In silico profiling of activating mutations in cancer

Keshav Patil¹, Krishna Suresh² and Ravi Radhakrishnan¹,²

¹ Department of Chemical and Biomolecular Engineering; ² Department of Bioengineering, University of Pennsylvania, Philadelphia PA

In silico Oncology for Personalized Medicine

- Progress in genomics has raised expectations in personalized cancer research which means it is becoming increasingly feasible to personalize treatment according to the composition of patient’s genome.
- The advancement in the sequencing techniques and subsequent large scale sequencing projects have generated copious data on somatic mutations in cancer.
- The majority of these somatic mutations incur little or no functional consequence on tumor progression (neutral passengers), while few (driver mutations) provide a selective advantage to cancer cells thereby making it necessary to distinguish driver mutations from the large number of neutral mutations.
- Profiling the mutations in the tumor cells and targeting the associated signaling pathways guides in rationalized targeted drug design therapy aimed at personalized cancer treatment.
- Kinases are frequently mutated proteins in cancer that account for 2% of all the mutations in the COSMIC database taking into account clinical data subject to whole genome sequencing only.
- Mutations in the structural domains that activate kinases and upregulate cell proliferation are well represented among known oncogenic driver mutations. Many have been clinically observed and experimentally verified, but determining a priori which mutations in a given kinase are activating (facilitated through computational techniques) is still quite challenging but necessary, as exhaustive experimentation to study the mechanism of each mutation is inefficient and expensive.
- Existing in silico methods mostly use sequence-based methods such as SIFT or additionally information from 3D crystalized structures like PolyPhen-2 for predicting whether a mutation is deleterious to protein function.

Mining Molecular/ Mutation Data from Cancer Atlases

Computational Methods: Enhanced Molecular Dynamics (MD), Machine Learning

Capturing the transition from inactive form of the anaplastic lymphoma kinase (ALK) to its active form (see Fig. on left) involves unavailable simulation times. This is resolved using enhanced MD techniques.

In metadynamics, an external history-dependent bias potential is constructed in the space of a few selected degrees of freedom (q), called collective variables (CVs). This potential is built as a sum of Gaussian kernels deposited along the trajectory in the CV’s space. This enables to accelerate the system along CVs in a single simulation. The free energy landscape from metadynamics is depicted in Fig. to the left. Four zones are identified for further analysis.

Results

Hbond occupancy difference (Fig. left) is computed for 45 ALK mutations that we studied to classify as activating or not. Hbond occupancy as a single feature has a balanced accuracy of 81% as determined from comparison with experiment (Fig. left).

Ongoing Work

To compute the H-bond occupancies and other distinguishing biophysical properties for large datasets of mutated protein structures, it is required that the operation is streamlined with automation; so we have developed an open source computational platform: BioPhysCode (see Fig. left) that essentially facilitates executing and analyzing large numbers of trajectories.

Conclusion and future work

- Methods that employ biophysical properties as features could boost the prediction performance if used in conjunction with biochemical and sequence based features.
- The mutations classified as activating need to be investigated in the presence of inhibitor drugs for reduced deleterious activity which thereby would help oncologists in personalizing therapy.

References


Acknowledgements

We thank Dr. Jin Park, the labs of Mark Lemmon and Yael Mosse for experimental support. The research is supported by NIH, ERC, XSEDE.