SUPPLEMENTARY MATERIAL:
Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers

Jüri Reimand\textsuperscript{1,*}, Gary D. Bader\textsuperscript{1,*}

October 31, 2012

The Donnelly Centre, University of Toronto, Toronto, Canada.

* Corresponding Author:
Jüri Reimand (Juri.Reimand@utoronto.ca);
Gary D. Bader (Gary.Bader@utoronto.ca)
Supplementary Figure 1: Histogram of mutations in the kinase-substrate phosphorylation network. Most kinase-substrate interactions involve at least one pSNV in the phosphosite or kinase domain of interacting proteins.
Supplementary Figure 2: Histogram of topological features of proteins shows that phospho-mutated proteins tend to have central positions in the kinase-substrate interaction network. Bar height shows proportion of genes with phospho-specific mutations (pSNVs; red) and without pSNVs (black). Top panel displays protein degree in the kinase-substrate network (number of interaction partners of a protein), and bottom panel shows betweenness centrality (number of shortest network paths passing through a protein).
Supplementary Figure 3: Comparison of gene ranking by mutation frequency (total number of missense SNVs) and significance from ActiveDriver predictions (-log10 p-value). Phosphomutation-based model of statistical significance reveals many cancer genes not discovered by frequency-based ranking (dark blue bars). At the level of ActiveDriver significance cutoff (FDR \( p \leq 0.05 \), 58 genes), we found 9 known cancer genes that were not covered by an equivalent set of genes ranked by mutation frequency (median mutation-based rank of these 9 genes is 319). Gene ranking by mutation frequency also includes known cancer genes that are not found by ActiveDriver, however some of these are not phosphoproteins or involve no pSNVs.
Supplementary Figure 4: ActiveDriver results compared with global analysis of mutation significance. We used binomial statistics to evaluate the significance of missense point mutations in genes in comparison to genome-wide missense mutation rates. Tests were conducted and corrected for multiple testing for all analysed genes, separately for each cancer type. Y axis shows the log ratio of gene-specific and genome-wide missense mutation rates. Genes shown in blue have less mutations than expected according to genome-wide average. Genes with asterisks indicate significant p-values of enrichment (FDR $p \leq 0.05$). Only genes considered significant by ActiveDriver are shown. These results demonstrate that only a minority of ActiveDriver results can be recovered by a global estimation of mutation significance. ActiveDriver analysis is therefore complementary to standard, frequency-based predictions of cancer drivers.
Supplementary Figure 5: Top 50 significant GO terms from pSNV analysis highlight hallmarks of cancer. Terms are ranked according to number of covered cancer samples.