Copy number evolution with weighted aberrations in cancer

Ron Zeira and Benjamin J. Raphael

Presented by Mrinmoy
Copy number aberrations (CNAs)

- Deletion or amplification of large genomic regions
- Source of somatic mutation in many cancer type
CNPs and Events

- Copy number profile, \( C = (c_1,c_2,\ldots,c_n) \)
  Vector of non-negative integers

- Events, \( e = (i, j, \tau) \),
  \( 1 \leq i \leq j \leq n, \tau \in \{+1, -1\} \)
CNPs and Events

\[
\begin{array}{ccccccc}
  c_0 & \ldots & c_{i-1} & c_i & \ldots & c_j & c_{j+1} & \ldots & c_n \\
\end{array}
\]

\[
( i, j, \tau )
\]

\[
\begin{array}{ccccccc}
  c_0 & \ldots & c_{i-1} & \max(C_{i+\tau}, 0) & \ldots & \max(c_j + \tau, 0) & c_{j+1} & \ldots & c_n \\
\end{array}
\]
CNT & CND

- Copy number transformation from CNP S to CNP T:
  
  \[ E = (e_1, e_2, ..., e_l) \]
  
  such that \( e_l(...(e_1(S)) = T \)

- Copy Number Distance, \( d(S,T) = \min\{E:E(S)=T\} \) \(|E|\)
- CNT is not a true distance
- \( d(S,T) = \infty \) if \( s_i = 0 \) for any \( 1 \leq i \leq n \)
Phase

- $E = (E_1, E_2, ..., E_n)$

  **Phase** - $E_i$ where (1) $E_i$ is a subsequence of $E$, (2) all the events in $E_i$ has the same type, and (3) adjacent segments have different types.

- $\text{op}(E_j, i) = | \{ (l, r, \tau) \in E_j | l \leq i \leq r \} |$

  Change in segment $i$ by events in phase $E_j$

- CNT $E$ from $S$ to $T$ is **phase-bounded** provided if -

  $\text{op}(E_j, i) \leq B$ where $B = \max( \max(S), \max(T) )$
Ordered CNT

- Ordered CNT, $E = (E_-, E_+)$
  All deletions come before all amplifications

- If $d(S, T) < \infty$, then there exists an ordered phase-bounded CNT $E$ s.t. $E(S) = T$
Semi-ordered CNT

- Which one is more probable?
Semi-ordered CNT

- Semi-ordered CNT:
  \[ E = (E_1, E_2, E_3) \]
  s.t. \( \tau(E_1 \cup E_3) = -1, \tau(E_2) = +1 \)
  and \( E_1(S_i) = 0 \) if \( t_i = 0 \)

- Why?
  - Richer space of transformations
  - Still tractable
Problems with CND

CND considers all events equally.

Problems -

- CNAs of different length occurs at different rates
- Length dependent uncertainty in real data
Weighted CNT Model

- Event weight function, \( w: \{1..n\} \times \{1...n\} \times \{+, -\} \rightarrow \mathbb{R} \)
  
  Takes as input an event \( e \), and outputs its weight

- Weight can change based on position, length and type of CNA

- Weight of CNT \( E \):

  \[
  w(E) = \sum_{e \in E} w(e)
  \]
Weighted CNT Model
Minimum weight semi-ordered CNT

- **Problem Statement**: Given a source CNP S, a target CNP T and a weight function \( w \), find semi-ordered phase-bounded CNT \( E \) having a minimum weight \( W(E) \).
- If the weight of an event is the log of the probability of the event, then the problem becomes a Maximum Likelihood problem.

\[
E = \min_{\{E : E(S) = T \mid E \text{ is semi-ordered } \& \text{ phase bounded}\}}(-\sum_{e \in E} \log p_e)
\]
Minimum weight semi-ordered CNT Solution

\( x_{lk}^j \) = Number of events between l and k in phase j

Objective function -  
\[
\min \sum_j \sum_{l \leq k} w(l, k, j) x_{lk}^j
\]

Constraints -
\[
s_i \leq \sum_{l \leq i \leq k} x_{lk}^1 \quad 1 \leq i \leq n, \quad \text{if} \quad t_i = 0,
\]
\[
\sum_{l \leq i \leq k} x_{lk}^1 \leq s_i - 1 \quad 1 \leq i \leq n, \quad \text{if} \quad t_i > 0,
\]
\[
s_i - \sum_{l \leq i \leq k} x_{lk}^1 - x_{lk}^2 + x_{lk}^3 = t_i \quad 1 \leq i \leq n, \quad \text{if} \quad t_i > 0
\]

Minimum weight semi-ordered CNT Solution

Constraints -

\[ s_i - \sum_{l \leq i \leq k} x_{lk}^1 - x_{lk}^2 + x_{lk}^3 = t_i \quad 1 \leq i \leq n, \quad \text{if} \quad t_i > 0, \]

\[ \sum_{l \leq i \leq k} x_{lk}^j \leq B \quad 1 \leq i \leq n, \quad j \in \{1, 2, 3\} \]

\[ 0 \leq x_{lk}^j \quad 1 \leq l \leq k \leq n, \quad j \in \{1, 2, 3\}. \]
Results on simulated data
Results on real data