CRISPR-Analytics (CRISPR-A): a platform for precise analytics and simulations for gene editing

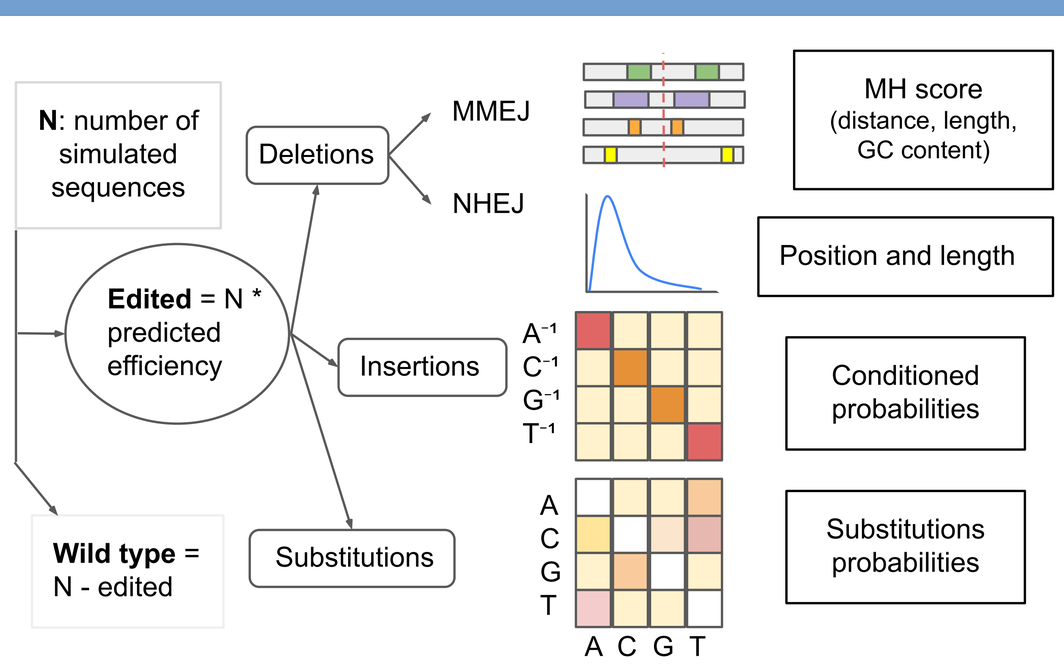
Marta Sanvicente-García,[1] Albert García-Valiente,[1] Socayna Jouide,[2] Jessica Jabara-Wallace,[1,3] Èric Bautista,[4] Marc Escobosa,[1] Avencia Sánchez-Mejías,[1,3] Marc Güell[1]

Synbio

Universitat Pompeu Fabra Barcelona

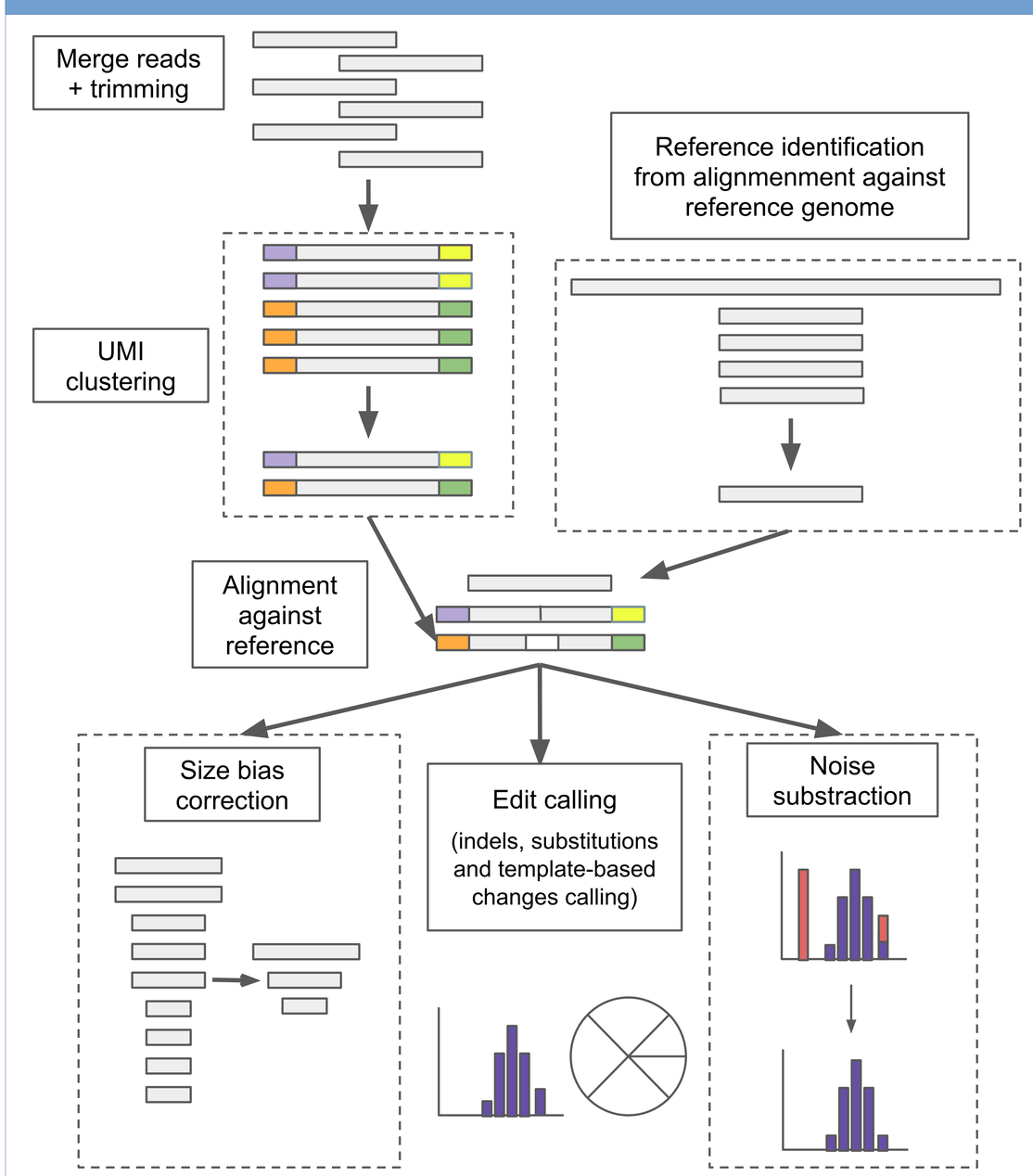
We have developed CRISPR-Analytics, CRISPR-A, a comprehensive and versatile genome editing **web application** tool, and a **nextflow pipeline** to give support to **gene editing** experimental design and analysis. This tool is ideal for analyzing highly sensitive cases such as clinical samples or experiments with low editing efficiencies. CRISPR-A is ideal to support multiple kinds of experiments such as double-stranded DNA break-based engineering, base editing (BE), primer editing (PE), and homology-directed repair (HDR).

SIMULATIONS PIPELINE



Simulation is based on the **fitting** of multiple parameters that describe the distribution of edits and its characteristics and proportions. Once the number of edited sequences is determined by the **protospacer predicted efficiency**, other **probability distributions** are applied to decide the number of each kind of edits (NHEJ deletions, MMEJ deletions, insertions and substitutions).

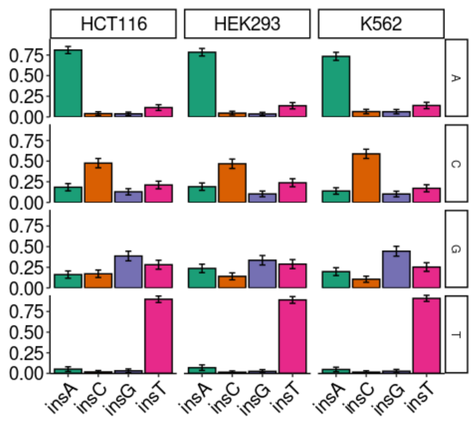
ANALYSIS PIPELINE

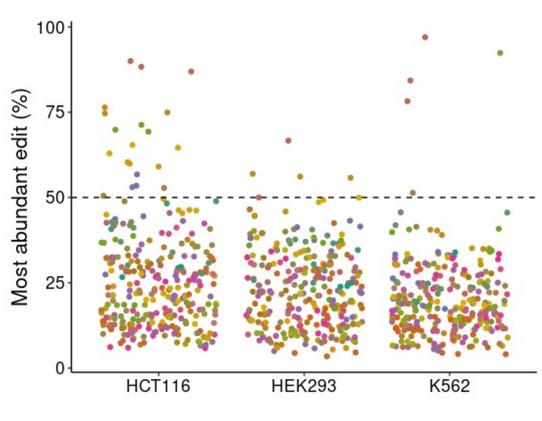


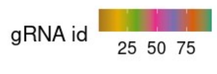
The analysis algorithm is composed of three mandatory steps:
1) reads **pre-processing** for quality assessment,
2) reads **alignment** against reference amplicon, and
3) **edit calling.**

In dashed line boxes, all processes that are optional: **UMI** clustering, reference discovery, size **bias correction**, and **noise subtraction** based in an empirical model from negative control samples.

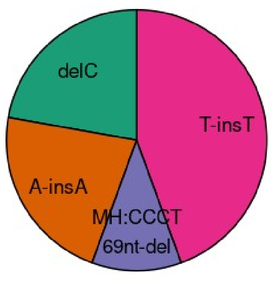
CASE EXAMPLE



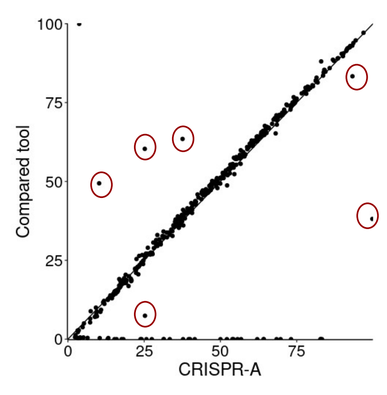


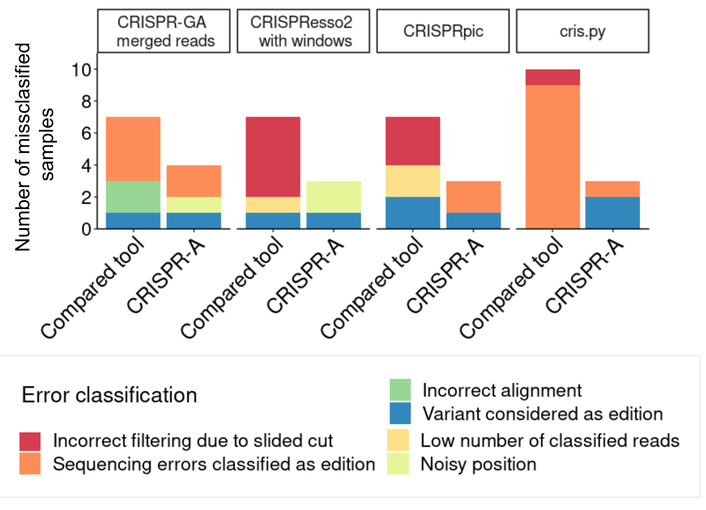


HCT116, HEK293 and K562 edited in 96 different targets. Percentages of **inserted nucleotides** in function of the free nucleotide at 3' in three different cell lines (top left). In the plot in the right side, we have represented tha **variants diversity** by samples; calculated through the percentage of the most abundant variant among all kinds of edits. In the pie chart (bottom) we can see the **nature of the variants with higher abundance**. All these variants that are above the 50% (dashed line in top-right plot).



BENCHMARCKING





Error characteritzation from the most discrepant values in t-cell edited samples (978 samples). CRISPR-A results are compared with the results of other tools for more distant results (red circles). Errors are classified in function of their source.

Reference:

Sanvicente-García, M., García-Valiente, A., Jouide, S., Jaraba-Wallace, J., Bautista, E., Escobosa, M., ... & Güell, M. (2022). CRISPR-Analytics (CRISPR-A): a platform for precise analytics and simulations for gene editing. bioRxiv.

Affiliations:

1. Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain
2. Faculty of Mathematics and Computer Science, University of Barcelona, Spain
3. Integra Therapeutics S.L., Barcelona, Spain
4. Technical University of Denmark, DK-2800 Lyngby, Denmark

Funding:

European Commission | Horizon 2020 European Union funding for Research & Innovation