Supplementary Materials for:

Bioinformatic Analysis of Topoisomerase IIa Reveals Interdomain Interdependencies and Critical C-Terminal Domain Residues

Clark E. Endsley, Kori A. Moore, Thomas D. Townsley, Kirk K. Durston, and Joseph E. Deweese

Supplementary Materials:

Data S1: Examples of spurious clusters derived from algorithm without entropy filtering (.xlsx file).

Data S2: PSICalc data output using TOP2A MSA with 105 species set at spread of 1, 0.1 entropy, and 56% non-insertion data (.xlsx file).

Data S3: PSICalc data output using TOP2A MSA with 105 species set at spread of 1, 0.11 entropy, and 56% non-insertion data (.xlsx file).

MSA: TOP2A 105 species MSA with human TOP2A as first sequence (.csv file). Note the MSA has been truncated to omit regions not found in human Top2A.

Figure S1 (in this document): Protein Alignment Between the CTD of TOP2A and TOP2B.

Table S1 (in this document): Example Clusters Compared between TOP2A and TOP2B CTD.

Entropy Calculation in PSICalc – Modified from scikit learn version under the BSD-3 Clause License to remove gaps in sequences as noted below. Full code in context available here: <https://github.com/jdeweeselab/psicalc-package/blob/main/psicalc/nmi.py#L163-L184>

def entropy(labels):

"""

Unlike the sci-kit entropy function, we are modifying it

to remove gap labels so they do not interfere with the

entropy result.

"""

# non-zero labels i.e. no gaps as gaps are always 0 in psicalc

nz\_labels = labels[labels != 0]

if len(nz\_labels) == 0:

return 1.0

labels\_u, label\_idx = np.unique(nz\_labels, return\_inverse=True)

if labels\_u[0] == 0:

label\_idx = label\_idx[label\_idx != 0]

pi = np.bincount(label\_idx).astype(np.float64)

pi = pi[pi > 0]

if pi.size == 0:

return 0.0

pi\_sum = np.sum(pi)

# log(a / b) should be calculated as log(a) - log(b) for

# possible loss of precision

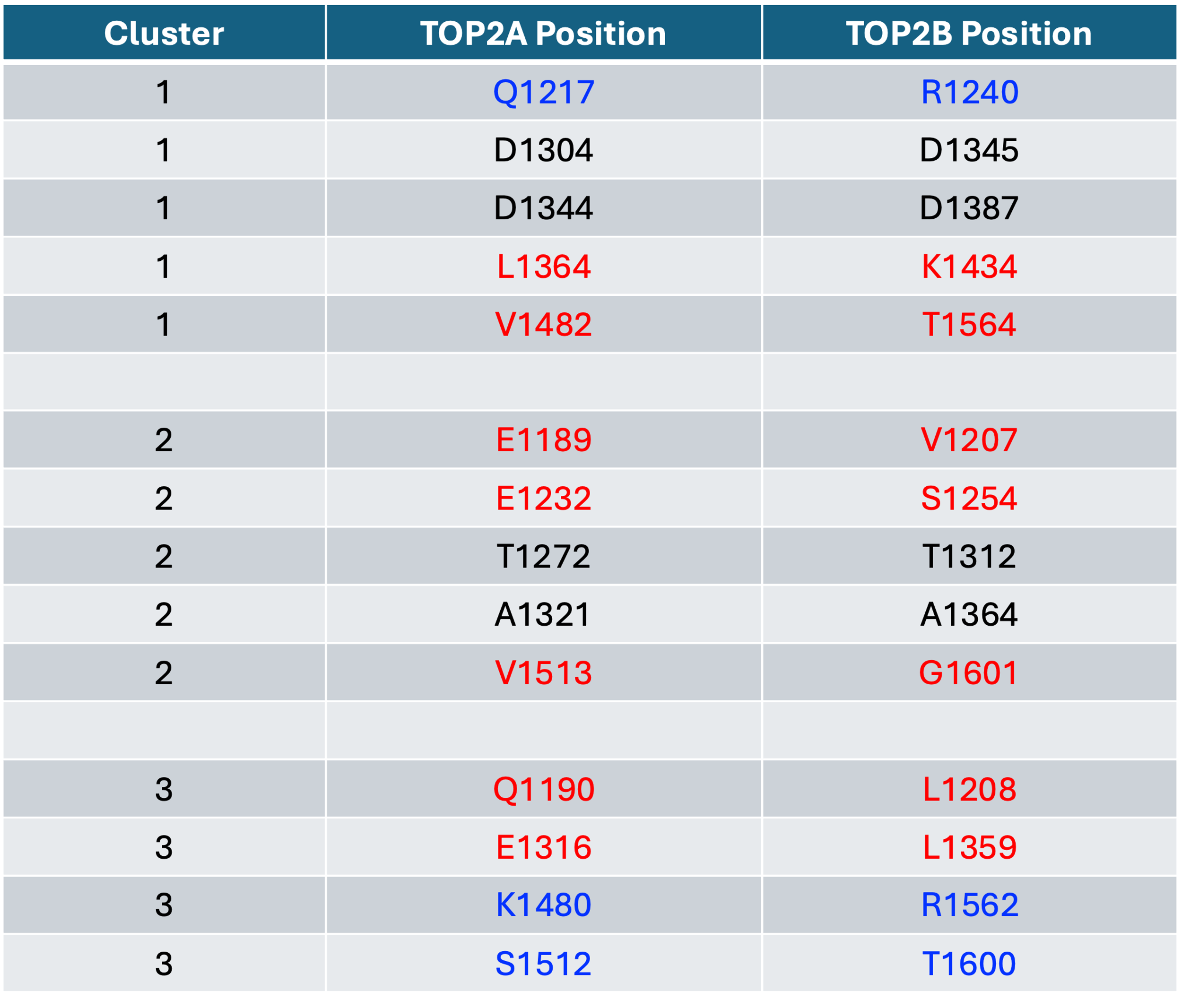
return -np.sum((pi / pi\_sum) \* (np.log(pi) - log(pi\_sum)))

Figure S1: Protein Alignment Between the CTD of Top2A (top) and Top2B (bottom). Aligned sequences (TOP2A NP\_001058 and TOP2B NP\_001317629) using Needleman-Wunsch in SnapGene.

A close up of a chart

Description automatically generated with medium confidence

Table S1: Example Clusters Compared between Top2A and Top2B CTD.



Red denotes differences; blue denotes similarities; black denotes no change. Clusters are from Top2A data and mapped to Top2B using alignment in Figure S1.