

a package to analyse associations between transcriptomics and 3D genomics

Projet Long
Hocine Meraouna

2018/2019

Introduction

- 3D genome organization role in regulation of biological processes and relationship with transcription
- chromosome conformation capture (3C) methods
- next generation sequencing application to transcriptomics
- Python package
- genome wide study associations genome 3D architecture, transcriptional activity and genome regulation

Material

- Python3
- pandas
- NumPy
- SciPy
- Matplotlib
- data files
 - The genes positions file
 - The genes expression file

Method

- genes positions and expression files, load into pandas data frames
- Euclidean distance matrix (scipy.spatial cdist function)
- Spearman correlation matrix (pandas corr function)
- sum correlation coefficient of n closest genes for each gene
- data frame containing with x, y z coord and the sum_corr
- 3D plot genes on x, y, z axis, coloured from blue to red following sum_corr (matplotlib.pyplot scatter function)

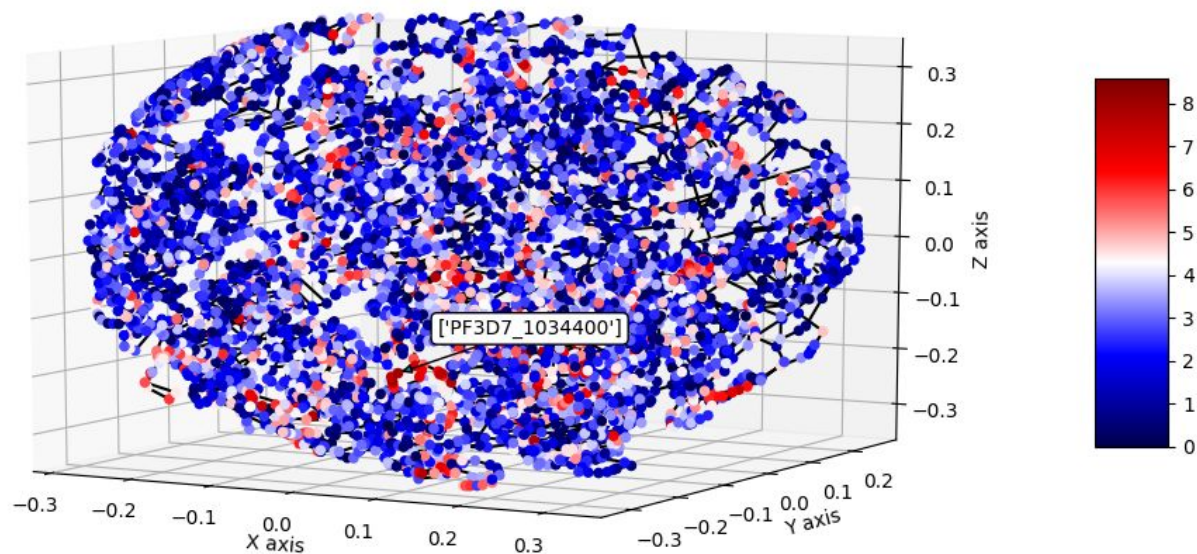
Result

- *plasmodium falciparum* genes expressions data (3 intraerythropoietic stages)
- genes position data (schizont stage)

Ferhat Ay et al. Genome research 2014

Result

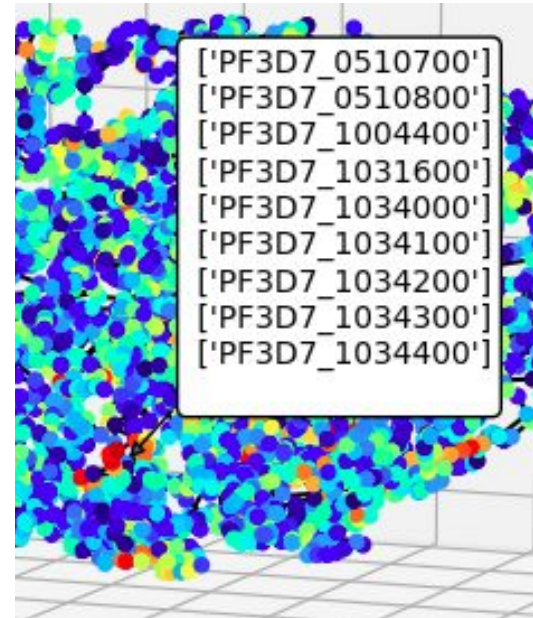
3D Transmap of profiles_Otto2010



region mostly coding for putative proteins

Discussion and Conclusion

- gives the expected result
- not fast enough (10~12 min)
- huge amount of genes less readability
- hover issue



THANK YOU FOR YOUR ATTENTION

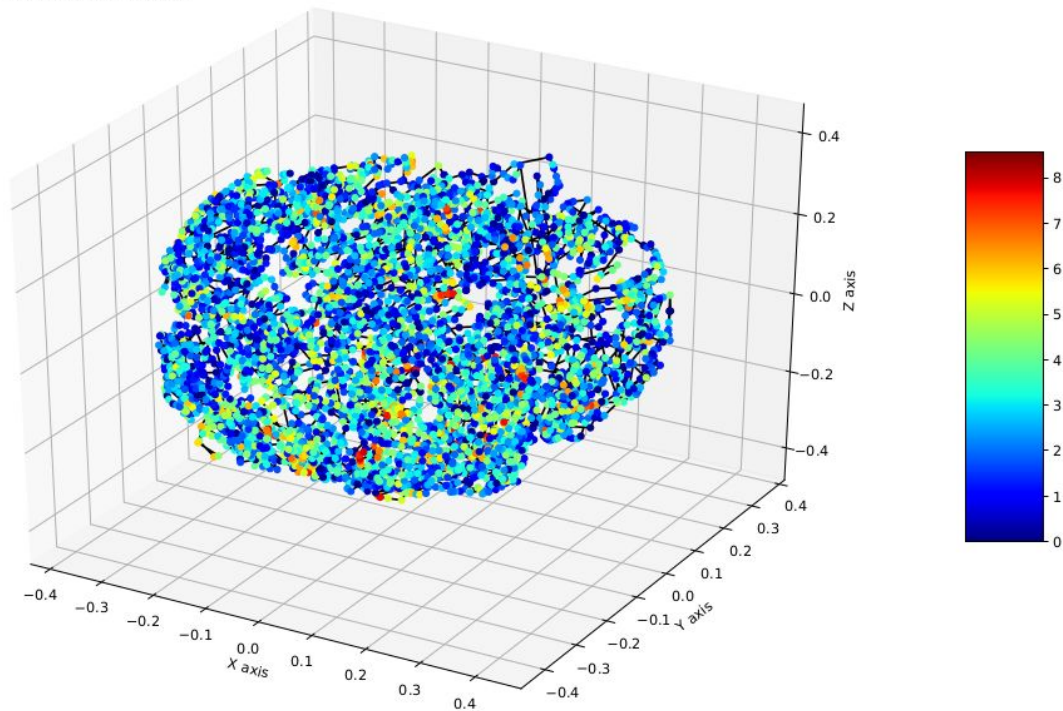
Supplementary

1	chr	X	Y	Z	
2	PF3D7_0100100	chr01	0.358894	0.120871	0.107271
3	PF3D7_0100200	chr01	0.356576	0.083654	0.093874
4	PF3D7_0100300	chr01	0.366940	0.097076	0.079452
5	PF3D7_0100400	chr01	0.374553	0.117058	0.065022
6	PF3D7_0100500	chr01	0.360451	0.115275	0.071775
7	PF3D7_0100600	chr01	0.353052	0.114340	0.075319
8	PF3D7_0100700	chr01	0.337844	0.112417	0.082601
9	PF3D7_0100800	chr01	0.318406	0.111964	0.093615
10	PF3D7_0100900	chr01	0.329376	0.129016	0.101694

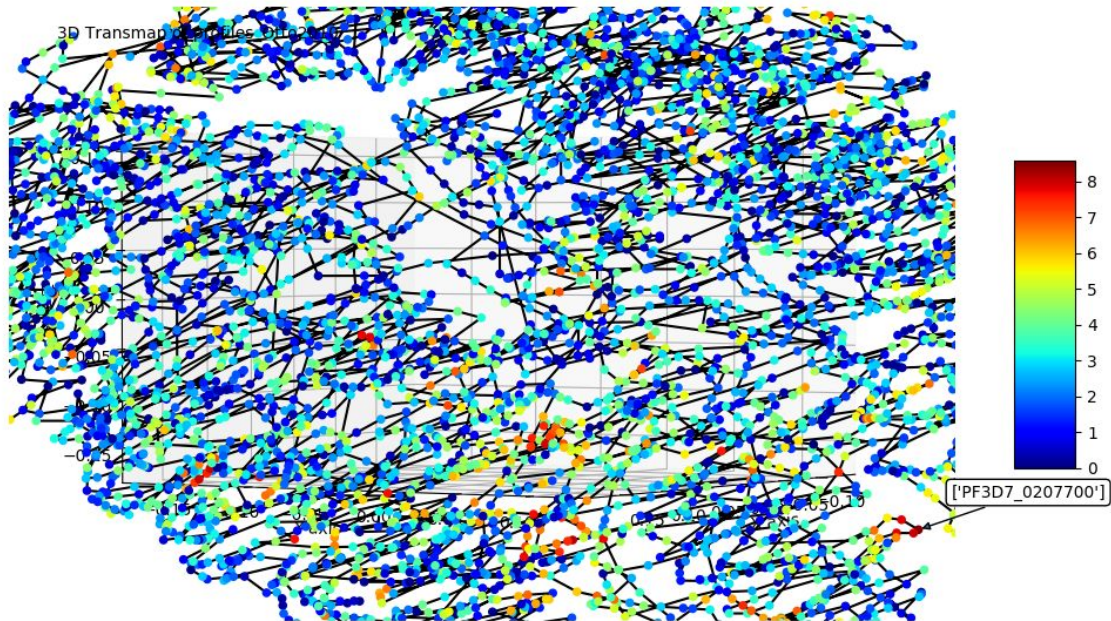
1	Hour0	Hour8	Hour16	Hour24	Hour32	Hour40	Hour48	
2	PF3D7_0100100	5.8692	10.0409	8.5659	6.1746	7.5189	11.3458	11.8587
3	PF3D7_0100200	5.3815	3.6215	1.6919	4.6601	1.2313	3.9779	3.8809
4	PF3D7_0100300	3.1295	5.0767	2.2089	1.1385	1.1082	1.5369	2.3447
5	PF3D7_0100400	3.7487	6.5621	6.6931	5.3734	7.0145	8.0469	6.9087
6	PF3D7_0100500	2.0807	0	0	6.593	0	0	7.1902
7	PF3D7_0100600	0.3236	0.2385	0.26	1.6604	0.2271	0.3335	2.9826
8	PF3D7_0100700	0	0	0	0	0.8827	0	
9	PF3D7_0100800	1.0132	0.6223	0.2713	1.4269	1.1849	4.35	2.9566
10	PF3D7_0100900	0.8852	0.4893	1.7781	3.0053	0.9317	2.9074	4.4356

Supplementary

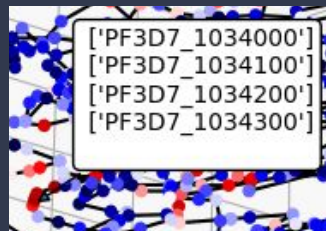
3D Transmap of profiles_Otto2010



Supplementary



Supplementary



Plasmodium falciparum 3D7: PF3D7_1034400

Help

Entry	PF3D7_1034400	CDS	T00095
Definition	(RefSeq) flavoprotein subunit of succinate dehydrogenase		
KO	K00234 succinate dehydrogenase (ubiquinone) flavoprotein subunit [EC:1.3.5.1]		
Organism	pfa Plasmodium falciparum 3D7		
Pathway	<p>pfa00020 Citrate cycle (TCA cycle)</p> <p>pfa00190 Oxidative phosphorylation</p> <p>pfa01100 Metabolic pathways</p> <p>pfa01110 Biosynthesis of secondary metabolites</p> <p>pfa01130 Biosynthesis of antibiotics</p> <p>pfa01200 Carbon metabolism</p>		
Brite	<p>KEGG Orthology (KO) [BR:pfa00001]</p> <p>09100 Metabolism</p> <p>09101 Carbohydrate metabolism</p> <p>00020 Citrate cycle (TCA cycle)</p> <p>PF3D7_1034400</p> <p>09102 Energy metabolism</p> <p>00190 Oxidative phosphorylation</p> <p>PF3D7_1034400</p> <p>Enzymes [BR:pfa01000]</p> <p>1. Oxidoreductases</p> <p>1.3 Acting on the CH-CH group of donors</p> <p>1.3.5 With a quinone or related compound as acceptor</p> <p>1.3.5.1 succinate dehydrogenase</p> <p>PF3D7_1034400</p> <p>BRITE hierarchy</p>		

Protein | Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial

Gene | PF3D7_1034400

Organism | Plasmodium falciparum (isolate 3D7)

Status | Unreviewed - Annotation score: ●●●○○○ - Protein inferred from homologyⁱ

Protein | Submitted name: **Sec1 family protein, putative**

Gene | PF3D7_1034000

Organism | Plasmodium falciparum (isolate 3D7)

Status | Unreviewed - Annotation score: ●○○○○○ - Protein inferred from homologyⁱ

Protein | Submitted name: **Thioredoxin-like associated protein 2, putative**

Gene | PF3D7_1034300

Organism | Plasmodium falciparum (isolate 3D7)

Status | Unreviewed - Annotation score: ●○○○○○ - Protein predictedⁱ

Supplementary

PF3D7_1034000 Sec1 family protein, putative [*Plasmodium falciparum* 3D7]

Gene ID: 810488, updated on 6-Apr-2018

PF3D7_1034100 conserved Plasmodium protein, unknown function [*Plasmodium falciparum* 3D7]

Gene ID: 8444951, updated on 6-Apr-2018

PF3D7_1034200 apicoplast ribosomal protein L27 precursor, putative [*Plasmodium falciparum* 3D7]

Gene ID: 810489, updated on 6-Apr-2018

PF3D7_1034300 thioredoxin-like associated protein 2, putative [*Plasmodium falciparum* 3D7]

Gene ID: 810490, updated on 6-Apr-2018

PF3D7_1034400 flavoprotein subunit of succinate dehydrogenase [*Plasmodium falciparum* 3D7]

Gene ID: 810491, updated on 6-Apr-2018