

# EspressoBar

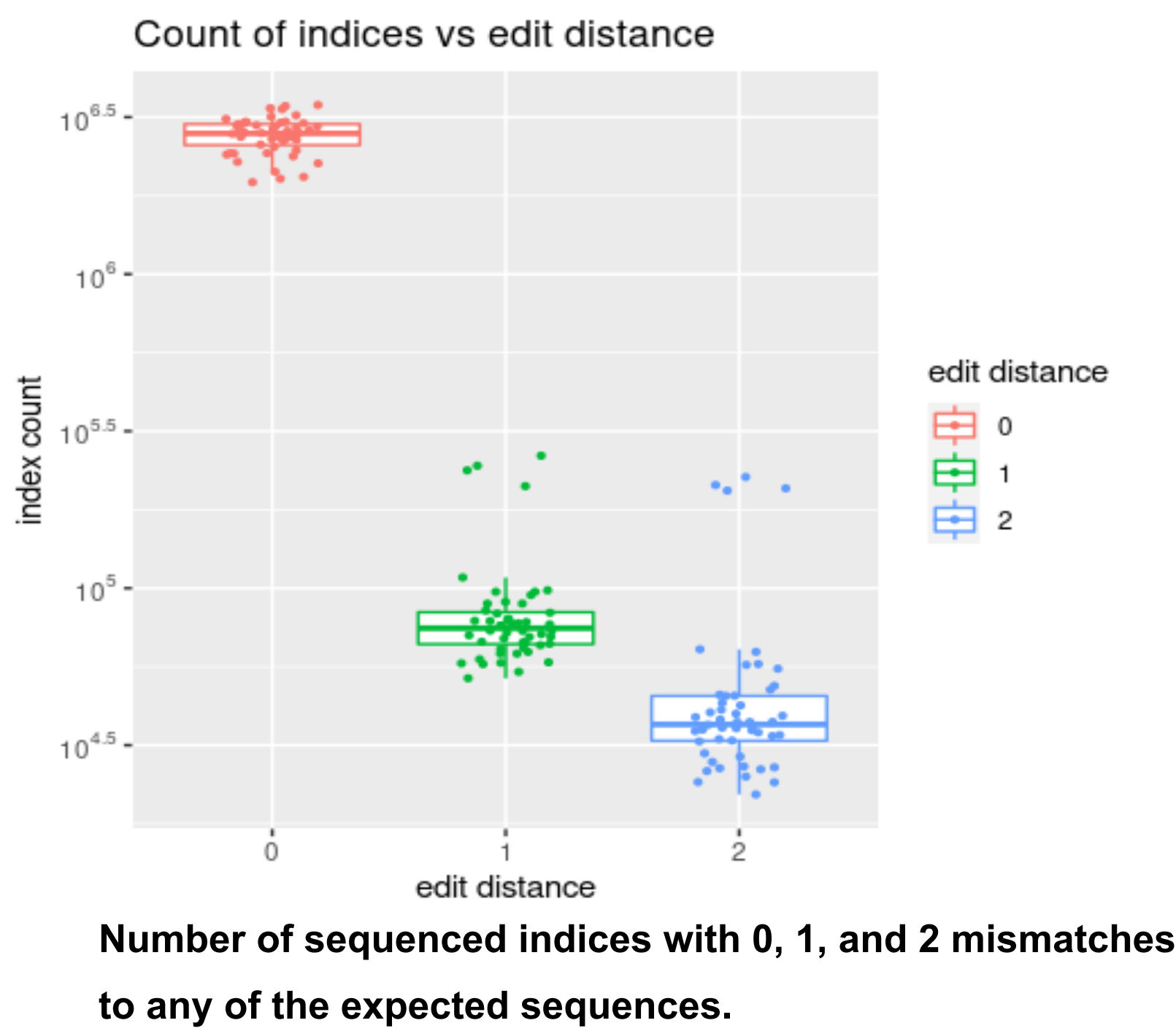
A web-interface for contamination aware design of mixed multiplexed Illumina sequencing experiments

Dilara Damar\*, Ashkan Ghassemi\*, Josephine Grützke, Burkhard Malorny, Simon H. Tausch

\* These authors contributed equally to this work

## Motivation

- Samples on Illumina runs can be multiplexed using unambiguous index sequences
- Cross contamination of samples in Illumina sequencing runs is a common problem
- Sequencing errors in index sequences may lead to cross-contamination
- Index sequences are used for Cluster detection

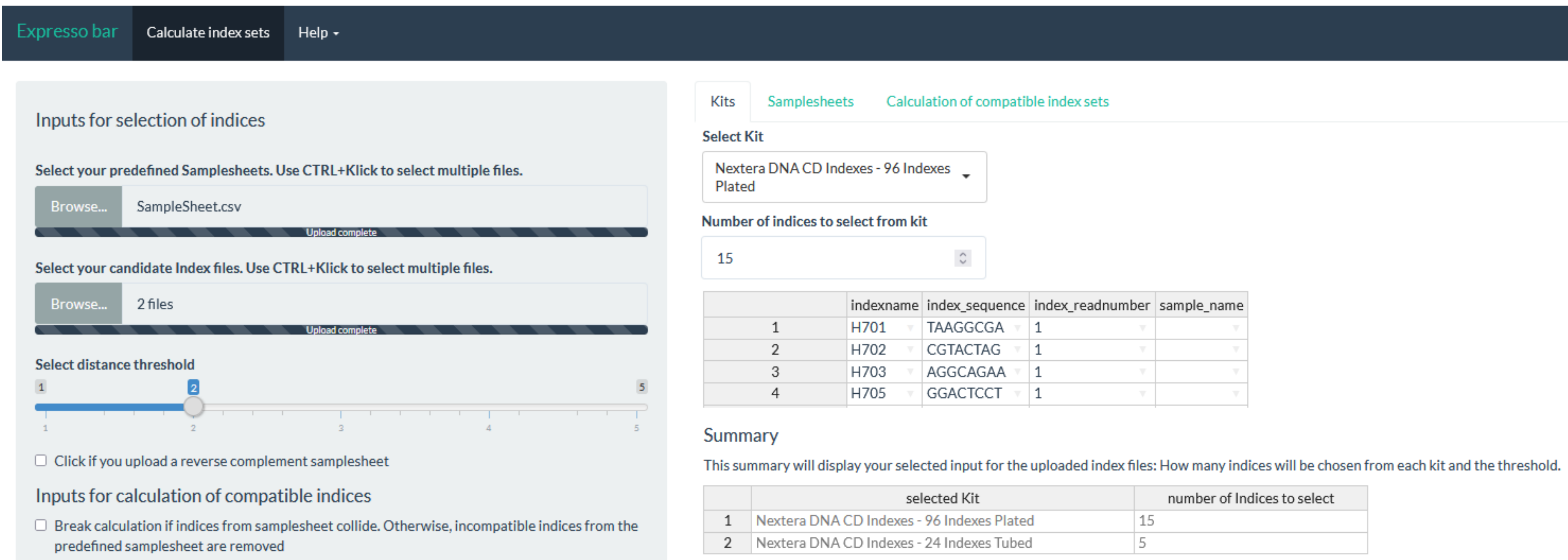


## Aim

- Combination of different library prep sets to optimize load factor of sequencing runs
- Select optimal set of indices for cross-contamination-free sequencing
- Maximize edit distance between indices to avoid misassignments
- Equally distribute signals over at each position
- Enable combination of paired- and single index kits, also with different lengths

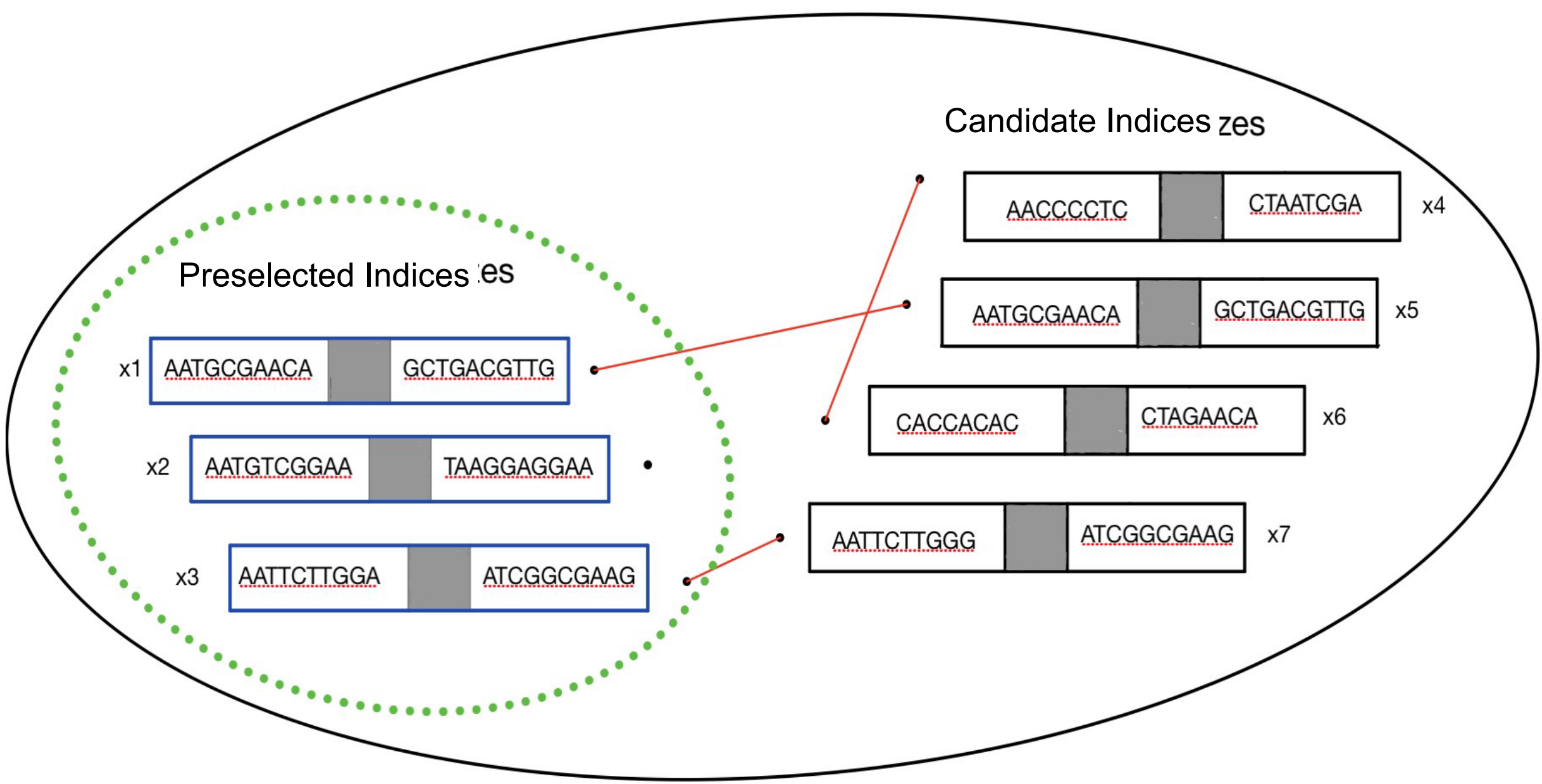
## Frontend: Shiny web application

- Easily usable:
  1. Upload Sample Sheets and candidate index sets
    - ✓ Conflicting indices from Sample Sheets are pointed out
  2. Calculate ideal combination of indices fulfilling all constraints
    - ✓ Distribution of bases per index position is visualized
  3. Download results as ready-to-use Illumina Sample Sheet



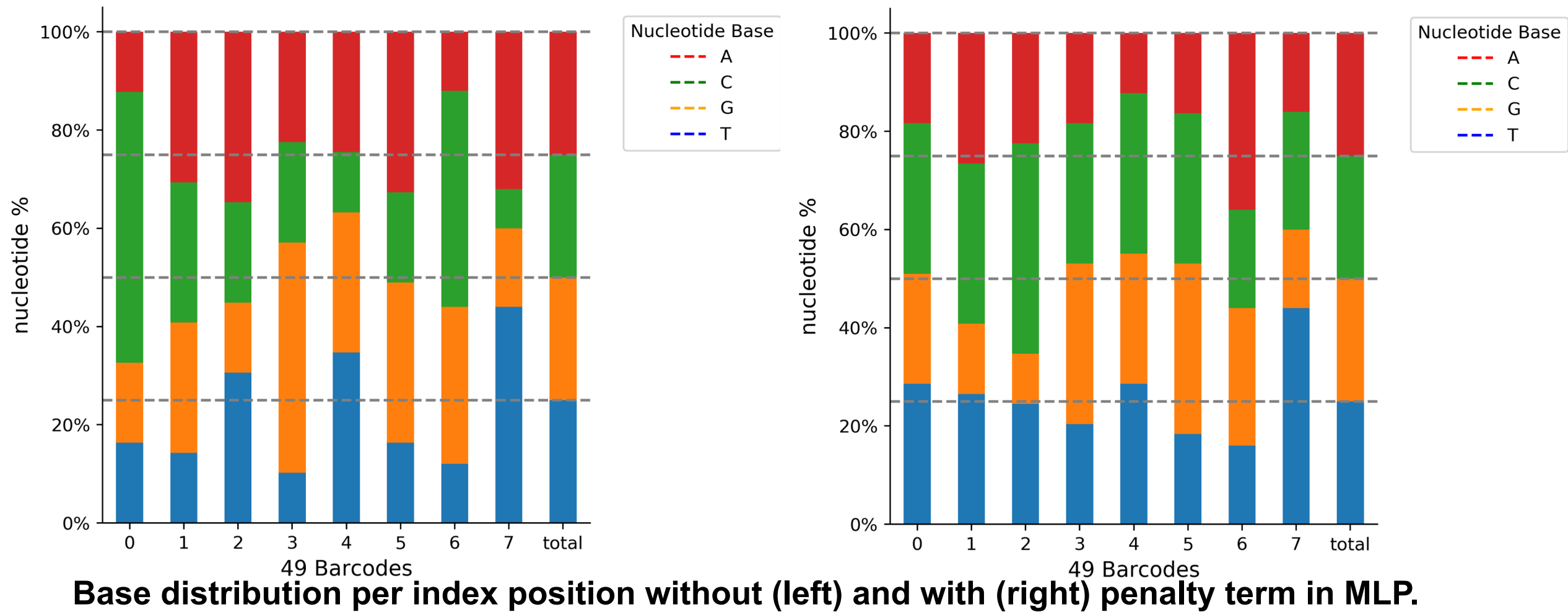
Screenshot of the Shiny web application available at <https://bfrbinfile.shinyapps.io/espresso-bar/>. In the left section, sample sheets and index files can be uploaded and options are set. On the right, the number of indices to be selected from each of the available kits is defined.

## Backend: Mixed Integer Linear Programming



## Conclusion

- If there is a feasible combination it will be found
- The selected set of indices will always be ideal



- Using Mixed Integer Linear Programming to define Maximal Independent Set of indices
- Constraints:
  1. Preselected indices must not be removed
  2. All indices must have a pairwise distance larger than the selected threshold
  3. The bases should be equally distributed at each position

- Code freely available on Gitlab ([https://gitlab.com/bfr\\_bioinformatics/espresso-bar](https://gitlab.com/bfr_bioinformatics/espresso-bar))
- Deployable as Docker container
- Backend also available as standalone Linux tool

- Also available as web application:

