Additional File 1 (Supplementary Methods)

1. Identification of perturbed hub genes across OSCC

The gene expression data was obtained from our previous study [1] that identified 1,652 genes differentially expressed between oral squamous cell carcinoma (OSCC) (355) tumor and healthy (131) samples with each sample class labeled as tumor and normal, respectively. The protein-protein interaction (PPI) data was compiled from following publicly available resources: Database of Interacting Proteins (DIP) [2], Human Protein Reference Database (HPRD) [3], Biological General Repository for Interaction Datasets (BioGrid) [4], International Molecular Exchange (IMEx) Consortium [5], contributed by IntAct [6] and MINT [7], and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database [8] (Table 1). While identifying non-redundant interactions, protein identifiers in each database were subsequently unified to approved HGNC official gene symbol; this unification resulted in the acquisition of 2,77,457 unique interactions among 15,794 human proteins. The data files were processed using in-house scripts.

During the identification of dysregulated gene pairs, a threshold value of 30 was selected for defining a gene 'hub'. A total of 1000 permutations were performed to determine the Benjamimi-Hochberg (FDR) [9] adjusted p-value, and the genes which had an FDR adjusted p-value less than 0.05 were considered.

2. Disease enrichment of the candidate genes

A random sampling was performed to test the probability using an existing method [10], where same number of known cancer genes was randomly picked; it was performed to estimate whether these known cancer genes included in the previous results were statistically significant. The whole procedure for this method is as follows: first genes equivalent to candidate disease gene numbers was randomly selected from the entire expression profiling gene set. Then the number of known cancer genes included in the random samples was counted and this whole random sampling procedure was performed for 10^5 times. Finally, the probability that one random sampling might contain a greater or equal number of known cancer genes than in study samples was defined as p-value of the candidate disease genes.

3. Selection of relevant genes by ensemble based feature selection

The quality of individual models was assessed using various standard parameters such as area under the receiver operating characteristic (ROC) curve, sensitivity (SE), specificity (SP), overall accuracy (Q), and Matthew's correlation coefficient (MCC) using the following equations:

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

$$Q = \frac{TP + TN}{TP + TN + FP + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}$$

where, the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) represent number of correctly predicted inhibitors, correctly predicted non-inhibitor, non-inhibitor wrongly predicted as inhibitor, and inhibitor predicted wrongly as non-inhibitor, respectively.

References

1 V. Randhawa and V. Acharya, *BMC Med. Genomics*, 2015, **8**, 39.

2 L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg, *Nucleic Acids Res.*, 2004, **32**, D449–51.

3 T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767–D772.

4 C. Stark, B.-J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski and M. Tyers, *Nucleic Acids Res.*, 2011, **39**, D698–704.

5 S. Orchard, S. Kerrien, S. Abbani, B. Aranda, J. Bhate, S. Bidwell, A. Bridge, L. Briganti, F. S. L. Brinkman, F. Brinkman, G. Cesareni, A. Chatr-aryamontri, E. Chautard, C. Chen, M. Dumousseau, J. Goll, R. E. W. Hancock, R. Hancock, L. I. Hannick, I. Jurisica, J. Khadake, D. J. Lynn, U. Mahadevan, L. Perfetto, A. Raghunath, S. Ricard-Blum, B.

Roechert, L. Salwinski, V. Stümpflen, M. Tyers, P. Uetz, I. Xenarios and H. Hermjakob, *Nat. Methods*, 2012, **9**, 345–50.

6 H. Hermjakob, L. Montecchi-palazzi, C. Lewington, S. Mudali, S. Kerrien, S. Orchard, M. Vingron, B. Roechert, P. Roepstorff, A. Valencia, H. Margalit, J. Armstrong, A. Bairoch, G. Cesareni, D. Sherman, R. Apweiler, C. Cb and U. T. Vergata, *Nucleic Acids Res.*, 2004, **32**, 452–455.

7 A. Chatr-aryamontri, A. Ceol, L. M. Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli and G. Cesareni, *Nucleic Acids Res.*, 2007, **35**, D572–4.

8 A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering and L. J. Jensen, *Nucleic Acids Res.*, 2013, **41**, D808–15.

9 Y. Benjamini and Y. Hochberg, *J. R. Stat. Soc.*, 1995, **57**, 289–300.

10 X. Liu, Z.-P. Liu, X.-M. Zhao and L. Chen, *J. Am. Med. Informatics Assoc.*, **19**, 241–8.