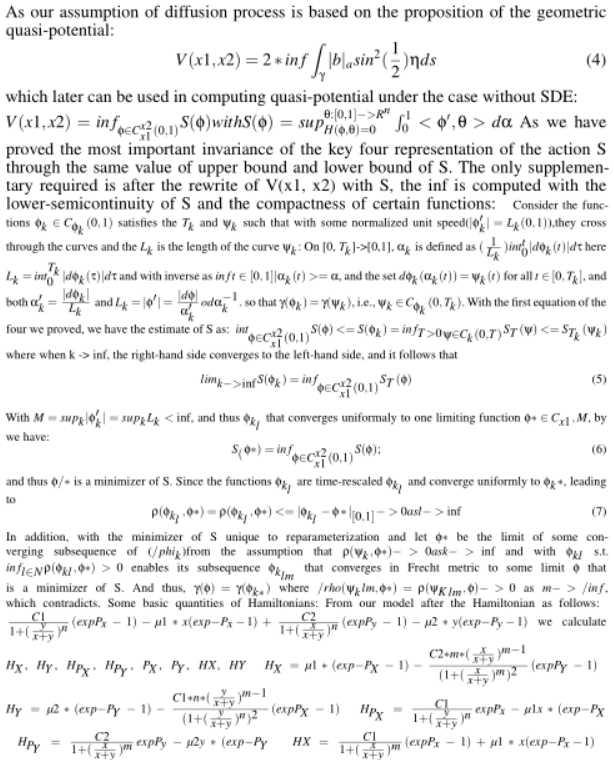
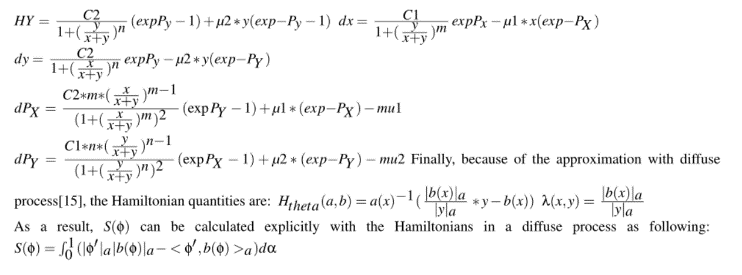
Append

A Please find supplementary data, code and experimental results in this link:

<https://github.com/dashboard>

B Proof





a. linear regression method on data

Y = B\*X=-0.00114 -6.69764 1.11999 4345.69955 3.25582

NA 0.33371

And the analysis summary :

#Coefficients: (1 not defined because of .66singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.00114 0.060137 88 0.001425

as.matrix(X)V2 1.11999 0.003344 66.79 0.0001425 \*\*

as.matrix(X)V4 3.25582 0.000215 271.46 0.0001425 \*\*

as.matrix(X)V5 -0.44712 0.000600 1.02 0.0001425 \*\*

as.matrix(X)V6 0.33371 0.287314 2e10\*\*2 0.0001425 \*\*

Residual standard error: 0.231 on 456 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 0.9946

F-statistic: 1.21e+02 on 2 and 456 DF, pvalue 254

The estimated effect of V2,V4, V5, V6 on the convert is

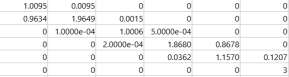
1.11999 3.25582 , -0.44712, 0.3371.

b.Auxiliary method

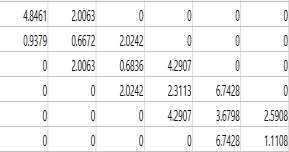
to resampled the makorv

transition matrix in the mRNA tto protein

switching model which is P1 before



And Augmented into P2:



h =0 0 0 0 0 0

p =[0.3667, 0.4350,0.2222, 0.1495,0.1030, 0.8709]

ci =[ -2.2811,1.0108; -1.8172, 0.9155;-1.7683, 0.5251;

-3.7605, 0.7645; -5.7797, 0.7328; -1.5520, 1.3584]

And whith those h all equals to 0, does not rejects the

P1 and P2 follows the same distribution.

C.Code

MaxStep = 50;

W = eye(size(z));

X = z;

v = var(X)+0.000000001;

for steps = 1:MaxStep

temp = ones(size(v))./v;

Vtemp = inv(X'\*inv(W)\*X+temp);

V = Vtemp\*eye(size(Vtemp))\*Vtemp';

L = chol(abs(V),'lower');

S = V\*X';

B = S\*inv(W)\*z;

% observations for normal with W

H = X.\*S;

CW = H./(W-H);

CW(isnan(CW))=0;

m = X.\*B;

m = m-CW.\*(z-m);

qtemp = CW.\*(CW+1);

qtemp(isnan(qtemp)) = 0;

% draw Z from truncated normal

q = sqrt(qtemp\*eye(size(qtemp))\*qtemp');

Z = X;

R = X;

%Z(:,i) = mvnpdf(X,mean(m,2)',q);

temp = mean(m,2)';

for i = 1:size(X,1)

Z(:,i) = normpdf(X(:,i),temp(i),q(i,i));

end

B.Auxiliary method to resampled the makorv

transition matrix in the mRNA to protein

switching model which is P1 before.

And Augmented into P2:

h =0 0 0 0 0 0

p =[0.3667, 0.4350,0.2222, 0.1495,0.1030, 0.8709]

ci =[ -2.2811,1.0108; -1.8172, 0.9155;-1.7683, 0.5251;

-3.7605, 0.7645; -5.7797, 0.7328; -1.5520, 1.3584]

And whith those h all equals to 0, does not rejects the

P1 and P2 follows the same distribution.

A. linear regression method on data, (Y = B\*X) :

-0.00114 -6.69764 1.11999 4345.69955 3.25582

NA 0.33371

And the analysis summary :

#Coefficients: (1 not defined because of .66singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.00114 0.060137 88 0.001425

as.matrix(X)V2 1.11999 0.003344 66.79 0.0001425 \*\*

as.matrix(X)V4 3.25582 0.000215 271.46 0.0001425 \*\*

as.matrix(X)V5 -0.44712 0.000600 1.02 0.0001425 \*\*

as.matrix(X)V6 0.33371 0.287314 2e10\*\*2 0.0001425 \*\*

Residual standard error: 0.231 on 456 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 0.9946

F-statistic: 1.21e+02 on 2 and 456 DF, pvalue 254

The estimated effect of V2,V4, V5, V6 on the convert is

1.11999 3.25582 , -0.44712, 0.3371.

This means that for every 1% increase in V2 on X, there is a

correlated 1.11% decrease in the incidence of Y. Similar to

V4, V5, V6

The standard errors for these regression coefficients are very

small, and the t-statistics are very large (66.79,271.46 , 1.02

and 200 , respectively). The p-values reflect these small

errors and large t-statistics. And for both parameters, there is

almost zero probability that this effect is due to chance.(\*\*

gives the variance in 0.001)

%update B

B = B +((Z-X)./W).\*S;

B(isinf(B)) = 0;

%update beta

beta = B + L\*T;

%observations for logistics

m = beta.\*X;

for i = 1:size(X,1)

temp =

makedist('Logistic','mu',mean(m(:,i)),'sigma',abs(std(

m(:,i))));

Z(:,i) = pdf(temp, Z(:,i));

R(:,i) = Z(:,i)- m(:,i);

end

%sampling lambda

Y = normpdf(Z,0,1);

Y = Y.^2;

Y = 1+(Y-sqrt(Y.\*(4\*R+Y)))./(2\*R);

lambda = Z;

for i = 1:6

Ztemp = R(:,i).\*Y(:,i);

Ztemp2 = R(:,i)./Y(:,i);

lambda(:,i) = Ztemp;

lambda(Z(:,i)>(ones(size(Y(:,i)))./(1+Y(:,i)))) =

if mean(lambda(:,i)) > 4/3

Z(:,i) = rightmost(Z(:,i),lambda(:,i));

else

Z(:,i) = leftmost(Z(:,i),lambda(:,i));

end

end

end

function Z = rightmost(U,lambda)

X = exp(-0.5\*lambda);

Z = X;

% squeezing

for t = 1:length(X)-1

Z(t) = Z(t) - (t+1)^2\*X(t)^(t+1)^2-1;

t = t+1;

Z(t) = Z(t) + (t+1)^2\*X(t)^(t+1)^2-1;

end

Z(Z<U)= 0;

End

function Z = leftmost(U,Lambda)

H = 0.5\*log(2)+2.5\*log(pi)-2.5\*log(Lambda)-

repmat(pi^2,6,1)./(2\*Lambda)+0.5\*Lambda;

lU = log(U);

X = exp(-pi^2/(2\*Lambda));

Z = X;

K = Lambda/pi^2;

% squeezing

for t = 1:length(X)-1

Z(t) = Z(t) - K(t)^(t^2-1);

t = t+1;

Z(t) = Z(t) + K(t)^(t^2-1);

end

Z((reshape(H,6,1)+reshape(log(Z),6,1))<resha

pe(lU,6,1))= 0;

End

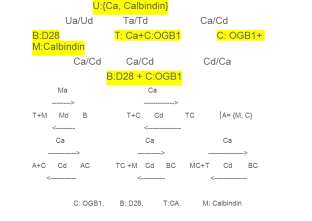
D.Previous Work

1.Markoc chain of CaMKII circuit with regards to cognitive systems diseases especially about the MC and GC networks around hippocampus and dendrate gyrus. Please see one of my current

work report following:

<https://www.overleaf.com/read/znrwdjxppyzs>

Mainly, I simplify one calciumodulin Activation by Calcium Transients of postsunaptic dendritic spines which finally constructed by 12 equations, first four as the clobe, nlobe binding to medium concentrated Ca and the second module as the similar clobe and nlobe but binding to high concentated Ca binding sites followed by the fast binding of kinase II and the indicator OGB1. As it is the simplified version, its only based on the four independent binding site.



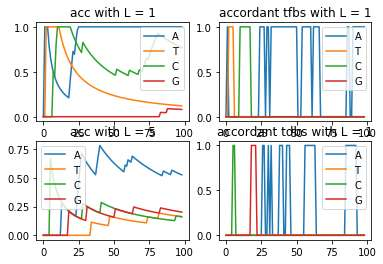
2. Cell reprogramming is usually a time lapse studied through gene expression data and DNA sequence data. Through computing the

reprogramming rate, it is shown that more reprogramming happen under the condition of inhibition of DNA methylation or the

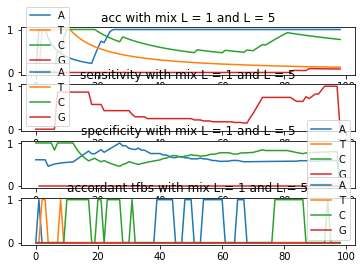
knowckdown of somatic transcription factors. In the first model, one fixed-variable-order Bayesian tree is constructed for the identification

of transcription factor binding sites(TFBSs), while in the second model, the focus is on the expression data of the cell and transcription factors. This is mainly based on the process that,

the promoters of ESCs can not only bounded by their own products but also activate other pluripotent genes and inhibit lineage specific

genes. Thus, a markov model is applied to induce the reconstruction of transcription regulation in embryonic stem cell states, by the ectopic expression of factors and reprogram the differentiated cells.

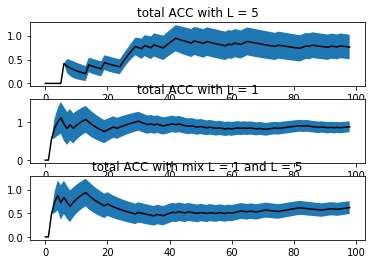
To realize the detection of binding site of DNA transcrition factor, wIth homogeneous order 1 and 5 VOM(0.65,10), I first get the binding sites with the highest accuracy prediction peaks.X = {A,T,C,G}, and thus the d = 4. And using the mix model with pruning on KL scaled by threshold c on log scaled odds , the bayesian tree is for specifically to predict those at binding site(pruned level less than 3).



For the foreground dataset, L = 1, and each DNA sequence from X is predicted directly through the one before it while using the most easiest critieria, chossing the one with largest probability. For the background dataset, L = 5, and each DNA sequence from X is predicted through the previous five sequences before it while using the frequency of the respective subsequencies combining Bayes Theory. And again, the highest peak frequency is chosen to be the biding site with its probability approximated:





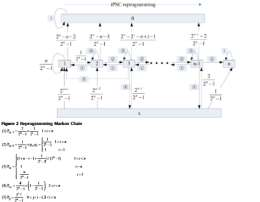
As the results, generally, the KL converges faster with L =1 and the mix model of order 1 and 5. In addition, the model is accumulated instead of step-wise, leading to the convergence not rising with the iteration. It is obvious that, on some specific TFBSs, the prediction is ven higher than others.

Algorithm: Stepwise Markov Ising model with lineage tree

1. The first step is to configure the epigenetic states of equencies through their expression, whehter "close" of "open" on any temporary cell state. (In our data, there are 64 cell states in all.)

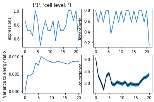
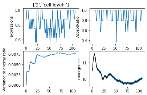
2. The second step is to check the final state of each lineage tree, as the expression of modules maybe conflicted with each other, influenced by other cell's transcriptional regulatory network which might lead to cell death(we denoted as 0) while $\epsilon$ denoted as the last state to the last states, which making 0 and n(66th) as the absorbing states.

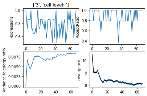
3. The transition matrix is built according to reference [2]:



Note that in the first step, The transition states predicted through the highest probability. And in the second step, either probability(0 or $\epsilon$ and further n) over 0.5 will lead to the end of the markov process.Basically, it is based on the Ising model, choosing direction

among {P1,P2,P3,P5} instead of the original random spin mchoise from {-1,1} and the final fate of the cell is predicted through either {0} or {n after $\epsilon$, i.e. P4}.





Among the 200 simulations, first level cell expression takes fewest amounts while the highest convergence.(Although the variance to energy ratio are quite similar for the three levels.). Generally, the Accept ratio is higher after the expression level goes to the lowest which is either when inhibition of DNA methylation exhists or the knowckdown of somatic transcription factors occurs.

[1]Novel Markov model of induced pluripotency predicts gene xpression changes in reprogramming Zhirui Hu,Minping Qian, Michael Q Zhang\*

[2]A. Gohr, J. Grau, S. Arviv1, A. Shmilovici, S. Posch, and I.