# PolyA Explorer requirement

In this project we build a visualization tool which graphically shows a part of genome. And I want you all to assume this as a satellite view of a google map for a straight one-dimension road with some establishments on it.

What we have in our database is start and end position of where these establishments on the straight road starts and where it ends. So, when a user searches for a gene we fetch the coordinates of all establishments that are part of this line.

With this analogy in mind, let me tell you what I mean by establishments. 5UTR, Intron, Exon, CDS are part of this genome, which would have a start and end point coordinates that help us place them at right place on this road. Along with the establishment we also would have few markings along this road. Like establishments we have seen, these markings would have start and end positions.

Like we search for a park on google map and see how it looks from satellite view, her we search for a gene. And we get all the start and coordinates of the establishments related to that gene. The way we represent these establishments in different colors and different shapes is the crucial part of the project. One more such important point would be zooming in and zooming out while we auto scale to conserve the positions of establishments and markings.
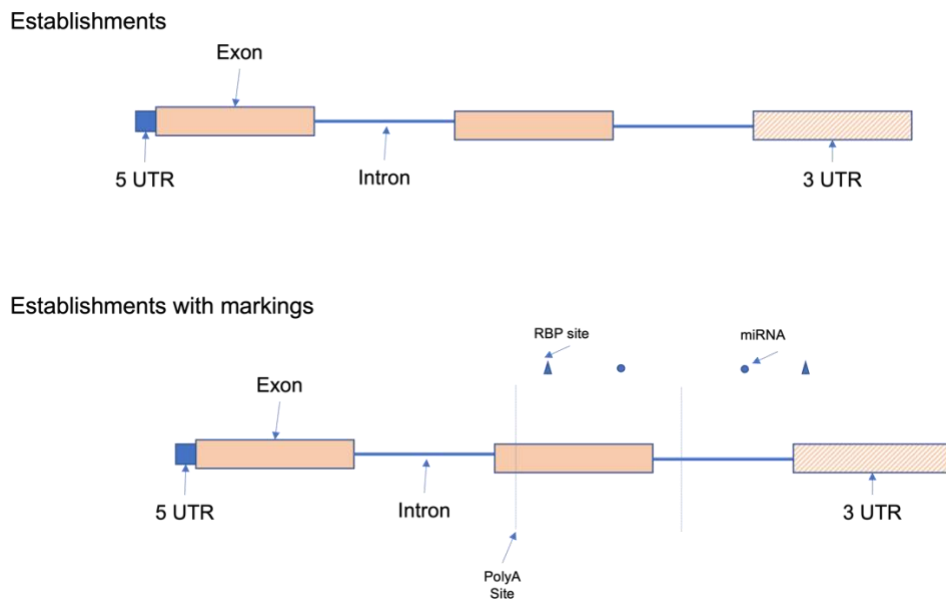

We provide some tabulated values based upon the positions of the markings. One such requirement we have now is for each polyA site (vertical line) count number of miRNA spots present to the right side of the line. A design complexity problem you would be solving here would be choosing the position of the markings. Right now, you are seeing the RBP sites and miRNA targets are on same line. What would be the case if we have two or more such markings at the same position? We use d3.js in angular to achieve all the above requirements.

Let's take a step back and see where all this data comes from. Obviously from database. And the database has been curated from files of different formats (one example of such format is .bed file). We use python scripts to take data from files and format it right and put into database (mongo db). You will be writing these curating scripts.

Along with the above we have one more visualization which we achieve with help of an angular library called igv viewer. See https://igv.org/app/ for demo. Give "atxn1" as input in second input box to see a view. It looks something like this.
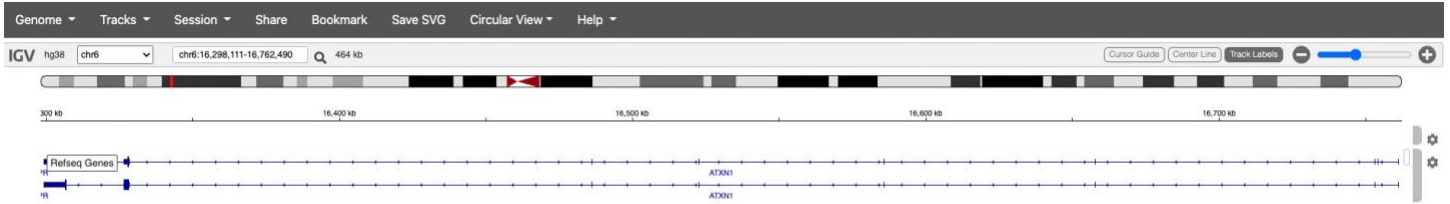
What's happening here is the data needed to draw those blue lines comes from a file. And the data gets queried with the gene input we have given and selects part of line that needs to be drawn here. To achieve this on our app, we get this file from either AWS S3 or local depending upon size of the file.
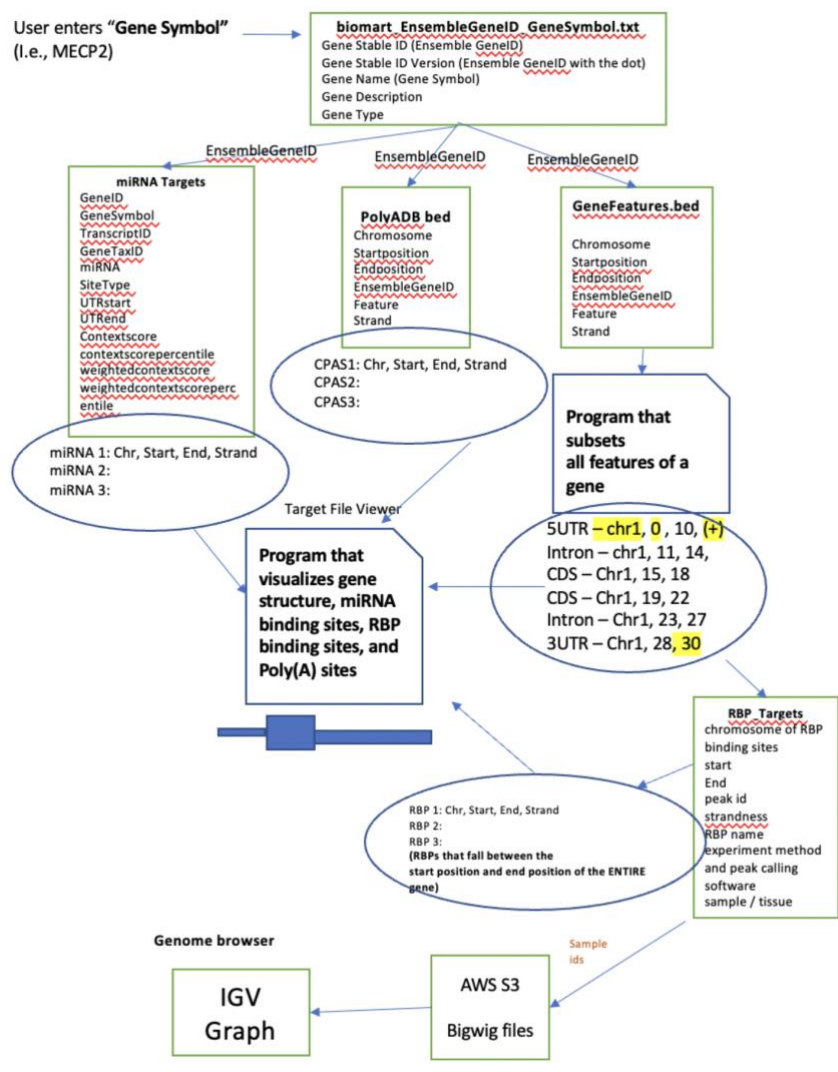
This above figure would help you understand the flow.