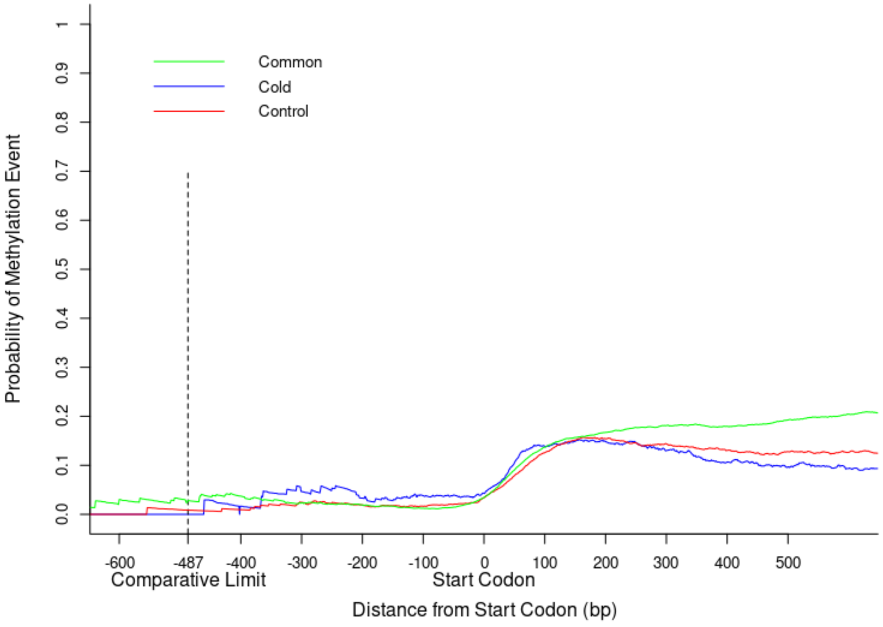
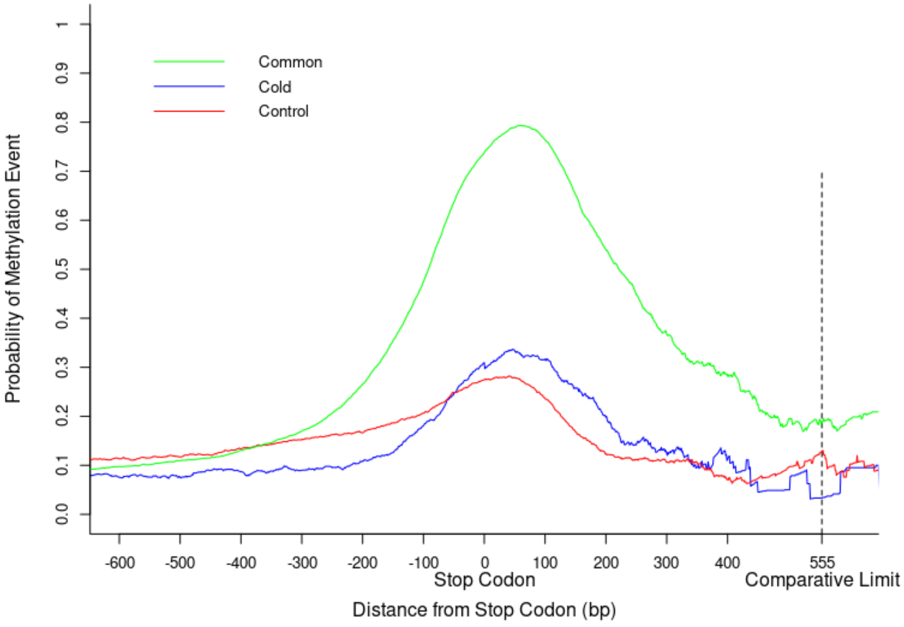
Probability vs. Normalized Distance

5'-UTR/CDS Interface (Methylated Transcripts Only)



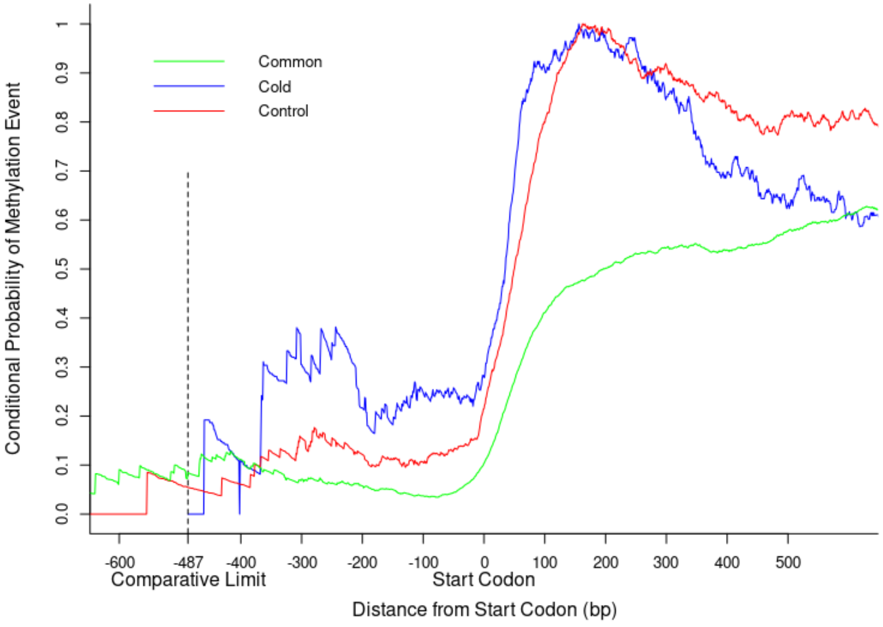
Probability vs. Normalized Distance

CDS/3'-UTR Interface (Methylated Transcripts Only)



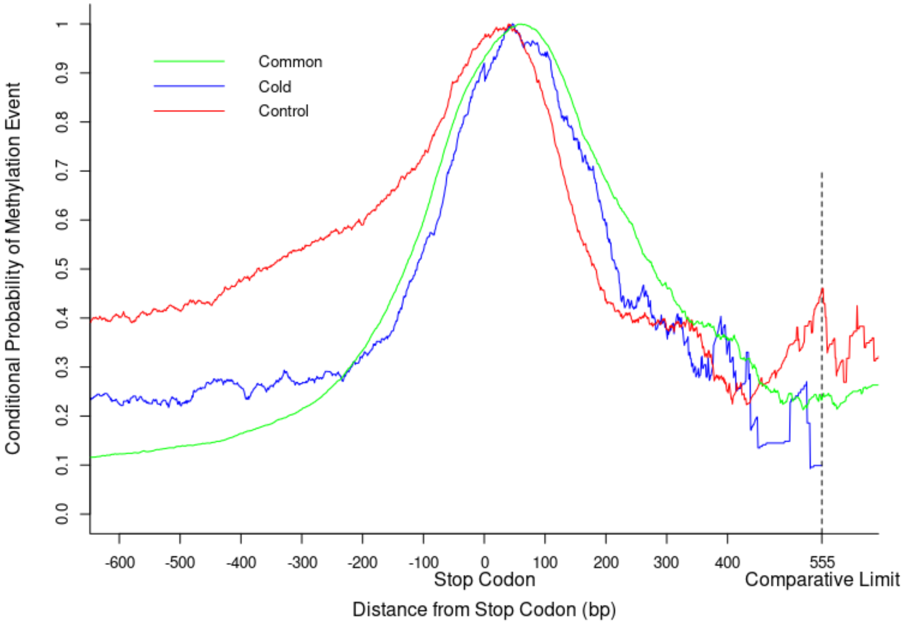
Conditional Probability vs. Normalized Distance

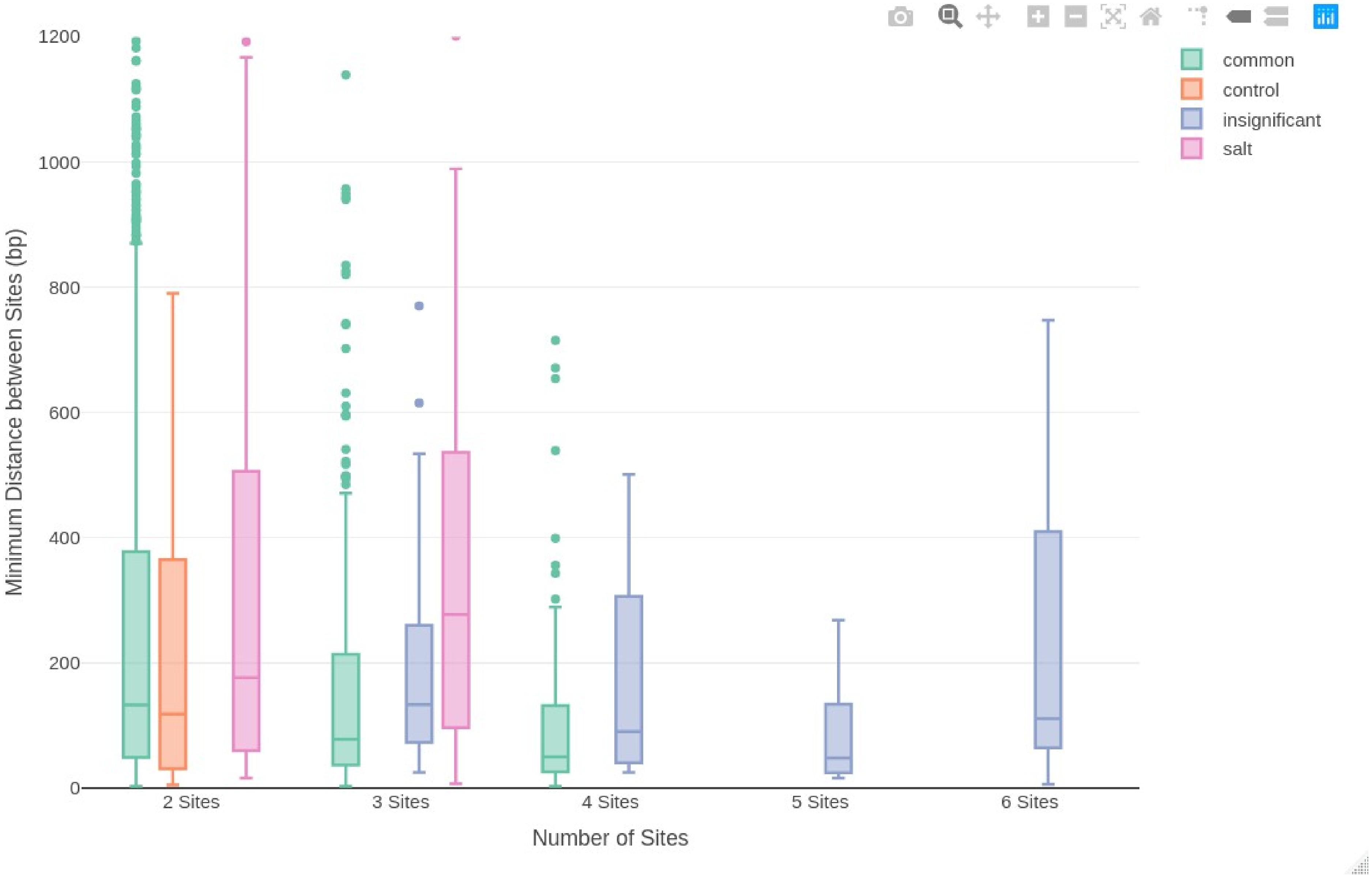
5'-UTR/CDS Interface (Methylated Transcripts Only)

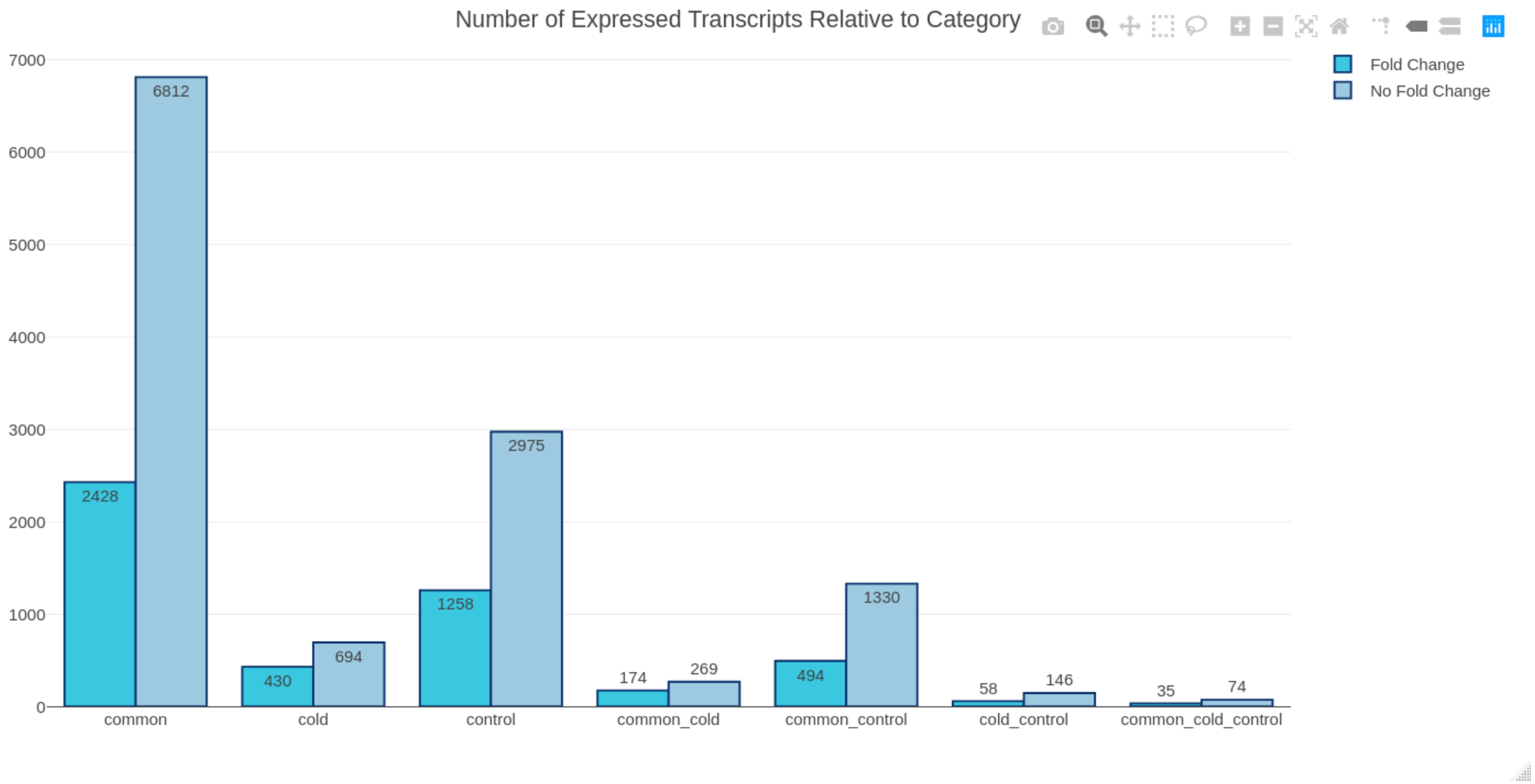


Conditional Probability vs. Normalized Distance

CDS/3'-UTR Interface (Methylated Transcripts Only)




1200
1000
800
600
400
200
0
Minimum Distance between Sites (bp)
2 Sites
3 Sites
4 Sites
5 Sites
6 Sites
Number of Sites
🟩 common
🟧 control
🟦 insignificant
🟪 salt



Number of Expressed Transcripts Relative to Category
Fold Change
No Fold Change
7000
6000
5000
4000
3000
2000
1000
0
2428
6812
430
694
1258
2975
174
269
494
1330
58
146
35
74
common
cold
control
common_cold
common_control
cold_control
common_cold_control

Interpretation of Results

The overall results in the CDS/3′-UTR region comport with previous studies. Namely, methylation sites are heavily concentrated around, largely separated by, this interface. However, our results show, under the Cross-Transcript Index-Event Probability Metric, that common transcripts are generally more likely to have sites at nearly all stop codon-downstream positions. Next most likely for this region are control-only transcripts, and lastly are condition-only transcripts.

Upstream of the stop codon, we see effectively no discernable methylation pattern within common and control transcripts. However, condition transcripts show two sites with pronounced methylation, and in particular, the second of these two sites is essentially equal in probability to the stop codon-downstream site for control as well. As to why these "camel humps" persist upstream remains to be studied.

Lastly, the 5′-UTR/CDS interface shows almost no methylation pattern for any of the transcript types.

The comparative limit is the point past which cross-type statistical comparisons cannot be made because at least one of the types does not have enough transcripts with that many nucleotides (up- or downstream).

Formal Definition of the Cross-Transcript Index-Event Probability Metric

Let T, the **transcript set**, consist of all transcripts where for a transcript $t \in T$ we have $t \subset \mathbb{N} \times \{N, Y\}$ with N and Y indicating lack of a methylation and presence of a methylation, respectively. Let $T_m \subseteq T$ consist of all methylated transcripts. That is, for each $t \in T_m$, we have at least one element in t which is of the form (a, Y) where $a \in \mathbb{N}$. Further, let T_m consist of the three **type sets** of *common*, *condition*, and *control*,

$$T_m = \{T_{common}, T_{condition}, T_{control}\}$$

where it is possible that $T_x \cap T_y \neq \{\varnothing\}$ for $x, y \in \{condition, control\}$.

Next, for each $t \in T_m$, define the **index component**, t_a, and the **response component**, t_b, to be the a-th and b-th components for any $(a, b) \in t$. For each $t \in T_m$, define some $q_t \in \mathbb{N}$, the **offset**. Define the **index normalization function** $\lambda : \; t \to \hat{t}$ as follows,

$$\lambda(t_a, t_b) = (q_t - t_a, t_b)$$

We will call the set $\hat{t}$ the **normalized space of** t. This function essentially moves the 0-th position of a set of natural numbers to (potentially) some other number. Denote by $\hat{T}_m$ the set of all methylated, normalized spaces of T.

Let $d \in \mathbb{Z}$ be some "well-behaved" distance (in index positions) from the datum (the 0 index) of some transcript $t \in \hat{T}_m$. By "well-behaved" we mean that necessarily $\min t_a \leq d \leq \max t_a$. Let $u \in \hat{T}_m$. u is said to be **within the radius of** t if and only if there exists some $u_a = d$ (by definition, t is within the radius of t). Let $R_t \subseteq \hat{T}_m$, the **radial set of** t, consist of all such transcripts within the radius of t,

$$R_t = \{u_1, u_2, \ldots, u_n\}$$

In other words, the radial set consists of all transcripts which also contain the index defined by d on t.

Next, define the **restriction of** u **by** d as follows,

$$u^d = \{(a, b) \in u \ : \ a = d\}$$

In essence, we want only the singular element in some u corresponding to the distance. Let R_t^d, consist of all such u restricted by d,

$$R_t^d = \{u_1^d, u_2^d, \ldots, u_n^d\}$$

We will call R_t^d the d-**restricted set**.

Finally, define the **Cross-Transcript Index-Event Probability** for t at d, $\mathbb{P}_t^d$, as the proportion of transcripts which contain a positive event at a given index,

$$\mathbb{P}_t^d = \frac{\sum\limits_{i=1}^{n} f(i)}{|R_t^d|}$$

where

$$f(i) = \begin{cases} 1 & \text{if } u_i^d|_b = Y \\ 0 & \text{if } u_i^d|_b = N \end{cases}$$

and $u_i^d|_b$ is the b-th component of u_i^d.