

# DREAM Single Cell Transcriptomics Challenge

MDC  
MAX DELBRÖCK CENTER  
FOR MOLECULAR MEDICINE  
IN THE HIGHLIGHTS ASSOCIATION

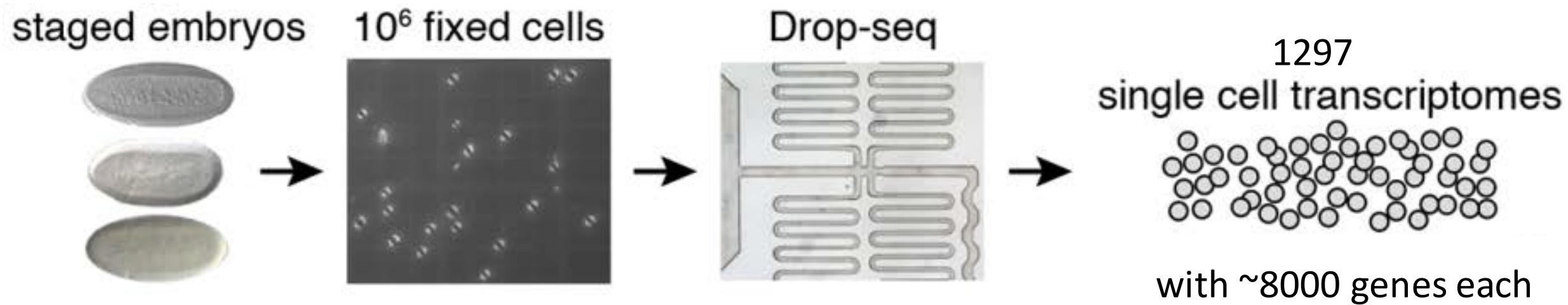
IBM

DREAM  
CHALLENGES  
*powered by Sage Bionetworks*

MEDIZINISCHE  
FAKULTÄT  
HEIDELBERG

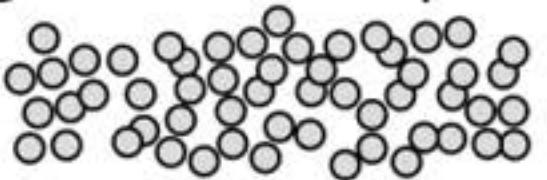
SageBionetworks

# Challenge dataset



# Challenge questions

single cell transcriptomes



Map  
→



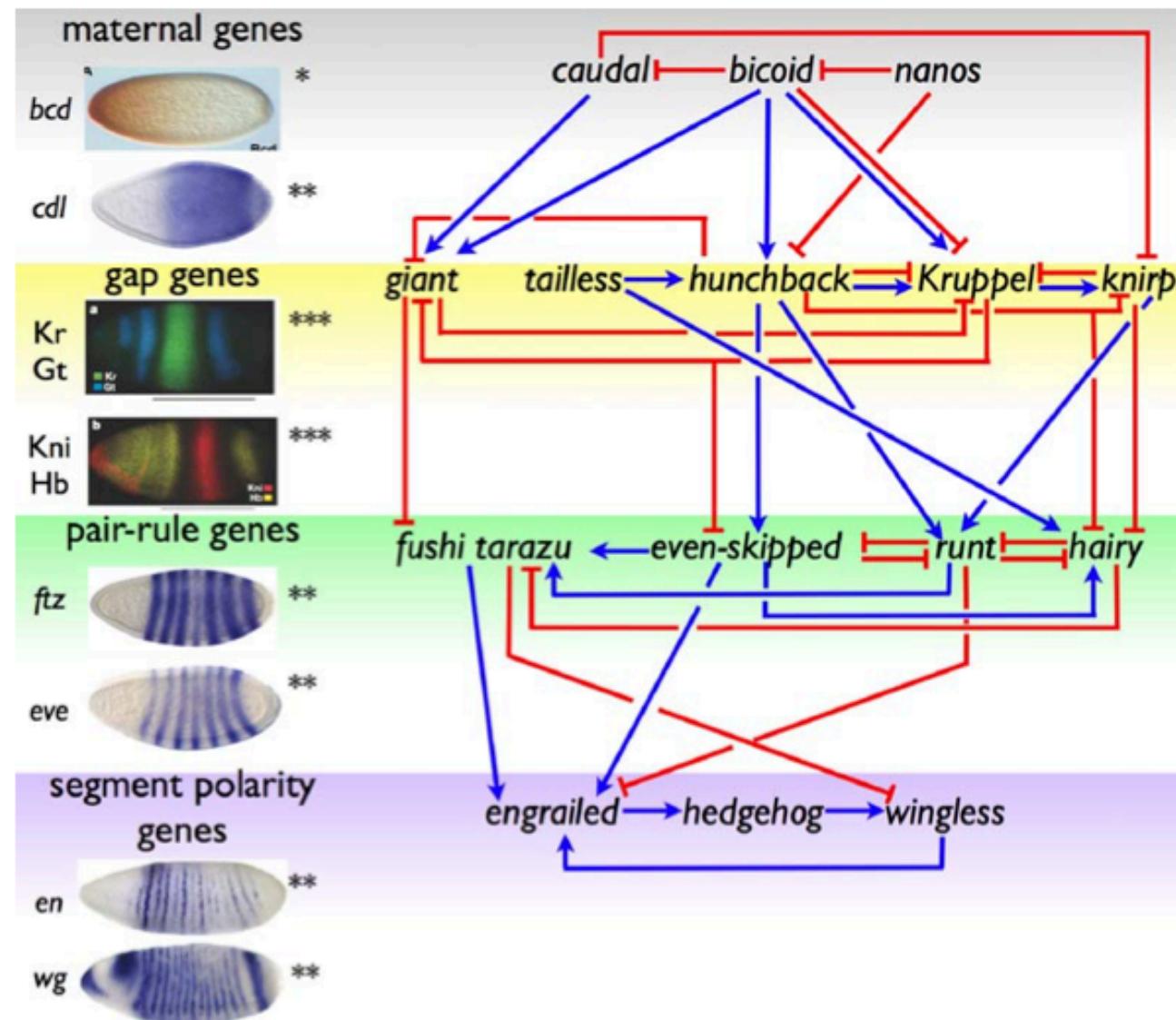
84 expression patterns

From BDNTP



3039 bins

# The Drosophyla embryo signaling network



## **Gene determinants of different tissues:**

**Dorsal ectoderm:** *Ance CG2162 Doc1 Doc2 egr peb tok ush zen*

**Neurectoderm:** *ac brk CG8312 l(1)sc mfas Ptp4E sog SoxN vnd*

**Mesoderm:** *CG9005 Cyp310a1 GEFmeso Itl Mdr49 Mes2 NetA ry sna stumps twi wgn zfh1*

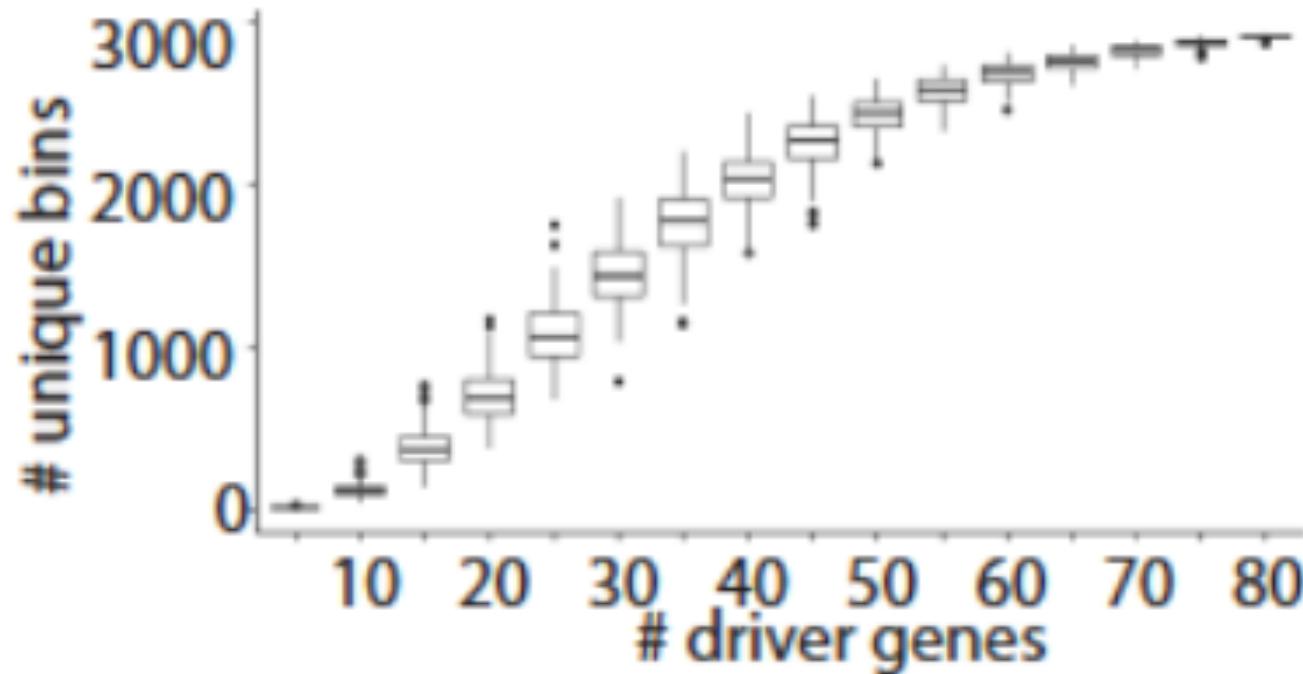
**Yolk cells:** *beat-IIIc CG8129 CG8195 Corp CNT1 sisA ZnT77C*

**Pole cells:** *Pgc*

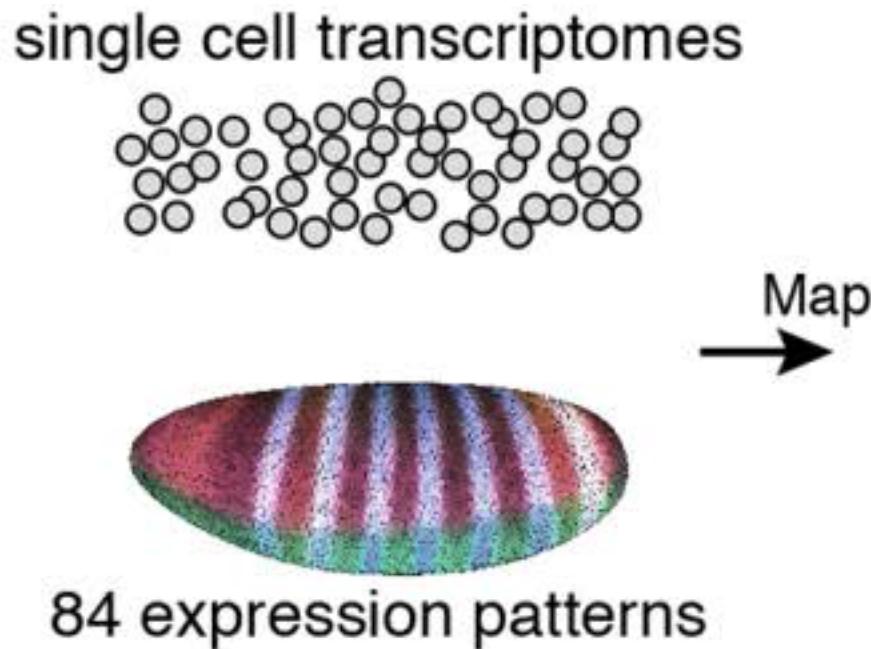
**Include biological information contained in the genes that determine embryonic patterning**

# Challenge formulation

There is no blind gold standard as the data has been published



# Challenge questions



using  
**Subch1:** 60 expression patterns  
**Subch2:** 40 expression patterns  
**Subch3:** 20 expression patterns

Participants need to predict the position of 1297 cells in the 3039 *Drosophila melanogaster* embryonic locations or bins for one half of a stage 6 -pre-gastrulation- embryo based on scRNAseq data.

# Challenge questions

Predict the position in the embryo of the 1297 cells using all the expression patterns from:

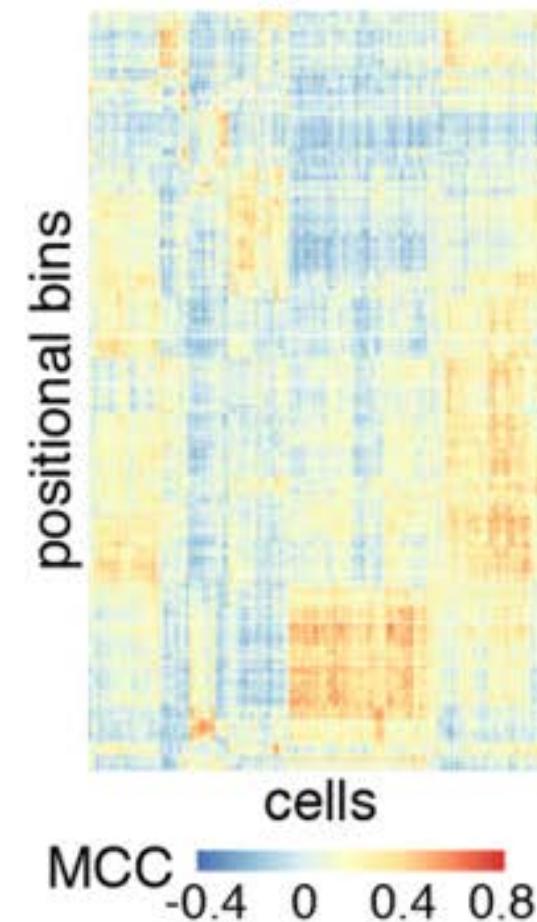
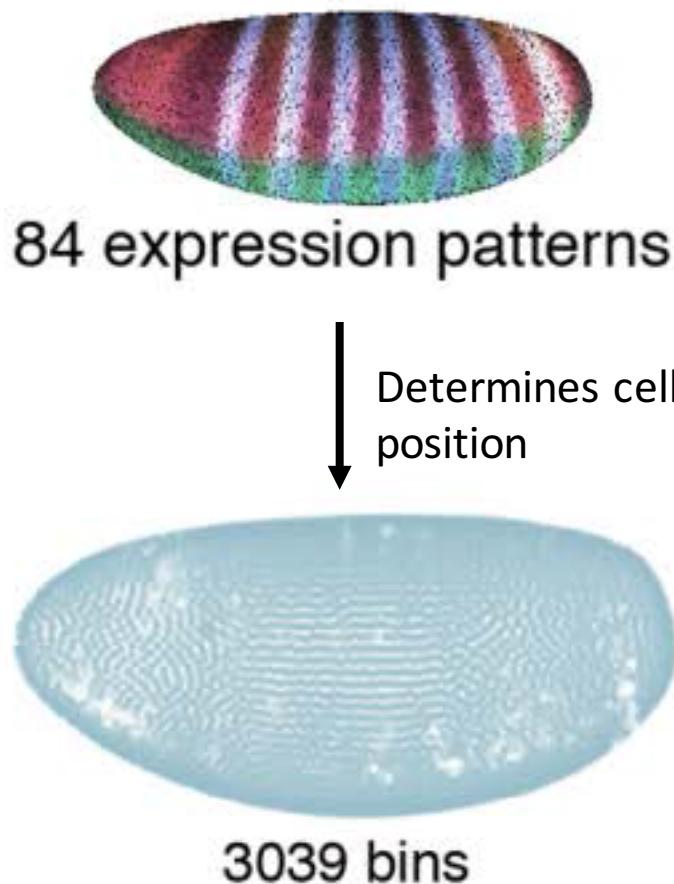
**1- the *in situ* hybridizations from 60 driver genes out of the 84 in subchallenge 1,**

**2- a subset of any expression patterns from 40 genes out of the 84 in subchallenge 2,**

**3- a subset of any expression patterns from 20 genes out of the 84 in subchallenge 3.**

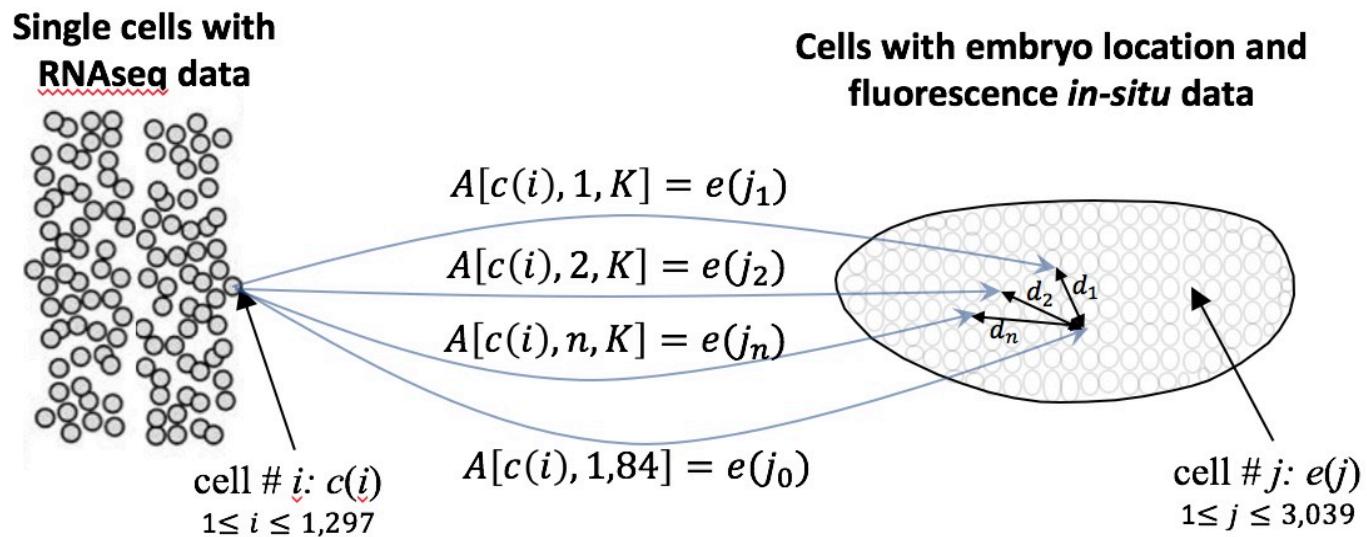
# Challenge Scoring

-Adequate prediction of 1297 cell positions a MCC is calculated at each of the 3039 positions between the binarized *in situ* localization values in for the 84 driver genes and the binarized 84 RNAseq values for the same 84 genes.



# Challenge Scoring

-Adequate prediction of 1297 cell positions a MCC is calculated at each of the 3039 positions between the binarized *in situ* localization values in for the 84 driver genes and the binarized 84 RNAseq values for the same 84 genes.



$A[c(i), n, K] = e(j_n)$  : Assigned  $n^{\text{th}}$  most likely location for cell  $c(i)$  in the embryo using information of only  $K$  genes from the *in-situ* data.

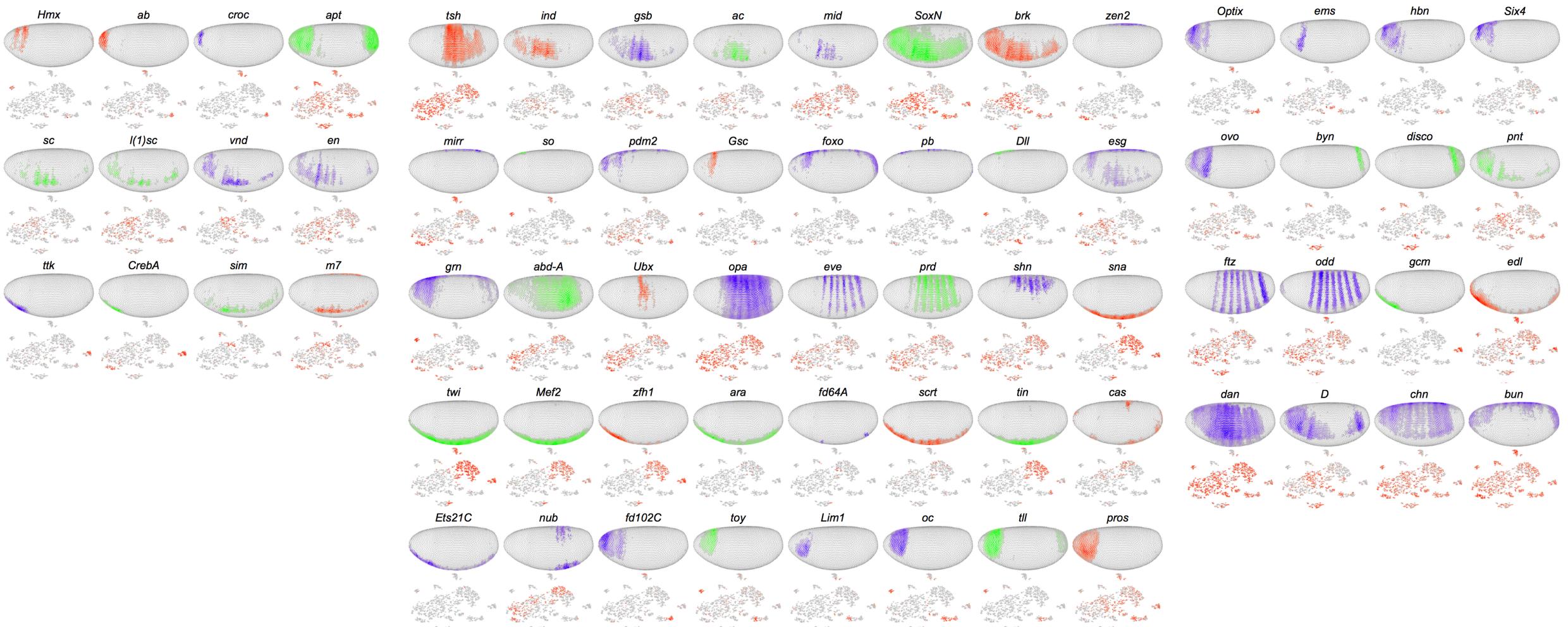
$j_1 \neq j_2 \neq \dots \neq j_n$  : The assigned  $n$  most likely locations in the embryo for cell  $c(i)$  should be different.

$\overline{d_k}(i) = \frac{1}{n} \sum_{l=1}^n d_l$  : Average distance of the  $n$  most likely locations in the embryo for cell  $c(i)$  using  $K$  *in-situ* genes and the “real” location

# Challenge Scoring

Scoring will be based on 2 characteristics:

-**Adequate prediction of gene patterns** calculated using the MCC for the 84 genes across 1297 cells.



# Challenge Scoring

Scoring will be based on 2 characteristics:

**-Adequate prediction of 1297 cell positions** a MCC is calculated at each of the 3039 positions between the binarized *in situ* localization values in for the 84 driver genes and the binarized 84 RNAseq values for the same 84 genes.

**-Adequate prediction of gene patterns** calculated using the MCC for the 84 genes across 1297 cells.

The scoring will have additionally weights that will take into account the number and the genes used to predict the spatial localization of the 1297 cells and the euclidean distance between the predicted position of the cell and the actual position as determined by the maximal MCC with the in situ from 84 driver genes.

# Challenge Leaderboard

We will allow one submission per sub-challenge to a leaderboard mid-way of the challenge.

**Date TBD**

# Challenge dates

**Challenge launch date** : September 1st 2018

**Webinar date:** September 14th 2018

**Final submission deadline** : November 21st 2018

**Final writeup deadline** : December 1st 2018

**Best performers** : December 8th 2018 at RSG/DREAM [conference](#)

# Challenge organizers

Name	Institution
Pablo Meyer	IBM RESEARCH
Gustavo Stolovitzky	IBM RESEARCH
Nikos Karaïskos	Max Delbrück Institute
Nikolaus Rajewsky	Max Delbrück Institute
Attila Gabor	Heidelberg University
Julio Saez-Rodriguez	Heidelberg University
Thomas Yu	Sage Bionetworks

