

# Davis McLearty: SCATER

Preprocessing, qc and  
normalization for  
SC RNA SEQ

150819

~~$C_i = O_i \cdot t_i$~~   
Counts of  $T_i$  is the  
#  $T_i$  times covered.

$$T_i = \frac{C_i}{\sum_{i=1}^n C_i}$$

we relevant  
for us!

~~$q_i = \frac{a_i \cdot t_i}{\sum_{i=1}^n a_i \cdot t_i}$~~

A fragment in my case is not a read  
or read pair, but a tuple which also  
contains the UM!

F is the collection of these (observed) tuples.

Z is the tuple-transcript assignment matrix  
(From RapMap)

T is the set of (known) transcripts and  
the counts of those transcripts ( $C_i$ )

Want to infer.



The equivalent of  $\eta$  for us  
(“nucleotide fraction”) should  
be  $\tau(C)$ , since  $l(t_i)$  “effectively”  
is just 1 (or read length, depending  
on how you view it).

$$P\{F | Z, \tau\} = \prod_{j=1}^N P\{f_j | Z, \tau\} =$$

$$= \prod_{j=1}^N \sum_{i=1}^M P\{f_j | t_i, z_{ij} = 1\}$$

$$P\{f_j | t_i\} \triangleq P\{f_j | t_i, z_{ij} = 1\}$$

Agreement model

$$P\{f_j | t_i\} = P\{a | f_j, t_i\}$$

a parameter of auxiliary model.  
Probability of having RapMap agreement with  $f_j$   
that it is taken from  $t_i$ .

For pseudotyping,  $P\{a | f_j, t_i\}$  is

the Coverage score. No output in  
RapMap: C  
RapMap outputs “all loci”.  
Let's simply  
assume that  
is  $\frac{1}{m}$ .  
Can we so  
on?



$$\text{Say } P\{f_j | t_i\} = \left(\frac{1}{N_H}\right) [f_j]$$

$\alpha$  estimated read counts.

For us  $\eta = \tau$ .

$$\hat{\mu}_i = \frac{\alpha_i}{\sum_j \alpha_j}$$

$\alpha \equiv$  integer  $\mathbb{N}(\mathbb{C})$

$$L\{\alpha | F, Z, \tau\} = \prod_{j=1}^N \sum_{i=1}^M \tilde{n}_i P\{f_j | t_i\}$$

~~$$\alpha_i = \#$$~~

$$P\{\alpha_i | f_j, t_i\} ?$$

$\sum_i \alpha_i$  is the total number of molecules.

~~$$\sum_i \alpha_i = \left| \{f_j\}_{f_j \in F} \right|$$~~

*Handwritten notes in red ink:*  
 This is the number of molecules  
 that are in the set F  
 and are in the set F  
 and are in the set F

$$\begin{bmatrix} a & A \\ a & A \\ a & B \\ b & A \\ b & B \\ b & B \\ b & A \end{bmatrix} \Rightarrow \begin{bmatrix} a & 2 \\ b & 3 \end{bmatrix}$$

$$P\{(n_j, m_j) | t_i\} = \left(\frac{1}{N_H}\right) [(n_j, m_j)]$$

$$\alpha_i = |\{n_j\}| | t_i$$

$$P\{\alpha_i | F, \tau, t_i\} = \frac{|\{n_j\}|}{\sum_{j \in F} |\{n_j\}|} \tau_i$$

$$\frac{|\{n_j\}|}{\sum_{j \in F} |\{n_j\}|} \tau_i$$

$\alpha_t$ : probability of selecting fragments from transcript  $t$ .

For a given molecule, we can only catch 1! The 3'/5' end one.

$$\text{Thus } \alpha_t = \frac{m_t}{\sum_{i \in T} m_i}$$

The number of molecules of  $t$  over the total # of molecules.

$$m_t = |\{u \mid u \in t\}|$$

↑  
Ambiguous!

$n$		$u_c$	$u_s$
3	<pre> _____ 0 _____ 0 _____ 0 </pre>	2	1.33
5	<pre> - - - 0 - - - 0 - - - 0 - - - 0 </pre>	2	1.66

159820

$A$  invertible  $\Rightarrow \text{Adj}(A) = |A| \cdot A^{-1}$

$$\frac{\partial |A|}{\partial a_{ij}} = \frac{\partial \sum_k A_{ik} (|A| \cdot A^{-1})_{ik}^T}{\partial a_{ij}}$$

$$(t, u) \cdot t = \frac{1}{m}$$

$$\frac{\partial |A|}{\partial a_{ij}} = \sum_k \frac{\partial A_{ik} \cdot (|A| \cdot A^{-1})_{ik}^T}{\partial a_{ij}}$$

$$\frac{d}{dt} \det(A(t)) = \sum_j (a_{ij}(t)) \frac{d}{dt} A(t)$$

Q