https://3DGENOMES.ORG

https://tinyurl.com/dir-upf-tutorial-3DG

PRE TUTORIAL

1. DOWNLOAD DATA TO YOUR HOME DIR
Go to https://tinyurl.com/dir-upf-tutorial-3DG

2. UNTAR DATA IN YOUR HOME DIR
>tar -xvf tutorial_3DG.tar.gz

3. SET GEM PATH

>PATH=\$PATH:/home/uXXXXX/tutorial_3DG/gem

TUTORIAL

1. ENTER THE TUTORIAL DIR >cd turorial 3DG

2. QUALITY PLOTS (seconds each command)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot

2. MAP (~8 minutes each command, ~15 minutes all)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII

3. PARSE (~5 minutes) >tadbit parse results --compress input --genome db/dm6.fa

4. FILTER (~4 minutes)

>tadbit filter results

4. NORMALIZE @100Kb (~2 minutes)
>tadbit normalize results -r 100000 --min count 100

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes)

>tadbit segment results -r 100000 -c 2L 2R 3L 3R 4 X

6. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]
>model_and_analyze.py --cfg modeling.cfg

Structure determination of genomes and genomic domains by satisfaction of spatial restraints

Marc A. Marti-Renom Structural Genomics Group (ICREA, CNAG-CRG)

http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu







Resolution Gap

Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

Know	edge								
to the second					IDM			$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	
								DNA length	
10 ⁰		10 ³			10 ⁶			10 ⁹	nt
								Volume	
10 ⁻⁹		10 ⁻⁶	10 ⁻¹	3		10 [°]		10 ³	μm³
								Time	
10 ⁻¹⁰	10 ⁻⁸	10 ⁻⁶	10 ⁻⁴	10 ⁻²		10 ⁰	10 ²	10 ³	S
								Resolution	
10 ⁻³			10 ⁻²				10 ⁻¹		μ



Hi-C experiment

Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.



Mapping & Filtering

Imakaev, M. V et al. (2012). Nature Methods, 9(10), 999–1003.

Raw reads































mapped read1









How much you normally map?

- 80-90% each end => 60-80% intersection
- ~1% multiple contacts
- Many of intersecting pairs will be lost in filtering...
- Final 40-60% of valid pairs
- One measure of quality is the CIS/TRANS ration (70-80% good)

https://tinyurl.com/dir-upf-tutorial-3DG

PRE TUTORIAL

1. DOWNLOAD DATA
Go to https://tinyurl.com/tutorial-3DG

2. UNTAR DATA
>tar -xvf tutorial_3DG.tar.gz

3. SET GEM PATH
>PATH=\$PÂTH:/home/uXXXX/tutorial_3DG/gem

TUTORIAL

1. ENTER THE TUTORIAL DIR
>cd turorial_3DG

2. QUALITY PLOTS (seconds each command)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot

2. MAP (~8 minutes each command, ~15 minutes all)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII

3. PARSE (~5 minutes)
>tadbit parse results --compress input --genome db/dm6.fa

4. FILTER (~4 minutes)
>tadbit filter results

4. NORMALIZE @100Kb and @10Kb (~2 minutes)
>tadbit normalize results -r 100000 -min count 100

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes) >tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X

7. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]
>Model_and_analyze.py -cnf modeling.cnf









Zooming in on genome organization. Zhou, X. J., & Alber, F. Nature Methods (2012)











https://tinyurl.com/dir-upf-tutorial-3DG

PRE TUTORIAL

1. DOWNLOAD DATA
Go to https://tinyurl.com/tutorial-3DG

2. UNTAR DATA
>tar -xvf tutorial_3DG.tar.gz

3. SET GEM PATH
>PATH=\$PÂTH:/home/uXXXX/tutorial_3DG/gem

TUTORIAL

1. ENTER THE TUTORIAL DIR
>cd turorial_3DG

2. QUALITY PLOTS (seconds each command)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot

2. MAP (~8 minutes each command, ~15 minutes all)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII

3. PARSE (~5 minutes)
>tadbit parse results --compress input --genome db/dm6.fa

4. FILTER (~4 minutes)
>tadbit filter results

4. NORMALIZE @100Kb and @10Kb (~2 minutes)
>tadbit normalize results -r 100000 -min count 100

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes) >tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X

7. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]
>Model_and_analyze.py -cnf modeling.cnf



TADs Chromosome 14



https://tinyurl.com/dir-upf-tutorial-3DG

PRE TUTORIAL

1. DOWNLOAD DATA
Go to https://tinyurl.com/tutorial-3DG

2. UNTAR DATA
>tar -xvf tutorial_3DG.tar.gz

3. SET GEM PATH
>PATH=\$PÂTH:/home/uXXXX/tutorial_3DG/gem

TUTORIAL

1. ENTER THE TUTORIAL DIR
>cd turorial_3DG

2. QUALITY PLOTS (seconds each command)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot

2. MAP (~8 minutes each command, ~15 minutes all)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII

3. PARSE (~5 minutes)
>tadbit parse results --compress input --genome db/dm6.fa

4. FILTER (~4 minutes)
>tadbit filter results

4. NORMALIZE @100Kb and @10Kb (~2 minutes)
>tadbit normalize results -r 100000 -min count 100

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes) >tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X

7. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]
>Model_and_analyze.py -cnf modeling.cnf

Got normalized Hi-C maps?



http://3DGenomes.org http://www.integrativemodeling.org





Model representation and scoring

Harmonic

 $d = d_0$ - - - 1 $d < d_0$ $d > d_0$

$$H_{i,j} = k \left(d_{i,j} - d_{i,j}^0 \right)^2$$

$$\begin{cases} if \ d_{i,j} \ge d_{i,j}^{0}; & ubH_{i,j} = k \left(d_{i,j} - d_{i,j}^{0} \right)^{2} \\ if \ d_{i,j} < d_{i,j}^{0}; & ubH_{i,j} = 0 \end{cases}$$

Harmonic Lower Bound

$$\begin{cases}
if \ d_{i,j} \le d_{i,j}^{0}; \quad lbH_{i,j} = k(d_{i,j} - d_{i,j}^{0})^{2} \\
if \ d_{i,j} > d_{i,j}^{0}; \quad lbH_{i,j} = 0
\end{cases}$$



From 3C data to spatial distances



Neighbor fragments







Parameter optimization





Parameter optimization



Contact map



Optimization of the scoring function



Model analysis: clustering and structural features



Accessibility (%)



Interactions

Angle







https://tinyurl.com/dir-upf-tutorial-3DG

PRE TUTORIAL

1. DOWNLOAD DATA
Go to https://tinyurl.com/tutorial-3DG

2. UNTAR DATA
>tar -xvf tutorial_3DG.tar.gz

3. SET GEM PATH
>PATH=\$PÂTH:/home/uXXXX/tutorial_3DG/gem

TUTORIAL

1. ENTER THE TUTORIAL DIR
>cd turorial_3DG

2. QUALITY PLOTS (seconds each command)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII --qc_plot
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII --qc_plot

2. MAP (~8 minutes each command, ~15 minutes all)

>tadbit map results --fastq FASTQs/Control_rep4.1_10M.fastq --read 1 --index db/dm6.gem --renz DpnII
>tadbit map results --fastq FASTQs/Control_rep4.2_10M.fastq --read 2 --index db/dm6.gem --renz DpnII

3. PARSE (~5 minutes)
>tadbit parse results --compress input --genome db/dm6.fa

4. FILTER (~4 minutes)
>tadbit filter results

4. NORMALIZE @100Kb and @10Kb (~2 minutes)
>tadbit normalize results -r 100000 -min count 100

5. SEGMENT THE GENOME @100Kb A/B + TADs (~2 minutes) >tadbit segment results -r 100000 --fasta db/dm6.fa -c 2L 2R 3L 3R 4 X

7. MODEL (~20 minutes)

[PROBLEM: cPicke module not installed... we will walk through the process and see the final result]
>Model_and_analyze.py -cnf modeling.cnf



David Castillo Yasmina Cuartero Irene Farabella Silvia Galan Mike Goodstadt Julen Mendieta Francesca Mugianesi Juan Rodríguez François Serra Paula Soler Aleksandra Sparavier Marco di Stefano

http://sgt.cnag.cat/www/presentations/



http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu