

# Estimation of the dependence between the proportion of zeros and distance between interacting regions

*John Stansfield, Mikhail Dozmorov*

## Contents

<b>Introduction</b>	<b>1</b>
<b>Proportion of zeros in individual matrices</b>	<b>1</b>
Chr 1 . . . . .	2
Chr 18 . . . . .	2
Chr 19 . . . . .	3
Summary . . . . .	3
<b>Proportion of zeros between matrices</b>	<b>3</b>
Chr 1 . . . . .	3
Chr 18 . . . . .	4
Chr 19 . . . . .	4
Summary . . . . .	4

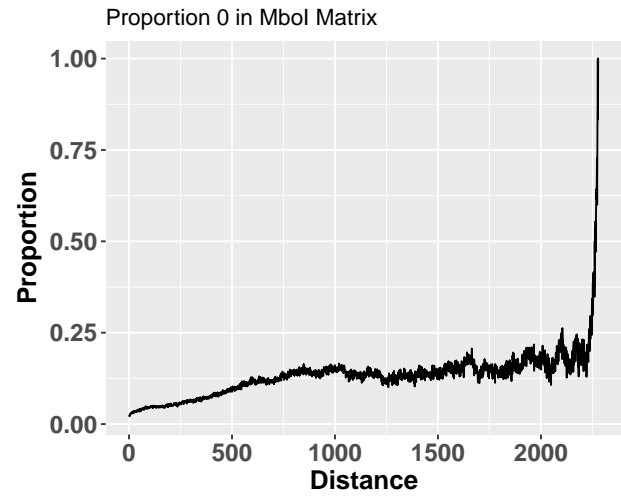
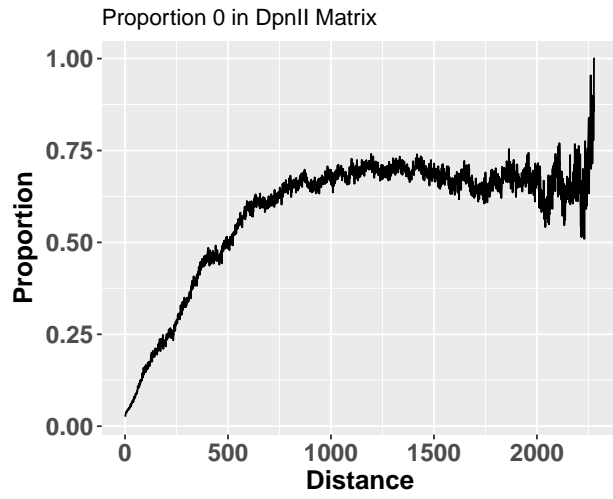
## Introduction

To estimate the distribution of the proportion of zeros vs. distance, real Hi-C data from Gm12878 cell line were used (Supplementary Table 1). The first dataset was obtained with the MboI restriction enzyme, while the second dataset was obtained with the DpnII enzyme. Data from chromosomes 1, 18, and 19 at 100kb resolution were used. The matrices were used in a sparse upper triangular matrix format (see `HiCcompare-vignette.Rmd` for details).

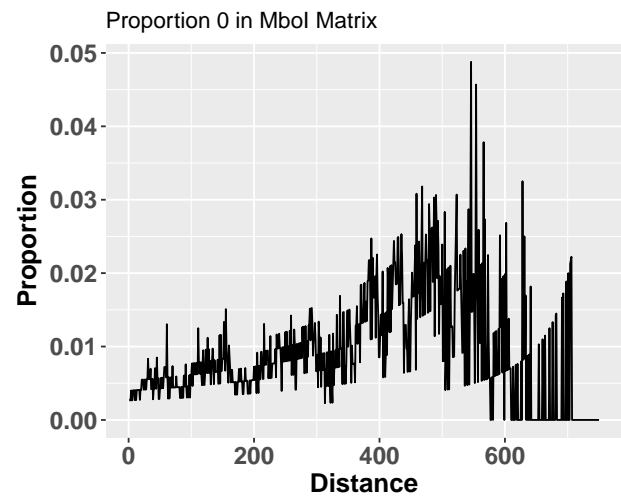
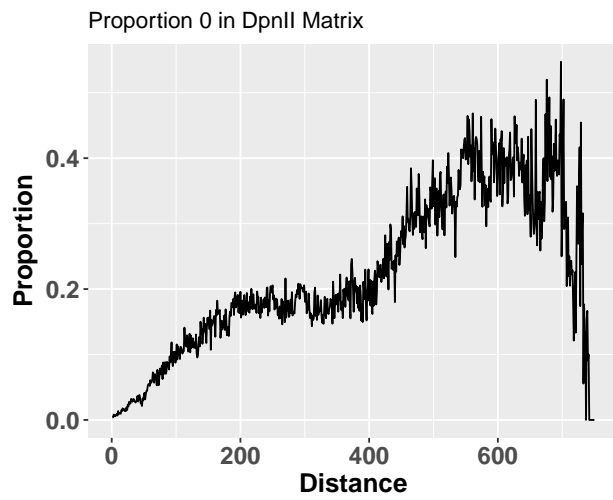
## Proportion of zeros in individual matrices

Here the proportion of zeros in each matrix are plotted for chromosomes 1, 18, and 19.

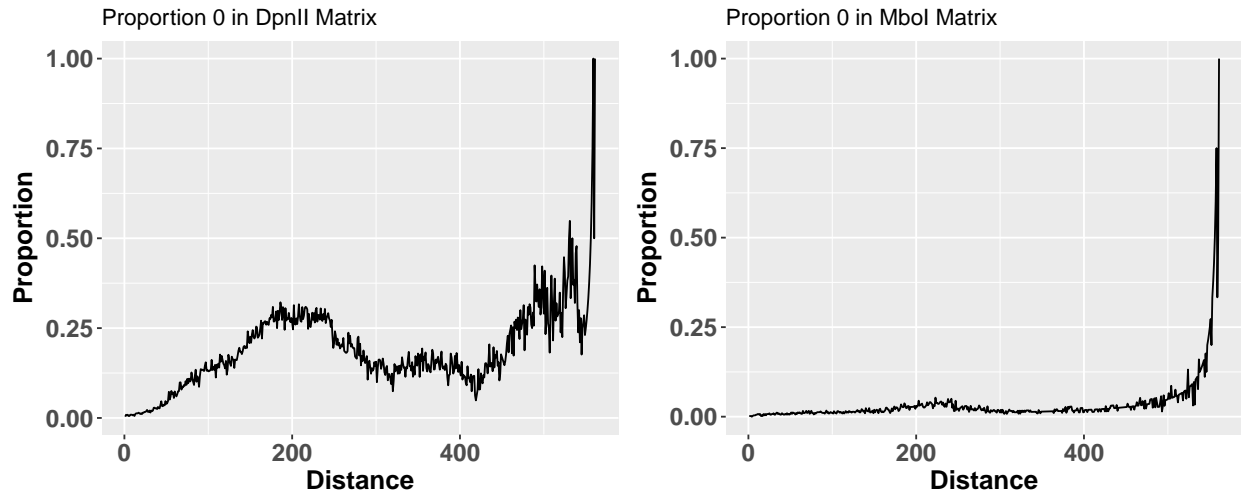
## Chr 1



## Chr 18



## Chr 19



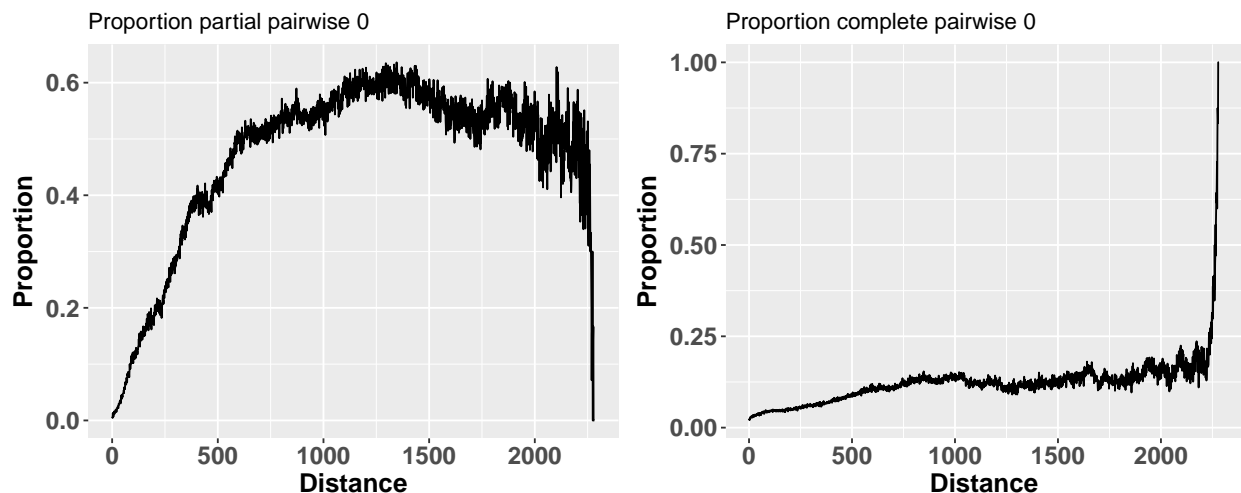
## Summary

The proportion of zeros in the individual matrices increases with distance. This increase does not seem to follow a consistent trend across chromosomes or restriction enzymes. The proportion of zeros for one unit distance compared to the next unit distance can vary considerably.

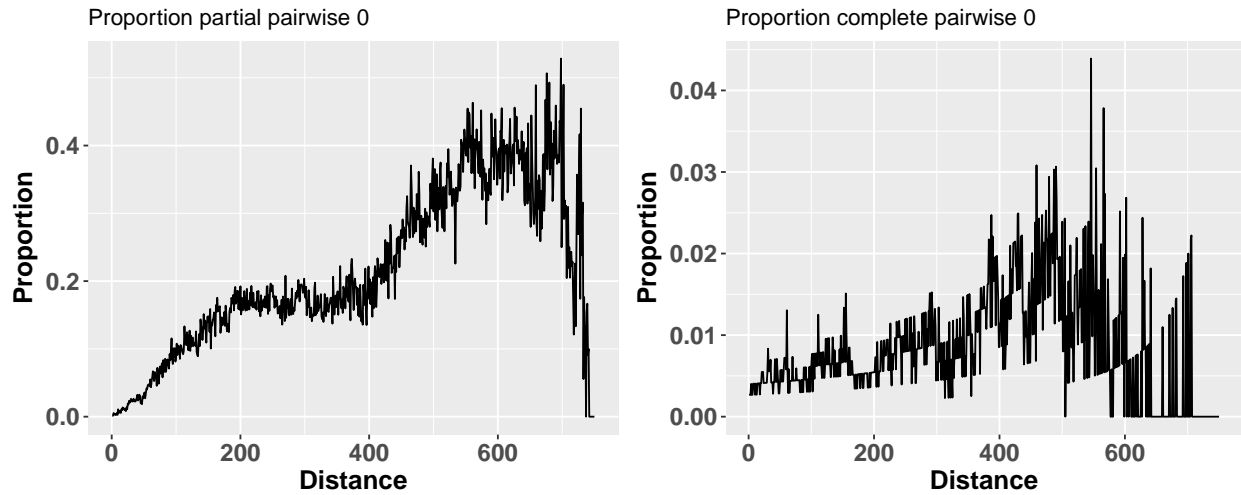
## Proportion of zeros between matrices

When representing matrices on an MD plot, some pairs of interaction frequencies may be completely zero (“complete zero pairs”), while some may have one zero (“partial pairwise zeros”). Distribution of such pairs across distances follows approximately the same pattern of zeros as in individual matrices. The data used here is the same as the above section, but now the pairwise zeros are being compared between the DpnII and MboI matrices.

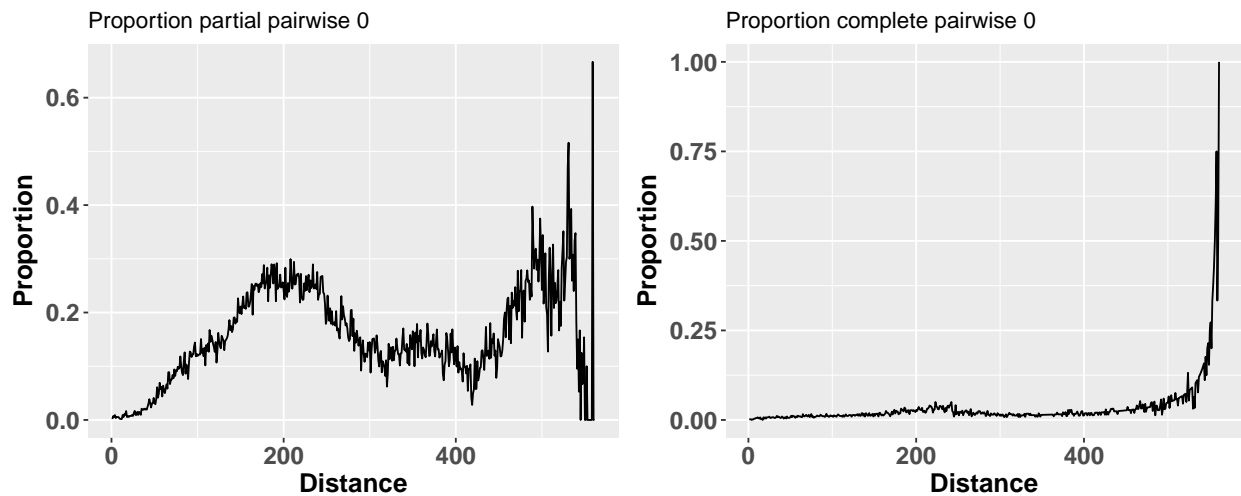
## Chr 1



## Chr 18



## Chr 19



## Summary

The proportion of partial pairwise zeros also tends to increase with distance until the furthest points of the matrices are reached at which point the proportion begins to drop. The partial pairwise zeros do not follow a consistent pattern of increase and decrease over the different chromosomes tested and there is a large amount of variance between unit distances. The proportion of complete pairwise zeros also tends to increase with distance and spike up towards the very furthest distances. This trend is again not very consistent in its pattern over the different chromosomes tested.