I-CONVEX: De Novo Transcriptome Sequencing from Long Reads

Sina Baharlouei

Daniel J. Epstein Department of Industrial Engineering

Meisam Razaviyayn

ISE, USC

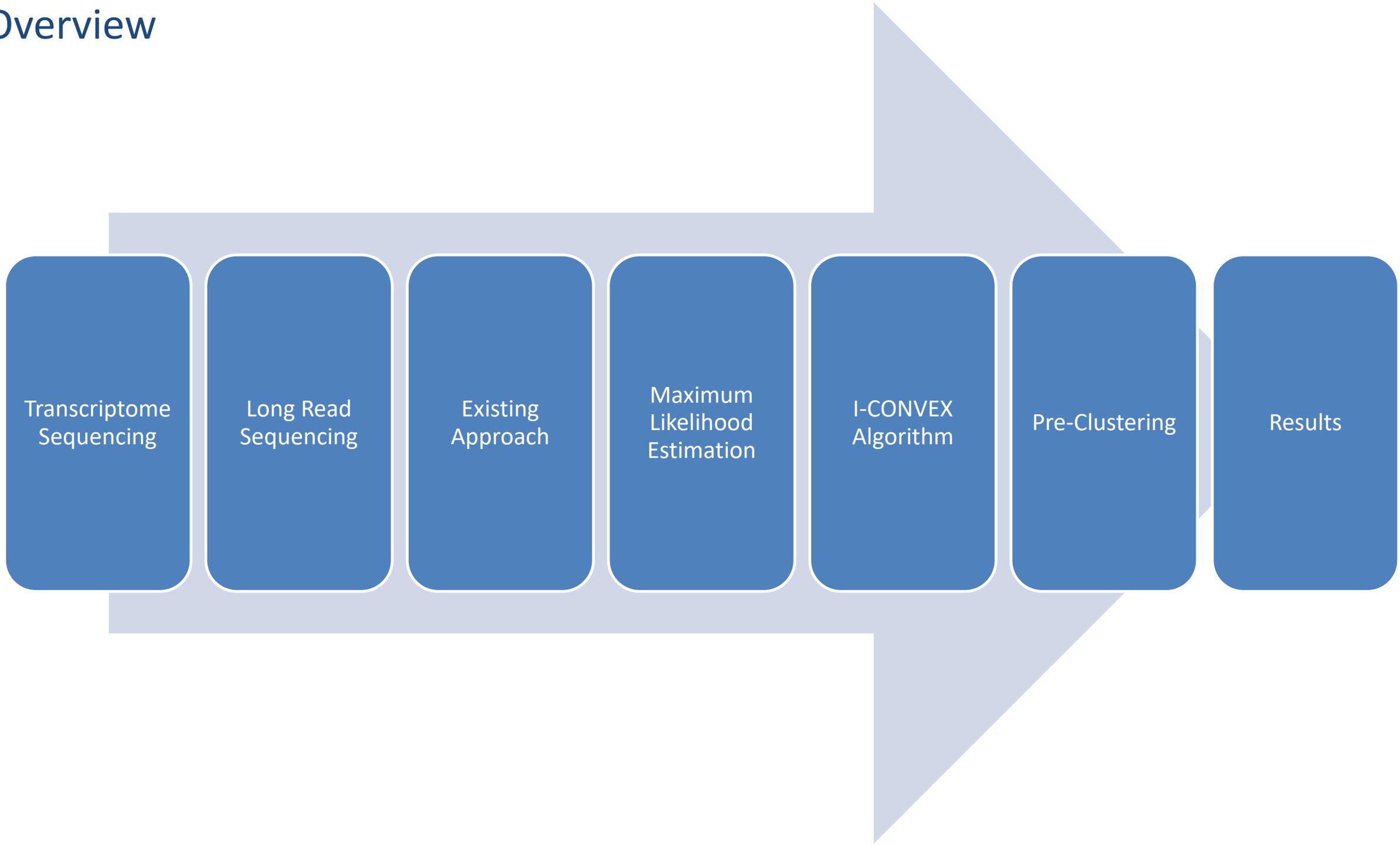
Elizabeth Tseng

PacBio

David Tse

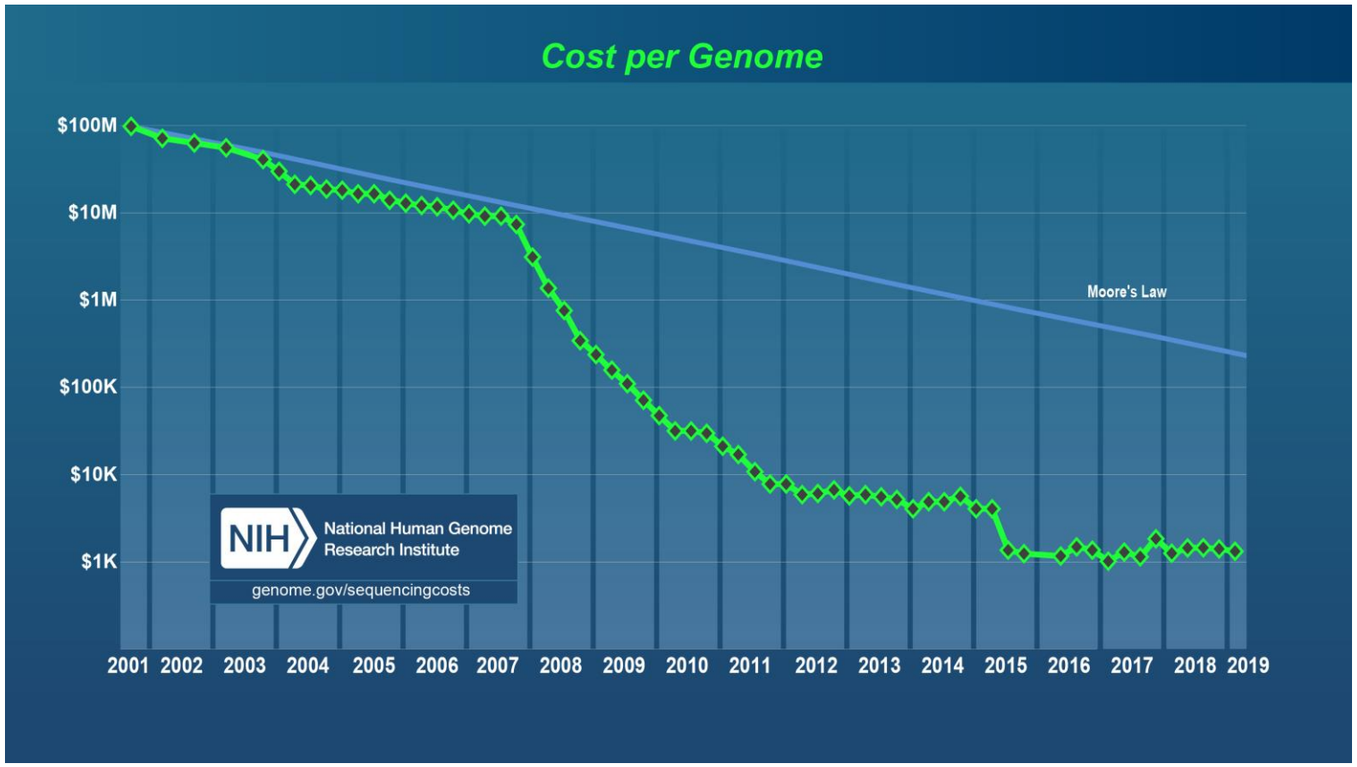
EE, Stanford

Overview


Transcriptome Sequencing
Long Read Sequencing
Existing Approach
Maximum Likelihood Estimation
I-CONVEX Algorithm
Pre-Clustering
Results

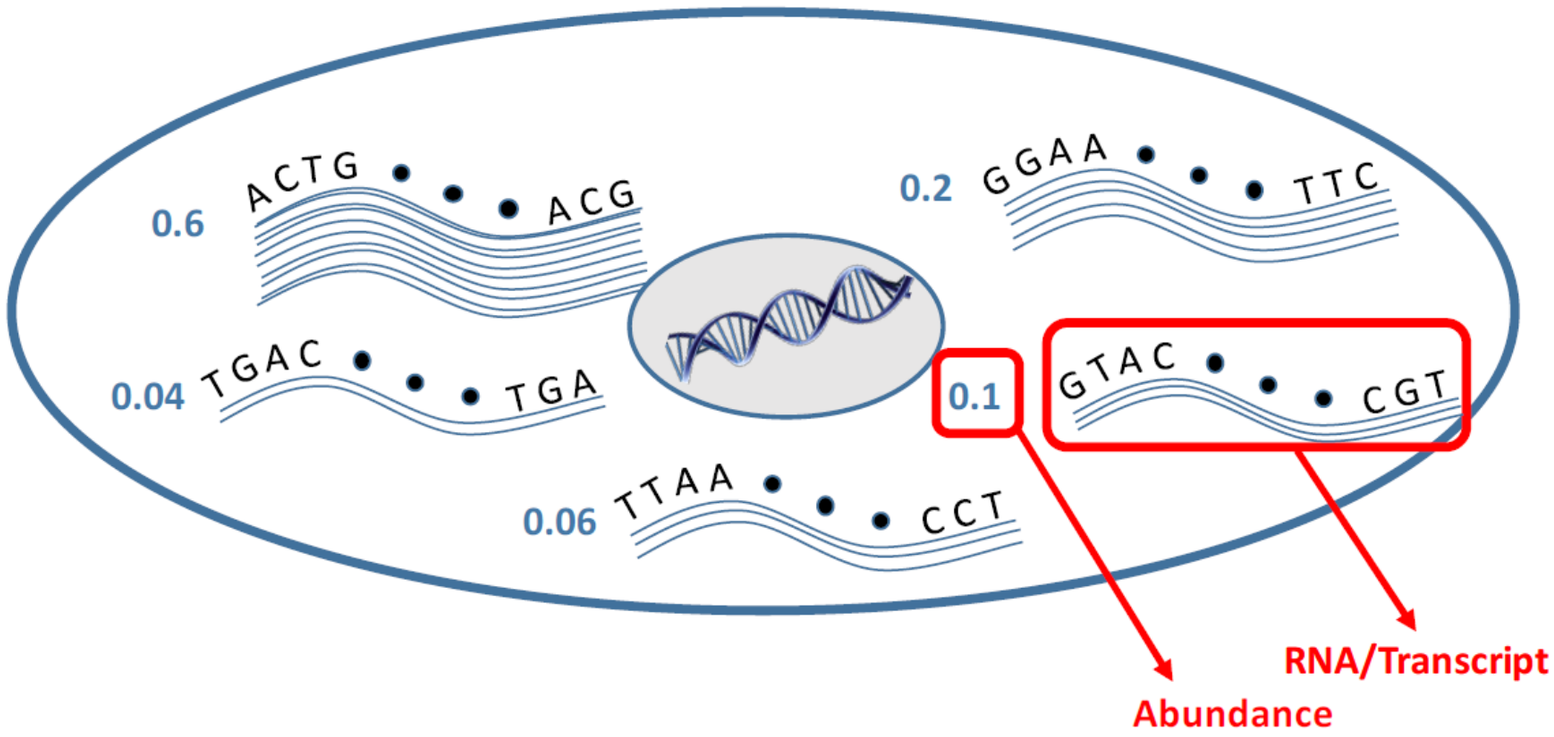
What is Genome Sequencing?

- The process of determining the complete DNA/RNA sequence of an organism
- Complicated and expensive
- Fast progress



Transcriptome Sequencing and its Importance

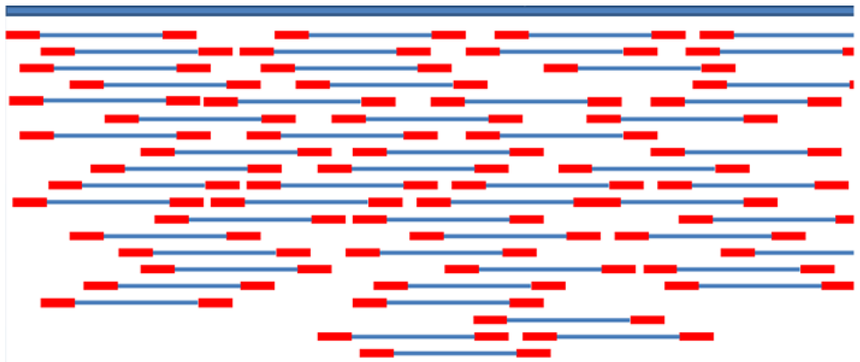
- Can fundamentally change our understanding of organisms and diseases
  - Snapshot of cell state (health, type, ...)
  - Personalized medicine
  - Understanding and detection of diseases such as cancer, Alzheimer, Diabetes
  - Virology, Immunology



How to Read a Single Sequence?

Short read Sequencing

(Illumina)



✓ Low error in fragments

✗ No unique assembled sequence

Long read Sequencing

(PacBio)

✓ Statistically identifiable

✗ High percentage of errors

Transcriptome Recovery Problem

Transcripts

ACCGAAGTAATCCTTATGAATCAGAT

GAACGTACCTGCTGTAGCTTCA

GTCCTACAGCTACTCCTGCACTACTC

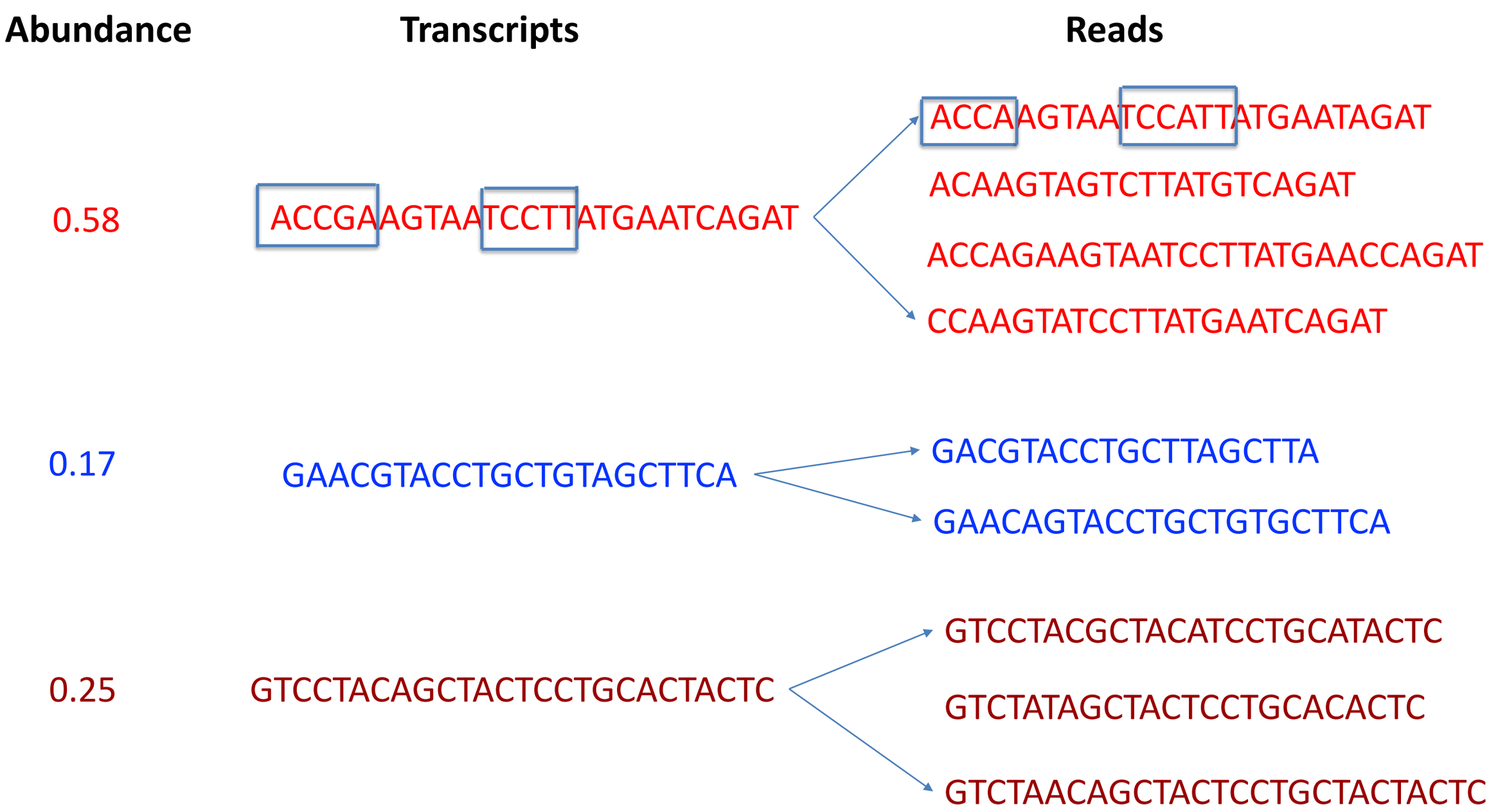
Transcriptome Recovery Problem

Abundance	Transcripts
0.58	ACCGAAGTAATCCTTATGAATCAGAT
0.17	GAACGTACCTGCTGTAGCTTCA
0.25	GTCCTACAGCTACTCCTGCACTACTC

Transcriptome Recovery Problem

Abundance	Transcripts	Reads
0.58	ACCGAAGTAATCCTTATGAATCAGAT	ACCAAGTAATCCATTATGAATAGAT ACAAGTAGTCTTATGTCAGAT ACCAGAAGTAATCCTTATGAACCAGAT CCAAGTATCCTTATGAATCAGAT
0.17	GAACGTACCTGCTGTAGCTTCA	GACGTACCTGCTTAGCTTA GAACAGTACCTGCTGTGCTTCA
0.25	GTCCTACAGCTACTCCTGCACTACTC	GTCCTACGCTACATCCTGCATACTC GTCTATAGCTACTCCTGCACACTC GTCTAACAGCTACTCCTGCTACTACTC

Transcriptome Recovery Problem



Transcriptome Recovery Problem

Abundance	Transcripts	Reads
0.58	ACCGAAGTAATCCTTATGAATCAGAT	ACCAAGTAATCCATTATGAATAGAT ACAAGTAGTCTTATGTCAGAT ACCAGAAGTAATCCTTATGAACCAGAT CCAAGTATCCTTATGAATCAGAT
0.17	GAACGTACCTGCTGTAGCTTCA	GACGTACCTGCTTAGCTTA GAACAGTACCTGCTGTGCTTCA
0.25	GTCCTACAGCTACTCCTGCACTACTC	GTCCTACGCTACATCCTGCATACTC GTCTATAGCTACTCCTGCACACTC GTCTAACAGCTACTCCTGCTACTACTC

Transcriptome Recovery Problem

Abundance	Transcripts	Reads
0.58	ACCGAAGTAATCCTTATGAATCAGAT	ACCAAGTAATCCATTATGAATAGAT ACAAGTAGTCTTATGTCAGAT ACCAGAAGTAATCCTTATGAACCAGAT CCAAGTATCCTTATGAATCAGAT
0.17	GAACGTACCTGCTGTAGCTTCA	GACGTACCTGCTTAGCTTA GAACAGTACCTGCTGTGCTTCA
0.25	GTCCTACAGCTACTCCTGCACTACTC	GTCCTACGCTACATCCTGCATACTC GTCTATAGCTACTCCTGCACACTC GTCTAACAGCTACTCCTGCTACTACTC

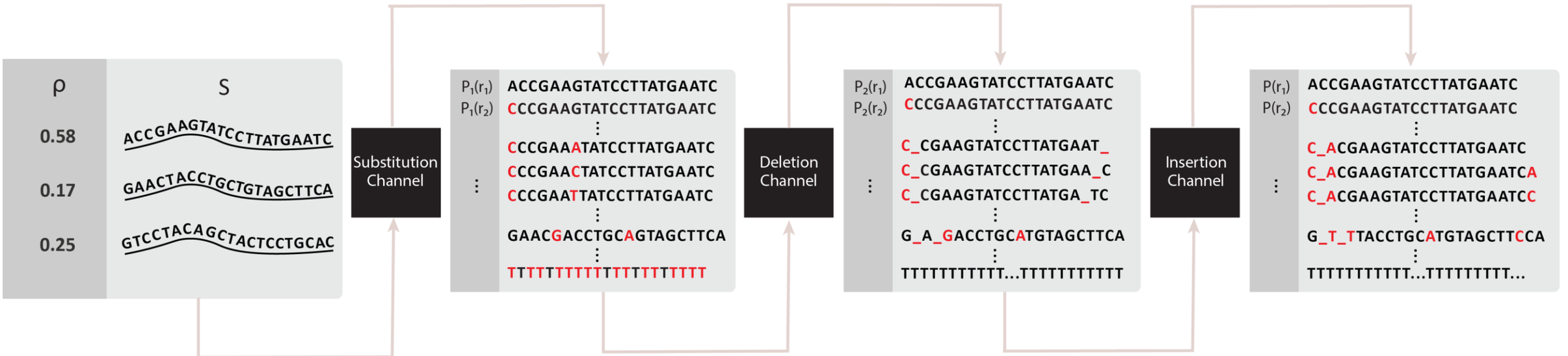
How?

Identifiability of Long Read Sequencing Problem

- A family of distributions $P = \{P_\theta | \, \theta \in \Theta\}$ is identifiable if it satisfies the following condition:

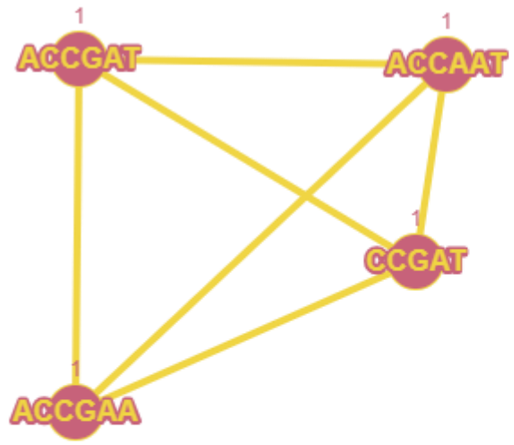
$$P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2 \quad \theta_1, \theta_2 \in \Theta$$

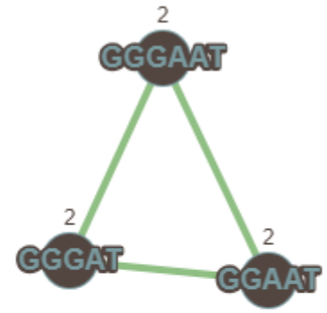
- What are the parameters and distributions in long read sequencing?



Idea of Clustering

- Need to compute pairwise distances to create a similarity graph
- Edit distance (Largest Common Subsequence)
  - Time Complexity: $O(N^2L^2)$
    - N: Number of reads
    - L: Length of reads





- For a dataset with 1M reads and average length 10000:
  - Need 10^{20} Operations!
  - No statistical guarantee for optimality of this method

Sahlin, Kristoffer, et al. "Deciphering highly similar multigene family transcripts from Iso-Seq data with IsoCon." *Nature communications* 9.1 (2018): 1-12.

Gordon, Sean P., et al. "Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing." *PloS one* 10.7 (2015): e0132628.

Reminder (Dynamic Programming for Computing Edit Distance)

$$L(m,n) = \min \begin{cases} L(m-1,n-1) + \delta(m,n) \\ L(m,n-1) + 1 \\ L(m-1,n) + 1 \end{cases}$$

$$L(i,0) = L(0,i) = 0$$

```
I N T E - N T I O N
| | | | | | | | | |
- E X E C U T I O N
```

-	0	E	X	E	C	U	T	I	O	N
0	0	1	2	3	4	5	6	7	8	9
I	1	1	2	3	4	5	6	6	7	8
N	2	2	2	3	4	5	6	7	7	7
T	3	3	3	3	4	5	5	6	7	8
E	4	3	4	3	4	5	6	6	7	8
N	5	4	4	4	4	5	6	7	7	7
T	6	5	5	5	5	5	5	6	7	8
I	7	6	6	6	6	6	6	5	6	7
O	8	7	7	7	7	7	7	6	5	6
N	9	8	8	8	8	8	8	7	6	5

A Maximum Likelihood Estimation

$$P(r;\rho,t) = \prod_{i=1}^{N} P(r_i;\rho,t) = \prod_{i=1}^{N} \left(\sum_{j=1}^{m} \alpha_{ij}\, \rho_j \right)$$

- **r** = {r_1, r_2, …, r_N} : Set of noisy reads
- **t** = {t_1, t_2, …, t_m} : Set of unknown transcripts
- **ρ** = {ρ_1, ρ_2, …, ρ_m}: Set of abundances of unknown transcripts
- α_{ij}: Probability of observing the read r_i from transcript t_j

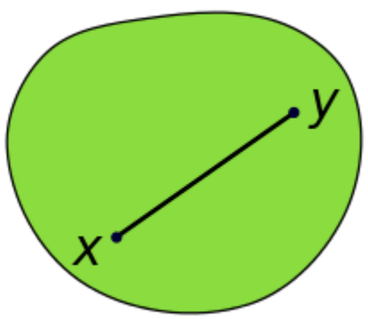
Is it a concave problem?

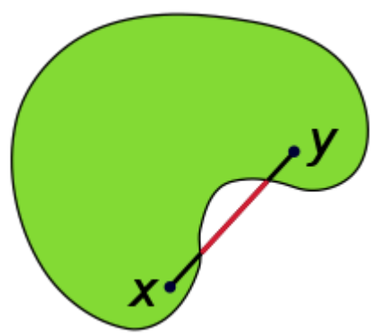
$$\hat{\rho}_{ML}, \hat{t}_{ML} = \underset{\rho,t}{\operatorname{argmax}} \sum_{i=1}^{N} \log\left(\sum_{j=1}^{m} \alpha_{ij}\rho_j \right) \quad s.t \quad \sum_{j=1}^{m} \rho_j = 1\,, \quad \rho_j \geq 0$$

A Brief Review of Convex Optimization

➢ Definition of convex set:

$$\mathcal{X} \text{ is convex} \Leftrightarrow \quad \forall x, y \in \mathcal{X}, 0 \leq \alpha \leq 1 : \alpha x + (1-\alpha) y \in \mathcal{X}$$



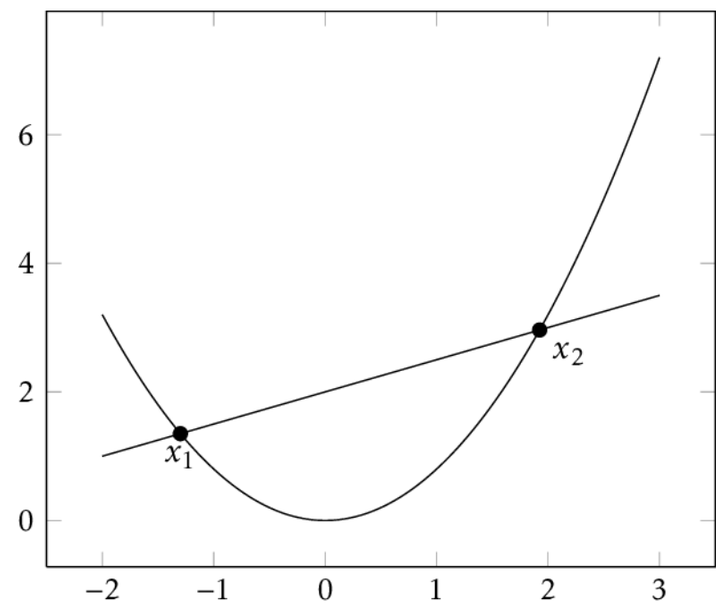


➢ It makes sure we stay inside the set by choosing a point between any two given points.

A Brief Review of Convex Optimization

- Definition of convex function:

$$f\Big(\alpha x + (1-\alpha)y\Big) \leq \alpha f(x) + (1-\alpha)f(y) \forall \quad x, y \in D_f, 0 \leq \alpha \leq 1$$



- First-order definition:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in D_f$$

- Second-order definition:

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in D_f$$

- How many of these functions are convex?

$-log(x)$	$x^T x$	$x^T Q x$	$a^T x$	$\|A - xx^T\|_F^2$	$\max(f_1, f_2)$
$f_1 + f_2$	$f_1 - f_2$	$\min(f_1, f_2)$	$\log\left[e^{x_1} + \cdots + e^{x_n}\right]$		

A Brief Review of Convex Optimization

- Convex problems:

$$\min_x f(x) \quad \text{s.t} \;\; x \in \mathcal{X}$$

- Function f must be a convex function.

- Set $\mathcal{X}$ must be a convex set.

- First-order optimality condition: $\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in D_f$

- Second-order optimality condition: $\nabla^2 f(x^*) \succeq 0$

- All local minima of a convex problem are **global** minima.

  - We can find local minima of convex problem and we are done!

A Brief Review of Convex Optimization

- Methods for solving convex optimization problem:
  - Gradient-free approaches (Genetic, Particle Swarm)
  - First-order methods (gradient descent, acceleration method of Nesterov)
  - Second-order methods (Newton method)
  - Majorization-Minimization
  - Block Coordinate Descent and ADMM
- Convergence rates:

	General Convex	Strongly Convex
GD	$\mathcal{O}(\frac{LD^2}{\epsilon})$	$\mathcal{O}(\kappa \log(\frac{1}{\epsilon}))$
Nesterov	$\mathcal{O}\left(\sqrt{\frac{LD^2}{\epsilon}}\right)$	$\mathcal{O}(\sqrt{\kappa} \log(\frac{1}{\epsilon}))$

Concavity of Maximum Likelihood

Is it a concave problem?

$$\hat{\rho}_{ML}, \hat{t}_{ML} = \underset{\rho,t}{\operatorname{argmax}} \sum_{i=1}^{N} \log\left(\sum_{j=1}^{m} \alpha_{ij}\rho_j\right) \quad s.t \quad \sum_{j=1}^{m} \rho_j = 1\,, \quad \rho_j \geq 0$$

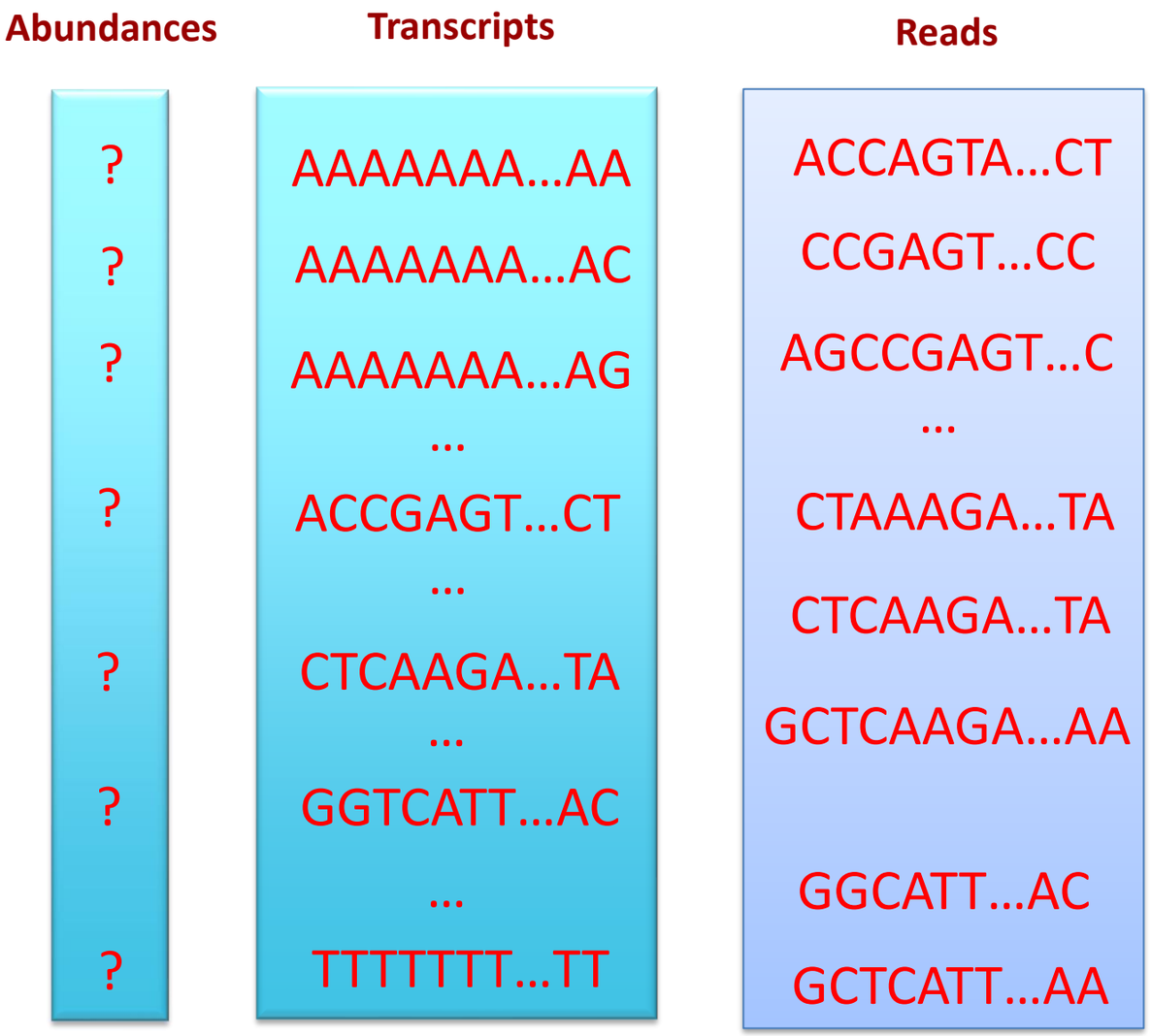
Concavity of Maximum Likelihood

Is it a concave problem?

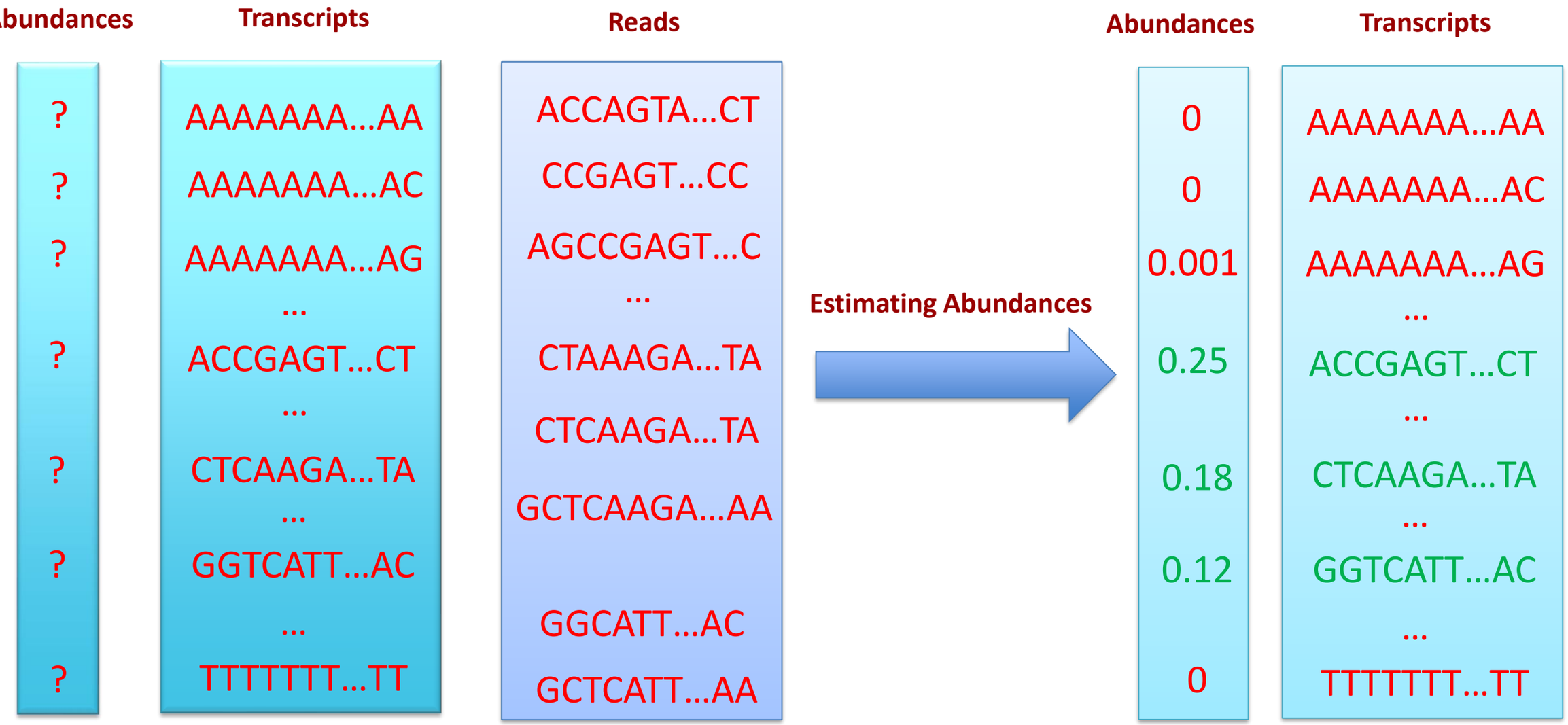
$$\hat{\rho}_{ML}, \hat{t}_{ML} = \underset{\rho,t}{\operatorname{argmax}} \sum_{i=1}^{N} \log\left(\sum_{j=1}^{m} \alpha_{ij}\rho_j\right) \quad s.t \quad \sum_{j=1}^{m} \rho_j = 1\,, \quad \rho_j \geq 0$$

- What if the set of transcripts is given?
- We can consider all possible transcripts
- Estimating the abundances and keeping non-zero ones.
- Is this approach efficient?

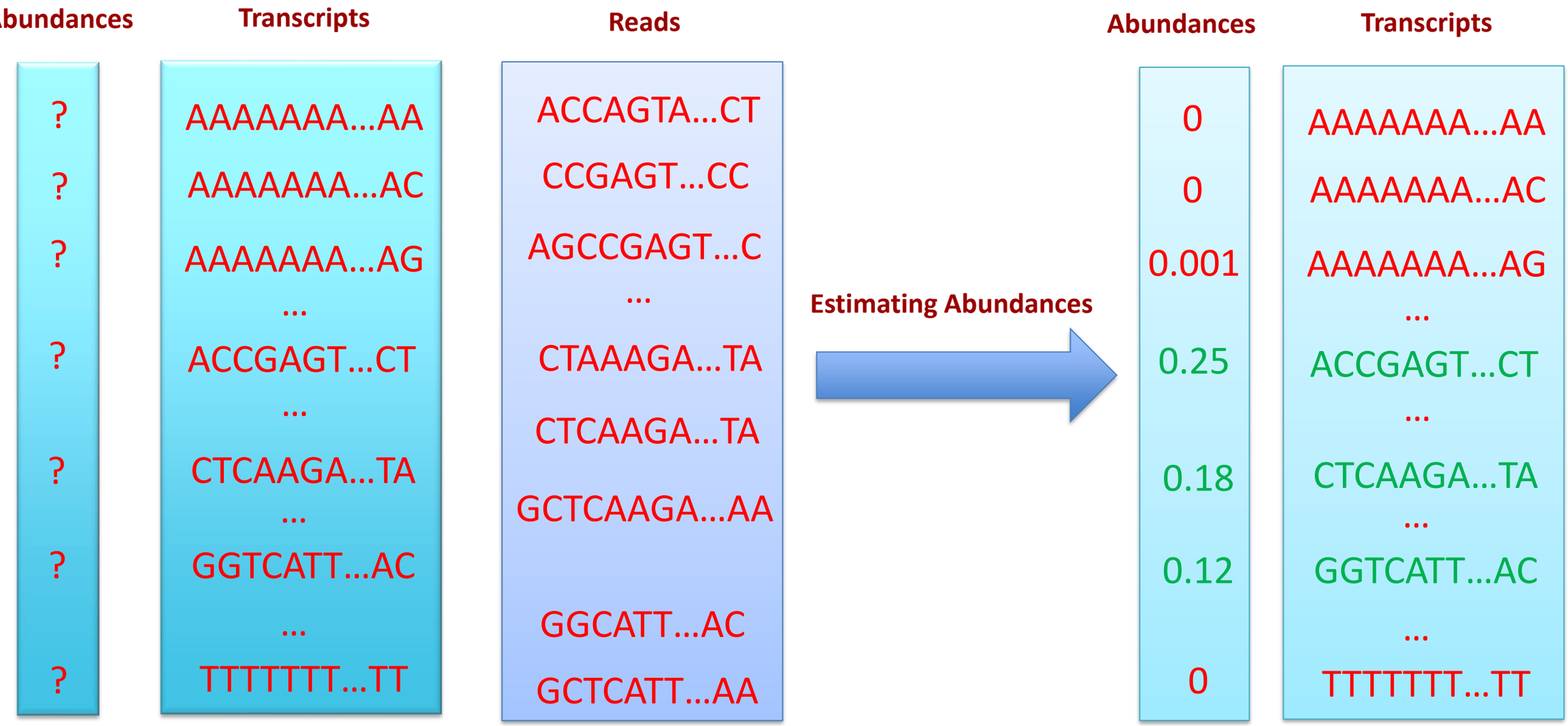
A Brute Force Approach



A Brute Force Approach

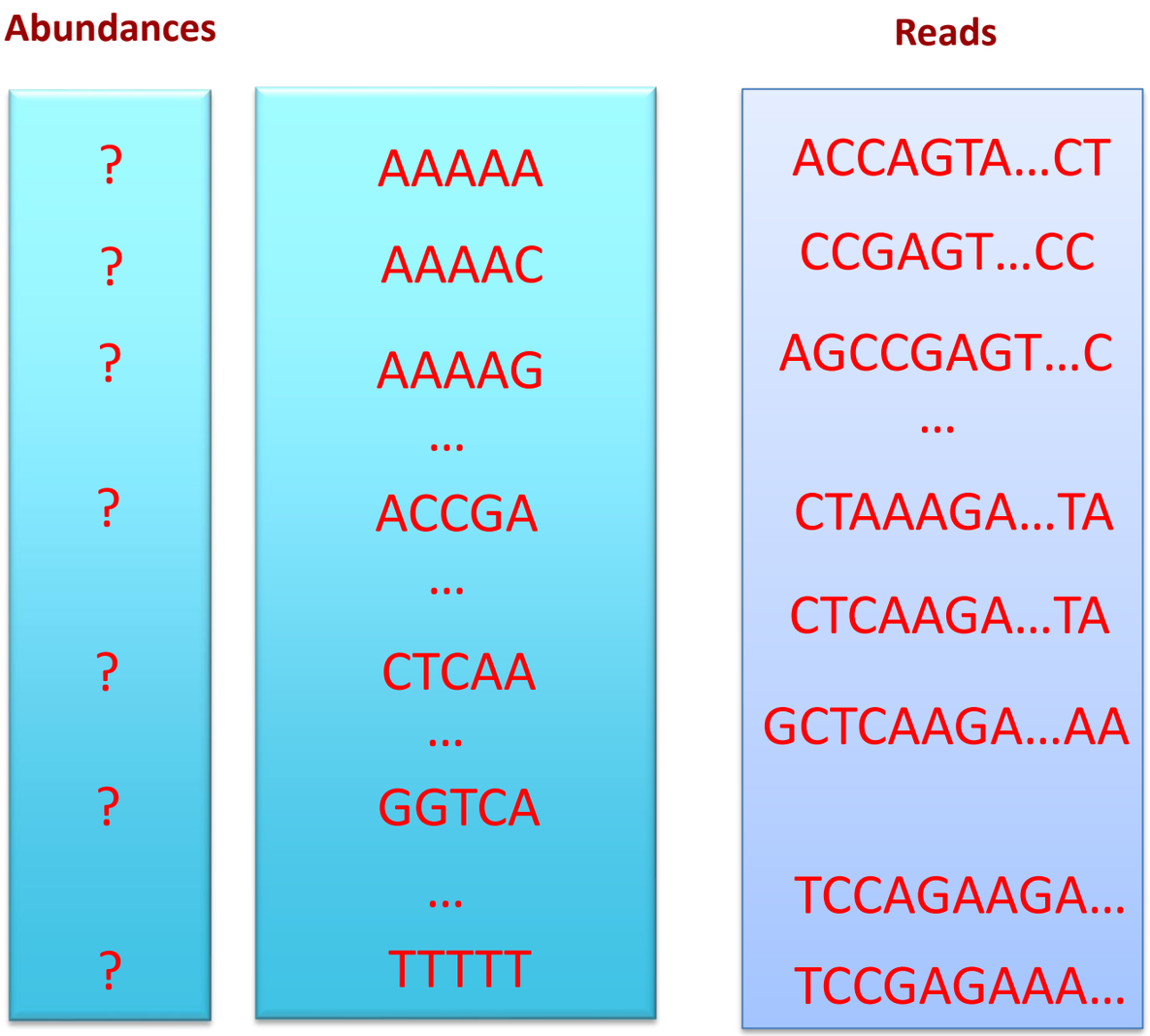


A Brute Force Approach



- Total number of transcripts with length L: $\mathbf{4^L}$

I-CONVEX Algorithm



I-CONVEX Algorithm

Abundances	Prefixes with length 5	Reads
0	AAAAA	ACCAGTA…CT
0.001	AAAAC	CCGAGT…CC
0.002	AAAAG	AGCCGAGT…C
	…	…
0.25	ACCGA	CTAAAGA…TA
	…	CTCAAGA…TA
0.19	CTCAA	GCTCAAGA…AA
	…	
0.16	GGTCA	TCCAGAAGA…
	…	TCCGAGAAA…
?	TTTTT	

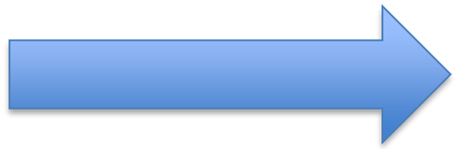
I-CONVEX Algorithm

Abundances	Prefixes with length 5	Reads
0	AAAAA	ACCAGTA…CT
0.001	AAAAC	CCGAGT…CC
0.002	AAAAG	AGCCGAGT…C
	…	…
0.25	ACCGA	CTAAAGA…TA
	…	
0.19	CTCAA	CTCAAGA…TA
	…	
0.16	GGTCA	GCTCAAGA…AA
	…	
?	TTTTT	TCCAGAAGA…
		TCCGAGAAA…

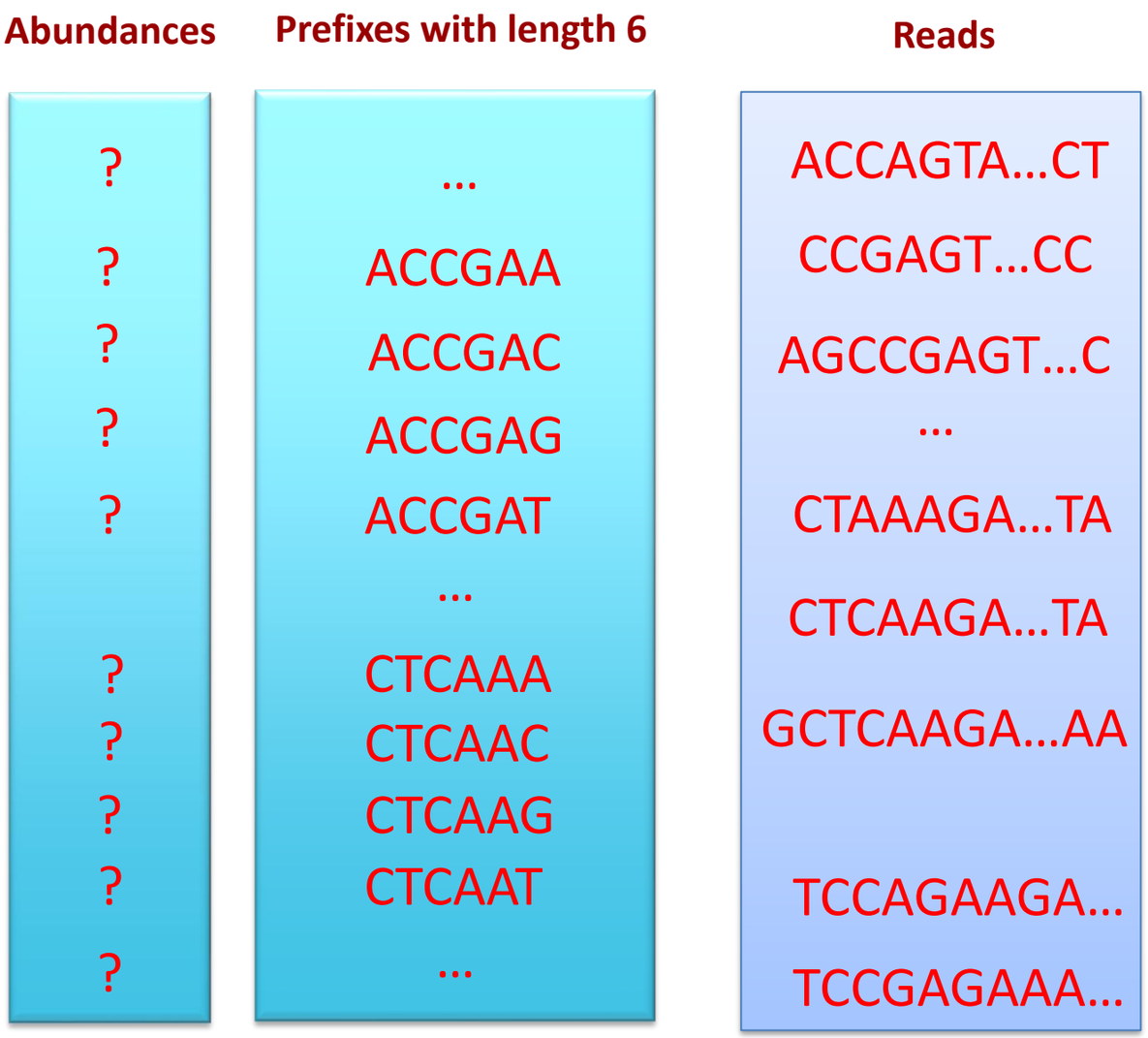
I-CONVEX Algorithm

Abundances	Prefixes with length 5	Reads
0	AAAAA	ACCAGTA…CT
0.001	AAAAC	CCGAGT…CC
0.002	AAAAG	AGCCGAGT…C
	…	…
0.25	ACCGA	CTAAAGA…TA
	…	
0.19	CTCAA	CTCAAGA…TA
	…	
0.16	GGTCA	GCTCAAGA…AA
	…	
?	TTTTT	TCCAGAAGA…
		TCCGAGAAA…

Extend by one base



I-CONVEX Algorithm



I-CONVEX Algorithm

Abundances	Prefixes with length 6	Reads
0	…	ACCAGTA…CT
0.05	ACCGAA	CCGAGT…CC
0.04	ACCGAC	AGCCGAGT…C
0.27	ACCGAG	…
0.01	ACCGAT	CTAAAGA…TA
	…	CTCAAGA…TA
0.02	CTCAAA	GCTCAAGA…AA
0.01	CTCAAC	
0.19	CTCAAG	TCCAGAAGA…
0.03	CTCAAT	TCCGAGAAA…
?	…	

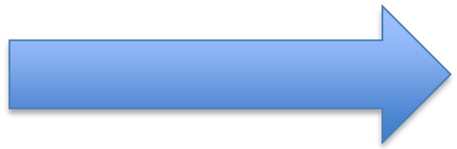
I-CONVEX Algorithm

Abundances	Prefixes with length 6	Reads
0	…	ACCAGTA…CT
0.05	ACCGAA	CCGAGT…CC
0.04	ACCGAC	AGCCGAGT…C
0.27	ACCGAG	…
0.01	ACCGAT	CTAAAGA…TA
	…	CTCAAGA…TA
0.02	CTCAAA	GCTCAAGA…AA
0.01	CTCAAC	
0.19	CTCAAG	TCCAGAAGA…
0.03	CTCAAT	TCCGAGAAA…
?	…	

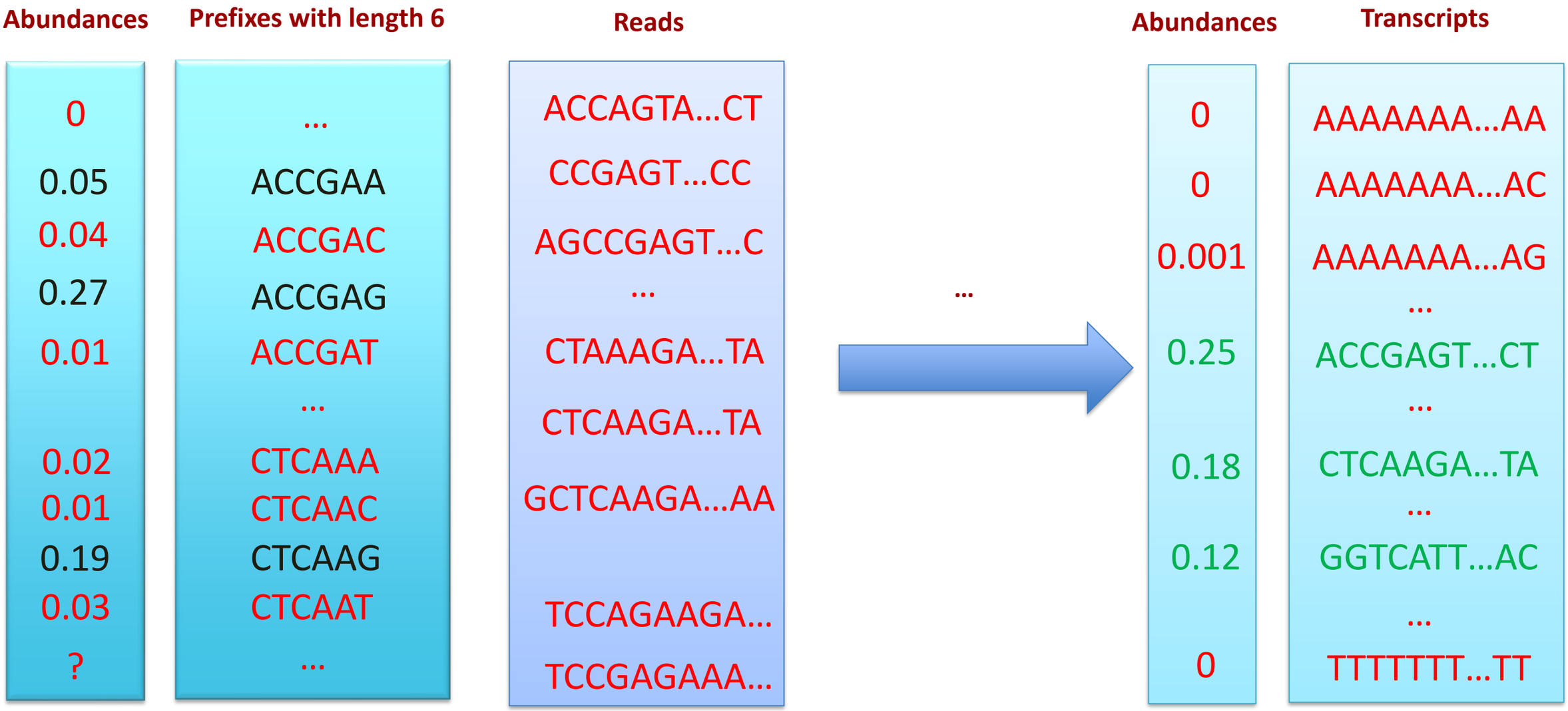
I-CONVEX Algorithm

Abundances	Prefixes with length 6	Reads
0	…	ACCAGTA…CT
0.05	ACCGAA	CCGAGT…CC
0.04	ACCGAC	AGCCGAGT…C
0.27	ACCGAG	…
0.01	ACCGAT	CTAAAGA…TA
	…	CTCAAGA…TA
0.02	CTCAAA	GCTCAAGA…AA
0.01	CTCAAC	
0.19	CTCAAG	TCCAGAAGA…
0.03	CTCAAT	TCCGAGAAA…
?	…	

Extend by one base …

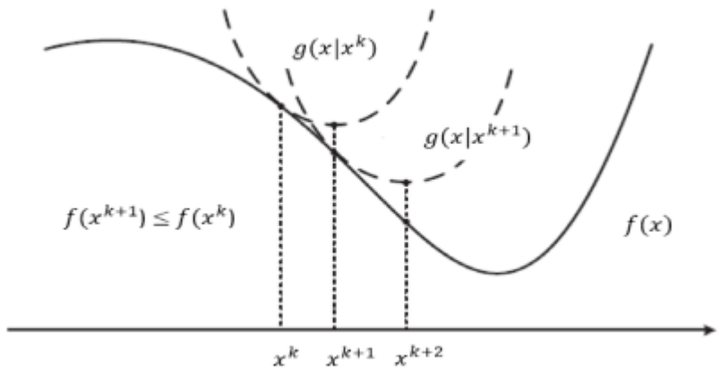


I-CONVEX Algorithm

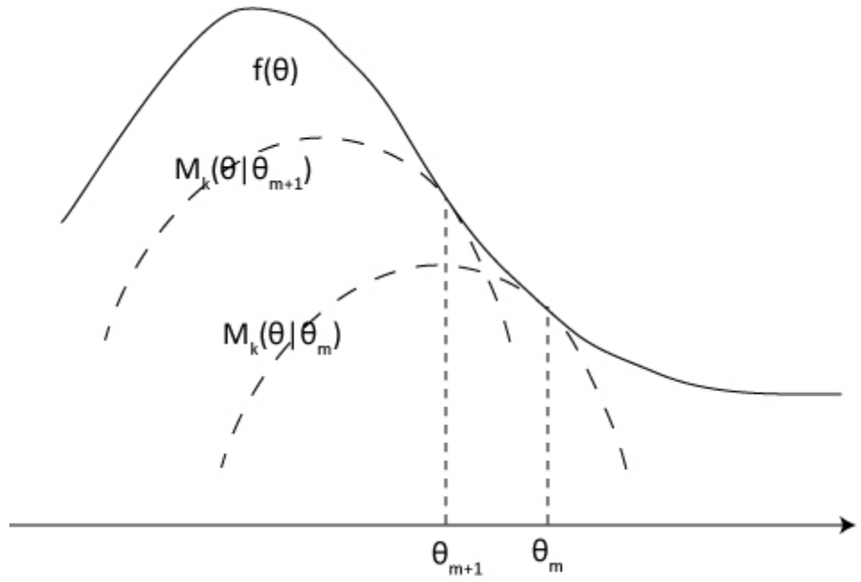


How to Solve the Abundance Estimation Problem?

- Majorization-Minimization (MM)
  - Find a tight local upper-bound at each iteration
  - Find the minimum value of the upper-bound
  - Repeat until a convergence criteria is satisfied.



- Tight local upper-bound at point x_k:
  - Equal at that point: $f(x_k) = g(x_k)$
  - Equality of gradients: $\nabla f(x_k) = \nabla g(x_k)$
  - Upper-bound: $f(x) \leq g(x) \quad \forall x \in D_f$
- Non-increasing sequence of points (why?)



How to Solve the Abundance Estimation Problem?

$$\hat{\rho}_{ML} = \underset{\rho}{\operatorname{argmin}} - \sum_{i=1}^{n} \log \Big(\sum_{j=1}^{m} \alpha_{ij} \rho_j \Big) \quad \text{subject to} \quad \rho \geq 0, \quad \sum_{j=1}^{m} \rho_j = 1$$

- Minimizing a tight local upper-bound at each iteration:

$$\rho^{t+1} = \underset{\rho}{\operatorname{argmin}} - \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\alpha_{ij} \rho_j^t}{\sum_{j'=1}^{m} \alpha_{ij'} \rho_{j'}^t} \log \left(\frac{\rho_j}{\rho_j^t} \right) - \sum_{i=1}^{n} \log \Big(\sum_{j=1}^{m} \alpha_{ij} \rho_j \Big)$$

$$\text{subject to} \quad \rho \geq 0, \quad \sum_{j=1}^{m} \rho_j = 1$$

- Lagrangian function:

$$L(\rho, \lambda) = - \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\alpha_{ij} \rho_j^t}{\sum_{j'=1}^{m} \alpha_{ij'} \rho_{j'}^t} \log \left(\frac{\rho_j}{\rho_j^t} \right) - \sum_{i=1}^{n} \log \Big(\sum_{j=1}^{m} \alpha_{ij} \rho_j \Big) + \lambda \Big(\sum_{j=1}^{m} \rho_j - 1 \Big)$$

- Dual Problem:

$$\max_{\lambda} \ \min_{\rho} L(\rho, \lambda)$$

How to Solve the Abundance Estimation Problem?

$$L(\rho,\lambda) = -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\alpha_{ij}\rho_j^t}{\sum_{j'=1}^{m}\alpha_{ij'}\rho_{j'}^t}\log\left(\frac{\rho_j}{\rho_j^t}\right) - \sum_{i=1}^{n}\log\Big(\sum_{j=1}^{m}\alpha_{ij}\rho_j\Big) + \lambda\Big(\sum_{j=1}^{m}\rho_j - 1\Big)$$

- Stationarity condition:

$$\nabla L_\rho(\rho,\lambda) = 0$$

$$-\sum_{i=1}^{n}\frac{\alpha_{ik}\rho_k^t}{\sum_{j'=1}^{m}\alpha_{ij'}\rho_{j'}^t}\frac{1}{\rho_k^*} + \lambda^* = 0, \quad \forall k = 1,\ldots,m.$$

- Complementary slackness: $\lambda^*(\sum_{k=1}^{m}\rho_k^* - 1) = 0$

- Thus:

$$\lambda^* = n$$

$$\rho_k^{t+1} = \frac{1}{n}\sum_{i=1}^{n}\frac{\alpha_{ik}\rho_k^t}{\sum_{j'=1}^{m}\alpha_{ij'}\rho_{j'}^t} \quad \forall k = 1,\ldots,m.$$

Add Sparsity Regularizer:

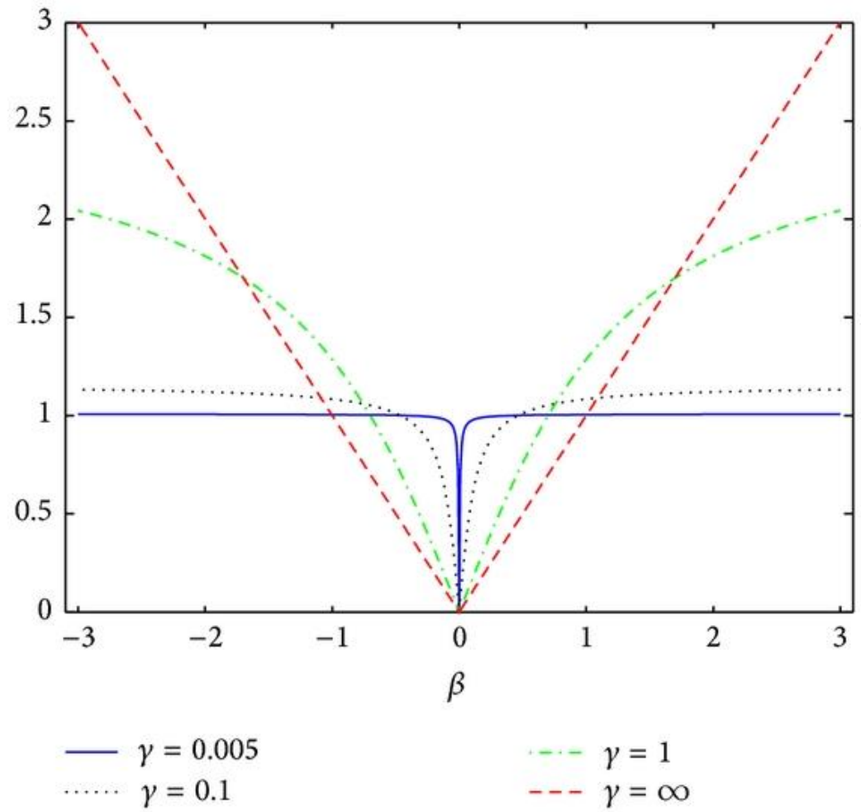
$$L(\rho,\lambda)=-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\alpha_{ij}\rho_j^t}{\sum_{j'=1}^{m}\alpha_{ij'}\rho_{j'}^t}\log\left(\frac{\rho_j}{\rho_j^t}\right)-\sum_{i=1}^{n}\log\Big(\sum_{j=1}^{m}\alpha_{ij}\rho_j\Big)+\lambda\Big(\sum_{j=1}^{m}\rho_j-1\Big)$$

- For adding sparsity, we need to approximate ℓ_0 regularizer:
- Can we use ℓ_1 regularizer?
- Thus:

$$\rho^{t+1}=\underset{\rho}{\operatorname{argmin}}-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{\alpha_{ij}\rho_j^t}{\sum_{j'=1}^{m}\alpha_{ij'}\rho_{j'}^t}\log\left(\frac{\rho_j}{\rho_j^t}\right)-\sum_{i=1}^{n}\log\Big(\sum_{j=1}^{m}\alpha_{ij}\rho_j\Big)$$
$$\text{subject to}\quad \rho\geq 0,\quad \|\rho\|_q\leq 1$$

- Update rule:

$$\rho_k^{t+1}=\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\alpha_{ik}\rho_k^t}{\sum_{j'=1}^{m}\alpha_{ij'}\rho_{j'}^t}\Big)^{\frac{1}{q}},\quad \forall k=1,\ldots,m.$$



Parallelization:

- Assume that the reads are distributed over different cores:
  - Updating the abundances for each core:

$$\rho_j^l \leftarrow \frac{1}{n} \sum_{i \in R_l} \frac{\alpha_{ij}\rho_j}{\sum_{k=1}^{m} \alpha_{ik}\rho_k} \quad \forall j = 1, \ldots, m,$$

  - Aggregating the computed abundances:

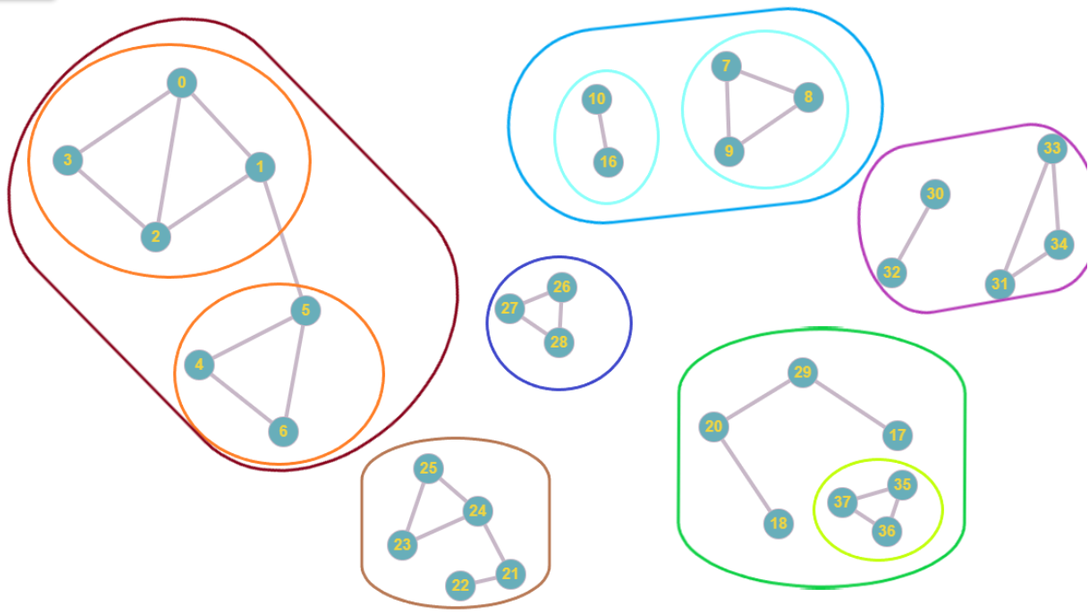
$$\rho_j \leftarrow \left(\sum_{l=1}^{c} \rho_j^l \right)^{\frac{1}{q}} \quad \forall j = 1, \ldots, m.$$

Pre-clustering Module:

- O(NL min{C, m})
  - N: Number of Reads
  - L: Maximum Length of the reads
  - m: Number of Transcripts
  - C: Number of comparisons per read at each iteration

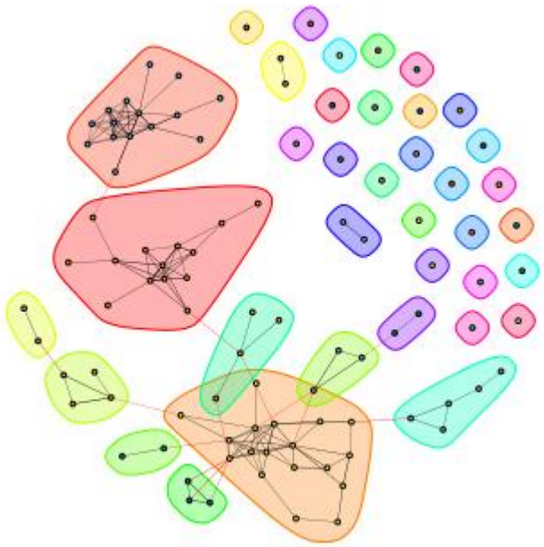
How to Improve it?

- Pre-clustering the reads



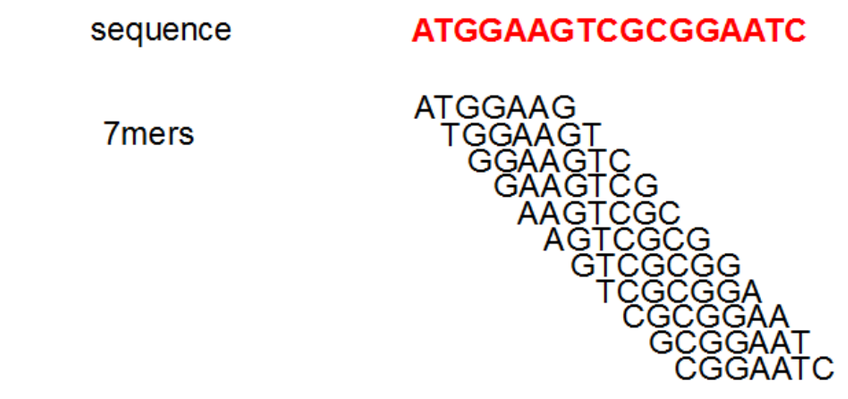
How to Do Pre-clustering Efficiently:

- Reduce the dimension of the reads with a hash function called MinHash:
  - Project the reads to a lower dimension space
  - The distance between the reads are preserved with high probability.
  - Needs O(NL) operations.
  - Some of the detected candidate pairs are false positive.



Pre-clustering Module:

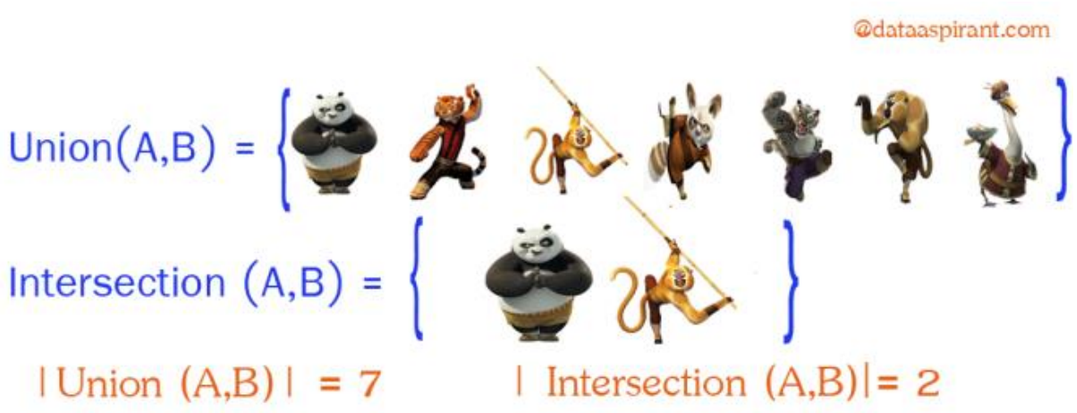
- K-mer representation
- *Min Hash*: A hash function for transformation of data
  - Reduce the dimension (length) of the reads
  - Preserve the Jaccard similarity between the reads
  - Determine the candidate pairs in O(NL)
- Some of the candidate pairs are false positive.



Jaccard Similarity J (A,B) = | Intersection (A,B) | / | Union (A,B) |

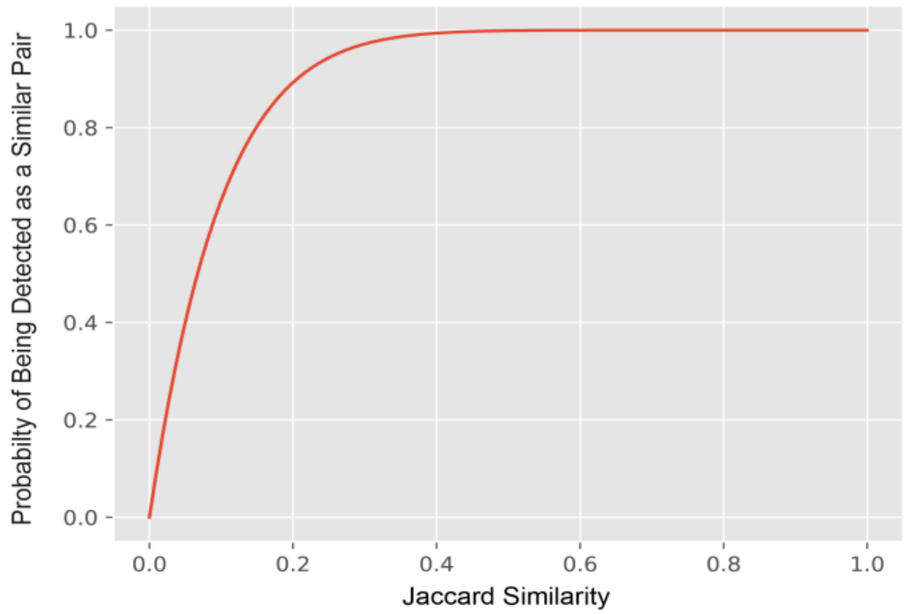
= 2 / 7

= 0.286



Connection of Jaccard Similarity and MinHash:

- It is easy to show that the probability of having equal MinHash with respect to a permutation for two given k-mer sets is their Jaccard similarity.
- Number of matches between two MinHash signatures with length *h:*
  - A binomial distribution with *h* trials
  - What is the expectation of number of matches?
  - What is the variance of number of matches?
- LSH: probability of mapping to the same bucket with
  - *b* bands
  - *d* rows



Connection of Jaccard Similarity and MinHash:

- Generating a training dataset by imitating the sequencing procedure.
- Training a Convolutional Neural Network using generated dataset.
- Validating Candidate Pairs to eliminate false positives.

Layer	Input	Filter Size	Filters	Stride	Activation
Convolution 1	8X400	8X8	32	1	Relu
Max Pool 1	8X400X32	2X2	-	2	-
Convolution 2	8X100X32	5X5	64	1	Relu
Max Pool 2	8X100X64	2X2	-	2	-
Fully connected 1	4X50X64		50		Relu
Fully connected 2	50		1		Linear

Results:

➢ Recovery accuracy on SIRV datasets:

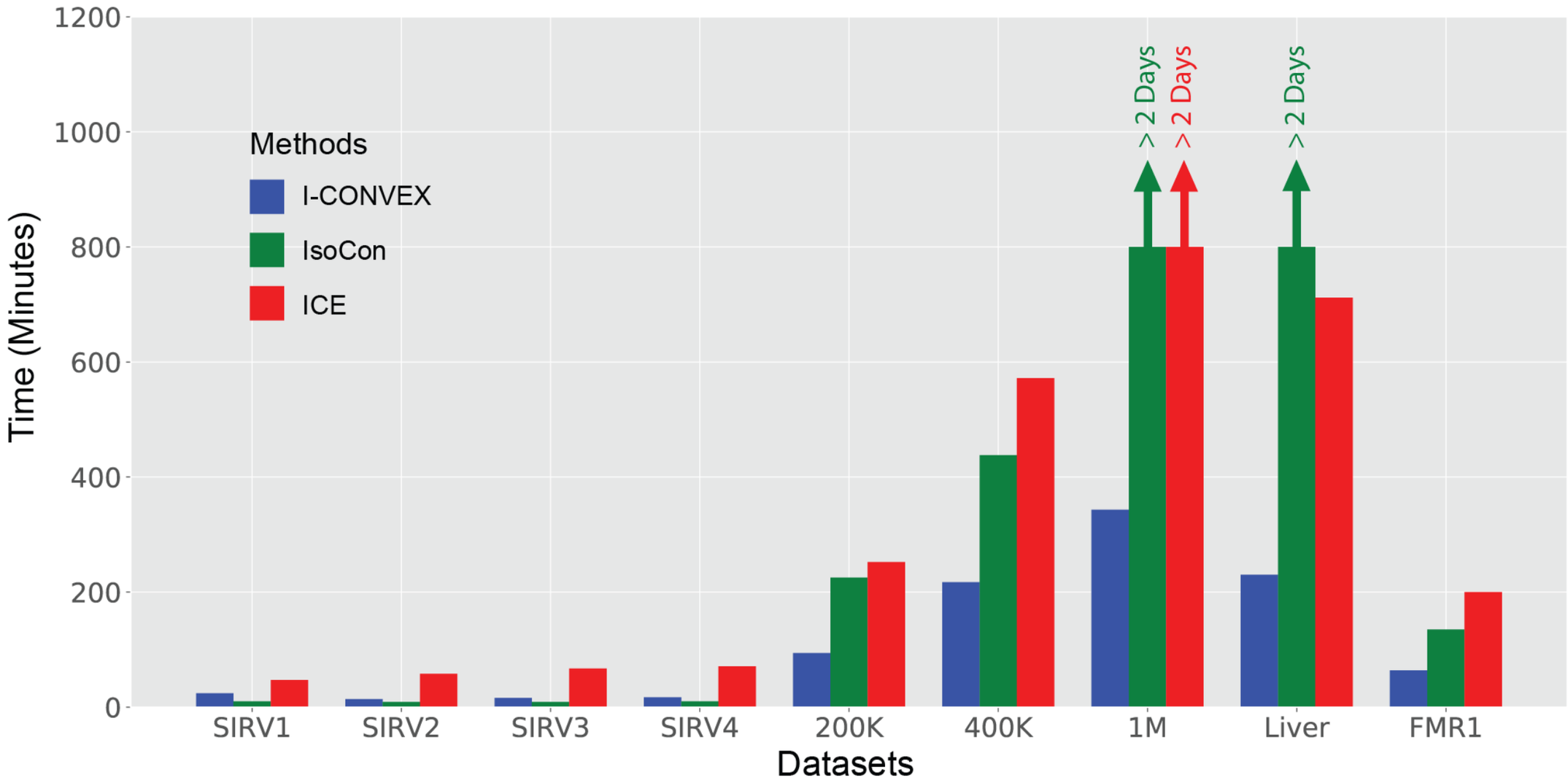
	SIRV1			SIRV2			SIRV3			SIRV4		
	Recall	Precision	F-score	Recall	Precision	F-score	Recall	Precision	F-score	Recall	Precision	F-score
I-CONVEX	95.65%	27.61%	**0.42**	95.65%	21.29%	**0.34**	95.65%	16.25%	**0.27**	88.40%	15.13%	**0.25**
ICE	97.10%	7.46%	0.13	95.65%	5.11%	0.09	94.20%	9.48%	0.17	79.71%	7.23%	0.13
IsoCon	95.65%	14.60%	0.25	97.10%	11.71%	0.20	97.10%	9.43%	0.17	92.75%	9.07%	0.16

➢ Recovery accuracy on synthetic datasets:

	200K			400K			1M		
	Recall	Precision	F-score	Recall	Precision	F-score	Recall	Precision	F-score
I-CONVEX	93.4 %	96.88 %	**0.95**	97.0 %	99.18 %	**0.98**	97.0 %	98.18 %	**0.98**
ICE	74.2 %	8.68 %	0.15	80.6 %	5.92 %	0.11	No Results After Two Days		
IsoCon	98.4 %	75.11 %	0.85	98.8 %	51.94 %	0.68			

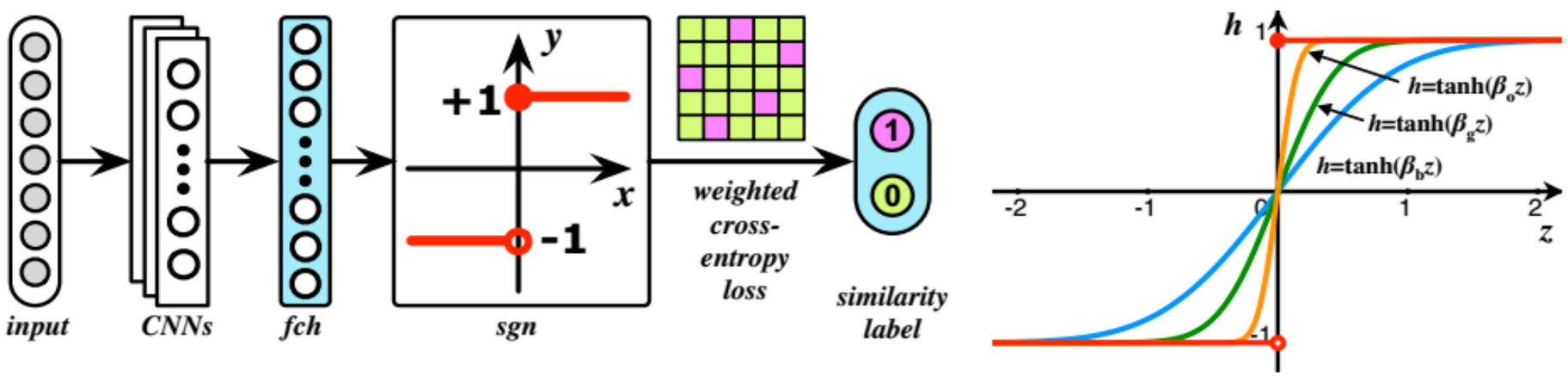
Time Comparison:

- Time comparison of I-CONVEX on real and synthetic datasets:



Potential Challenges:

- Can we use neural networks for hashing sequences?



- Sample complexity: How many samples we need to recover transcripts?
  - Lower-bound: proved by information theory tools (Fano's inequality)
  - I-CONVEX: can recover transcripts with minimal number of samples (in progress …)

Cao, Zhangjie, et al. "Hashnet: Deep learning to hash by continuation." *Proceedings of the IEEE international conference on computer vision*. 2017.

References (Long-read Sequencing):

- Roberts, Adam, and Lior Pachter. "Streaming fragment assignment for real-time analysis of sequencing experiments." *Nature methods* 10.1 (2013): 71-73.

- Rajaraman, Anand, and Jeffrey David Ullman. *Mining of massive datasets*. Cambridge University Press, 2011.

- Kannan, Sreeram, et al. "Shannon: an information-optimal de Novo RNA-Seq assembler." *BioRxiv* (2016): 039230.

- Amarasinghe, Shanika L., et al. "Opportunities and challenges in long-read sequencing data analysis." *Genome biology* 21.1 (2020): 1-16.

References (Optimization):

- Nesterov, Yurii. *Introductory lectures on convex optimization: A basic course*. Vol. 87. Springer Science & Business Media, 2013.

- Ben-Tal, Aharon, and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Society for industrial and applied mathematics, 2001.

- Hong, M., Razaviyayn, M., Luo, Z. Q., & Pang, J. S. (2015). A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing. *IEEE Signal Processing Magazine*, *33*(1), 57-77

- Nouiehed, Maher, Jong-Shi Pang, and Meisam Razaviyayn. "On the pervasiveness of difference-convexity in optimization and statistics." *Mathematical Programming* 174.1-2 (2019): 195-222

- Baharlouei, S., Nouiehed, M., Beirami, A., & Razaviyayn, M. R\'enyi Fair Inference. *International Conference on Learning Representations (ICLR), 2020.*