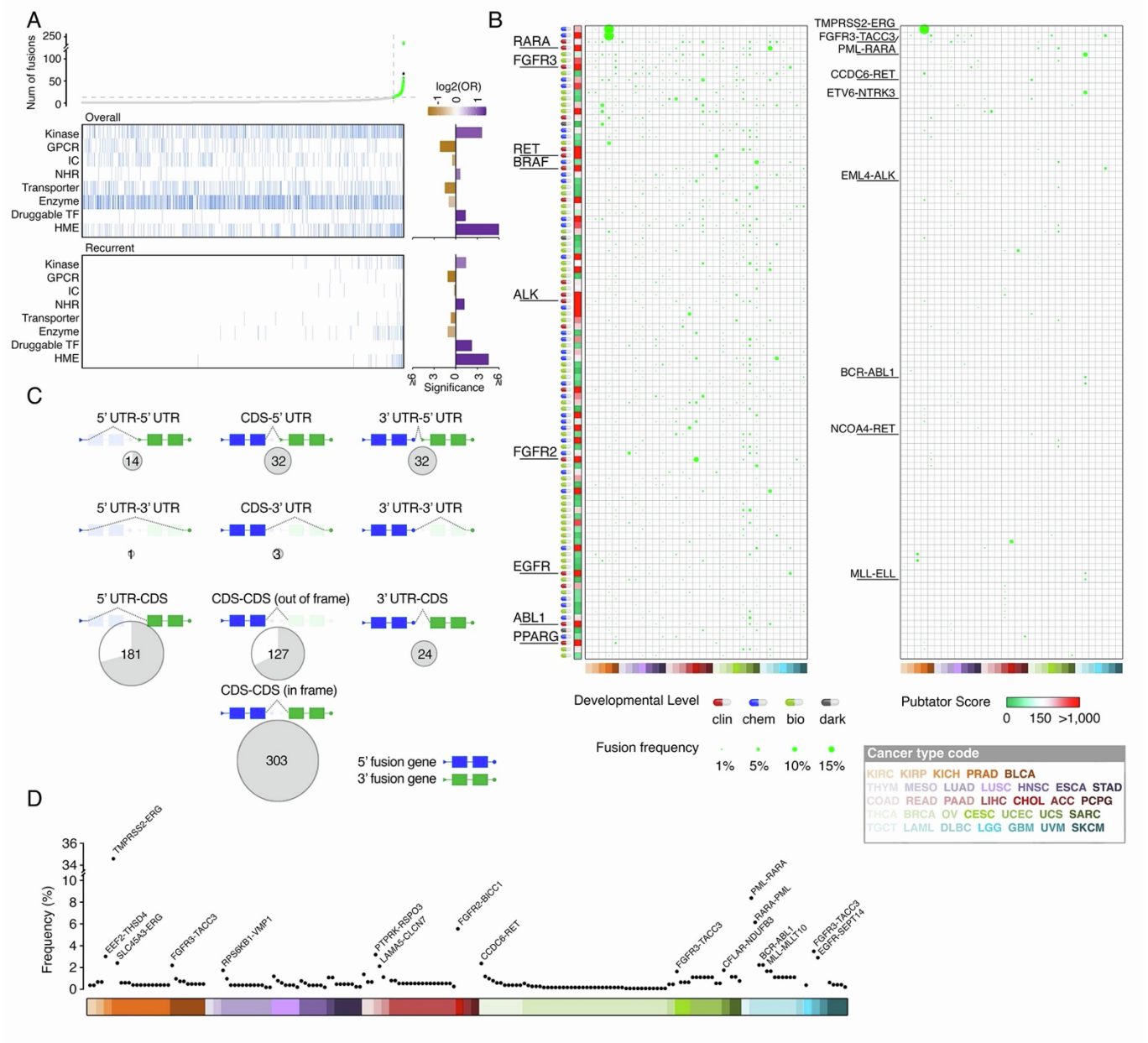

Figure S6. Transcript fusions of PDGs across cancers. Related to STAR Methods (transcript fusion data collection and analysis).

A. Scatterplots show distribution of the numbers of transcript fusion events of all protein-coding genes, which were arranged in ascending order of fusion event numbers. A cutoff determined by the Elbow method was applied to define the fusion events whose numbers were higher than background. Heatmaps show PDGs with fusion events (upper: all fusion events, lower: recurrent fusion events), displayed by gene families and in the same order as the scatterplots. Bar plots (right) show enrichment of fusion PDGs in the corresponding gene families. Purple: enriched; orange: depleted. **B.** Bubble plot showing the top 100 PDGs with the most frequent fusion events (left) and their corresponding fusion pairs (right) across cancers. Size of bubbles: frequencies of fused PDGs or fusion pairs. Heatmap (left) shows the PubTator scores. Green: <150 (understudied genes); red: >150. Target development level of each gene is indicted by color codes. Red: Tclin; blue: Tchem; green: Tbio; grey: Tdark. **C.** Illustration of 10 fusion pattern types identified in the PDGs. Different parts of 2 fused genes are linked by dotted lines, with possible lost and integrated parts shown as faded and solid in color, respectively. Size of circles: number of recurrent fusion events; grey: fusion pairs with the CDS of PDGs; white: fusion pairs without the CDS of PDGs. **D.** Frequencies of possibly targetable recurrent fusions in each

cancer type. Fusions with frequencies greater than 2% in a given cancer type are marked. Bottom color bar indicates cancer types.

Supplementary Figure S7

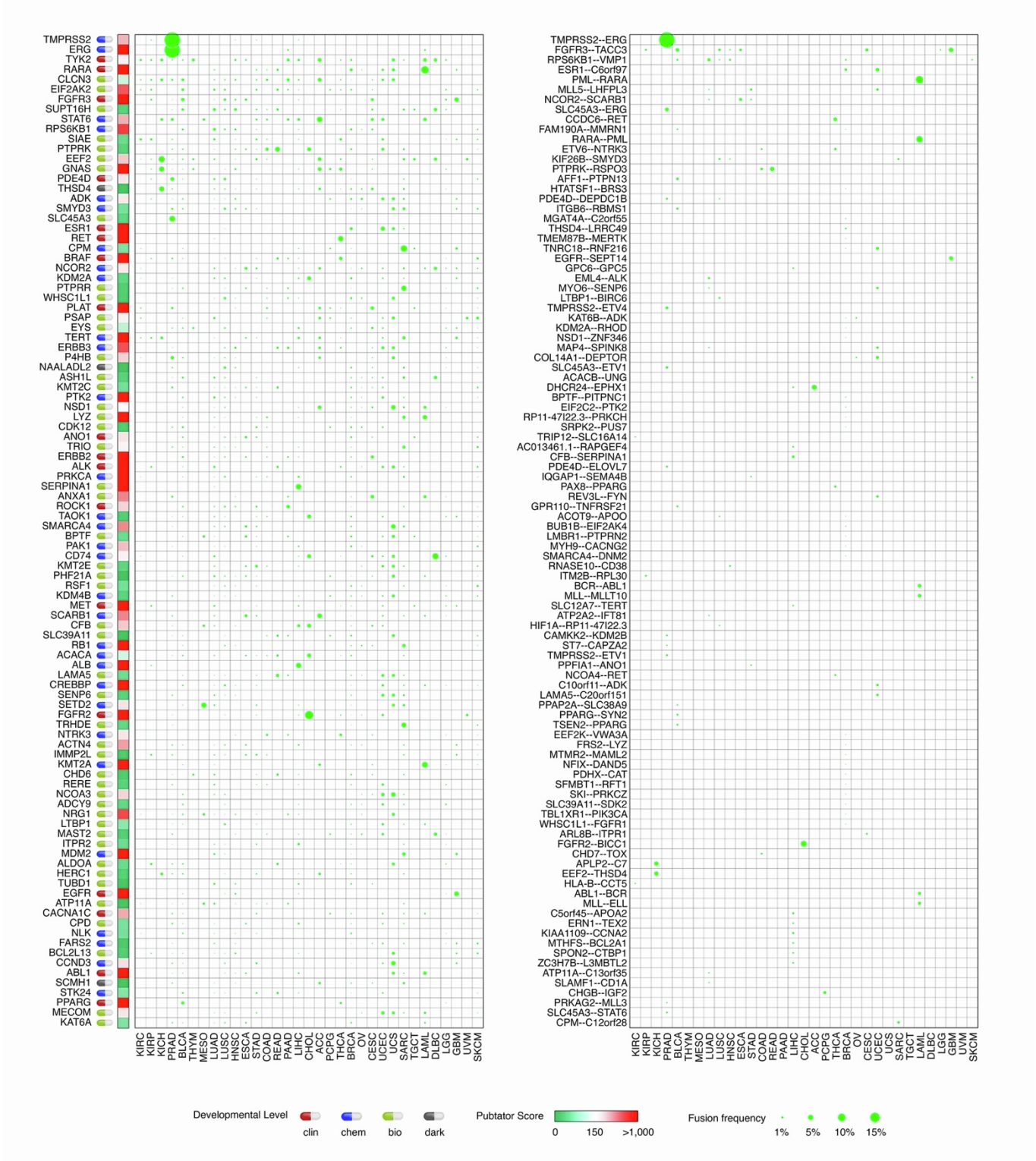

Figure S7. Top 100 PDGs with the most frequent fusion events and their corresponding fusion pairs across cancers. Related to STAR Methods (transcript fusion data collection and analysis).
Details for Figure S6B. Size of bubbles: frequencies of fused PDGs (left) or fusion pairs (right). Heatmap (left) shows the PubTator scores. Green: <150 (understudied genes); red: >150. Target development level of each gene is indicted by color codes. Red: Tclin; blue: Tchem; green: Tbio; grey: Tdark.

Supplementary Figure S8

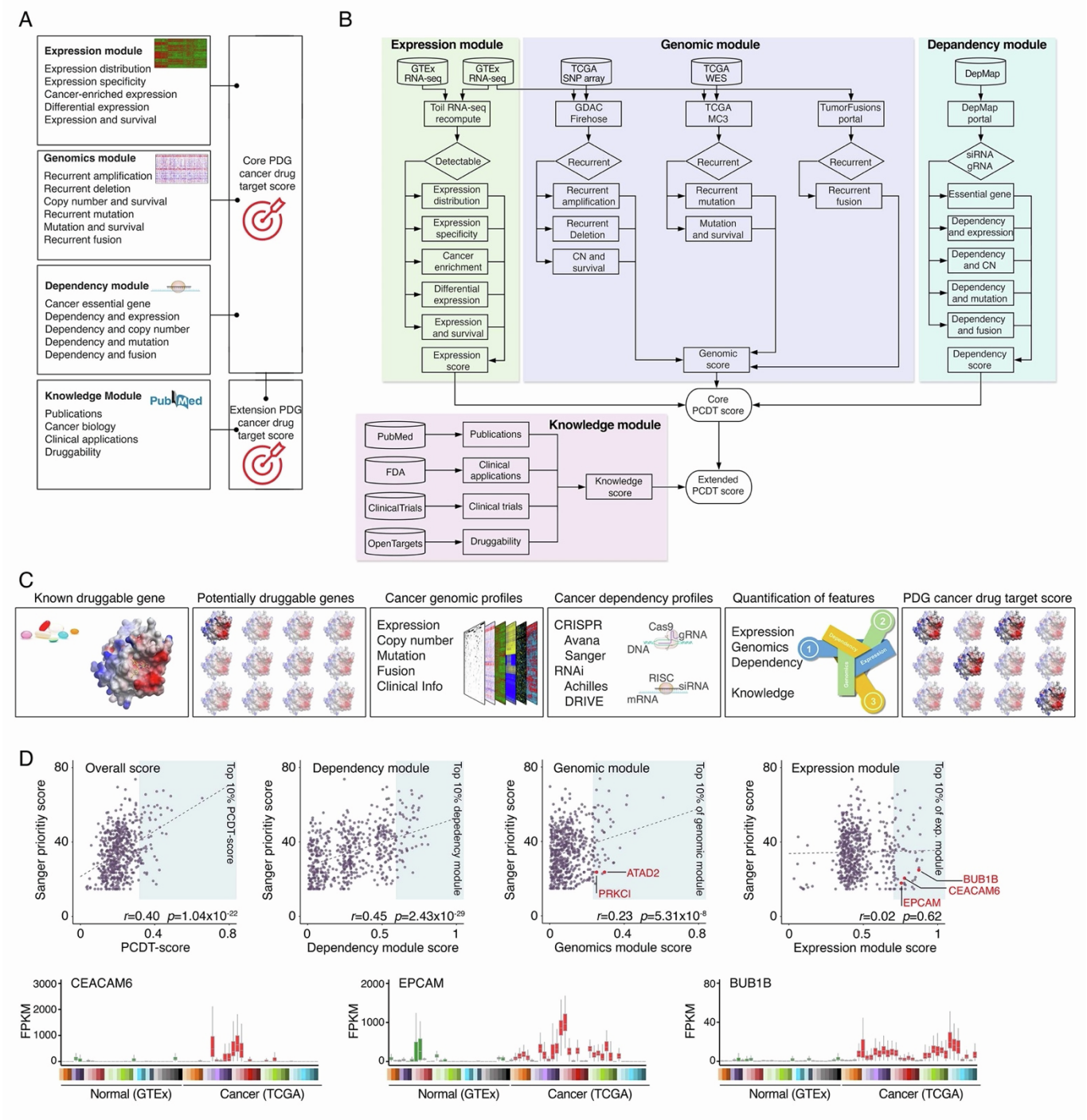

Figure S8. Estimation of the PDG cancer drug target score (PCDT-score) for PDGs. Related to Figure 6.

A. Cancer-related features were quantified and identified by systematically integrating expressional, genomic, functionomic and pharmacological profiles of PDGs across cancers. Three modules were used to estimate the core PCDT-score and a knowledge-based module was further incorporated to generate the extension PCDT-score. **B.** Workflow of estimation of the PCDT-score for PDGs. **C.** Schematic illustrates the strategy for discovering PDGs in the human genome and analyzing their expression, genomic alterations, and cancer dependencies. **D.** Correlation between the PCDT-score and the pan-cancer target priority score. Upper panel: Scatter plot showing the correlations between the pan-cancer target priority score developed by the Sanger and the core PCDT-score as well as its different modules. Lower panel: Expression levels of *CEACAM6*,

EPCAM and *BUB1B* across normal (GTEx) and tumor (TCGA) specimens. Cancer types in which the caPDGs were identified are labeled by colors.