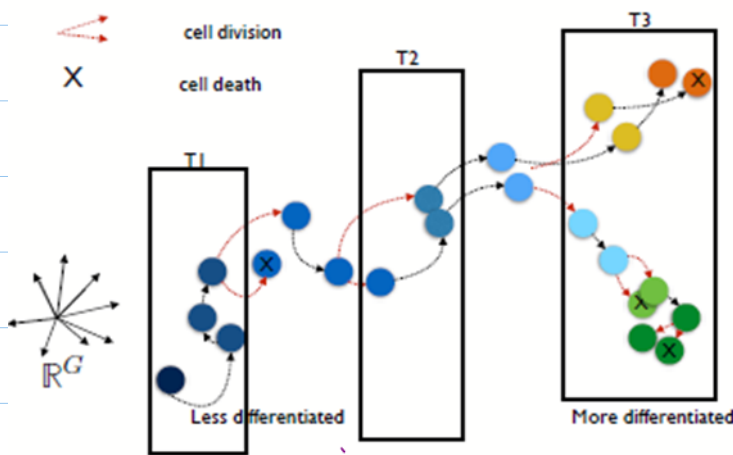


سیستم تراد است روی داده های ژنی است

درست است که DNA بیان است اما در حد سلول ژن ها متفاوت است

ما می خواهیم بین تمام ژن ها در حد سلول بیان شده اند و ...



این پروسه به صورت stochastic انجام می شود  
 =D  
 diffusion  
 drift نه به سمت خاصه می رود  
 و تویه یا ژن سلول ها

درست نه برای در diffusion و drift استفاده می شود dynamical inference

=> each method have these 3 component

- Drift (needs energy)
- Diffusion (stochastic)
- Birth-Death

اینه این method ها برای هم هستند = validation به ما می ده

Methods

- Dynamical Inference
- Data Integration and latent spaces
- Regulatory Networks: perturbation —> control of cell fates

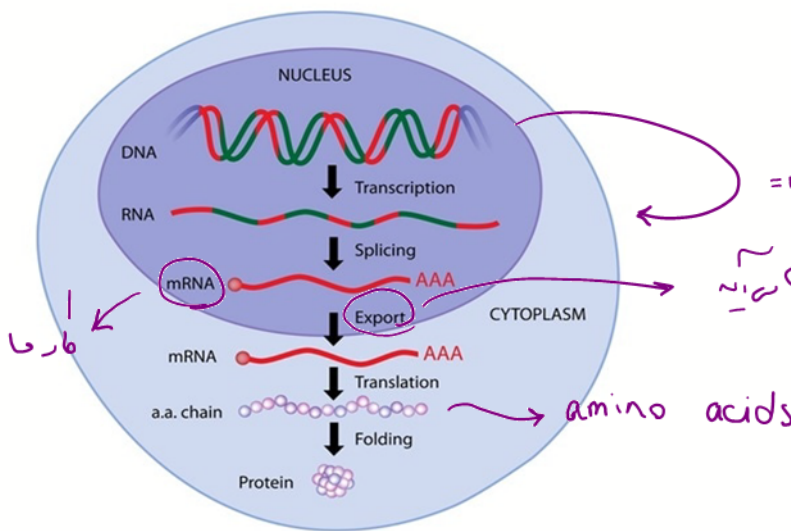
حلونه به فزین ارتباط دارد

فزین علم اندازه گیری است و از روش های آن  
میتوان استفاده کرد.  
فزین مشاهده را به تئوری وصل میکنند اما باید دانست  
علم را هم دانسته باشیم

## Topics:

- Intro, single-cell NGS measurements and data quantification
- Preprocessing
- Dimension reduction
- Dynamical inference (Langevine, Fokker-planck, OT, etc.)
- Gene regulatory networks
- Learning theory (optimisation, identifiability, etc.)
- Maximum Likelihood and Expectation Maximisation algorithm
- Variational methods (PPCA, VAE, etc.) and generative models

پروسه بیولوژی:



این بخش به مرخواه فعال  
شود

از سبیل بیرون می آید

پارچه

معمولاً مرئخ نیست DNA های میان است

انسان در موش 20k ژن دارند

سیتر اکسین و وی Transcriptomics است چون اندازه گیری ساده تری دارد

قابل تغییر است

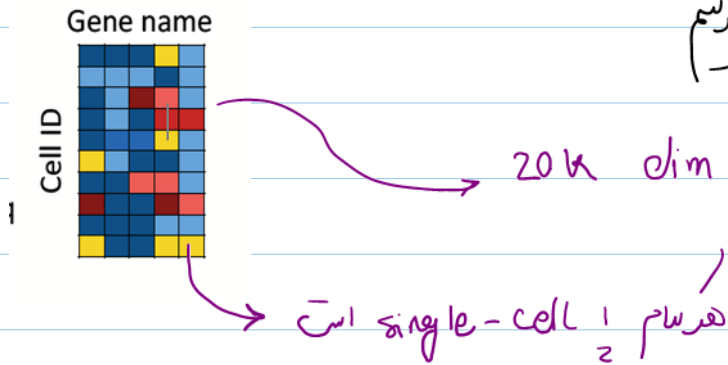
ص 15 کو cluster کا توضیح دے۔ کس B خون کے groupic ہے۔

ص ۱۶ = sequence زنجیره‌ها به ترتیب داخل file - q fast مورد

در گس‌ر Read Alignment به زنجیره Reference نیاز داریم چقدر جاها وجود ندارد

بعد از این گس‌ر و در bam file می‌سوز

ص ۱۷ بعد از این کاری به فایلیش می‌کنیم



در این فضای ۲۰ هزار بعد ۲۰ هزار هر زنجیره داشتن یا نداشتن است  $2^{20k}$  = sample سیر

نسبت است =  $\frac{\text{Signal}}{\text{Noise}}$  سیر ratio توصیف است

= نیاز خواهیم داشت ۲۰ هزار را به خاطر فصل dim بالا نمی‌داریم = dim reduction

روش unsupervised هست

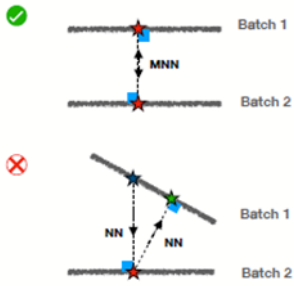
در مسکن کار کردن در بعد بالا :

PCA (محاسبه به خطی) اگر  $N \gg D$  و  $D$  به  $10^5$  نرسد فعلی است مورد

واقعی نیست

این کارها Artifact می‌سازد

گاهی می‌توان از خواص بعد با استفاده کرد ← ص 21



$\Rightarrow MNN = \Rightarrow CLT$  بعد با این رابطه می‌توانیم

Network پیدا می‌کند و برای این Network

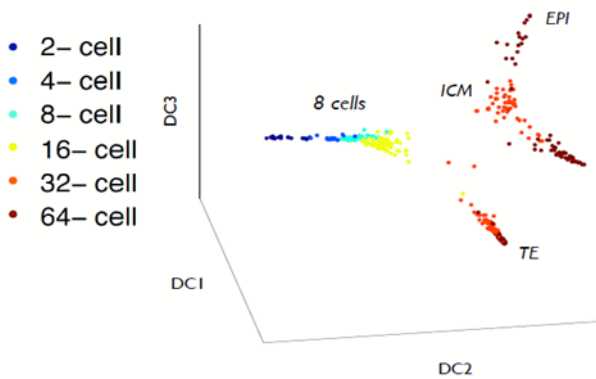
برای ریشه از curse این dim با رها کنیم، هر سطر می‌بینیم کم شود

Feature selection. → What we are studying?

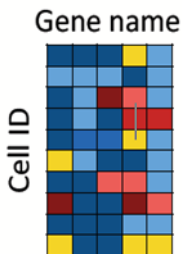
→ Avoid high-dim effects.

Example of good selection of features:

ص 23



$\Rightarrow 48$  gene → back ground into



→ From this → dim reduction →



این‌ها express شده اند در NB (Negative Binomial) می‌باشد

$\Rightarrow$  باید احتیاط از ZINB (Zero-Inflated Negative Binomial) می‌باشد با احتیاط از NB

و در این مورد truth گوییم نه ایم transformation می‌باشد.

Human cell Atlas → پس استاندارد بودن  
مازهم تفاوت می‌باشد