# Initial BigOnc Functionality

### Scope and Description

The goal of this project is to provide insight into effective oncological treatment for cancer causing mutations based on a variety of genetic data from multiple sources.

To that end we (Tristan Carland, Amarnath Gupta and Bill West) met to discuss an initial development project that should be able to demonstrate efficient and scalable behavior from the data repository, the use of remote data accessed via API or direct database calls, the ability to create, store and use custom functions for filtering (i..e. interval matching, etc.) and ranking (mutation matching, K-Fold scoring, etc.).

This process is being described from the contents of diagram 1 which is a whiteboard representation of our initial project discussions.

### Design Guidelines (biological design guidelines need to be added here)

**November**

* + Fast and scalable database solution
  + Reuses of data currently in the public domain (NIF, UCSC Genome Browser) via API calls or direct database connection
  + Incorporate custom filtering and ranking into the middle tier as required

**Future**

* + Workbench for creating and adding custom functions
  + Site security and HIPPA privacy compliance
  + Website and API documentation

### Website Description

The web interface will consist of a number of pages but we will only deal with two in this paper and will only provide a high level description for one will providing just a brief explanation of the other.

The Researcher Workbench will provide direct access to patient and genetic data related to that patient to allow researcher to perform custom filtering and scoring of these data. This is not within the scope of the initial project. However direct access to the data will be made available to the team and any fruitful functionality will be built into a Java rules class to be used on the Clinicians site.

The Clinicians page will be available in a prototype format for November. The following provides a detailed description of this site as mapped out in the whiteboard sessions and illustrated in diagram 1.

### Clinicians Page

The inputs for the Clinicians page will be a variety of variation detecting outputs from whole genome expression exercises. Initially, we have identified the Variant Calling Format (VCF), MuTect and ReSeq files for this purpose. In addition, various patient metadata will be stored as well containing such information as age, gender, type of disease, haplotype, etc.

Once the data is stored the clinicians will be able to select from a number of custom and stored filtering and rankings to create a dataset which best fits the nature of the case and the data.

### Initial Filters Discussed:

* Mutations located within an oncogene
* Any aberrant expression on the target oncogene
* Any aberrant expression near an oncogene
* Is the mutation in an oncogene factor
* Is the mutation near and oncogene factor

### Rankings

Certain scores and ranks may be used to order the dataset for instance allele frequency or K-fold score.

### Web Services and Direct Access to Remote Repositories

Three initial sources have been identified as sources for reliable, normalized and curated data. Access to these data will be accomplished via Java restful web services in the case of the Neurological Information Framework (NIF) .

For data related to gene and gene location data we can extract these data directly from the UCSC Genome Browser via a jdbc connection.

Finally, the NCBI database has been identified as the best source for treatment data. This area of data will require more research regarding not only the drug treatment data but how effective it might be for the targeted mutation.

### Remote Data Storage Policy

Once a decision is made to make a specific treatment recommendation, all data retrieved from remote sources related to that recommendation will be stored in the local database associated with that patient for auditing purposes.

Data associated to variant sites not considered germane to the treatment decision will not be stored in BigOnc.

Diagram 1.

