The Association of Indels with Meiotic Recombination Sites in Maize

Nikita Sajai and Wojtek P. Pawlowski
School of Integrative Plant Science, Cornell University

Abstract

The processes occurring during meiotic recombination, from the initiation of DNA double-strand breaks (DSBs) to the completion of crossing-over (CO) formation, have numerous opportunities for inaccuracy. In particular, several types of missteps during recombination, including the formation of double DSBs, could lead to genome deletions. To explore this phenomenon in maize, we examined the occurrence of indels at recombination sites. To do it, we mapped indels to meiotic DSB hotspots and CO sites and measured the presence and degree of enrichment of indels of different sizes. We assessed the indel generation potential of meiotic recombination by calculating three measures: indel overlap by recombination sites, recombination site overlap by indels, and indel density at recombination sites. We found substantial enrichment of small indels (1-50 bp) at CO sites, providing the first strong evidence of mutagenicity of meiotic recombination in plants. Small indel density decreased in regions 2 kb upstream and downstream from the recombination sites, implying that the indels were generated in a localized fashion. Indels have the potential to alter gene function and plant phenotype, marking their importance as a source of genetic diversity. Thus, understanding mechanisms of indel origination and their relationship to the recombination mechanism is important for efforts that exploit genetic diversity for crop improvement.

Overall Question: To what extent is recombination in maize mutagenic and what factors control this mutagenicity?

Background

- COs are key sources of genetic variation[2]
  - Associated with active chromatin
  - Frequent in gene promoters and terminators
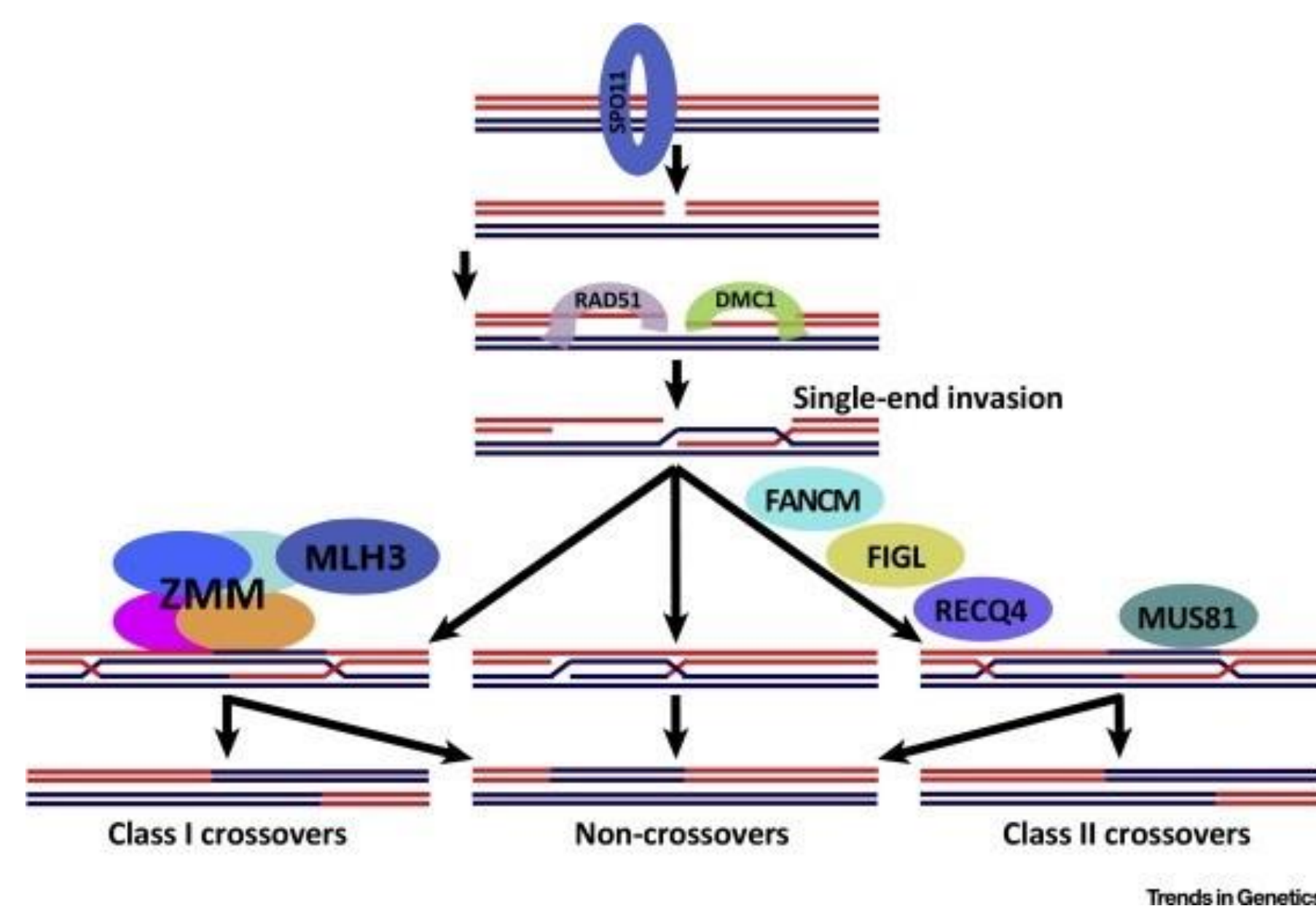 - Suppressed around the centromere

Diagram of the meiotic pathway showing key proteins involved at each step. (Zelkowski *et al.*, 2019)

- **Potential sources of indels during recombination**:
  - Double DSBs
  - Non-Homologous End-Joining Mediated DSB Repair
  - Microhomology-Mediated DSB Repair
- Indels may alter phenotype, marking their importance as a source of genetic diversity

Data

Recombination Sites: ~3100 DSB hotspots, ~30000 empirical COs, ~300 empirical CO hotspots, ~60000 predicted CO sites, ~32000 MLH3 hotspots identified at the diplotene stage of maize meiocytes

Indels: Small indels (1-50 base pair (bp) indels), medium indels (100-500 bp), large indels (500 bp-50kb).

Methodology

Characterizing Indel Distribution
Distribution plots using ChromoMap in R
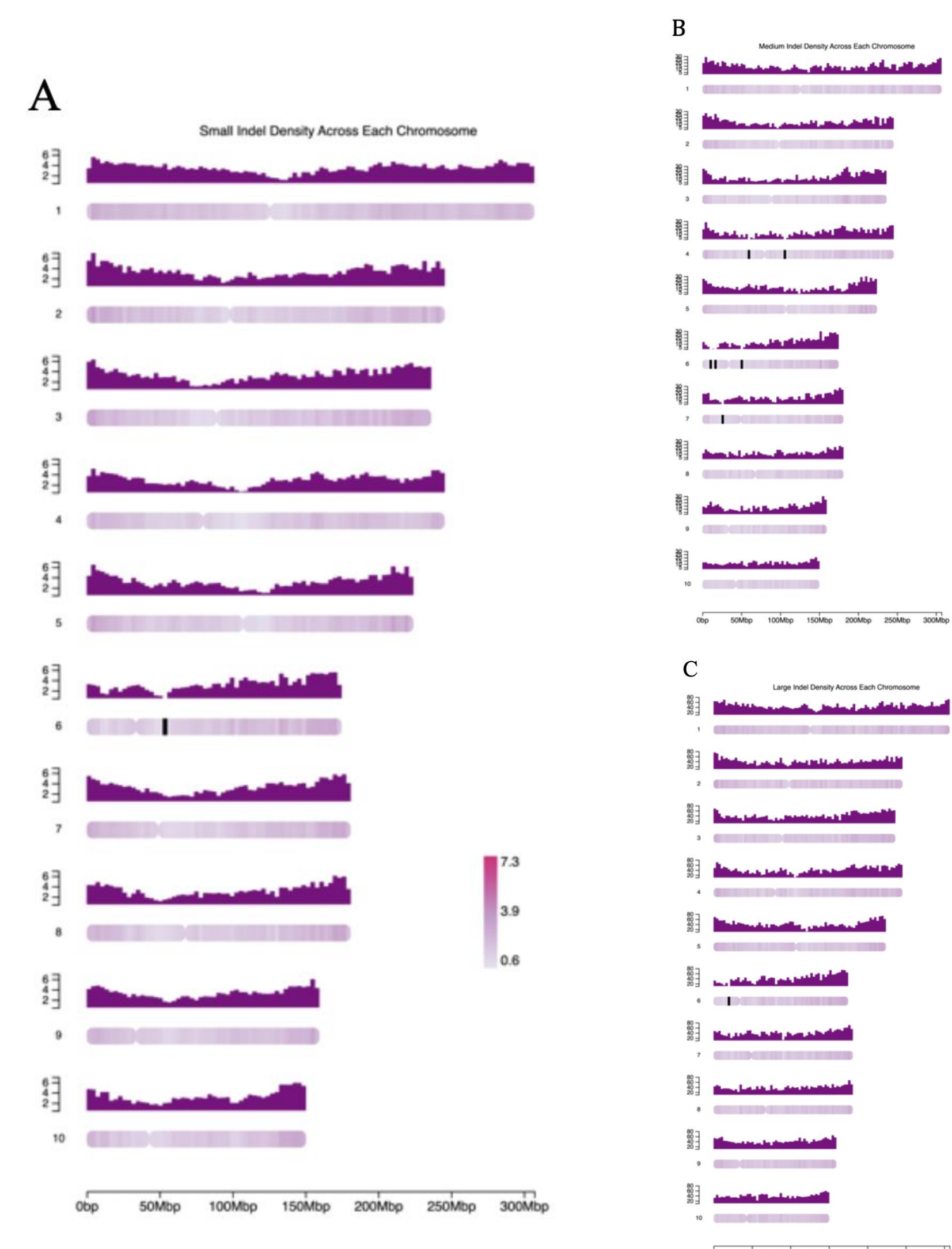
Figure 1: Indel density distribution across chromosomes 1-10. Density was calculated as the number of indels per megabase (Mb). Shading across chromosome indicates indel density. Small indel density scaled down 1000 fold (7.3 represents 7300 indels).

Analyzing Indel Presence at Recombination Sites
Used *bedtools intersect* to assess three measures:

$$\text{Indel overlap (\%)} = \frac{\text{\# of indels that intersect with each interval (CO/DSB)}}{\text{total \# of indels in dataset}} * 100\%$$

$$\text{Feature overlap (\%)} = \frac{\text{\# of features (CO/DSB) that intersect with indels}}{\text{total \# of features in dataset}} * 100\%$$

$$\text{Indel Density (indels/bp)} = \frac{\text{\# of indels that intersect with each feature}}{\text{size of feature (bp)}}$$

Indel size	Dataset	Percent of Indels in Site(s) (%)	Percent of Indels in Site(s) (-) (%)	P-value	Percent of Site(s) Containing Indel(s) (%)	Percent of Sites (-) Containing Indel(s) (%)	P-value
Small Indels	DSB	0.380	0.013	<2E-16	92.582	85.873	0.116
	COs	27.618	10.727	2.49E-9	93.710	89.874	9.51E-4
Medium Indels	DSB	0.359	0.261	0.0291	3.288	2.404	0.0119
	COs	36.136	12.485	6.12E-11	12.777	9.735	0.446
Large Indels	DSB	1.142	0.035	3.40E-16	21.78	24.04	0.709
	COs	0.00709	0.00335	1.79E-11	36.778	38.114	0.799

Table 1: Summary measures (percent of indels intersecting with recombination sites and percent of recombination sites intersecting with indels) for small (1-50bp), medium (100-500bp), and large (500bp-50kb) indels.

Results
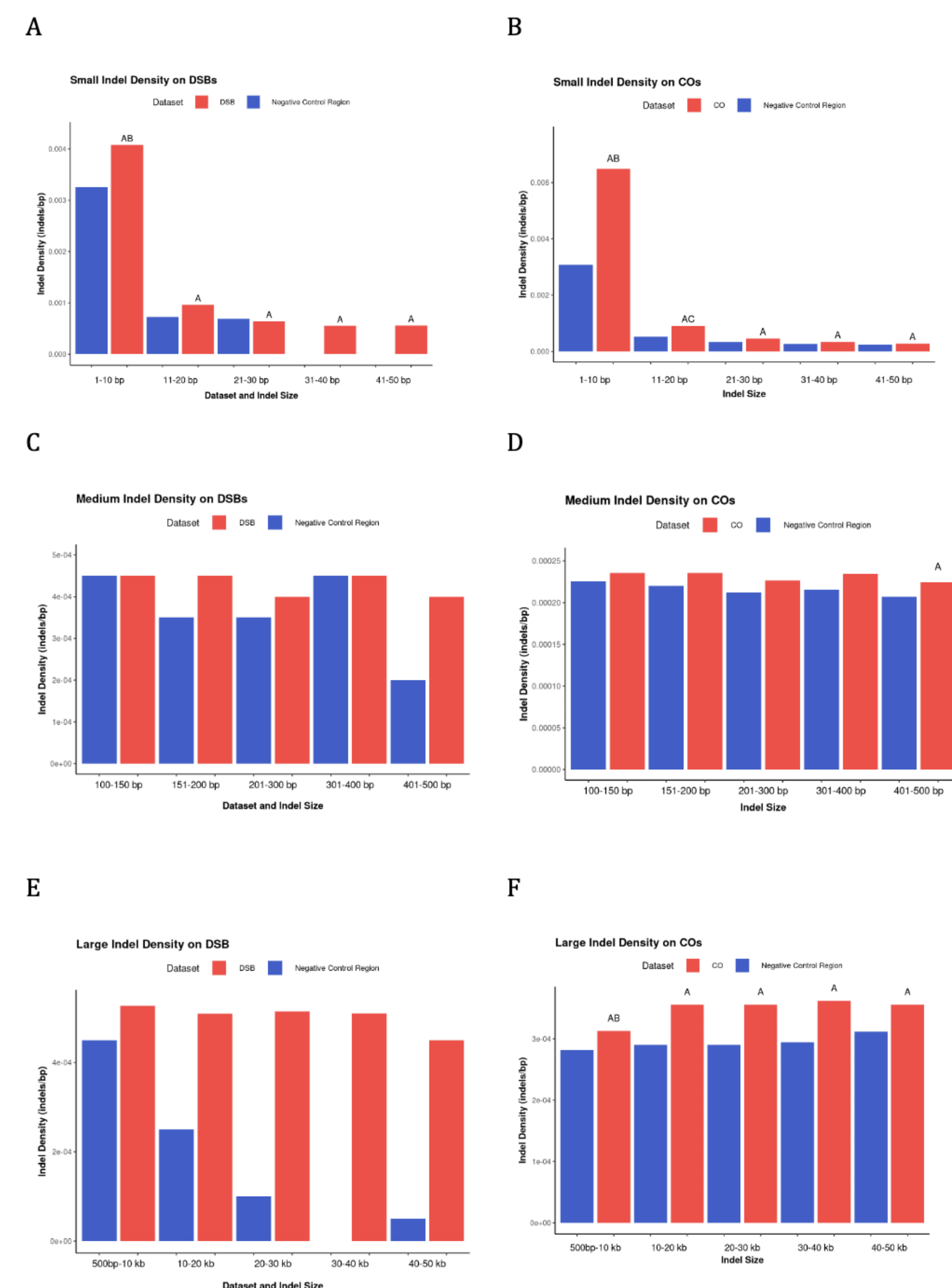
Figure 2: Comparison of indel density of different indel size groups at CO sites and negative control sites. Letters above the bars indicate a statistically significant group at *P* < 0.05. Different letters indicate different statistical groups (i.e. A represents difference between the CO sites and control)
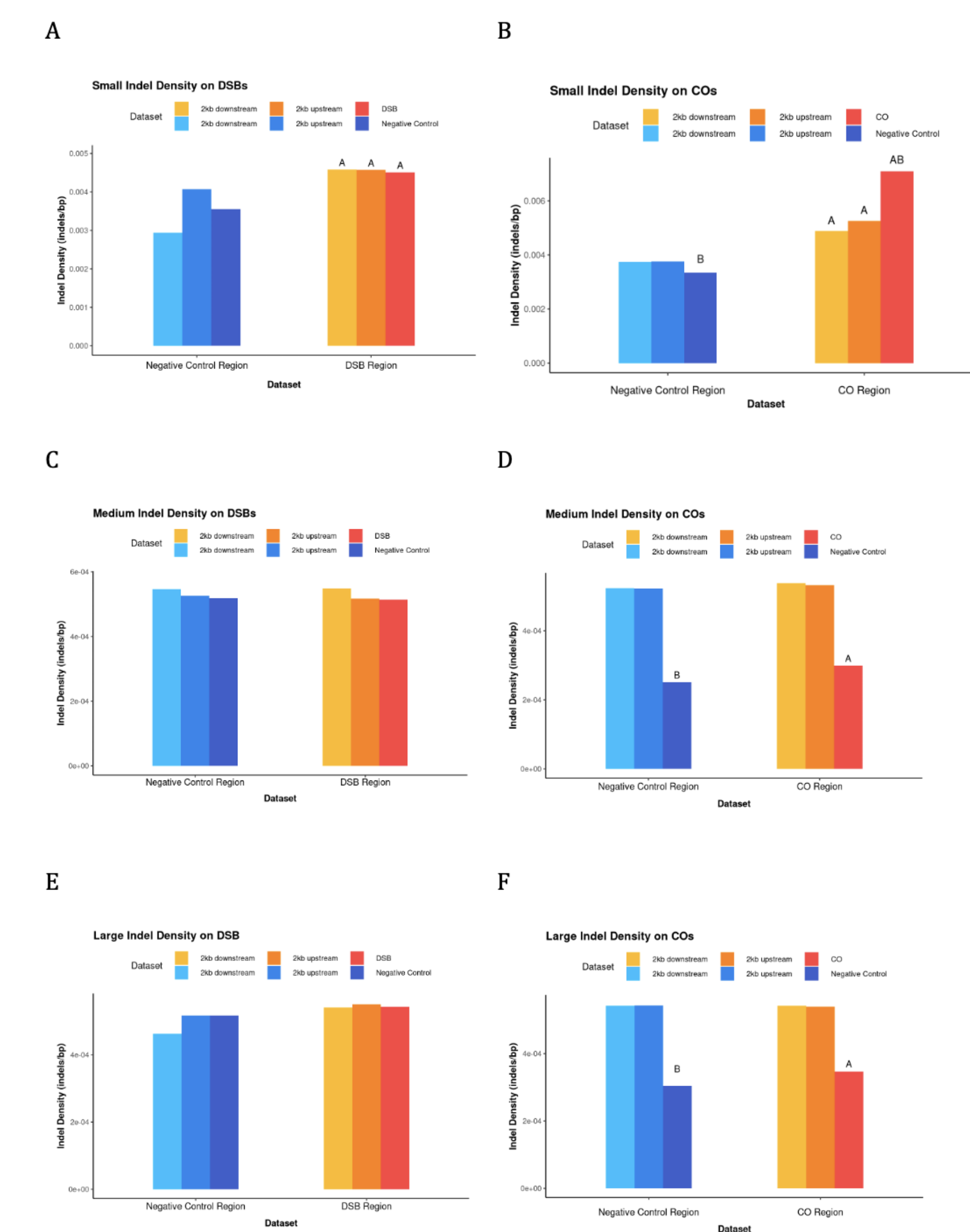
Figure 3: Comparison of small, medium, and large indel density at CO and negative control sites and their respective 2kb upstream and downstream regions. Letters above the bars indicate statistically significant differences at P < 0.05.

Conclusions

Where do indels lie?
- U-shaped pattern of indel density (Fig 1)

Is recombination mutagenic?
- Enrichment of small indels (1-10 bp) at recombination sites, indicating the mutagenic effect of recombination. (Fig 2)
  - More indels ate CO hotspot and predicted CO sites than in 2kb up- and downstream regions. (Fig 3)
- CO hotspots have the greatest increase in small indel density (Fig 4)

What fraction of indels are at recombination sites and what fraction of recombination sites overlap with indels?
- Fewer than 50% of all indels were located at recombination sites (Tb 1)
- Nearly all CO hotspots, COs, and DSBs intersected with at least one small indel

	CHG Methylation	P value for CHG site methylation	CHH Methylation	P value for CHH site methylation	CG Methylation	P value for CG site methylation
DSB hotspots with indels	49.469	<2e-16	1.414	0.0867	67.847	<2e-16
DSB hotspots without indels	70.271		1.152		85.406	
CO hotspots with indels	16.419	<2e-16	1.306	<2E-16	23.612	<2e-16
CO hotspots without indels	30.347		2.516		44.763	
COs with indels	39.057	<2e-16	6.559	<2e-16	54.923	<2e-16
COs without indels	72.919		1.481		88.410	
Predicted COs with Indels	23.768	<2e-16	1.719	2.7e-05	34.191	<2e-16
Predicted COs without indels	57.394		1.596		72.326	

Table 2: CG, CHG, and CHH methylation levels (%) at indels intersecting DSB hotspots, COs, CO hotspots, and predicted COs compared to recombination sites that do not intersect with indels.

What explains the elevated indel density in CO regions?
- Decrease in CG and CHG site methylation levels
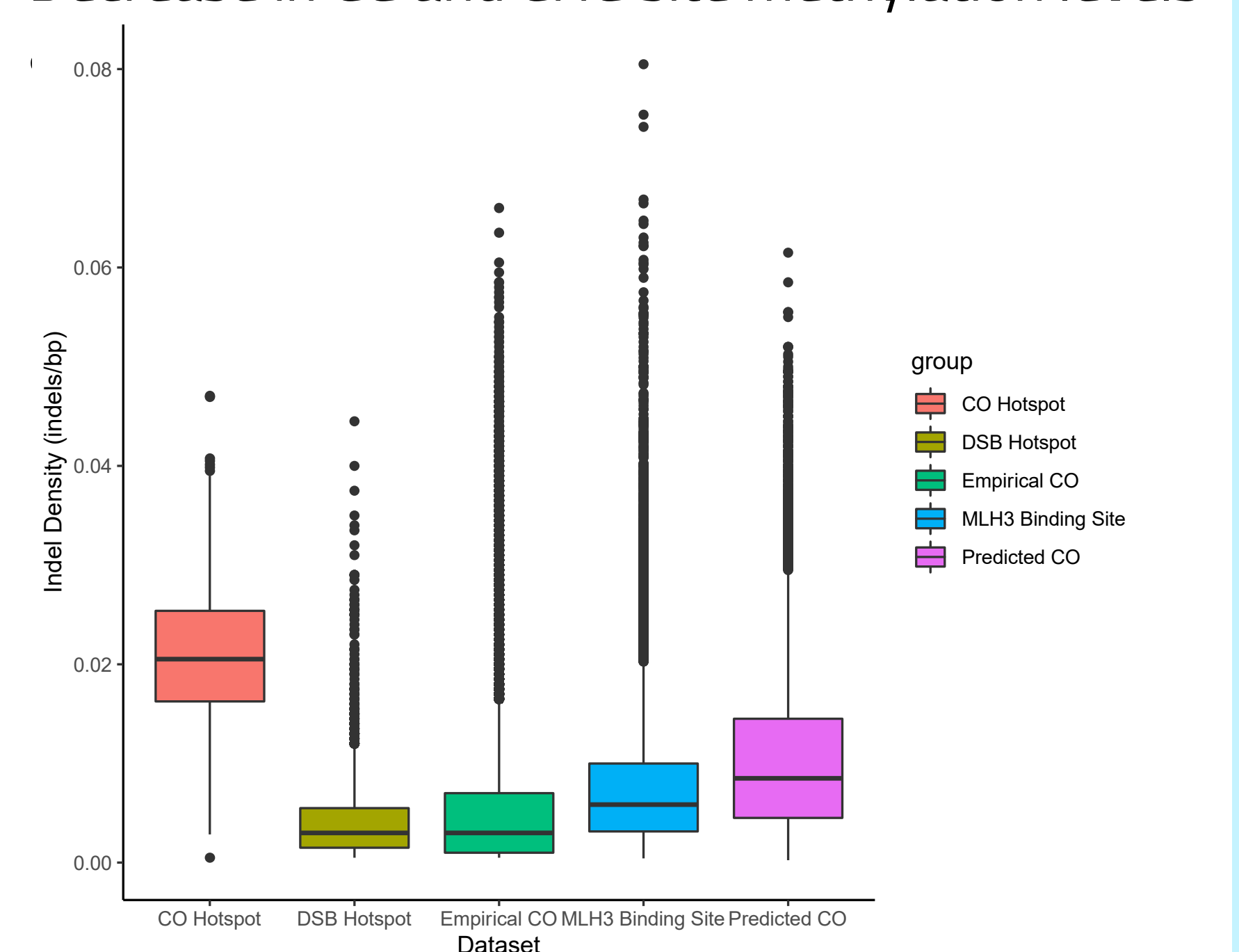
Figure 4: Summary of small indel density among all recombination sites, including DSB hotspots, empirical COs, predicted COs, CO hotspots, and MLH3 hotspots.

Acknowledgements