CRISPR-GA

# Overview and Motivation:

CRISPR/cas9 is a method for genome engineering developed over the last years (Mali et al, Science 2011). This technology has revolutionized human genome engineering and opened countless possibilities to basic science, synthetic biology and gene therapy. Albeit the enormous potential of these tools, their performance is far from perfect. It is still necessary to carry out a posterior careful analysis of the gene editing experiment. However, there are no computational tools for genome editing assessment yet, and current experimental tools lack sensitivity and flexibility. In 2014, I developed a website (CRISPR­GA, http://crispr-­ga.net ) to easily analyse genome engineering editing results based on next generation sequencing data based only on with three mouse clicks (Guell et al, Bioinformatics 2014). CRISPR Genome Analyzer provides a report for the locus selected, which includes a quantification of the edited site and the analysis of the different alterations detected. However, current representations are informative but lack interactivity. Three static png are presented to the user providing the localization of gene edits and overall efficiency (see Fig. 1). Some more dynamic information could be presented to help the users to understand the results of their experiments.

# Provide an overview of the project goals and the motivation for it.

Four types of information are presented in CRISPR­-GA:

* Nature of gene edits (insertions, deletions, homologous recombination (HR), non­homologous end joining (NHEJ)).
* Localization of gene edits
* Quantification of the percentage of HR and NHEJ
* Source data files containing the exact sequencing reads containing gene alterations

Currently all this information is presented fragmented and static. I would like to generate an integrated and dynamic representation. The users could interact with the results and at the same time browse the original raw data that generated those reads. These are the the benefits:

● User interaction

● Integrated representation

● Help to link results and experimental cause

# Related Work

I would like to provide a tool that is dynamic and that provides multiple visualizations of the the results. I think that homework 3 had several of those features.

Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

# Data: Source, scraping method, cleanup, etc.

The data is collected from the CRISPR­-GA output after every analysis ([see documentation](http://crispr-ga.net/documentation.html)). NHEJ results are provided in XML format ([see link to XML](http://54.80.152.219/images/1428114971A912968177results.xml)), reads mapping data with PSL format ([see link to PSL),](http://www.ensembl.org/info/website/upload/psl.html) and reads are provided in fastq format.

The data output from CRISPR-­GA will require minimal processing as it is structured and

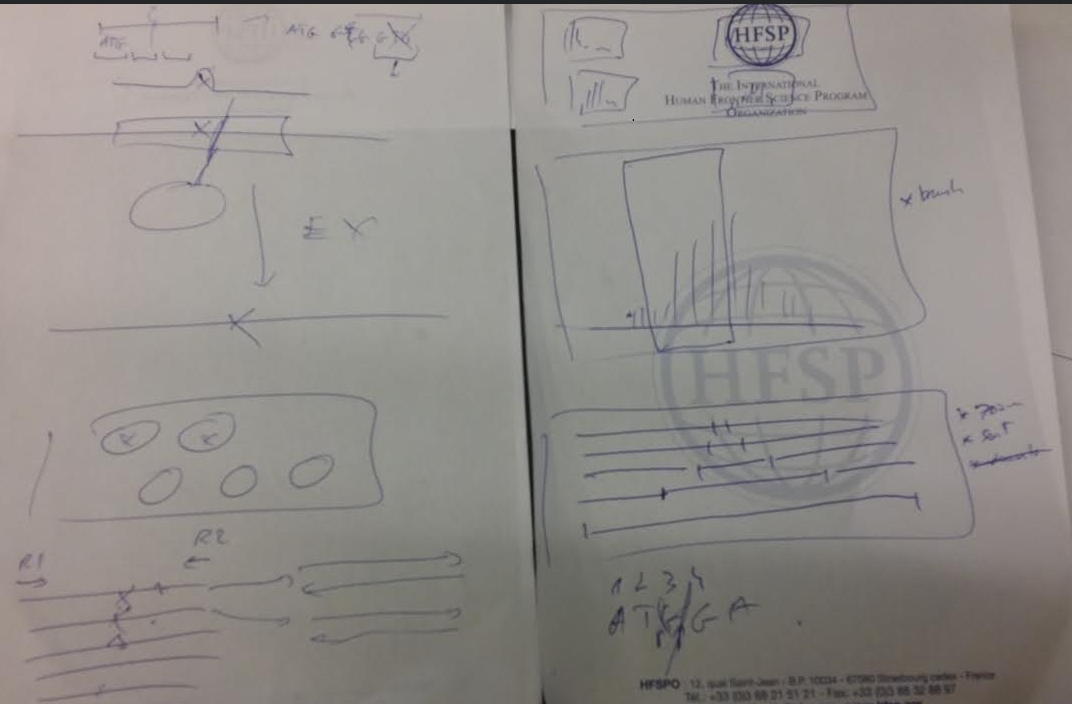
provided with standard formats (XML, PSL, fastq).

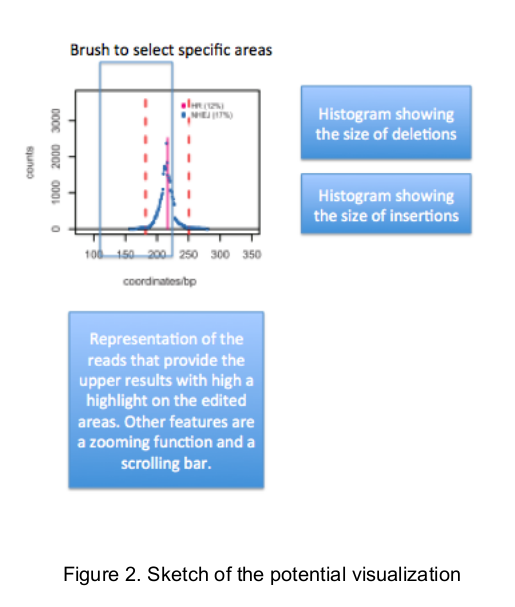
Data processing will be implemented in R statistical language, and transformed into json

using the R library rjson

Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

# Design Evolution:

These are the initial sketches and selected payout.



# Implementation

* Data preparation has been done. XML and tabular format have been transformed to JSON (figure 3).
* Main display from homework 3 has been used as starting point
* Central display has been implemented with brushing function