ALiBERO

“hands-on” session

ICM-UGM 2016

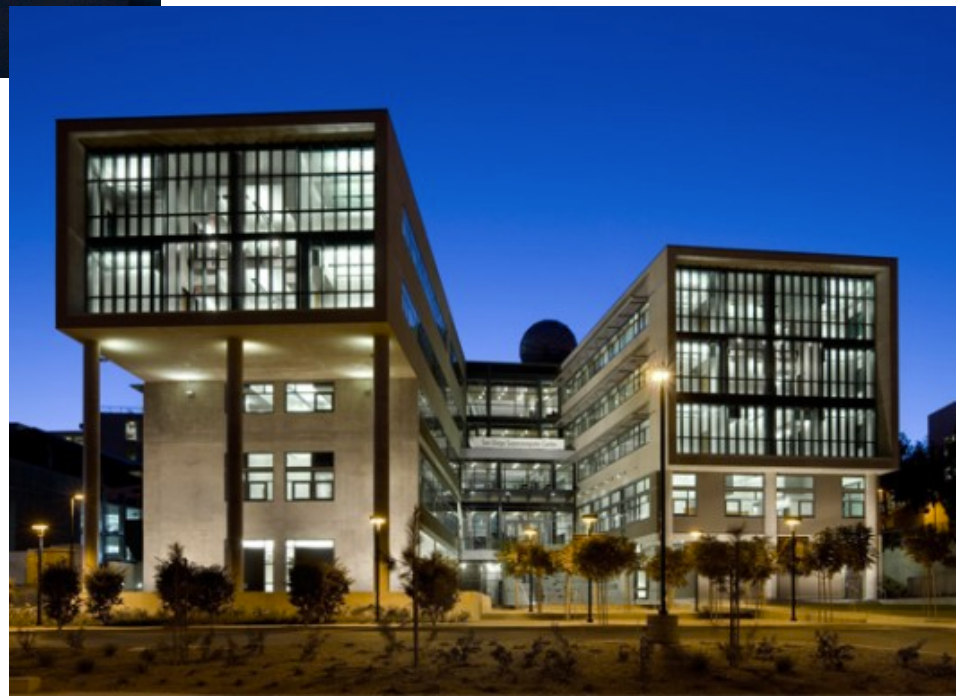
Manuel Rueda

mrueda@scripps.edu

TSRI (2 yr)

UCSD

SSPPS – SDSC (3 yr)



I have **not** used **ICM**…

…in 3 years

ALiBERO: evolving a team of complementary pocket conformations rather than a single leader

Authors Manuel Rueda, Max Totrov, Ruben Abagyan

Publication date 2012/9/17

Journal Journal of chemical information and modeling

Volume 52

Issue 10

Pages 2705-2714

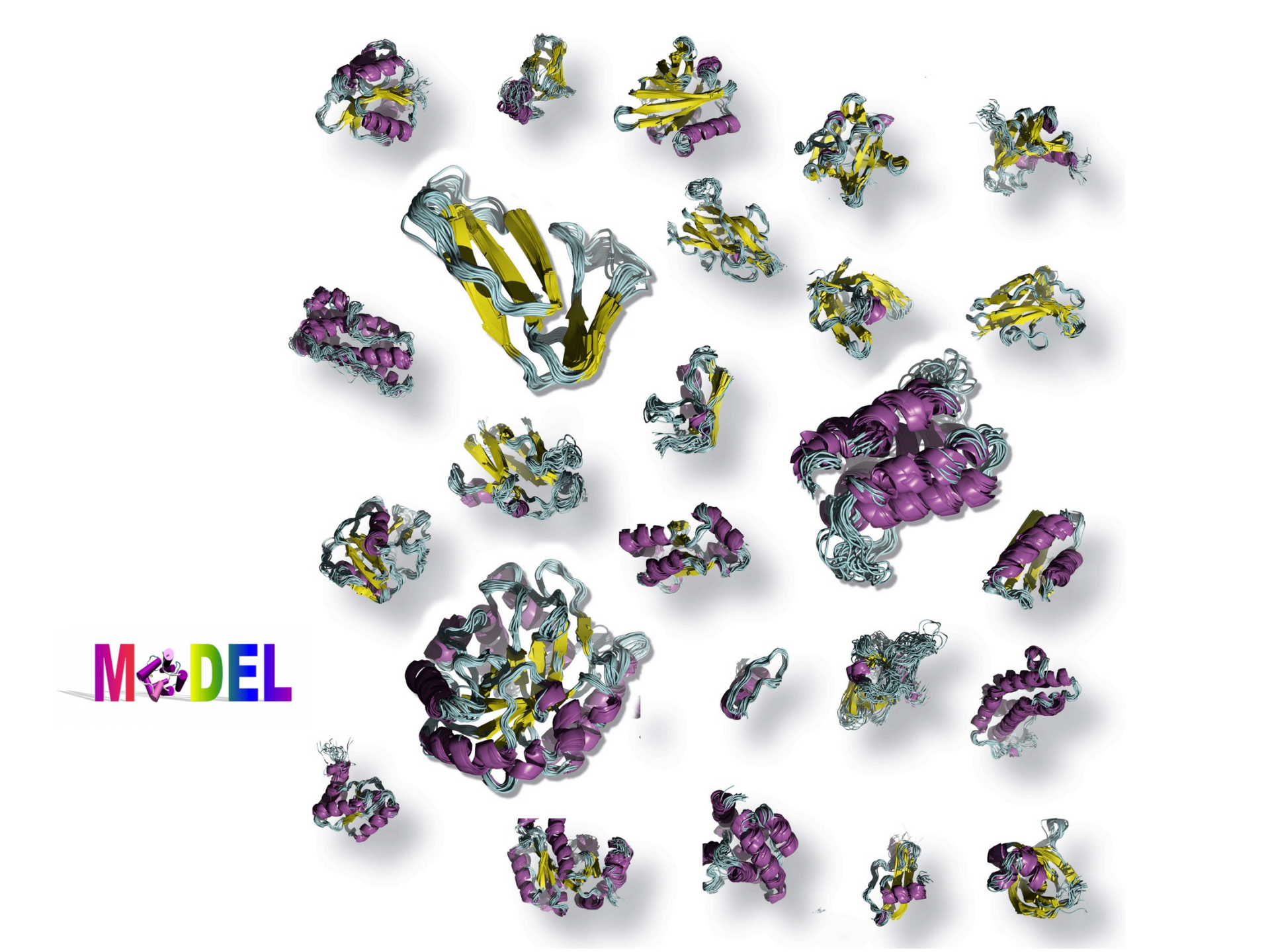
Publisher American Chemical Society

Description Docking and virtual screening (VS) reach maximum potential when the receptor displays the structural changes needed for accurate ligand binding. Unfortunately, these conformational changes are often poorly represented in experimental structures or homology models, debilitating their docking performance. Recently, we have shown that receptors optimized with our LiBERO method (Ligand-guided Backbone Ensemble Receptor Optimization) were able to better discriminate active ligands from inactives in flexible-ligand VS docking ...

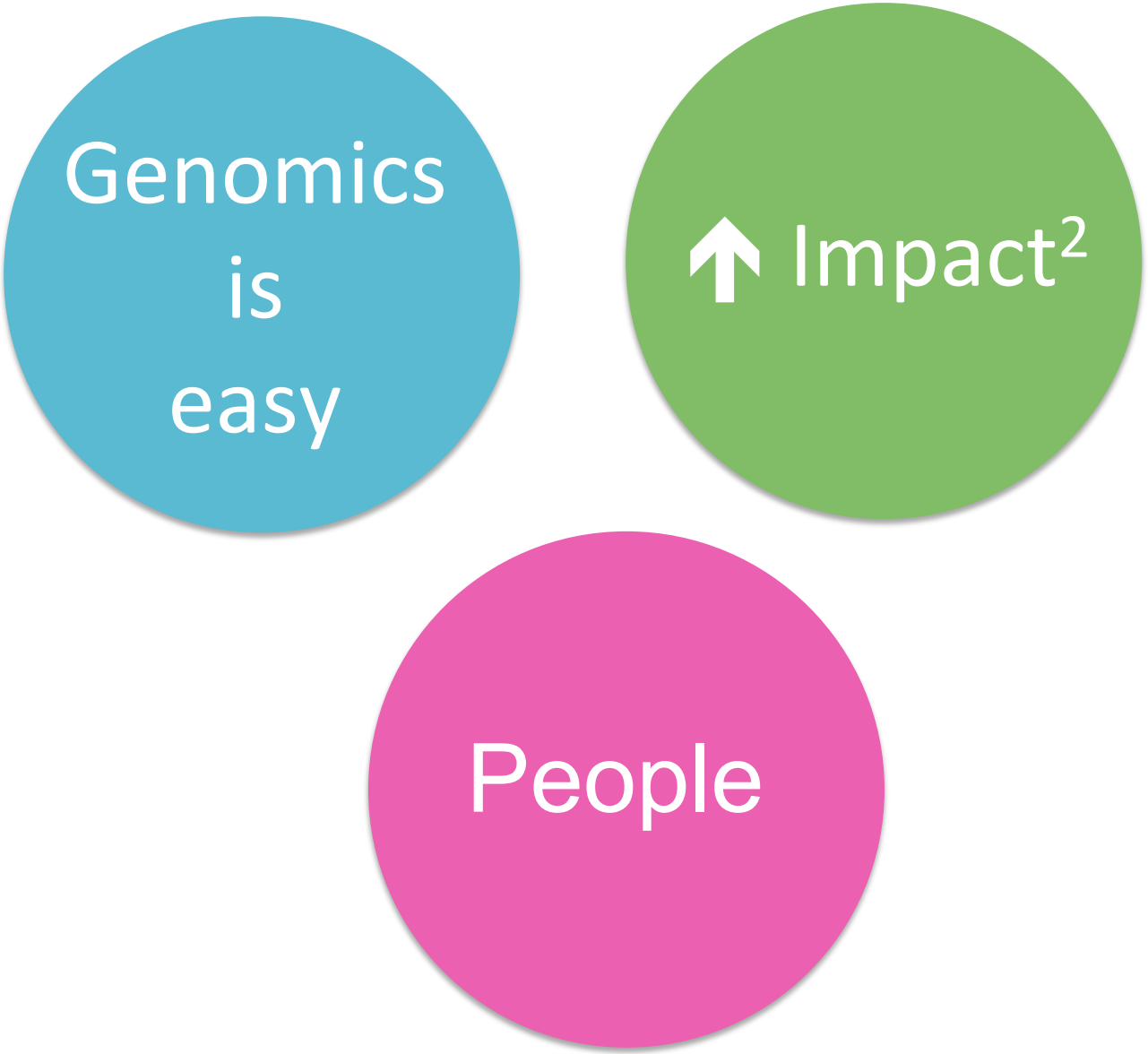

NO
PARKING
ANYTIME

molsoft
molecules *in silico*


BETONDEMONTAGETE
LOVE & PEACE
IN
WEST & EAST
Test
West


MODEL

Differences between Modeling – Genomics



It’s very easy to forget the past

I need to grab column X
from dbSNP for 250 ids

~ 600 lines of code
Results in *only* 45 min

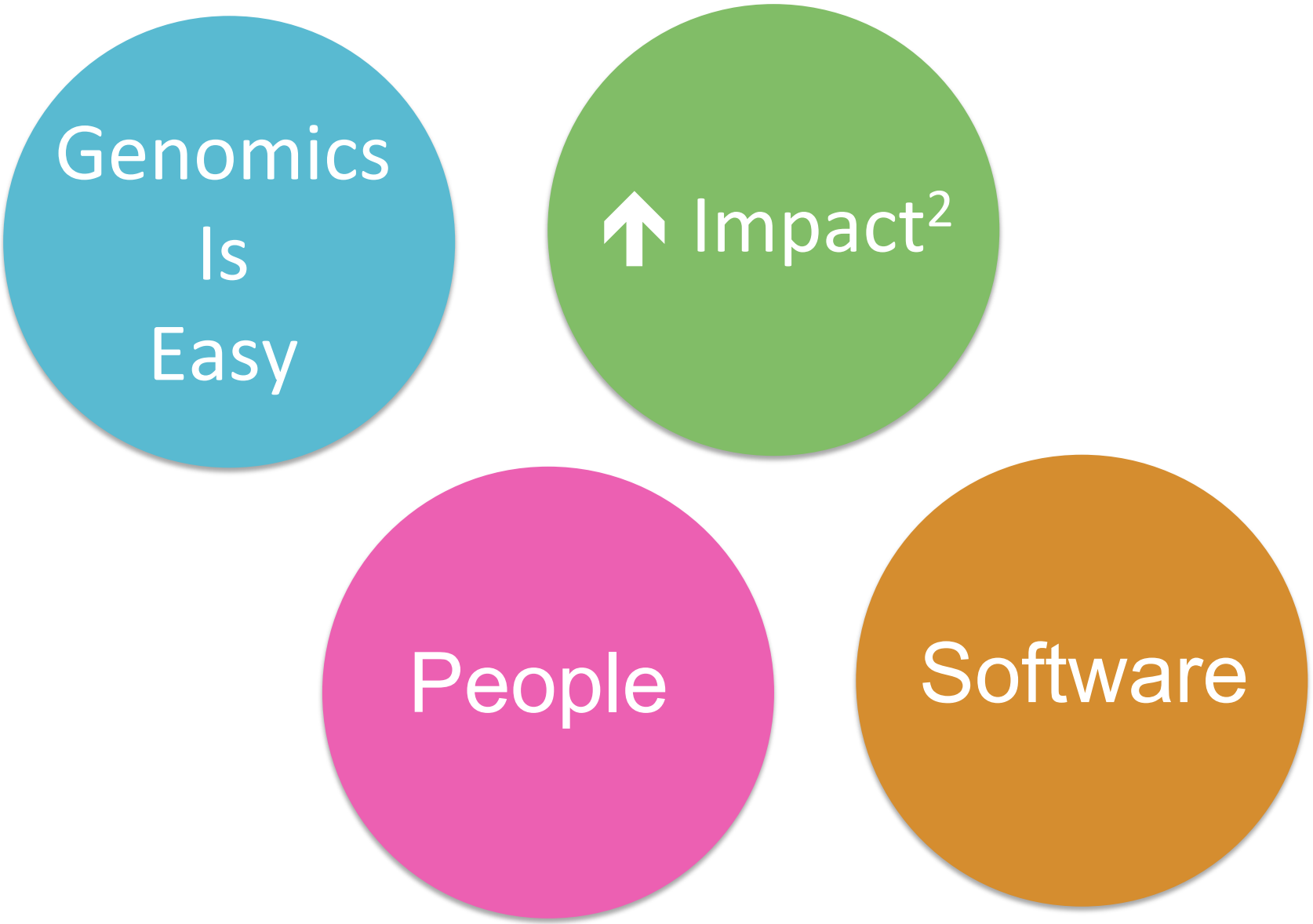
Python + R
Github

Import BioPython
BioConductor

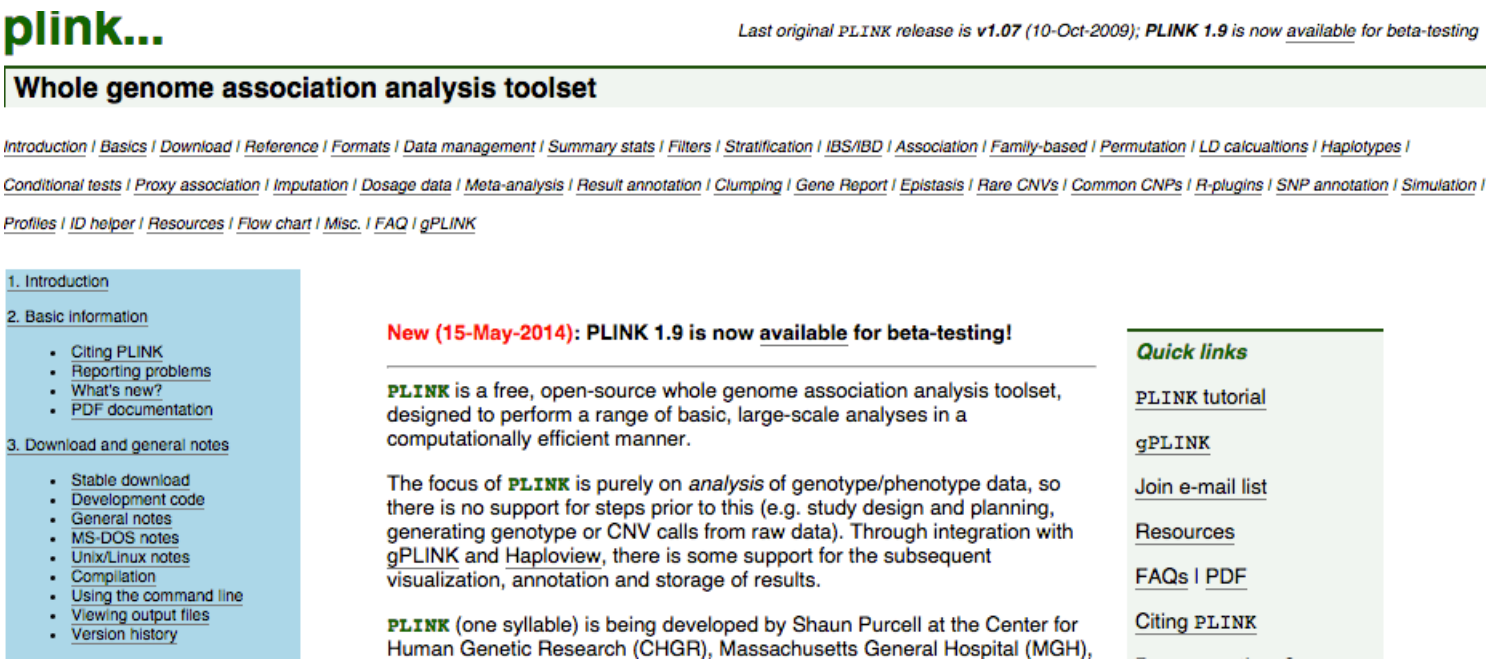
Connect MongoDB
somewhere

Elastic search /
Hadoop somewhere

Differences between Modeling – Genomics



PLINK software has 10K cites



plink...

*Last original PLINK release is **v1.07** (10-Oct-2009); **PLINK 1.9** is now available for beta-testing*

Whole genome association analysis toolset

Introduction | Basics | Download | Reference | Formats | Data management | Summary stats | Filters | Stratification | IBS/IBD | Association | Family-based | Permutation | LD calcualtions | Haplotypes | Conditional tests | Proxy association | Imputation | Dosage data | Meta-analysis | Result annotation | Clumping | Gene Report | Epistasis | Rare CNVs | Common CNPs | R-plugins | SNP annotation | Simulation | Profiles | ID helper | Resources | Flow chart | Misc. | FAQ | gPLINK

1. Introduction
2. Basic information
   - Citing PLINK
   - Reporting problems
   - What's new?
   - PDF documentation
3. Download and general notes
   - Stable download
   - Development code
   - General notes
   - MS-DOS notes
   - Unix/Linux notes
   - Compilation
   - Using the command line
   - Viewing output files
   - Version history

New (15-May-2014): PLINK 1.9 is now available for beta-testing!

PLINK is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner.

The focus of **PLINK** is purely on *analysis* of genotype/phenotype data, so there is no support for steps prior to this (e.g. study design and planning, generating genotype or CNV calls from raw data). Through integration with gPLINK and Haploview, there is some support for the subsequent visualization, annotation and storage of results.

PLINK (one syllable) is being developed by Shaun Purcell at the Center for Human Genetic Research (CHGR), Massachusetts General Hospital (MGH),

Quick links

PLINK tutorial

gPLINK

Join e-mail list

Resources

FAQs | PDF

Citing PLINK

(free, open source)

ALiBERO

Automated Ligand-guided Backbone Ensemble Receptor Optimization

Blind prediction of adenosine A_{2A} receptor structure with ligand

What is ALiBERO?

- A computational method that iteratively selects the **combination of pockets** that maximize a fitness function (e.g., AUC)
- **Perl script** (command line) + **ICM**-VS
  - Connected via icm scripts
  - Results come as graphical .icb

What do I need to run it?

- Linux workstation
- Alibero script (GitHub: mrueda)
- ICM-VS license
- Optional
  - Linux cluster

When should I try it?

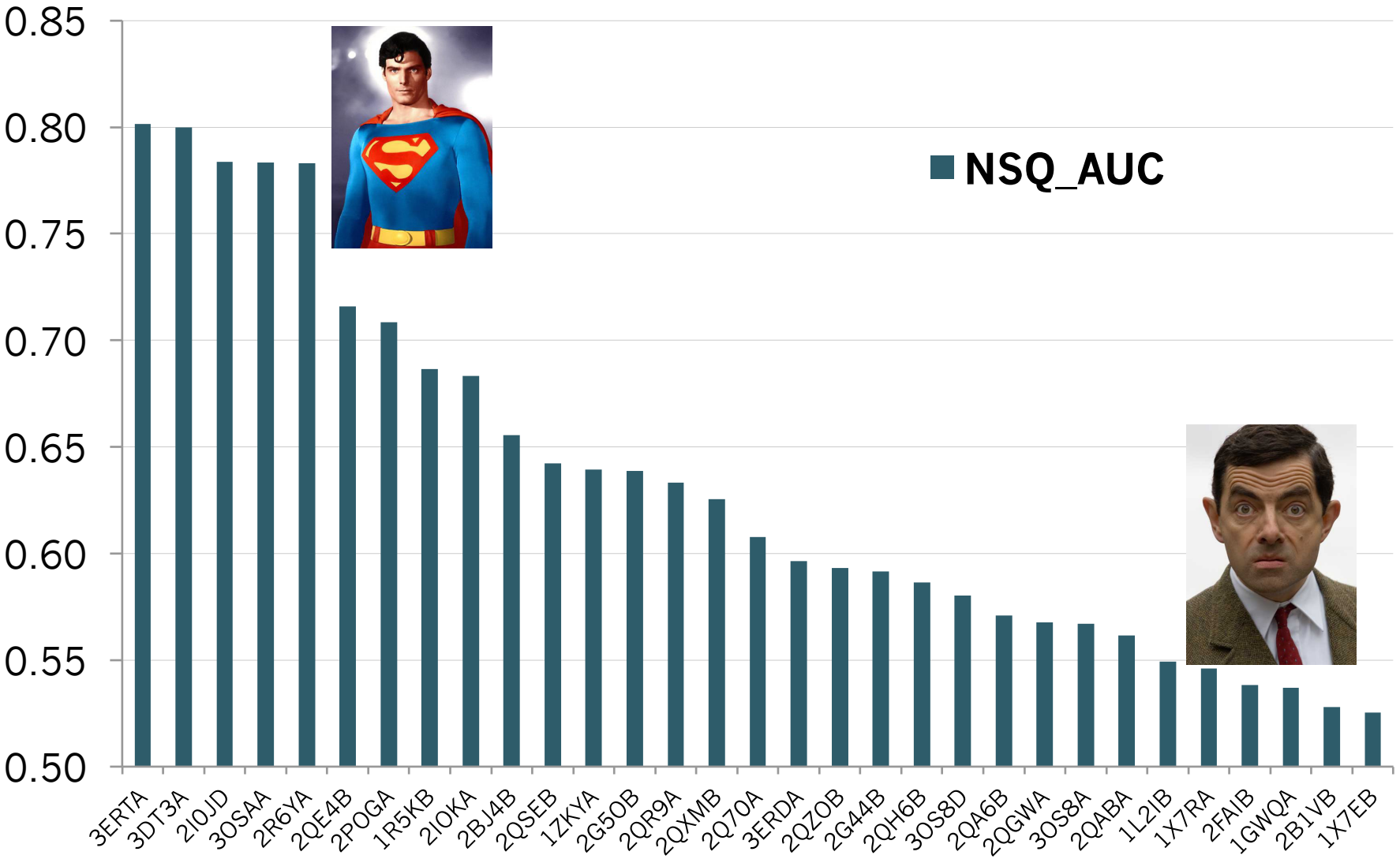
- You have experimental data for a few *actives* (2D is fine) and…

⟶ ***Sampling route***: your pocket(s) display bad recognition (AUC, scores) of known actives

  - *Non-sampling route*: you have multiple pockets and want to select the optimum ensemble for VS
  - both

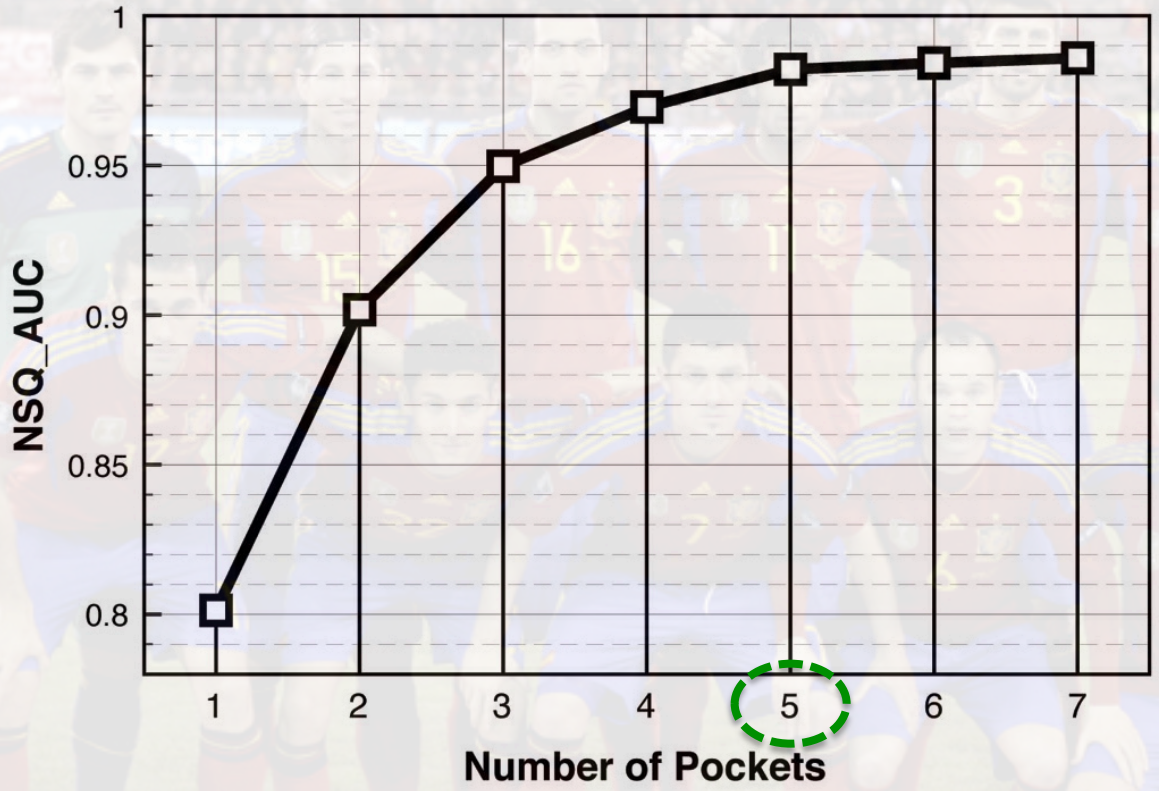
*via Normal Mode Analysis and Monte Carlo side chain refinement

Holo individual performance is hard to predict



A posteriori

Team-players selection method

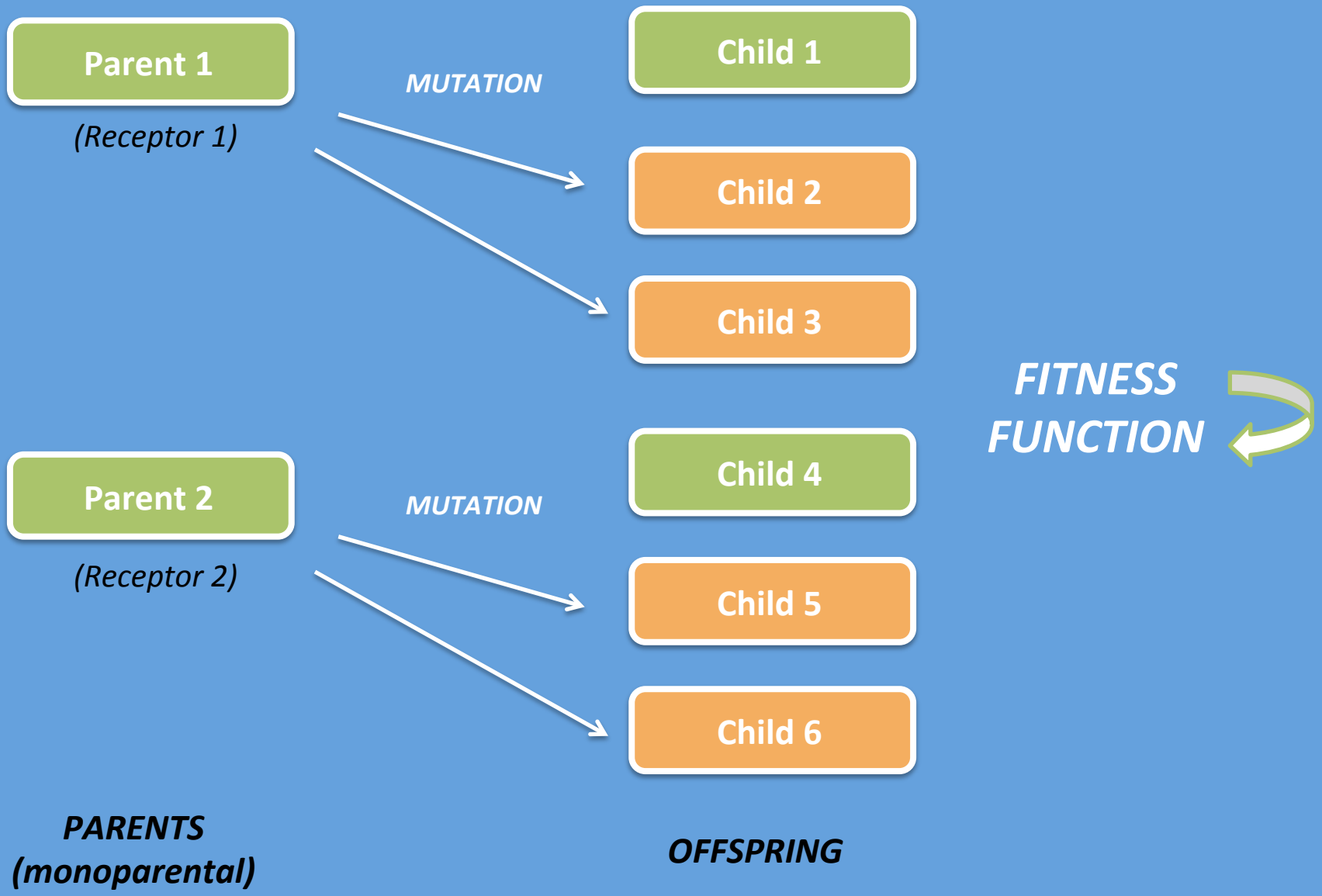


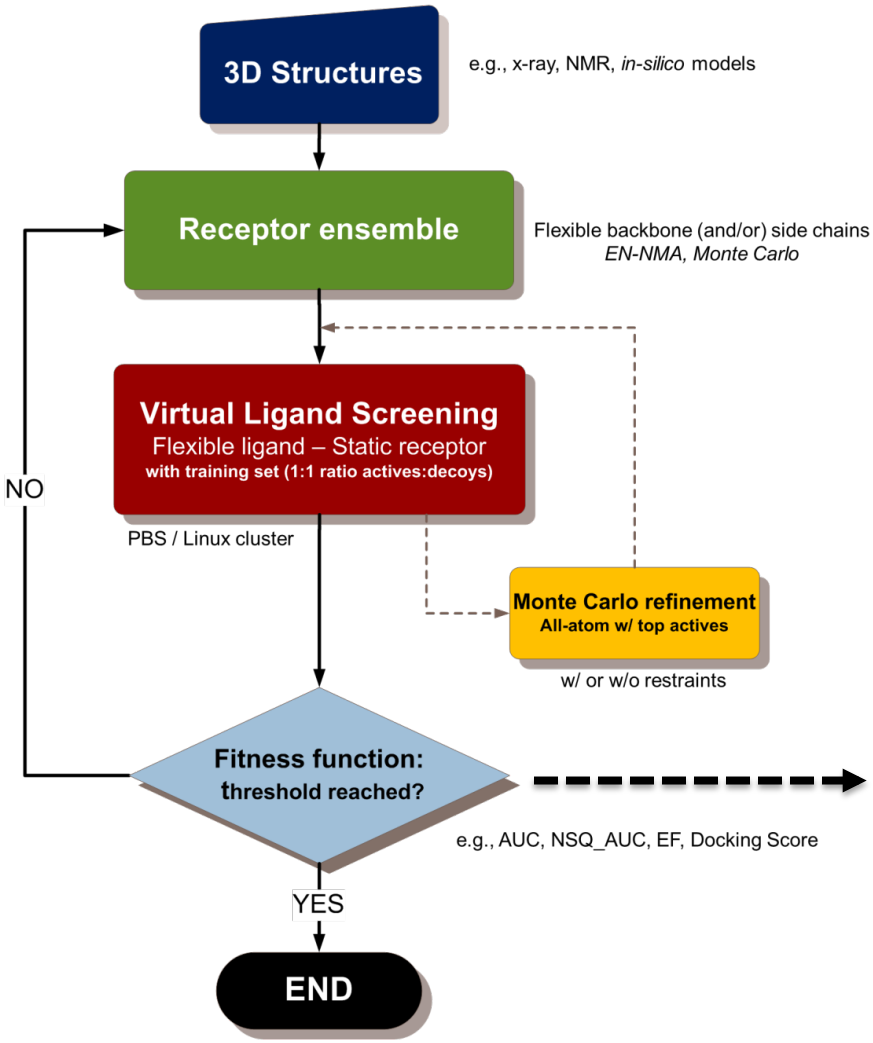
C(31,5)
169,911

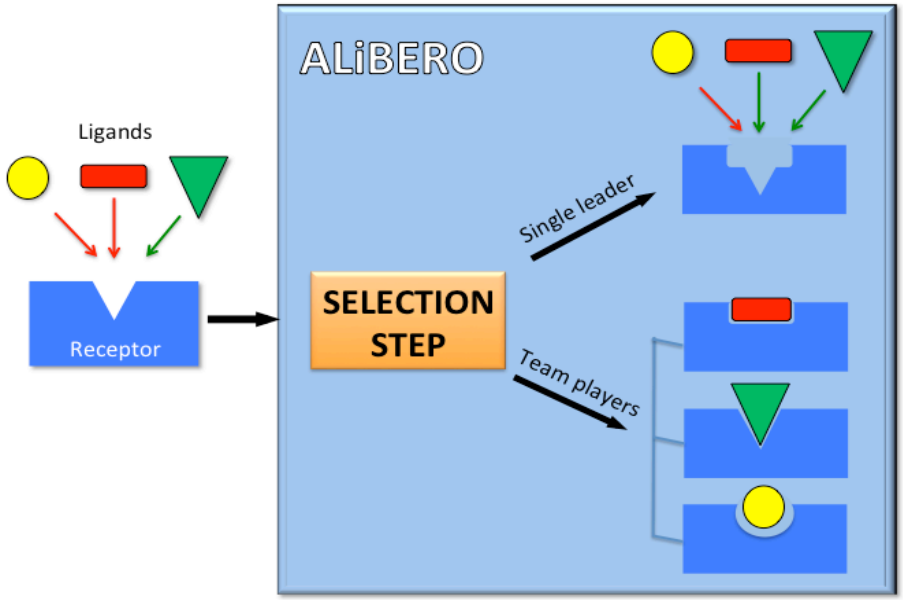
Our method
558

- Rueda. M, et al. ALiBERO: evolving a team of complementary pocket conformations rather than a single leader. *J Chem Inf Model*. 2012 Oct 22;52(10):2705-14.

ALiBERO is implemented as an Evolutionary Algorithm







- Rueda. M, et al. ALiBERO: evolving a team of complementary pocket conformations rather than a single leader. *J Chem Inf Model*. 2012 Oct 22;52(10):2705-14.

What is the expected performance?

Training Set
AUC: 80-90%
NSQ_AUC: ≥60%
Homology model

Test Set
NSQ_AUC
15-20%

Great for predicting binding modes

OK, now the bad news…

Overfitting happens

Prospective VS sensitive to false positive team players ≤ 5

What files do I need?

- One or multiple converted **receptors** embedded as objects in an .icb file
  - I recommend deleting residues not involved in the pocket
- ICM scripts (provided)
- An .sdf file with the **ligands** to be docked
  - The file must contain a column named "Active" with 1=actives and 0=decoys

Where is ALiBERO exe?

/pro/alibero/alibero

(latest version)

How do I run ALiBERO?

- $path/alibero –i *config_file* –n *MRCs* [–options]

  -v

  -help

  -man

  -verbose

  -debug

What is the flag –i?

- Configuration file - **sampling route**

```
inputicb      INPUT/RECEPTORS/aliberoMicro.icb
nligands      41
sdf           INPUT/LIGANDS/v10actives_w31Decoys.sdf
macrodir      INPUT/MACROS      # dir with ICM scripts
projdir       ESR1_HUMAN_test
function      nsaplus           # NSQ_AUC + score
refinement    on                # restraints in macro
mrc           3                 # team size
```

What is the flag –n?

- Number of “children”
  - Two modes:
    - Desktop (slow when –n > ncpu)
    - Cluster (recommended when –n 100)
      - Note that PBS.pm must be updated if outside Abagyan’s lab

ALiBERO has 5 fitness functions

- AUC (auc)
- NSQ_AUC (nsa)
- Average Score for ½ actives (score)
- **NSQ_AUC+ (nsaplus)←recommended**
- Consistency of binding mode for actives (con)

What do I need to modify?

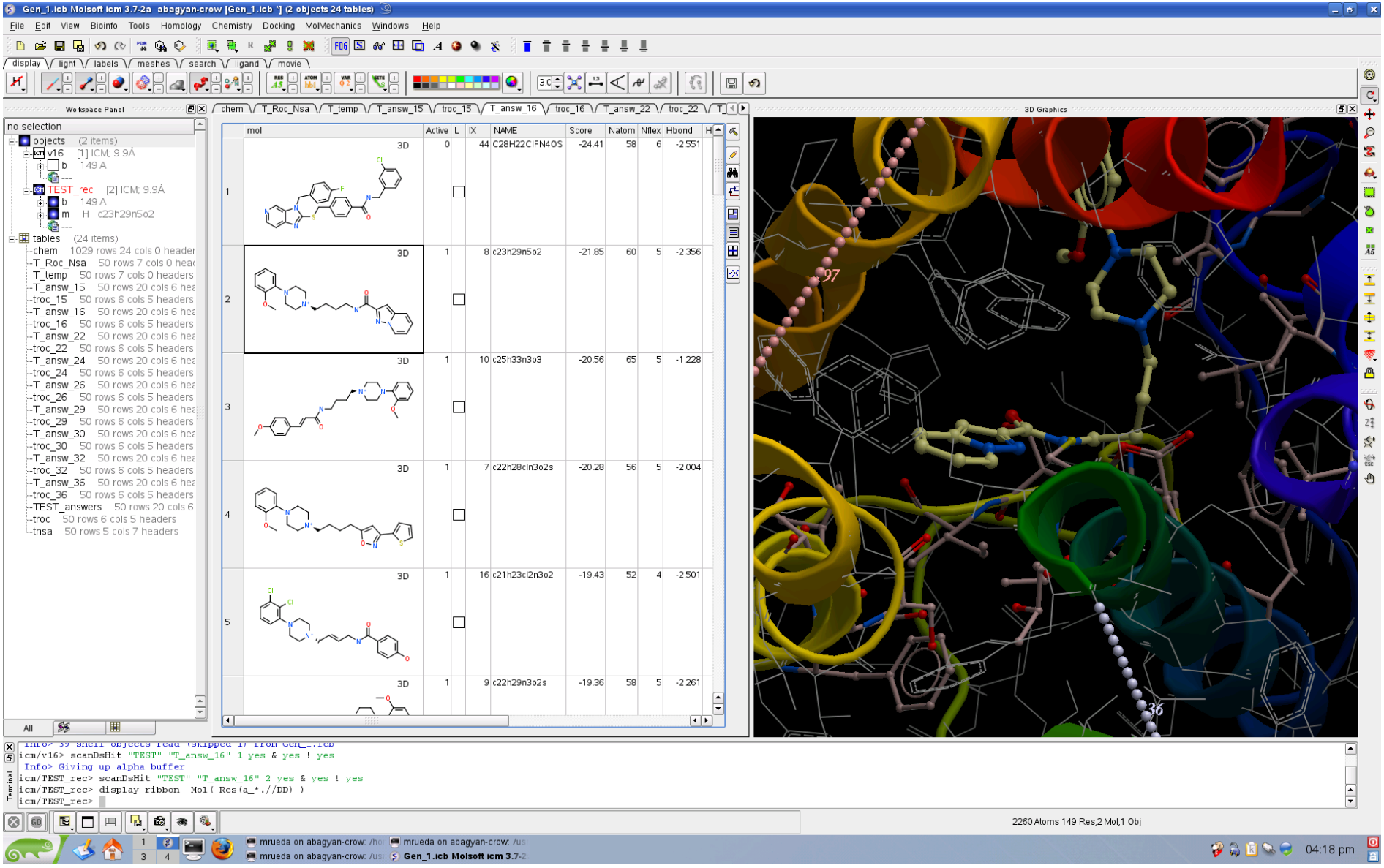
- Configuration file (parameters)
- *asel* for pocket definition
  - @ MakeDock.icm
- Drestraints (if any)
  - @Refine_Hitlist.icm

Is there any test I can run?

- Yes, just copy this folder:

  /pro/alibero/test

  locally, then modify the paths inside the *.in file


Gen_1.icb Molsoft icm 3.7-2a abagyan-crow [Gen_1.icb *] (2 objects 24 tables)
File Edit View Biointo Tools Homology Chemistry Docking MolMechanics Windows Help
display light labels meshes search ligand movie
Workspace Panel
no selection
objects (2 items)
v16 [1] ICM; 9.9Å
b 149 A
TEST_rec [2] ICM; 9.9Å
b 149 A
m H c23h29n5o2
tables (24 items)
chem 1029 rows 24 cols 0 header
T_Roc_Nsa 50 rows 7 cols 0 hea
T_temp 50 rows 7 cols 0 headers
T_answ_15 50 rows 20 cols 6 hea
troc_15 50 rows 6 cols 5 headers
T_answ_16 50 rows 20 cols 6 hea
troc_16 50 rows 6 cols 5 headers
T_answ_22 50 rows 20 cols 6 hea
troc_22 50 rows 6 cols 5 headers
T_answ_24 50 rows 20 cols 6 hea
troc_24 50 rows 6 cols 5 headers
T_answ_26 50 rows 20 cols 6 hea
troc_26 50 rows 6 cols 5 headers
T_answ_29 50 rows 20 cols 6 hea
troc_29 50 rows 6 cols 5 headers
T_answ_30 50 rows 20 cols 6 hea
troc_30 50 rows 6 cols 5 headers
T_answ_32 50 rows 20 cols 6 hea
troc_32 50 rows 6 cols 5 headers
T_answ_36 50 rows 20 cols 6 hea
troc_36 50 rows 6 cols 5 headers
TEST_answers 50 rows 20 cols 6
troc 50 rows 6 cols 5 headers
tnsa 50 rows 5 cols 7 headers
All
chem T_Roc_Nsa T_temp T_answ_15 troc_15 T_answ_16 troc_16 T_answ_22 troc_22 T
mol Active L IX NAME Score Natom Nflex Hbond H
1 3D 0 44 C28H22ClFN4OS -24.41 58 6 -2.551
2 3D 1 8 c23h29n5o2 -21.85 60 5 -2.356
3 3D 1 10 c25h33n3o3 -20.56 65 5 -1.228
4 3D 1 7 c22h28cln3o2s -20.28 56 5 -2.004
5 3D 1 16 c21h23cl2n3o2 -19.43 52 4 -2.501
3D 1 9 c22h29n3o2s -19.36 58 5 -2.261
3D Graphics
97
36
Info> 39 shell objects read (skipped 1) from Gen_1.icb
icm/v16> scanDsHit "TEST" "T_answ_16" 1 yes & yes ! yes
Info> Giving up alpha buffer
icm/TEST_rec> scanDsHit "TEST" "T_answ_16" 2 yes & yes ! yes
icm/TEST_rec> display ribbon Mol(Res(a_*.//DD))
icm/TEST_rec>
Terminal
2260 Atoms 149 Res,2 Mol,1 Obj
mrueda on abagyan-crow: /ho
mrueda on abagyan-crow: /us
mrueda on abagyan-crow: /us
Gen_1.icb Molsoft icm 3.7-2
04:18 pm