FUNCTIONAL ANALYSIS OF
TYROSINE KINASE MUTANTS IN
CANCER

By
Prashant Bharadwaj Kalvapalle

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science and Engineering

Baltimore, Maryland
August, 2016

© 2016 Prashant Kalvapalle
All Rights Reserved
Abstract
Cancer or uncontrolled cell division which can invade into other tissues is today among the leading causes of morbidity and mortality worldwide. Genomic alterations in protein coding genes are a major feature of cancerous cells. But in general, any tumor has many mutations that arise spontaneously during cell division, of which very few have a causal impact on oncogenesis (1). More than 100-fold variation of such background mutation rate between different individuals, different regions of the genome and different cancers accentuated by mutated DNA repair pathways, exposure to mutagens, chromosomal abnormalities, all of which affect cancer cells make it very difficult to statistically tease apart causative driver mutations from inert passenger mutations even with the most recent advances in sequencing to identify somatic mutations in tumor cells (2). Hence there is a need to incorporate experimental data involving functional implications of mutations to pin point the driver mutations and characterize the causal mechanism (3).

The goal of this project is to develop an experimental workflow to interrogate biochemical functional implications of mutations in coding regions in order to characterize their driver potential. We have chosen to work with protein tyrosine kinases, owing to their importance in cell signaling, acute involvement in many cancers, and the promise of existing therapeutic successes targeting these pathways. We test for protein activity in high throughput on Human Proteome (HuProt) arrays, covering 19,000 human proteins and splice variants. This approach is helpful to know the catalytic activity and substrate preference of different mutant kinases, giving valuable insights into driver like potential depending on the pathways it acts on and the variation between the wild type enzyme and the mutant. Success with this platform will enable deployment of the same pipeline to other protein families with known biochemical functions and potentially to clinical applications in personalized medicine.
Thesis Committee

Joel S. Bader, Ph.D (Primary advisor)
Professor, Biomedical Engineering
Interim Director, High Throughput Biology Center
Johns Hopkins University

Heng Zhu, Ph.D (Secondary advisor)
Professor, Pharmacology and Molecular Sciences
Johns Hopkins University

Kevin Yarema, Ph.D (reader)
Professor, Biomedical Engineering
Johns Hopkins University
Acknowledgements

I owe my little contribution towards the vast body of knowledge to Professor Joel Bader, Dr. Hee Sool Rho (post doc) and Dr. Heng Zhu for funding, guiding and helping me on various aspects of the project.

I want to thank all the Zhu lab members Cedric, Pinaki, Devlina, Guanda, Shuang, Guong and Yi Hwang for making my stay in the lab enjoyable. Special thanks to Cedric and Pinaki for valuable discussions on the project and Guanda for making sure my thesis is complete!

My stay at Hopkins wouldn’t be complete without mentioning all the amazing friends and awesome time spent with them including the BME folks, and also my great roommates Shravan, Lijo, Jacinth and Simar who kept my morale high, made sure I was well fed and provided endless entertainment and enriching discussions on life.

I wish to dedicate the thesis to my parents but for whose support, inspiration and motivation I would never have made it to Hopkins.
# Contents

Abstract: ............................................................................................................................... ii

Thesis Committee: ................................................................................................................. iii

Acknowledgements: .............................................................................................................. iv

Contents: ................................................................................................................................... v

List of Tables: .......................................................................................................................... vi

List of Figures: ........................................................................................................................ vi

Introduction: ............................................................................................................................ 1

Background and Literature: ................................................................................................. 4

  Targeted therapy .................................................................................................................. 4

  Driver mutations identification ......................................................................................... 5

  Protein arrays ...................................................................................................................... 6

Materials and Methods: ......................................................................................................... 9

  Cloning ................................................................................................................................ 9

  Expression and protein purification .................................................................................... 10

  Verification - Protein expression and activity ................................................................. 10

  HuProt- Kinase assay ......................................................................................................... 11

  Data analysis ...................................................................................................................... 12

Results: ................................................................................................................................ 14

Discussion: ............................................................................................................................. 24

Appendix: Protocols ................................................................................................................ 26

  1) Quick change mutagenesis: ......................................................................................... 26

  2) Yeast Cell Preparation ................................................................................................. 27

  3) Yeast Transformation .................................................................................................... 28

  4) Protein Expression and Purification ............................................................................. 29

  5) Kinase reaction ............................................................................................................... 32

Appendix: Mutants ................................................................................................................... 33

Bibliography: ........................................................................................................................ 35

Curriculum Vitae: .................................................................................................................. 37
List of Tables

Table 1: Number of hits: Jak2 batch .................................................................23
Table 2: Number of hits: Fgfr2 batch ...............................................................23
Table 3: Mutagenesis reaction mixture ............................................................26
Table 4: PCR thermocycling conditions ............................................................26
Table 5: LiAc/TE mixture .................................................................................27
Table 6: Peg mixture for yeast transformation ..................................................28
Table 7: Kinase reaction mixture .......................................................................32
Table 8: List of mutant notation .......................................................................33

List of Figures

Figure 1: BsrG I digestion of destination clones ...............................................14
Figure 2: Western blot - All purified kinases .......................................................15
Figure 3: Western blot - purified Jak2 and Fgfr2 kinases ....................................16
Figure 4: Comparison of replicates of Negative Controls with Jak2 batch ...........16
Figure 5: Comparison of replicates - Jak2 arrays ............................................17
Figure 6: Jak2 vs Negative control .................................................................17
Figure 7: J1 vs Jak2 ......................................................................................18
Figure 8: J2 vs Jak2 ......................................................................................18
Figure 9: J3 vs Jak2 ......................................................................................19
Figure 10: J4 vs Jak2 .....................................................................................19
Figure 11: Comparison of replicates – Fgfr2 arrays .........................................20
Figure 12: Fgfr2 vs Negative control ...............................................................20
Figure 13: F1 vs Fgfr2 .................................................................................21
Figure 14: F2 vs Fgfr2 .................................................................................21
Figure 15: F3 vs Fgfr2 .................................................................................22
Figure 16: F4 vs Fgfr2 .................................................................................22
Introduction

When a normal somatic cell accumulates random mutations during cell division, by damage or by other processes, certain mutations may confer a selective advantage to the cell and may predispose the clones of this cell to shift balance between the rate of cell division and cell death and to acquire more mutations and eventually give rise to a metastatic tumor. This rate of mutation may be exacerbated by environmental factors like diet, lifestyle, and exposure to mutagens like UV light, smoking etc. which may increase the natural mutation rate and hence increase the risk of cancer. Similarly, any inherited germline mutation may make it easier to acquire new mutations, for example BRCA1 (DNA repair gene) mutations, in addition to the possibility of increasing cell fitness directly (2).

Most cancers’ origins can be traced to a relatively sparse number of somatic mutations in the genome that confer a selective growth advantage and help drive the transformation of the cells harboring them and hence are called driver mutations (1). These mutations generally affect critical cellular processes like cell division, adhesion to tissue extracellular matrix, metabolism, apoptosis, epigenetics, DNA repair etc. Disruption of normal processes can lead to increasing proliferative pathways or resistance to cell death signals which eventually tips the balance of the rate of [Cell division – Cell death] upwards, a hallmark of cancers.

Different cancer sub-types in different individuals may be driven by different driver mutations and hence identification of causal driver mutations can be immensely useful to develop diagnostic protocols and targeted therapies in the era of precision and personalized medicine (1).
Since mutations arise stochastically during cell division, most of them have no functional consequence or selective advantage for cells \(^1\). These mutations termed as ‘passengers’ substantially outnumber the drivers contributing to a large background of somatic mutations occurring even in non-cancerous cells \(^1\). One way to reduce the statistical signal from the mutational background is to estimate the frequency of mutations occurring in individual genes. Genes in cancer cells that harbor many somatic mutations that occur more frequently than random may give an evolutionary advantage and hence may be undergoing selection. But the random mutation background varies more than 100-fold between different individuals and also between different regions of the genome even in normal cells, and cancer cells are expected to have more variations in the background depending on the cancer type and environmental trigger \(^4\). This makes the definitive identification of causal mutations statistically challenging. And hence there is a need for improving the methods of identification of such mutations from personal genome or exome sequencing in a timely and cost effective manner. Experimental data to find functional effects of various mutants can be very useful in this context to inform the computational methods and add additional filtering to predict the drivers with more accuracy.

Among the well-studied mutations which enhance cell division, mutations in proteins involved in transducing signals from outside the cells to the inside, eventually into the nucleus to trigger cell division or other cancer related processes, are over-represented as cancer drivers. This is also evidenced with many cancers frequenting tissues responsive to hormonal growth cues like breast, prostate etc. Most of the signaling happens with only slight modification to the proteins, for example an attached phosphate or sugar which switches the protein into an active state and triggers downstream pathways. Hence the mutant versions of these are often found to be highly enriched in many cancer genomes.
Protein kinases which modify target proteins by adding phosphate groups have been especially well studied being one of the earliest identified cancer related pathways and have also been the targets of very successful targeted anti-cancer therapies like Gleevec and Herceptin. Kinases are also highly enriched among known driver mutations.

The functional effect of these proteins in the cell depends on the

- **Substrate selectivity**: Identity of the target proteins that are phosphorylated

- **Catalytic Activity**: Rate of phosphorylation of each target protein

Together these decide which pathways the mutant kinase interacts with and hence which cellular processes it can influence which might give a clue as to its potential for being a 'driver'. Both of these phenotypes could be found using in-vitro experiments involving the protein kinase of interest and the target proteins with proper assay conditions.

In a way this approach is a key step of a multiscale systems biology analysis, starting from the genotype going through molecular phenotype (protein level), to the pathway level phenotype to cellular phenotype and ultimately to the tissue and organism level phenotype.

Our approach is to use functional protein microarrays, which have individually expressed and purified human proteins spatially separated and spotted on a glass slide in native confirmation. The current work uses HuProt arrays marketed by CDI laboratories, which has 19,000 unique human proteins purified from yeast (*Saccharomyces cerevisiae*).

This platform could also be useful for testing drug libraries for kinase inhibitors, especially handy for drugs that target protein-protein (kinase-substrate) interaction sites or allosteric sites within the kinase domain.
Background and Literature

Targeted therapy
Before the knowledge of genetic or biochemical basis of cancers, the only available treatment modalities were surgical removal of tumor and chemotherapy. Chemotherapy uses cytotoxic drugs targeting DNA synthesis and cell division which indiscriminately kill any rapidly dividing cell. Along with cancer cells, rapidly dividing cells include healthy cells lining the gut, mouth, hair follicles, skin etc. The broad activity contributes to a lot of side effects and limit the possible drug dosing. Since not all cancer cells are likely to be killed, there may be large numbers of relapses surviving tumor cells. Interestingly drug resistance has also been observed in the relapsed tumors (5).

A better approach would be more selective, going after the unique features of cancer cells and the pathways that are causally implicated in oncogenesis. Imatinib Mesylate was the first rationally developed and targeted therapy for cancer and has demonstrated tremendous clinical and market success in the treatment of chronic myeloid leukemia (CML) and gastrointestinal stromal tumors (GIST). It inhibits the activity of BCR-Abl kinase in CML and c-Kit or platelet derived growth factor receptor-a in GIST, which are receptor tyrosine kinases with constitutive activity due to mutations or gene fusions that caused hyper-proliferative phenotypes (6).

By a similar approach epithelial growth factor receptor (EGFR) targeting drugs have varying clinical response on lung cancer depending on the tumor’s dependence primarily on mutant EGFR. If there are additional driving mutations in the tumor, it is noticed that inhibitors reduce the EGFR activity in cells but cannot shrink the tumor (7).
It was also noticed that the identity of genetic alterations observed in melanomas depended on the levels of sun exposure and also location of the tumor site. The mutations often included BRAF or N-RAS or a copy number change in cyclin-dependent kinase 4 (CDK4) or cyclin D1 (CCND1) all of which are along the same pathway (8).

Hence for therapy to be targeted to the cancer in the specific patient, the exact identity of the proteins and the activating mutations that drive tumor progression are essential for administering a therapy that is tailored to a patient.

**Driver mutations identification**

By definition, driver mutations are necessary for the development of cancer implying that one of the earliest cells that start the tumor would have acquired the primary driver mutation and all its clones are likely to have the mutation due to the founder effect. Hence the frequency of the driver mutant is likely to be high in a population of tumor derived cells (late stage tumor).

Such mutations can be identified by sequencing the exome from late stage primary tumors and by looking at the frequency of occurrence of each variant at every position in the exome. A plot of such frequency distributions, at each sequence position would look like a hilly landscape with a few mountains (high frequency mutants) and many smaller hills. The mountains though easily identifiable, do not yet account for all the possible driver mutants which means that lower frequency mutations (with small fitness advantages) might also cause cancer by acting in combination, for example acting along the same pathway result in a larger fitness advantage like a driver (9). There have also been attempts to include pathway information while classifying drivers and such methods can identify more drivers than with frequency information alone (9).
In order to identify if a particular variant is a driver, statistical methods are used to estimate the likelihood of the observed frequency of a variant arising from a passenger mutation using and empirical or theoretically calculated (10) frequency distribution of passenger mutations. Since it is physically impossible to empirically estimate the frequency of non-selected passenger mutations as it would require hypothetical cells which have been exposed to the same environment and mutagens as a real tumor without becoming cancerous, it is approximated by other measures like the rates of synonymous mutations (no change in the protein sequence) etc (11). For low frequency mutations the effectiveness of the classification is very sensitive to the accuracy of the estimate of passenger mutation frequency and hence the confidence in the classification is limited by high variability of passenger mutation across tissue types, individual genes and from tumor to tumor depending on the number of generations, intrinsic mutability etc. (10)(11).

Hence there is a need to incorporate additional information regarding the mutants that cannot be derived from the mutation frequency alone. Such information could be from bioinformatics approaches: evolutionary conservation of residues, domain architectures, pathways etc. Direct wet lab data provides important additional information: gene and protein expression, protein-protein interactions and functional information from in vitro, cell culture or model organism experiments (10).

**Protein arrays**
Protein arrays which were inspired from DNA microarrays can be extremely useful in performing in vitro experiments involving protein-protein interactions or enzymatic assays on a large set of individual proteins in high throughput.
Protein arrays that have functionally active, non-denatured proteins are referred to as functional microarrays. To make such arrays the proteins have to be purified in native confirmation in a high throughput manner (in 96 deep well plates) using conditions in which most of the proteins purify effectively without denaturation. Stringent quality controls regarding the purity, concentration, solubility etc. have to be set in order to ensure batch to batch reproducibility in the arrays. Since the actual quantity of protein on the array is very small (on the order of nano-litres), many microarrays can be printed from the same batch of proteins, and protein arrays use at-least 3 orders of magnitude less protein per assay compared to the micro-litre scale of typical liquid phase reactions.

A main advantage of protein arrays is that much of the cost and effort goes into purifying and arraying the individual proteins and there is minimal effort on the part of the researchers who use the array to perform functional assays. Also most of the existing infrastructure for DNA microarrays including chip scanners, fluorophores and analysis algorithms can also be used for protein microarrays.

The current work uses HuProt arrays V3.0 (CDI laboratories), which has 19,000 unique human proteins and splice variants purified using a yeast expression system (Saccharomyces cerevisiae).

The advantage of using these arrays is the high coverage of most of the human proteome (~90% coverage) and adaptability of the method. It is important to work with the whole proteome because mutants might have activity on non-native substrates which might have implications for driver activity. These would be missed by following a common practice of looking at only a small set of known substrates. Also using full length proteins in native confirmation is a closer mimic of native conditions inside the cell than a library of small
peptides, because the conformational epitopes needed for recognition might be missing in
the peptide. The HuProt screening approach could also be used for drug screening assays
especially when the drug binds to sites other than the catalytic site of the kinase.

Possible caveats include the fact that we can only look at binary interactions. Mutants whose
interactions with scaffolding proteins regulate protein activity, for example the activation by
an upstream kinase or other post translational modifications like O-GlcNAcylation which
could change activity would go unnoticed. Interactions that are otherwise not possible
because of differential expression or compartmentalization would also show up as positive
spots on the array and might give misleading targets. Another concern is that extracellular
and secreted proteins are likely to be under-represented in the collection or have low
expression or improper folding\(^{(12)}\).

Some of these issues are addressed with advances in Mass Spectrometry methods (critically
cost reductions) using phosphor-tyrosine enriched protein extracts from cell cultures which
can then be quantitatively analyzed. This might give a more realistic picture of what the
mutant does inside the cell and used in tandem with microarray approach might yield more
robust data for exploratory purposes. But due to the inherent high throughput nature of the
microarray technology, it is more amenable to translation as a diagnostic tool rather than
being limited to research or laboratory settings.
**Materials and Methods**

As a proof of concept of the method, we studied the effectiveness of a HuProt assay to detect differential activity of protein-coding mutations kinases with known driver ability. There are 148 annotated protein tyrosine kinases based on gene ontology molecular functions annotations. Of these, 64 tyrosine kinases were known to be amenable to purification and activity on HuProt arrays. These were considered for the pilot study to eliminate the need to optimize purification conditions. When filtered by known cancer driver ability identified in a previous work \(^2\), we were left with 6 protein tyrosine kinases that are known drivers and have established activity on HuProt arrays. Mutations in these kinases occurring with a high frequency were hand-picked from Catalogue of Somatic Mutations in Cancer (COSMIC) database. We narrowed down to 53 different mutations encompassing 6 different tyrosine kinases: FGFR2, JAK2, KIT, MET, RET and MAP2K1.

**Cloning**

Entry clones having full length gene in pDONR vector for the 6 tyrosine kinases were ordered from Invitrogen’s Ultimate Human ORF collection stored at the High-Throughput Biology (HIT) center at Johns Hopkins University.

These entry clones were subjected to site directed mutagenesis using Agilent’s Quikchange lightning kit \(^{14}\) with desired mutagenic primers and DPN1 digested to eliminate methylated native vector and subsequently transformed into E. coli. (For details refer- Appendix: Protocols -1)

The clones were verified using restriction digestion using BsrGI (New England BioLabs) and the inserts were also Sanger sequenced to verify the presence of expected mutation.
Mutant genes were subsequently transferred into an in-house yeast expression vector PegH-A under a GAL1 promoter with an N-terminal GST tag, using Invitrogen Gateway cloning system - LR recombinase \(^{(15)}\) and the resulting vector was restriction digested using BsrG1 to verify proper insertion. Verified clones were transformed into the auxotrophic baker’s yeast \((Saccharomyces cerevisiae)\) strain Y258. (For details refer- Appendix: Protocols -2, 3)

**Expression and protein purification**

Yeast clones that grew under auxotrophic uracil deficient selection plates were grown in glucose deficient media with raffinose as the carbon source and then induced using galactose.

Yeast were lysed using glass beads under violent agitation using paint shaker at 4°C with protease inhibitors to minimize protein degradation and phosphatase inhibitors (Orthovanadate, sodium fluoride) to preserve native phosphorylation that might be required for kinase activity.

Proteins were purified using glutathione-sepharose affinity chromatography in 96 well filter plate format as described previously \(^{(12)}\) and eluted with 40mM Glutathione solution. (For details refer- Appendix: Protocols -4)

**Verification - Protein expression and activity**

Purified kinases were subjected to a kinase reaction at 30°C for 60 mins in a reaction buffer with 1mM ATP, 10mM MgCl2 and 1mM MnCl2 along with other components to mimic cellular conditions. Expression and activity of the purified proteins were measured by western blotting using Mouse anti-GST (Abcam) and rabbit anti-phospho-Tyrosine (p-Tyr-1000, Cell Signaling Technology) antibodies respectively \(^{(16)}\).
HuProt- Kinase assay

Owing to the costs of the protein arrays, we decided to start with two kinases for proof of concept validation. We chose two kinases which had maximum substrates identified for WT protein on previous HuProt experiments, FGFR2 and JAK2. Subsequent HuProt experiments were carried out only on the 5 mutants of JAK2 and 5 mutants of FGFR2 and their wild type counterparts for reference, together with negative controls lacking added kinase to detect background phosphorylation and autocatalytic activity.

Protein arrays called HuProt were purchased from CDI laboratories. The HuProt arrays have spots of more than 19,000 purified proteins and splice variants covalently immobilized on a Super Epoxy2 glass slide using a contact-less spray printer from Arrayjet.

Initially wildtype kinases were tested on HuProt V2.0 arrays (CDI laboratories) using the same protocol as the western blots (fluorescent antibody detection) and slides were also pre-treated with a bacterial tyrosine phosphatase (recombinant Yersinia YopH) to clean any pre-existing phosphorylation. There was no significant increase in signal between the kinase and negative control chips, so we decided to implement a more sensitive assay using radio-labelled ATP.

To test for reproducibility, each kinase and negative control reaction was performed in duplicate. All the subsequent reactions were performed on HuProt V3.0 arrays (CDI laboratories) following the same procedure detailed in previous work\(^1\). (For details refer Appendix: Protocols -5)

HuProt arrays were blocked in Licor TBS blocking buffer for 1hr to saturate the epoxy slide surface’s covalent bonds. Arrays were then incubated with 120µl of reaction buffer (10mM MgCl2 and 1mM MnCl2, 150 mM Tris pH 7.5, 100 mM NaCl) including 30µl purified kinase,
55 nM Hot ATP [γP33] (20 µCi), 1 mg/ml BSA and sealed with a coverslip. Negative control chips had no kinase in them to identify proteins which auto-phosphorylate, so that only trans-phosphorylation due to the kinase of interest can be identified by comparison. Chips were incubated in a closed humidity chamber (with wet tissues for maintaining humidity) at 30°C for 60 mins. Chips were washed (three times for 5min each) stringently with 0.5% SDS to remove the kinase and then with TBS-T (3x 5 min). Chips were finally dipped in water and dried by centrifugation. Chips were serially arranged and taped onto a cassette and then exposed to radiographic film in a dark room for 2, 6 and 25 days.

The exposed films were developed and imaged in a desktop scanner at 3,200 dpi, cropped each individual chip to a separate image in Photoshop CS6 before importing into Genepix Pro 6.0 for analysis.

**Data analysis**

Chips were analyzed using Gene Pix pro software. Spots and grid were aligned on the images and median intensity values of foreground and local background around the spots were used for further analyses along the lines of previous work (17).

Pixels inside each feature in the grid were used to calculate the ‘foreground’ intensity value and we have used the median of all the pixels inside the feature as a representative foreground value. Pixels lying in a zone around the feature that is three times the feature radius are considered as background pixels and their intensities are used to measure the ‘background’ intensity and we have used the median of this value for analysis.

Data sets generated from Genepix pro as a Genepix Result (GPR) file were imported into R software (18) as data tables and Foreground values of different chips were used for
comparisons. Background values are less reliable because a saturated foreground can create a high background intensity over a large region.

Optimal exposures were selected for each kinase family for further analysis which was 6-day exposure for the Jak2 batch and 2-day exposure for the Fgfr2 batch.

Each spot is arrayed twice on each array, in order to identify legitimate spots from spurious signals and we have used the minimum of both the spots to represent the signal for that particular substrate. Also because each condition was replicated in duplicate, the minimum of each spot in both the duplicates was chosen as the final intensity for the spot. This process of using the minimum eliminates chip to chip effects due to differential washing, etc., that are possible in the process.

Signal from every spot was used to make the scatter plots comparing each mutant to the wild-type and the wild-type to the negative control.

In order to identify changes in substrate selectivity between the mutants and the wildtype, further analysis was done to look for substrates with significant phosphorylation signal that are above a given threshold (Mean + 3 Standard Deviations) and are hereafter referred to as ‘hits’. Since each array has its own threshold, we can eliminate effects of overall increase in signal in the whole array, due to changes in the absolute quantity of the kinases used in the assay.

Each duplicate array was treated individually while applying the threshold in order to look for hits common in both arrays (we assumed that common spots are more reproducible).
Results

The digestion shows three different fragments, 6.5 Kb, 2.7 Kb from the vector backbone and one which corresponds to the size of the insert.

Figure 1: BsrG I digestion of destination clones
Purified kinases were subjected to kinase reaction for 30 min and used for western blotting. Probing for Expression (GST - red colour) as well as Autophosphorylation activity (Phospho-tyrosine – green colour) to show co-localization at the same molecular weight band indicated on the left. (Letter notation: J = Jak2, F = Fgfr2, R = Ret kinase, M = Map2k1, K = Kit kinase, H = Met kinase; Mutations notation: numbering implies different mutants, d preceding the kinase letter indicates a deletion mutant) (Mutation identity- Appendix: Mutants)
Figure 3: Western blot - purified Jak2 and Fgfr2 kinases

Purified kinases were subjected to kinase reaction for 30 min and used for western blotting. GST (red colour) and phospho-tyrosine (green colour); molecular weight bands indicated.

Figure 4: Comparison of replicates of Negative Controls with Jak2 batch
**Figure 5**: Comparison of replicates - Jak2 arrays

**Figure 6**: Jak2 vs Negative control
Figure 7: J1 vs Jak2

Figure 8: J2 vs Jak2
Figure 9: J3 vs Jak2

Figure 10: J4 vs Jak2
Figure 11: Comparison of replicates – Fgfr2 arrays

Figure 12: Fgfr2 vs Negative control
Figure 13: F1 vs Fgfr2

Figure 14: F2 vs Fgfr2
Figure 15: F3 vs Fgfr2

Figure 16: F4 vs Fgfr2
Fig 4-16: Foreground Scatter plots: One to one comparisons of foreground intensity of all spots between WT, Negative control and various mutants. The median values of the respective axes are indicated by horizontal and vertical lines and a $y=x$ line is plotted to see differential phosphorylation. The raw_hits and known hit are indicative of substrates that were significantly phosphorylated by WT kinase in previous experiments by Newman et al, 2013 (17).

Table 1: Number of hits: Jak2 batch

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>171</td>
<td>59</td>
<td>180</td>
</tr>
<tr>
<td>Jak2</td>
<td>106</td>
<td>34</td>
<td>456</td>
</tr>
<tr>
<td>J1</td>
<td>212</td>
<td>141</td>
<td>545</td>
</tr>
<tr>
<td>J2</td>
<td>180</td>
<td>95</td>
<td>450</td>
</tr>
<tr>
<td>J3</td>
<td>227</td>
<td>79</td>
<td>200</td>
</tr>
<tr>
<td>J4</td>
<td>61</td>
<td>8</td>
<td>161</td>
</tr>
</tbody>
</table>

Table 2: Number of hits: Fgfr2 batch

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>139</td>
<td>7</td>
<td>38</td>
</tr>
<tr>
<td>Fgfr2</td>
<td>139</td>
<td>48</td>
<td>127</td>
</tr>
<tr>
<td>F1</td>
<td>111</td>
<td>31</td>
<td>99</td>
</tr>
<tr>
<td>F2</td>
<td>193</td>
<td>23</td>
<td>156</td>
</tr>
<tr>
<td>F3</td>
<td>134</td>
<td>28</td>
<td>123</td>
</tr>
<tr>
<td>F4</td>
<td>56</td>
<td>5</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 1 and 2: Number of hits: Comparison of hits among replicates (chips A and B). Three Standard deviations above Mean was used as a cut off for calling out hits. The number of hits that exceed the cut-off in both the replicates are designated as common.
Discussion

Advances in sequencing technologies along with the exponential growth in the number of gene variants found in tumors creates an opportunity to diagnose the genetic basis of various cancers and opens up possibilities for genome based personalized therapy. To best make use of the data, we have to filter out the important and causal mutations from other inert mutations. Hence there is a need for robust predictive tools for identifying the driver potential of mutations. Along with the frequency of occurrence of the mutation which indicates evolutionary selection, the precise biochemical function of a given mutant will provide important information about its potential for being a driver.

To profile functional variation among mutant proteins, we have used functional protein microarrays called HuProt arrays (CDI Laboratories) as a way to identify downstream activity of a given protein in comparison to its mutants. We have screened the activity of 2 wild type (WT) kinases Jak2 and Fgfr2 along with 4 mutants and a negative control each on HuProt arrays in duplicate and imaged the radio-assay readout.

An initial examination of the images suggests that most of the auto-phosphorylating spots have a much higher intensity than others even in the presence of the kinase, both WT and mutant. It was noticed that those signals were higher than the limit of the 16-bit image used ($65535 = 2^{16} - 1$). Due to high signal intensity the spread of such spots is significantly larger than the expected spot size of the protein arrayed on the slide. For example, one such bright region encompassed 16 spots in total. This signal bleed through has affected many spots around the auto-phosphorylating spot, creating a confusion as to which among those was contributing to the signal. Hence all such spots in the neighborhood were eliminated from
analysis (and treated as if they were all lighting up in the negative controls) while looking for substrate selectivity.

It was observed that the Fgfr2 batch of chips had a much higher background (visible to the naked eye on the film) compared to negative control and Jak2 batch even though same amount of ATP was used. Upon examination of the protocols it could be attributed to a subtle change in the washing protocol where washes with 0.5% SDS after washing with TBS-T caused the higher background in Fgfr2 batch because the kinases and non-specifically bound ATP are present for longer on the chips before SDS clears them.

Visual examination of the scatterplots suggests subtle differences between the mutants and wildtype kinases. Further analysis is required to establish the precise shifts in selectivity owing to the noise on the chips and inability to visually tell apart the signal from the noise.

If successful, this approach can be extrapolated to other protein families with known biochemical functions having established assays for activity and would also be handy for drug screening.
Appendix: Protocols

1) Quick change mutagenesis:

*Table 3: Mutagenesis reaction mixture*

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x Reaction buffer</td>
<td>3</td>
</tr>
<tr>
<td>Template (plasmid)</td>
<td>3 (~ 20 ng)*</td>
</tr>
<tr>
<td>Forward primer</td>
<td>1</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>1</td>
</tr>
<tr>
<td>dNTP</td>
<td>1.2</td>
</tr>
<tr>
<td>Quick solution</td>
<td>1</td>
</tr>
<tr>
<td>dH2O</td>
<td>19.2</td>
</tr>
<tr>
<td>Lightning enzyme (DNA polymerase)</td>
<td>0.6</td>
</tr>
<tr>
<td>total volume</td>
<td>27</td>
</tr>
</tbody>
</table>

* - (increased till 55 ng for some templates)

*Table 4: PCR thermocycling conditions*

<table>
<thead>
<tr>
<th>Thermocycling conditions</th>
<th>Temperature</th>
<th>Duration of step</th>
<th>Duration of step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95 °C</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td></td>
<td>95 °C</td>
<td>20 sec</td>
<td>18 cycles</td>
</tr>
<tr>
<td></td>
<td>60 °C</td>
<td>20 sec</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68 °C</td>
<td>3.5 min</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68 °C</td>
<td>5 min</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-4 °C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2) **Yeast Cell Preparation**

i. Grow 5 ml of Y258 strain yeast in YPD media overnight with light inoculum

ii. Add 45 ml YPD and grow for 2-3 hours more (OD$_{600}$ approx. 1.0)

iii. Spin down cells in 50 ml tubes at 2,000 rpm for 4 min.

iv. Resuspend in sterile water and spin down again

v. Make LiAc/TE mix and autoclave

\[ \text{Table 5: LiAc/TE mixture} \]

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$O</td>
<td>7.2 ml</td>
</tr>
<tr>
<td>10x Tris-EDTA (TE) buffer pH 7.5</td>
<td>0.9 ml</td>
</tr>
<tr>
<td>1M Lithium Acetate (LiAc)</td>
<td>0.9 ml</td>
</tr>
</tbody>
</table>

vi. Resuspend yeast in LiAc/TE mix, put back in the flask (rinsed with sterile water). Shake at 30 °C for 30 min.

vii. Add 60 µl beta-mercaptoethanol. Shake at 30 °C for 30 min.

viii. Use 100 µl of these cells per well for transformation
3) Yeast Transformation

i. Make PEG mix and autoclave

Table 6: Peg mixture for yeast transformation

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% Polyethylene Glycol 3350 (PEG)</td>
<td>400 ml</td>
</tr>
<tr>
<td>10x TE pH 7.5</td>
<td>50 ml</td>
</tr>
<tr>
<td>1M LiAc pH 7.5</td>
<td>50 ml</td>
</tr>
</tbody>
</table>

ii. Turn on water bath to 45 °C

iii. Dispense 100 µl of the yeast competent cells mix into deepwell plates, add required DNA and vortex briefly

iv. Cap the deepwells. Incubate at 30 °C for 30 min.

v. Add 250 µl of the PEG mix to cells. Vortex the deepwells immediately to resuspend cells

vi. Incubate deepwells at Room Temp for 15-30 mins.

vii. Heat shock at 45 °C for 15 min and cool on ice afterwards for 2 min. (Do not exceed 15 min, Y258 is temperature sensitive and transformation efficiency is adversely affected)

viii. Pellet the cells at 3,000 rpm for 5 mins, dump the supernatant and pat on a paper towel.

ix. Resuspend cells in sterile water by vortexing and spread the cells onto SC-Ura plates using glass beads.

x. Seal the plates using Parafilm (to prevent drying) and grow at 30 °C. for 2-3 days checking for colonies every day.
4) Protein Expression and Purification

a. Yeast Induction

i. (Grow seed): Grow yeast colonies picked up from the transformation plates and grow them in 96 well deep well plates with 1 ml of Ura−/Glucose media at 30 °C for 24 hrs in a 210 rpm shaking incubator.

ii. (Expansion): Add 25 µl of the seed to 8 ml of Ura−/Raffinose media in 12 channel plates in duplicates, grow at 30 °C overnight with 90 rpm shaking.

iii. O.D_{600} should be around 0.6 – 1.2. (discard cultures if > 1.2)

iv. (Induction): Add 400 µl of autoclaved 40% Galactose (2% final concentration)

v. Incubate at 30 °C for 4-6 hrs (depending on OD) with 90 rpm shaking.

b. Yeast Harvest (keep boxes on ice throughout and set centrifuge at 4 °C)

i. Spin the 12 channel boxes covered with clean PCR plates at 4,000 rpm for 3 min and discard supernatant.

ii. Add 200 µl of ice cold water and vortex (make sure there is no mixing between channels)

iii. Transfer contents to a 96 well deepwell box and chill on ice.

iv. Spin the 96 well boxes covered with PCR plates at 4,000 rpm for 5 min, discard supernatant by inverting boxes and pat on paper towels to eliminate all traces of water (pellets from this step can be used for purification right away or stored as suggested in the next step).
v. Dry the boxes and secure the top with adhesive aluminium foil (use roller for tight sealing) and place on ice until flash froze in liq. N\textsubscript{2} for 30 secs. Store on dry ice until stored at -80 °C.

c. **Protein purification** (work in cold room 4 °C as much as possible)

*(Preparation)*

i. Add 0.5 mm zirconia beads to each well of the frozen deepwell box kept on ice.

ii. Centrifuge required volume of glutathione beads (25 µl per well) at 2,000 rpm, resuspend in twice the volume of cold lysis buffer (without inhibitors) by inverting the tube up and down and centrifuge again. Repeat for 2 more times.

iii. In the meantime, seal the bottom of the filter plates with melted agarose such that it forms plugs in each well (for sealing) but can be removed easily.

iv. Add the washed glutathione beads into filter plates (25 µl per well) and leave it in cold room (on top of a 96 well deepwell box)

*(Lysis)*

i. Add 400 µl lysis buffer (with protease inhibitors, phosphatase inhibitors) to each well, cover tightly with a capmat (hammer it softly) and shake vigorously in a paint shaker. Follow cycles of 1 min shaking with 2 min of rest on ice for 6 times.

ii. Centrifuge deepwell boxes at 3,000 rpm for 5 min and collect the supernatant into 96 well filter plates (with glutathione beads in them)
*(if protein expression is expected to be very high, the lysis step can be repeated with fresh lysis buffer and volume of glutathione beads can be increased)*

**(Binding)**

i. Seal the top of filter plate with adhesive foil (tie the filter plate to the deepwell plate under it with a rubberband)

ii. Place the filter plate sideways and rock gently (up and down) on shaker in cold room at 4 °C for 1 hour.

iii. Remove agar plug with a nylon brush and spin down the filter plate (with the deepwell plate) at 2,000 rpm for 3 min, make sure the agar is off and take off the aluminium foil on the top.

**(Washing)**

i. Wash each well using 400 µl of ice cold wash buffer I (3 times) and then 400 µl of ice cold wash buffer II (3 times) (empty the deepwell plate on the bottom in between washes).

ii. Reclose the top with the same foil (or using a PCR plate) and Spin the filter plates at 2,000 rpm for 1 min to remove wash buffer completely.

*(At this step the filter plates can be sealed with aluminium foil – both top and bottom- and stored at -80 °C until the next step is started).*

**(Elution)**

i. Add glutathione to the elution buffer. (final conc. 40 mM)
ii. Remove the top foil momentarily.

iii. Add 80 µl of ice cold elution buffer to each well, seal the top using aluminium foil and shake vigorously for 30 mins in cold room.

iv. Remove bottom seal, place a 96 well PCR plate beneath the filter plate to collect the eluate and centrifuge at 4,000 rpm for 1 min.

v. Aliquot required volumes into multiple PCR plates, seal top with adhesive foil and store at -80 °C (label the plates well)

5) Kinase reaction

Table 7: Kinase reaction mixture

<table>
<thead>
<tr>
<th>Kinase reaction mixture: Master Mix (15 reactions)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reagent</strong></td>
</tr>
<tr>
<td>2x Kinase Reaction buffer</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>BSA (Bovine Serum Albumin)</td>
</tr>
<tr>
<td>MnCl2</td>
</tr>
<tr>
<td>DTT</td>
</tr>
<tr>
<td>Hot ATP</td>
</tr>
<tr>
<td><strong>Total volume</strong></td>
</tr>
</tbody>
</table>

* Add 30 µl protein along with 90 µl of this mix to each chip
## Appendix: Mutants

*Table 8: List of mutant notation*

<table>
<thead>
<tr>
<th>Kinase</th>
<th>Notation</th>
<th>DNA</th>
<th>Mutation</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGFR2</td>
<td>F</td>
<td>Wild-Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>F1</td>
<td>755C&gt;G</td>
<td>p.S252W</td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>F2</td>
<td>1647T&gt;A</td>
<td>p.N549K</td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>F3</td>
<td>1144T&gt;C</td>
<td>p.C382R</td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>F4</td>
<td>1124A&gt;G</td>
<td>p.Y375C</td>
<td></td>
</tr>
<tr>
<td>FGFR2</td>
<td>F6</td>
<td>758C&gt;G</td>
<td>p.P253R</td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>J</td>
<td>Wild-Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>J1</td>
<td>1849G&gt;T</td>
<td>p.V617F</td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>J2</td>
<td>2047A&gt;G</td>
<td>p.R683G</td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>J3</td>
<td>2049A&gt;T</td>
<td>p.R683S</td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>J4</td>
<td>1615_1616AA&gt;TT</td>
<td>p.K539L</td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>dJ1</td>
<td>1624_1629delAATGAA</td>
<td>p.N542_E543del</td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td>dJ2</td>
<td>1627_1632delGAAGAT</td>
<td>p.E543_D544del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K</td>
<td>Wild-Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K1</td>
<td>2447A&gt;T</td>
<td>p.D816V</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K2</td>
<td>1676t&gt;a</td>
<td>p.V559D</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K3</td>
<td>2466t&gt;a</td>
<td>p.N822K</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K4</td>
<td>1727T&gt;C</td>
<td>p.L576P</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K5</td>
<td>1669t&gt;c</td>
<td>p.W557R</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K6</td>
<td>1924a&gt;g</td>
<td>p.K642E</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K7</td>
<td>1676t&gt;c</td>
<td>p.V559A</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K8</td>
<td>1679T&gt;A</td>
<td>p.V560D</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K9</td>
<td>2446G&gt;T</td>
<td>p.D816Y</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K10</td>
<td>1961tg&gt;ct</td>
<td>p.V654A</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>K11</td>
<td>2446G&gt;C</td>
<td>p.D816H</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK1</td>
<td>1648_1674</td>
<td>p.K550_K558del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK2</td>
<td>1672_1677</td>
<td>p.K558_V559del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK3</td>
<td>1660_1674</td>
<td>p.E554_K558del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK4</td>
<td>1651_1662</td>
<td>p.P551_E554del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK5</td>
<td>1654_1665</td>
<td>p.M552_V555del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK6</td>
<td>1666_1671</td>
<td>p.Q556_W557del</td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>dK7</td>
<td>1663_1677</td>
<td>p.V555_V559del</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M</td>
<td>Wild-Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M1</td>
<td>370C&gt;T</td>
<td>p.P124S</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M2</td>
<td>171G&gt;T</td>
<td>p.K57N</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M3</td>
<td>167ag&gt;ca</td>
<td>p.Q56P</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Sample</td>
<td>Mutation</td>
<td>Amino Acid Change</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>----------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M4</td>
<td>371c&gt;t</td>
<td>p.P124L</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M5</td>
<td>607G&gt;A</td>
<td>p.E203K</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M6</td>
<td>362gc&gt;ct</td>
<td>p.C121S</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M7</td>
<td>1159c&gt;t</td>
<td>p.P387S</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M8</td>
<td>199G&gt;A</td>
<td>p.D67N</td>
<td></td>
</tr>
<tr>
<td>MAP2K1</td>
<td>M9</td>
<td>157T&gt;C</td>
<td>p.F53L</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H</td>
<td>Wild-Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H1</td>
<td>3703T&gt;G</td>
<td>p.Y1253D</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H2</td>
<td>2975C&gt;T</td>
<td>p.T1010I</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H3</td>
<td>3749T&gt;C</td>
<td>p.M1268T</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H4</td>
<td>3689A&gt;G</td>
<td>p.Y1248C</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H5</td>
<td>2908C&gt;T</td>
<td>p.R988C</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H6</td>
<td>504g&gt;t</td>
<td>p.E168D</td>
<td></td>
</tr>
<tr>
<td>MET</td>
<td>H7</td>
<td>796A&gt;G</td>
<td>p.R266G</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R</td>
<td>Wild-Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R1</td>
<td>2753tg&gt;ct</td>
<td>p.M918T</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R2</td>
<td>1900T&gt;C</td>
<td>p.C634R</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R3</td>
<td>1888t&gt;c</td>
<td>p.C630R</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R4</td>
<td>1901G&gt;A</td>
<td>p.C634Y</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R5</td>
<td>2647_2648GC&gt;TT</td>
<td>p.A883F</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>R6</td>
<td>1902C&gt;G</td>
<td>p.C634W</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>dR1</td>
<td>1894_1899delGAGCTG</td>
<td>p.E632_L633del</td>
<td></td>
</tr>
</tbody>
</table>
Bibliography


(14) Quickchange Lightning site directed mutagenesis kit (user manual):

(15) Invitrogen LR recombinase protocol:

(16) Licor, Near-Infrared Western Blot Detection protocol:
  https://www.licor.com/documents/fxc6evxvxbub4srkqv6i9yg46l7i0xz5


Curriculum Vitae

PRASHANT BHARADWAJ KALVAPALLE
Date of Birth – 06 Oct 1992
Place of Birth- Hyderabad, India

EDUCATION

<table>
<thead>
<tr>
<th>Year</th>
<th>Certificate</th>
<th>Institution</th>
<th>CGPA/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-2014</td>
<td>Bachelors of Technology in Biological Sciences and Bioengineering (BSBE)</td>
<td>Indian Institute of Technology (IIT), Kanpur, India</td>
<td>9.2/10</td>
</tr>
<tr>
<td>2014-2016</td>
<td>Masters of Science and Engineering in Biomedical Engineering (BME)</td>
<td>Johns Hopkins University, USA</td>
<td>3.87/4</td>
</tr>
</tbody>
</table>

ACADEMIC ACHIEVEMENTS

- Silver medal for best academic performance in BSBE department, IIT Kanpur.
- Received Academic Excellence Award twice from the Director, IIT Kanpur for outstanding academic performance in the academic years 2011-12 and 2012-13.
- Mitacs Globalink intern for a summer internship in McMaster University, Canada in 2012
- Selected among top 40 candidates from India for Khorana Scholars program'12 for internship
- Was among the top 0.5% in IIT JEE (Indian Institute of Technology, Joint Entrance Exam), 2010.
- Ranked among the top 0.6% in AIEEE (All India Engineering Entrance Exam).

KEY PROJECTS

Mapping human cancer driver mutations

Project guide: Prof. Joel Bader, Biomedical Engg., Johns Hopkins University (Master’s Thesis) (Jan’15 – Apr’16)
- Mutagenesis, Purification and In-vitro high throughput assay of tyrosine kinase mutants against human proteome arrays

Drug discovery for Tuberculosis against Carbonic anhydrase enzymes

Project guide: Prof. Balaji Prakash, BSBE dept., IIT Kanpur (B.Tech Project) (Jan’14 – Apr’14)
- In-silico high throughput virtual screening of library of molecules active in whole cell inhibition tests
**Induced In vitro Trans-differentiation of mouse Bone marrow Stromal cells into Neurons**  
*Project guide: Prof. Jonaki Sen, BSBE dept., IIT Kanpur (B.Tech Project) (July’13 – Dec’13)*  
- **Expressed and purified** ‘Noggin’ protein in a mammalian system DF1 cells  
- **Assayed** for inhibition of BMP signalling using Briter assay  
- Induced BMSC cells by adding Noggin protein and probed for Neuronal markers like MAP2 and βTubulin using **Immunocytochemistry**

**Discovery of novel signalling motifs in Scavenger receptor proteins using a Pattern finding algorithm and Multiple Sequence Alignments**  
*Project guide: Prof. Dawn Bowdish, Macrophage Lab, McMaster University, Canada (May’13 - July’13)*  
- Tried to develop an efficient **bioinformatics technique** for novel motif discovery in protein and DNA sequences  
  - Used a new software Aligned Pattern Clusters (APC), and Multiple Sequence Alignments in combination.  
- Validated the APC to give a minimum sensitivity of 54% and Positive Predicted Value of 60%, by using a well-studied Macrophage scavenger receptor protein - SRA as a standard.  
- Found motifs in a macrophage scavenger receptor protein MARCO (also called CD36), using the above tools. The lab is taking forward three of the major motifs to test for loss of phagocytic function mutations in the identified regions.  
  The project was **sponsored by Mitacs Globalink, Canada.**

**Crystallization, Structural and biochemical studies of yeast CaUAP1 and other Sugar Nucleotide Transferases**  
*Project guide: Prof. Balaji Prakash, BSBE, IIT Kanpur (May’12-July’12)*  
- **Cloned the gene** for CaUAP1 from the cDNA of yeast (Candida albicans)  
- **Purified the protein** using affinity chromatography and Gel exclusion chromatography.  
- Biochemically assayed the enzymatic activity using Malachite Green Assay.  
- Screened the protein along with substrate UTP and Ca\(^{2+}\), in multiple precipitant conditions for best crystal forming environment (used PEG ION screen and INDEX screens) and got positive crystals in Ammonium citrate tribasic condition.  
  Also tried to clone using the same techniques three other Sugar Nucleotide Transferases from *Pyrococcus furiosus, Pyrococcus horikoshi* and *Helicobacter Pylori.*

**Novel Method for Protein Crystallization**  
*Project guide: Prof. Balaji Prakash, BSBE, IIT Kanpur (Jan’13 – Apr’13)*  
We proposed a **novel method to crystallize proteins in micro-channels** in a gel by varying the pH and Precipitant concentration in orthogonal directions using simple diffusion. This will optimise
the existing methods by drastically lowering the amount of protein used and increasing the number of conditions screened.

**Literature Survey and Structural Bioinformatics for novel Drug Targets of Mycobacterium Tuberculosis**

*Project guide: Prof. Balaji Prakash, BSBE, IIT Kanpur*

(Jan’13 – Apr’13)

We as a team of 3 students, **conducted a Literature survey** to look for pathways Unique and essential to Mycobacterium’s survival and its virulence and then picked up the key protein of each pathway for further analysis of the Structure Function Relationships of the protein by studying the active site and comparing it with homologues using Chimera software. The final attempt was to come up with a targetable site in the protein and **propose a Drug molecule** which is already in clinical use.

We had narrowed down on a Mycobacterial secreted protein NdkA, which was found to be acting as an extraneous GTPase Activating protein for RAB 5, 7 and Rho GTPases, playing havoc with many process like Endosome maturation, Cytoskeletal polymerisations, hence presenting itself as an attractive drug target.

**Optimisation of Aptamer-mediated Targeted Drug Delivery Systems**

*Project guide: Prof. D S Katti, BSBE, IIT Kanpur*

(July’12 – Dec’12)

We have proposed an idea to utilize the ligand specificity and intricate confirmations of Aptamers (single stranded oligo nucleotides) to trigger a release mechanism by cleaving off an open-able lid in a Poly Lactic co Glycolic acid (PLGA) microsphere containing the drug by a self-cleaving ribozyme mechanism. This course project was done in a group of 5 students.

**SKILLS**

<table>
<thead>
<tr>
<th>LAB</th>
<th>Molecular cloning</th>
<th>Protein purification</th>
<th>Chromatography (FPLC)</th>
<th>Western Blotting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalian cell culture</td>
<td>Bioinformatics</td>
<td>Immunocytochemistry</td>
<td>Nanoparticle, Cryogel and Hydrogel fabrication</td>
<td></td>
</tr>
<tr>
<td>TECHNICAL</td>
<td>C, Python, R, Matlab</td>
<td>UCSF Chimera</td>
<td>Autodesk Inventor, AutoCAD</td>
<td>Adobe Photoshop, Gimp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Languages : English, Hindi, Telugu</td>
<td></td>
</tr>
</tbody>
</table>

**EXTRA-CURRICULAR ACTIVITIES**

- **Mentor**, STEM Achievement Baltimore Elementary schools program at Johns Hopkins University
- **Taught high school students** under a Non-Governmental organisation (NGO)- Shiksha Sopan
- Designed T-shirt and publicity fliers in the design team for Society of automotive engineers IIT Kanpur
- Played citywide league matches in my school cricket team and also won 2 inter-school tournaments