Preliminary exploration of geographic trends in *P. falciparum* relatedness on the Colombian coast

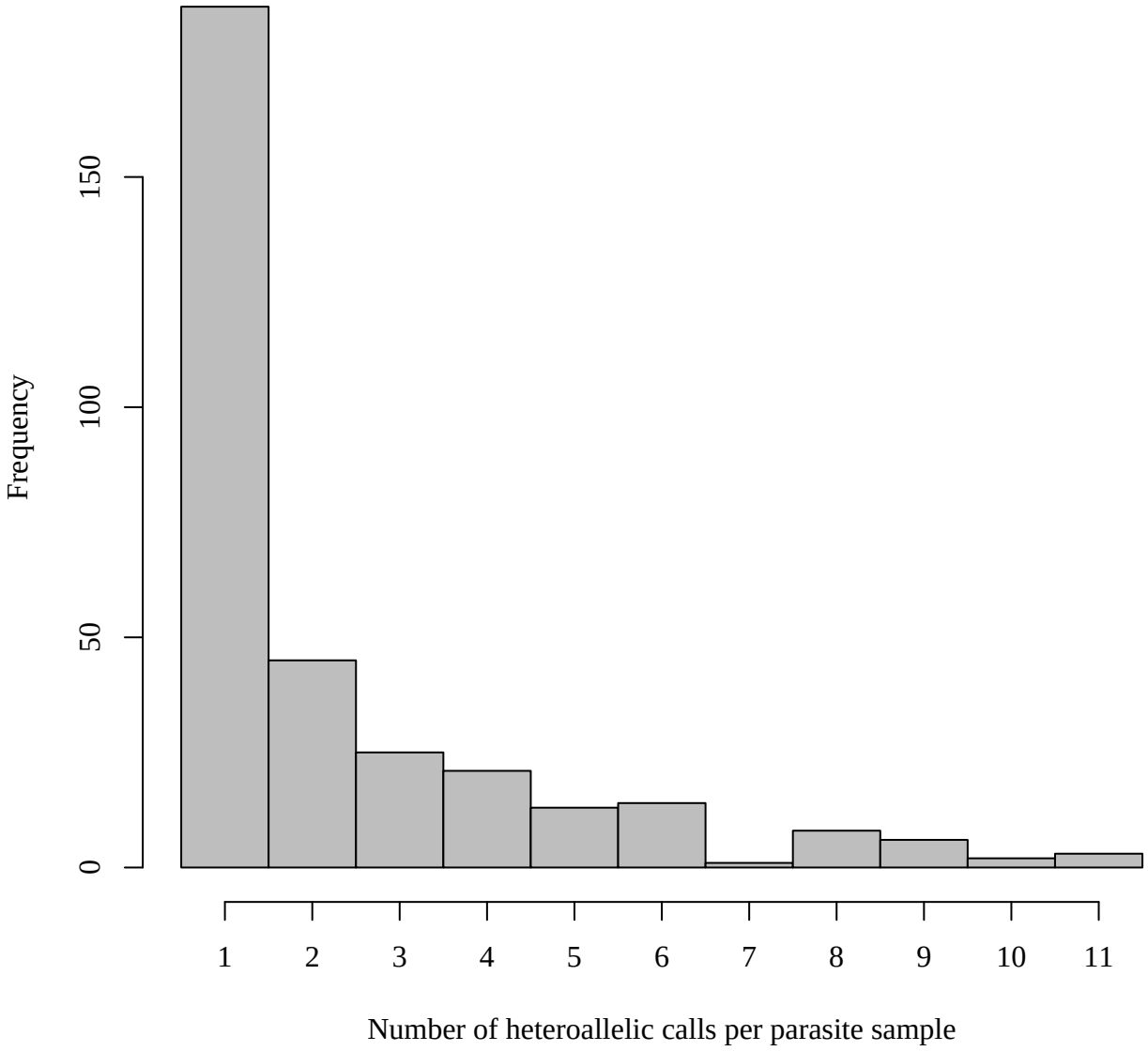
To-do list

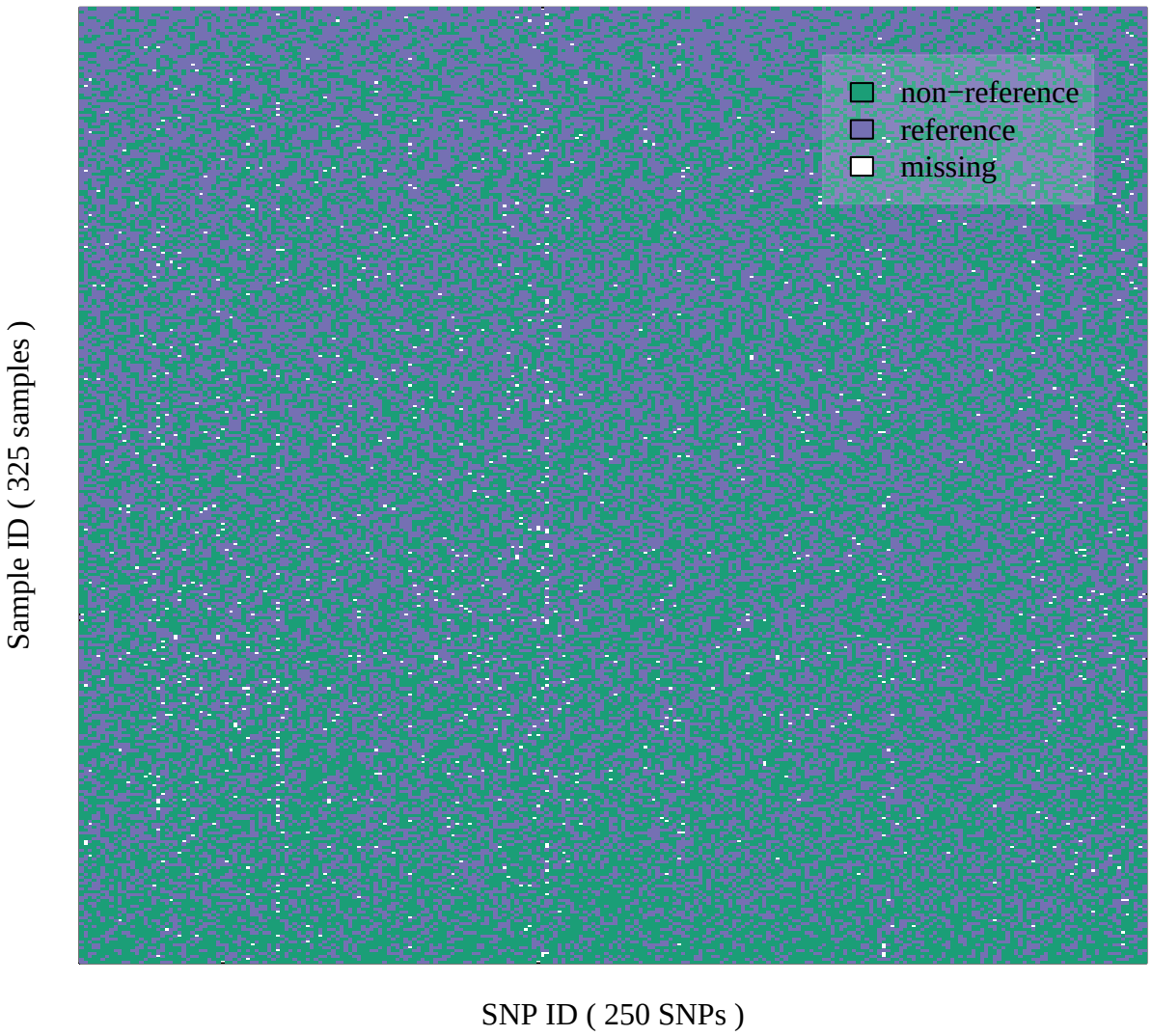
- All the 325 samples here are considered single-genotype (use all)
- Add fst analysis???
- put binned_residuals somewhere
- Why do I have only 250 SNPs not 307 as in XXX
- represent data by the 136 MLGs in the 325 samples
- check my results against the probabilities of sampling indentical MLGs over time intervals.
- What would happen if we built the tree based on IBD?
- p-values by perturbation

Introduction

In the following I have quantified the trend in parasite relatedness (based on identity by descent, IBD, estimated under a hidden Markov model, HMM, described elsewhere) with inter-city distance (km) and time between collection dates (weeks), with a view to assessing the potential for comparison with the Thai-Myanmar border, where, on average, the log-odds of relatedness decrease by 0.02 with every kilometer between collection sites and week between collection dates. I have treated calls labelled '–' as missing.

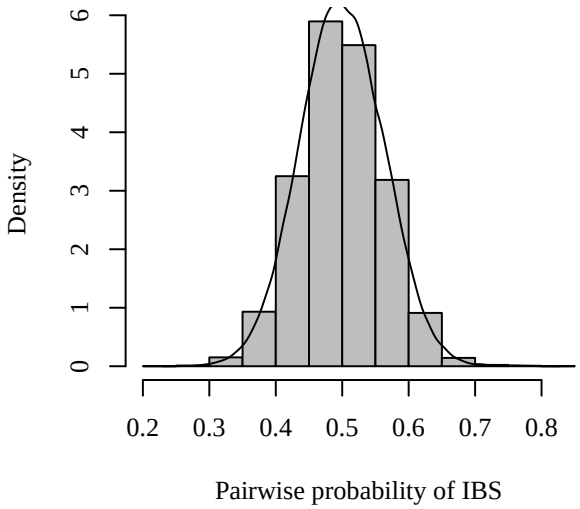
```
##
## Buenaventura        Guapi    Quibd\x97      Tad\x97         Tado
##           47           68           64           13            1
##       Tunaco
##          132
```

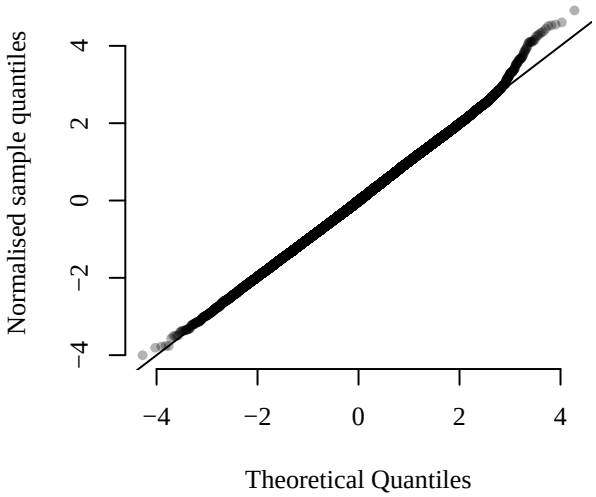

Frequency
0
50
100
150
1
2
3
4
5
6
7
8
9
10
11
Number of heteroallelic calls per parasite sample

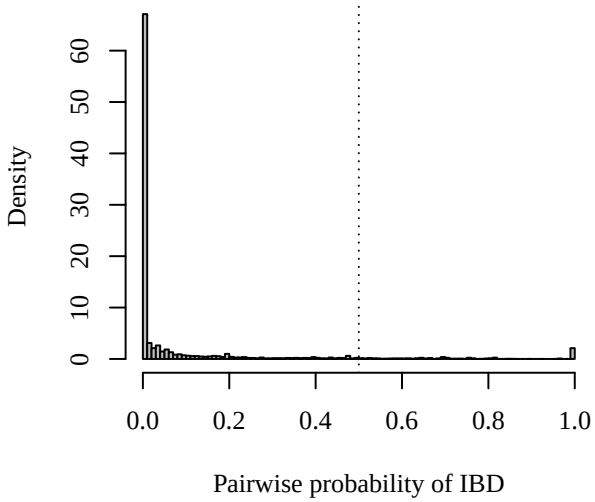


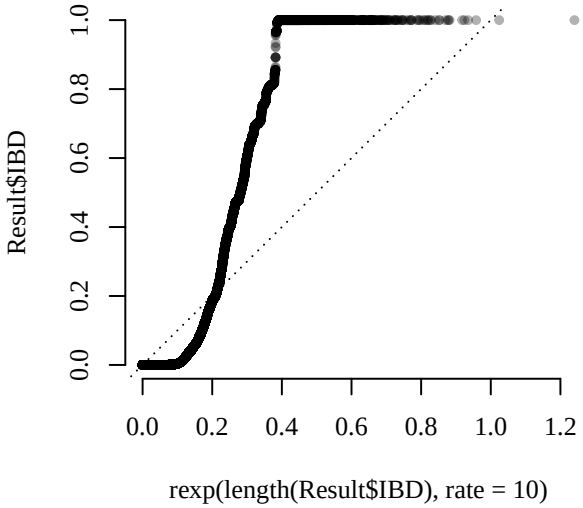
Pairwise identity by state and indentity by descent

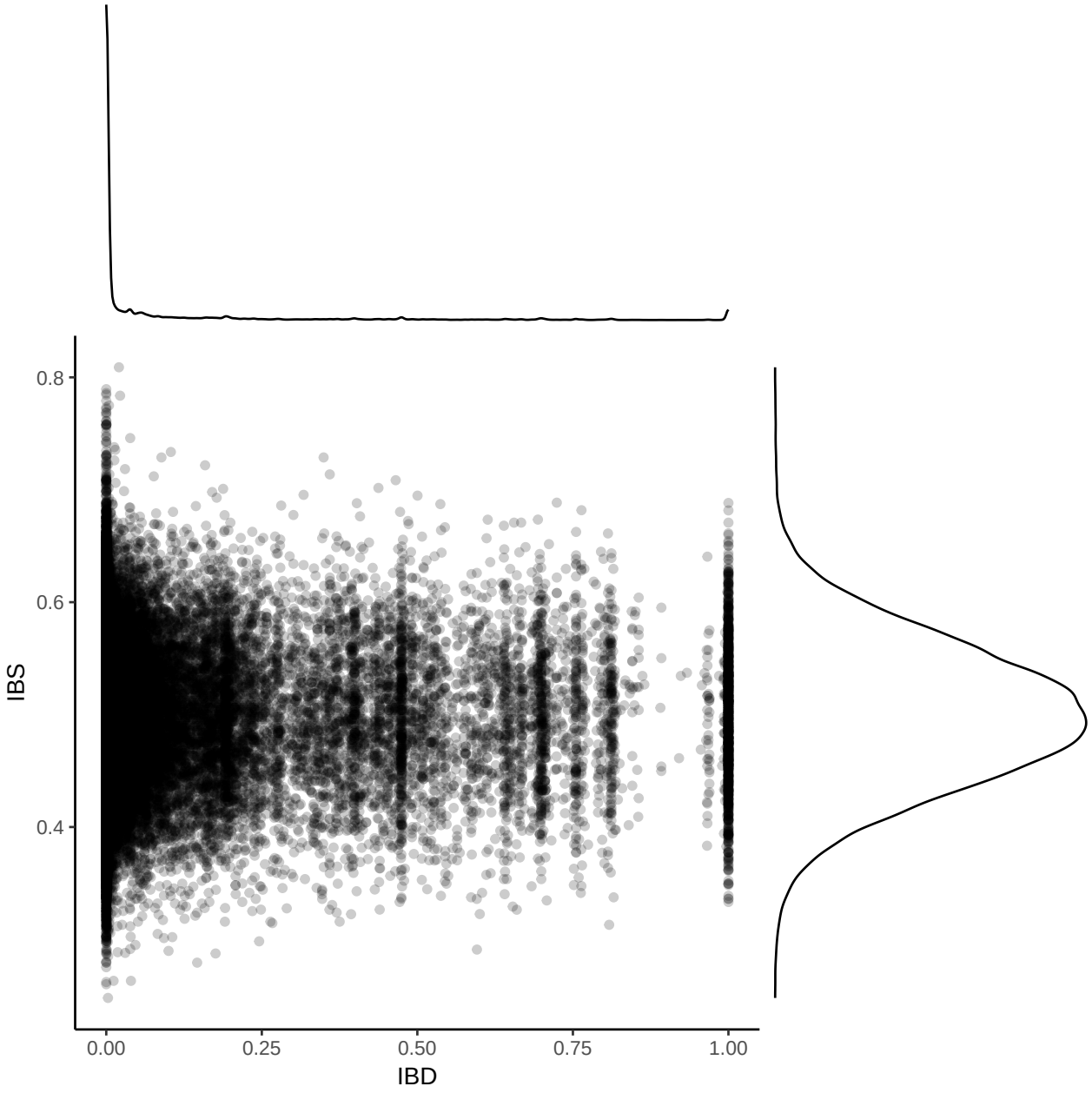
The empirical density of SNP-fraction identity by state (IBS), $\hat{\pi}_{\text{IBS}}$ for all pairwise sample comparisons appears to be Guassian, but with very slightly lighter tails on the left and heavier tails on the right. The empirical density of SNP-fraction IBD, $\hat{\pi}_{\text{IBD}}$ for all pairwise sample comparisons is positively skewed with some highly related parasites. They do not appear to be correlated.

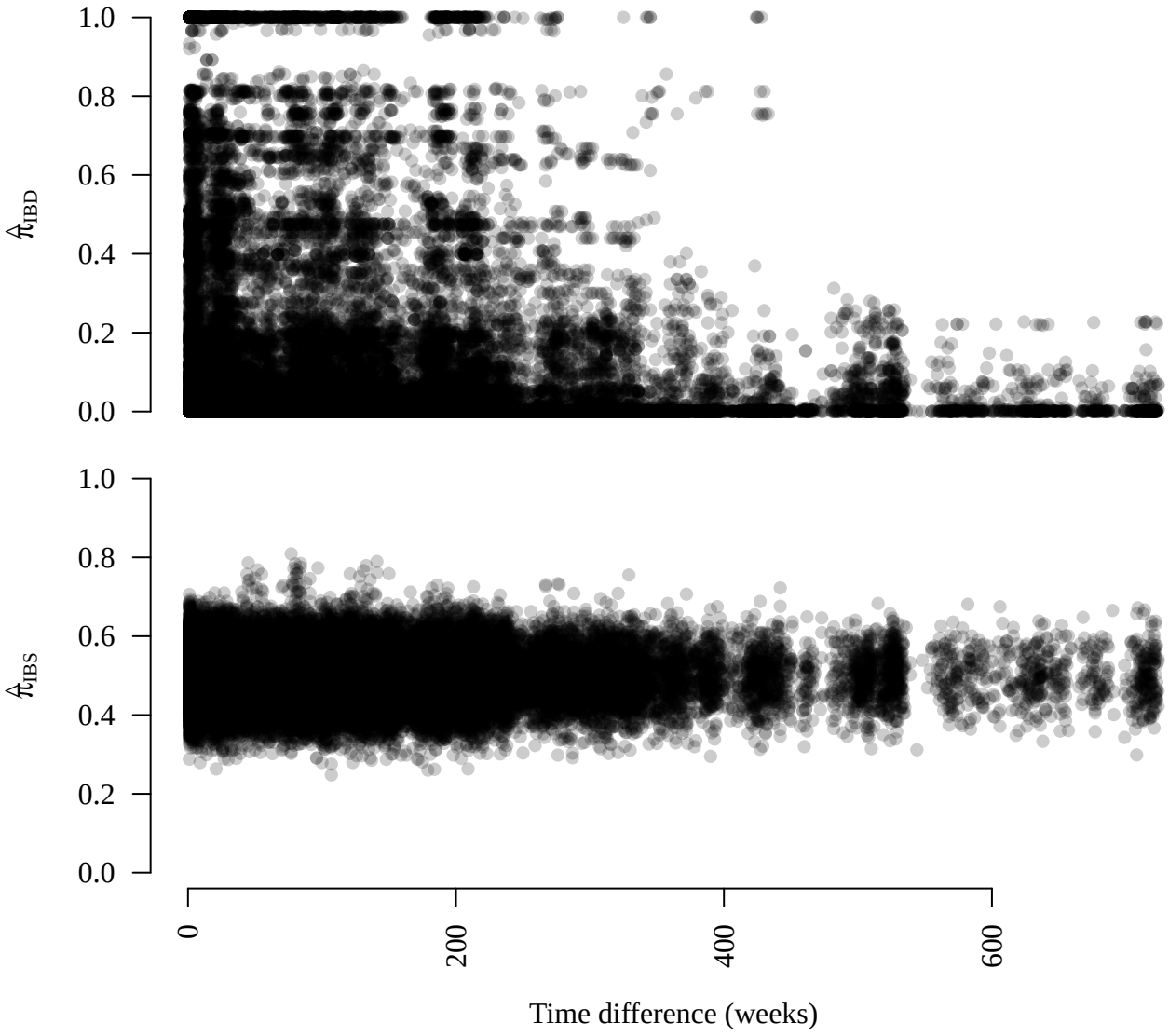

Density
0
1
2
3
4
5
6
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Pairwise probability of IBS

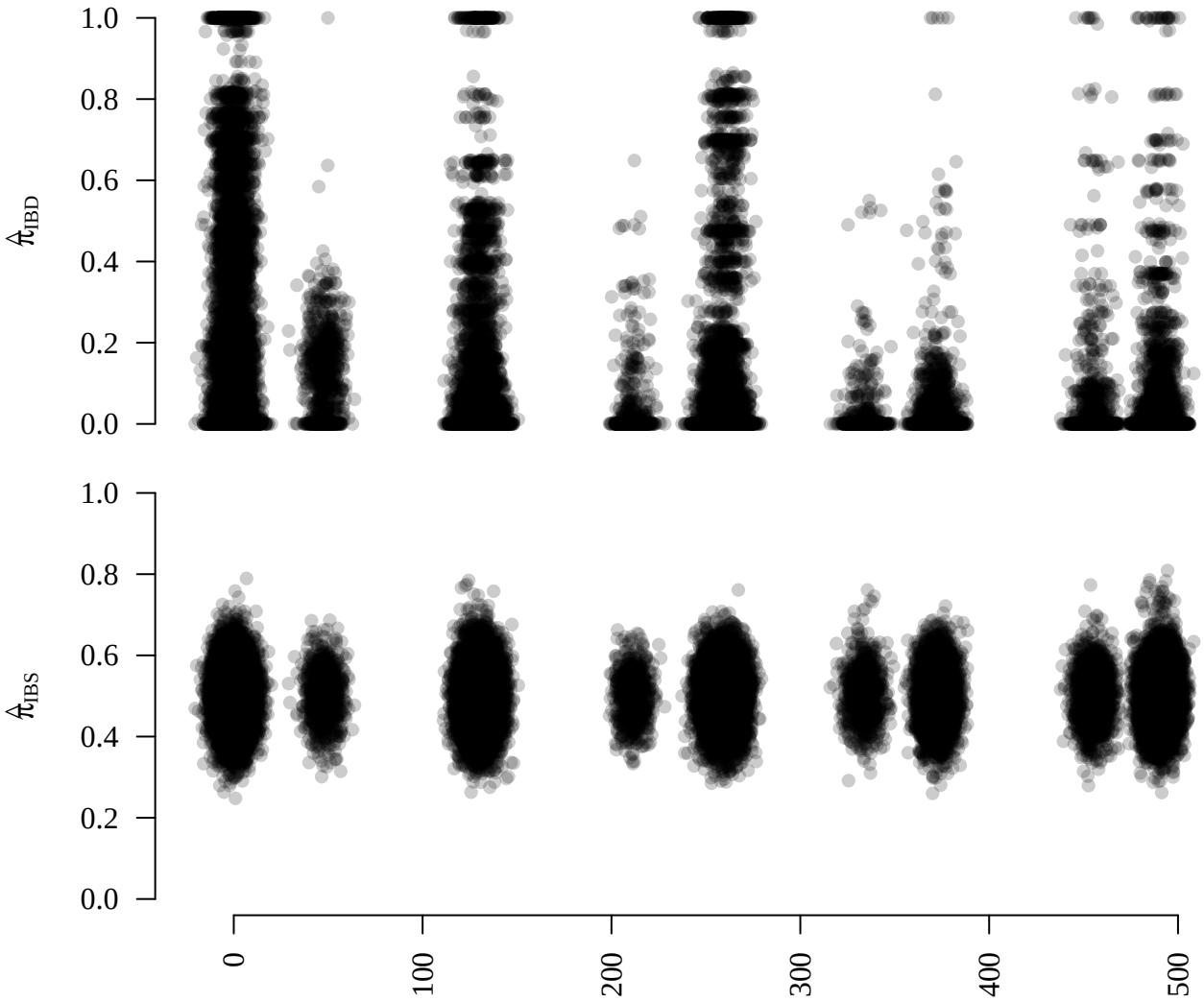

Normalised sample quantiles
−4
−2
0
2
4
−4
−2
0
2
4
Theoretical Quantiles


Density
0
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0
Pairwise probability of IBD


Result$IBD
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
rexp(length(Result$IBD), rate = 10)

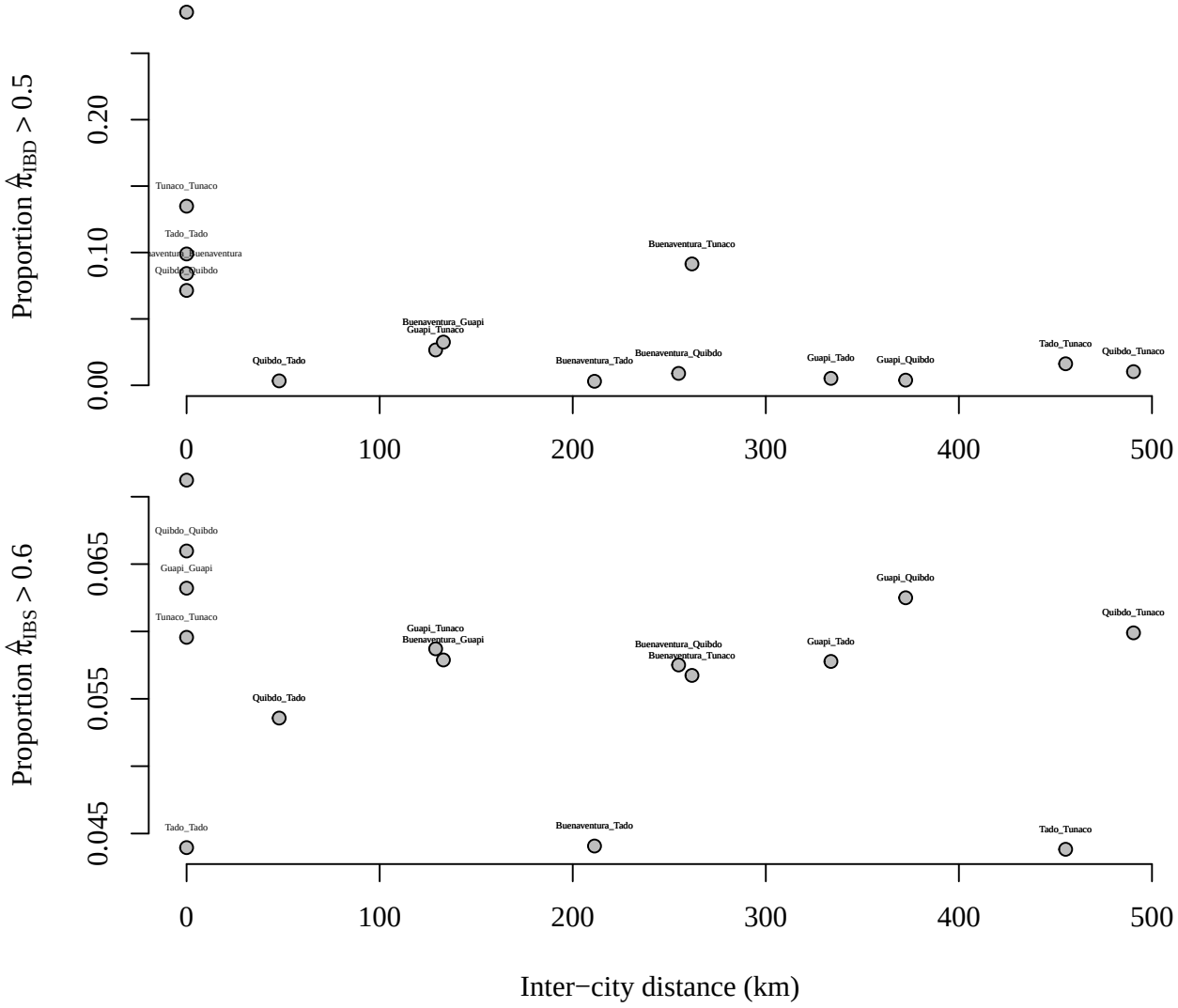

0.8
0.6
0.4
IBS
0.00
0.25
0.50
0.75
1.00
IBD

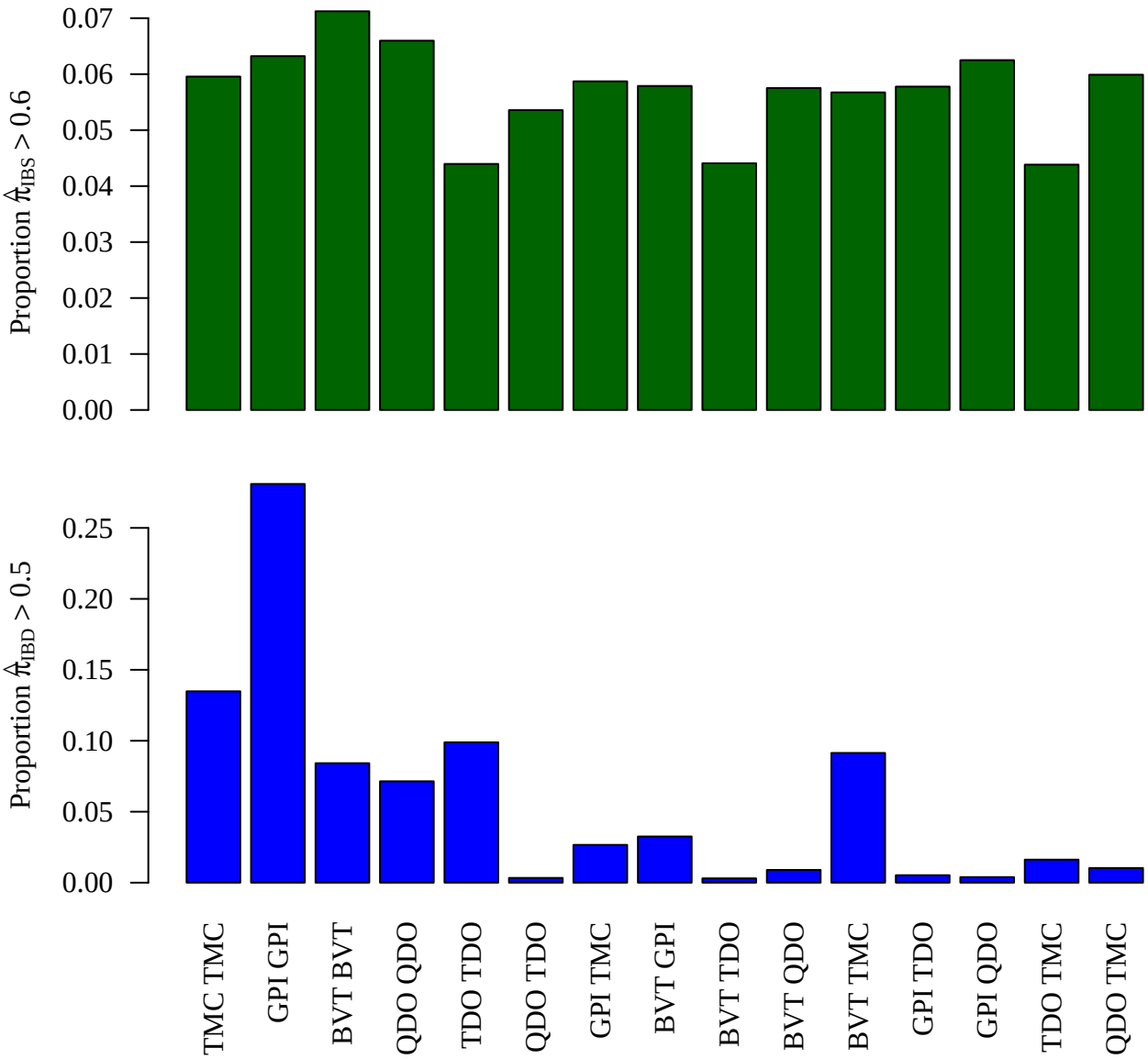

1.0
0.8
0.6
0.4
0.2
0.0
$\hat{\pi}_{IBD}$
1.0
0.8
0.6
0.4
0.2
0.0
$\hat{\pi}_{IBS}$
0
200
400
600
Time difference (weeks)

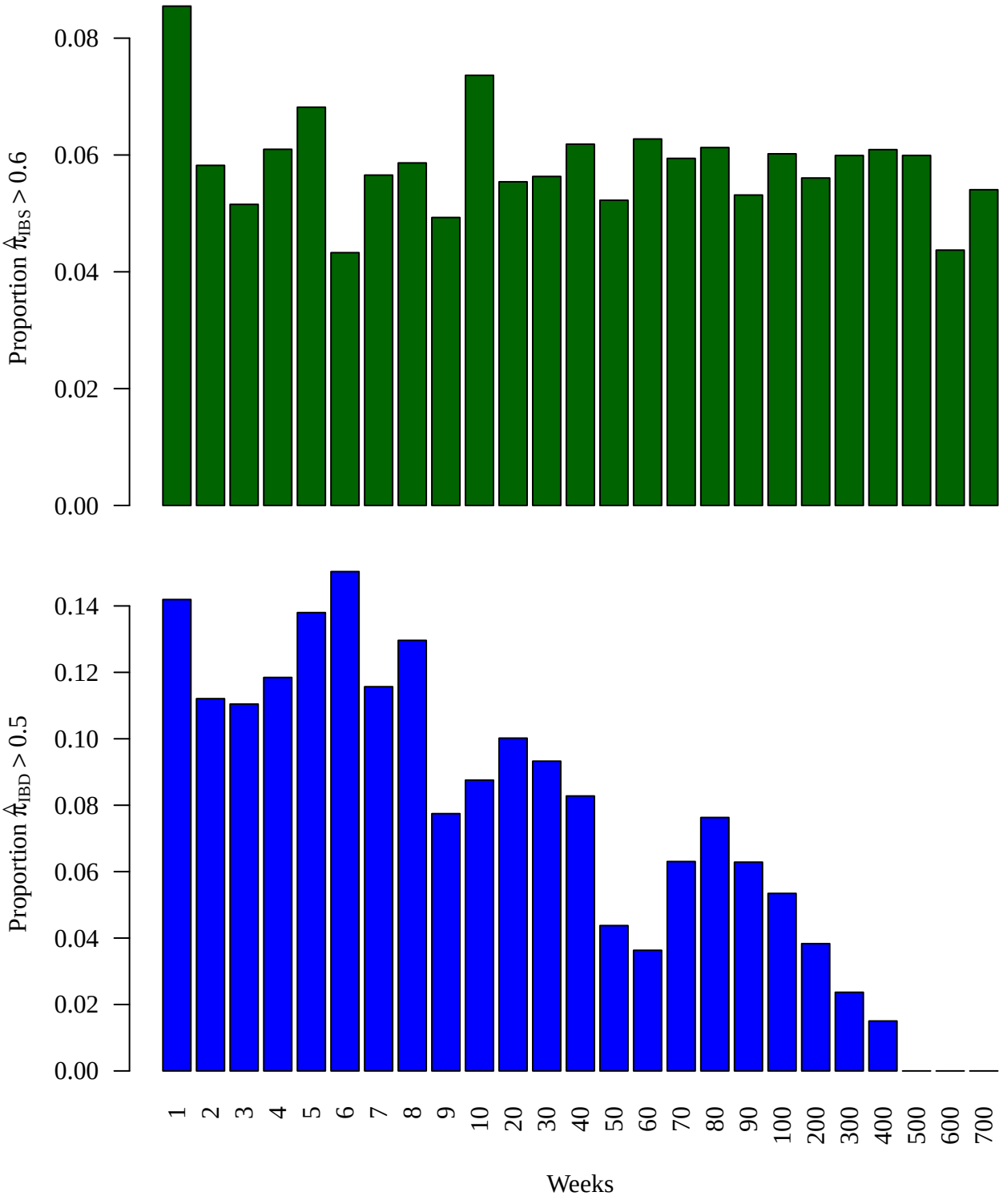


	M1_inter_IBD	M1_all_IBD	M2_all_IBD	M1_inter_IBS	M1_all_IBS	M2_all_IBS
(Intercept)	8.7e-02 (0.00)	1.6e-01 (0.00)	9.1e-02 (0.00)	5.0e-01 (0.00)	5.0e-01 (0.00)	5.0e-01 (0.00)
geo_dist	-1.1e-04 (0.00)	-2.8e-04 (0.00)	-1.1e-04 (0.00)	3.0e-06 (0.17)	2.9e-07 (0.85)	3.1e-06 (0.17)
time_dist	-5.3e-05 (0.00)	-1.4e-04 (0.00)	-8.9e-05 (0.00)	-3.5e-06 (0.19)	-4.6e-06 (0.05)	-4.1e-06 (0.08)
TunacoTRUE	NA	NA	8.6e-02 (0.00)	NA	NA	1.6e-03 (0.09)
GuapiTRUE	NA	NA	2.1e-01 (0.00)	NA	NA	2.2e-03 (0.15)
BuenaventuraTRUE	NA	NA	2.2e-02 (0.00)	NA	NA	7.3e-04 (0.72)
QuibdoTRUE	NA	NA	9.2e-02 (0.00)	NA	NA	7.7e-04 (0.62)
TadoTRUE	NA	NA	1.4e-01 (0.00)	NA	NA	6.5e-03 (0.33)

Propotions of highly related sample pairs




Proportion $\hat{\pi}_{IBS} > 0.6$
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Proportion $\hat{\pi}_{IBD} > 0.5$
0.25
0.20
0.15
0.10
0.05
0.00
TMC TMC
GPI GPI
BVT BVT
QDO QDO
TDO TDO
QDO TDO
GPI TMC
BVT GPI
BVT TDO
BVT QDO
BVT TMC
GPI TDO
GPI QDO
TDO TMC
QDO TMC



Regression

Before accounting for inter-state variability, regression coefficients suggest significant decrease of IBD with time and distance, but binned residuals suggest a poor fit:

	M1_inter_IBD	M1_all_IBD	M2_all_IBD	M1_inter_IBS	M1_all_IBS	M2_all_IBS

	M1_inter_IBD	M1_all_IBD	M2_all_IBD	M1_inter_IBS	M1_all_IBS	M2_all_IBS
(Intercept)	-2.61e+00 (0.000)	-1.57e+00 (0.000)	-2.62e+00 (0.000)	-2.80e+00 (0.000)	-2.73e+00 (0.000)	-2.78e+00 (0.000)
geo_dist	-1.92e-03 (0.000)	-5.27e-03 (0.000)	-1.94e-03 (0.000)	3.15e-05 (0.836)	-1.03e-04 (0.332)	4.36e-05 (0.774)
time_dist	-3.45e-03 (0.000)	-4.11e-03 (0.000)	-3.31e-03 (0.000)	2.89e-05 (0.874)	-1.39e-04 (0.380)	-1.55e-04 (0.334)
NA	NA	NA	NA	NA	NA	NA
TunacoTRUE	NA	NA	1.11e+00 (0.000)	NA	NA	4.00e-02 (0.549)
GuapiTRUE	NA	NA	1.76e+00 (0.000)	NA	NA	8.96e-02 (0.371)
BuenaventuraTRUE	NA	NA	7.02e-01 (0.000)	NA	NA	2.43e-01 (0.058)
QuibdoTRUE	NA	NA	4.18e-01 (0.000)	NA	NA	1.51e-01 (0.138)
TadoTRUE	NA	NA	4.64e-01 (0.194)	NA	NA	-2.96e-01 (0.565)

	geo_dist	time_dist
GLM_IBD1_all	-5.26e-03 (0.000)	-2.96e-03 (0.000)
GLM_IBD1_all	-5.30e-03 (0.000)	-3.31e-03 (0.000)
GLM_IBD1_all	-5.29e-03 (0.000)	-3.69e-03 (0.000)
GLM_IBD1_all	-5.47e-03 (0.000)	-3.61e-03 (0.000)
GLM_IBD1_all	-5.15e-03 (0.000)	-3.82e-03 (0.000)
GLM_IBD1_all	-5.27e-03 (0.000)	-4.11e-03 (0.000)
GLM_IBD1_all	-5.18e-03 (0.000)	-3.95e-03 (0.000)
GLM_IBD1_all	-5.11e-03 (0.000)	-3.61e-03 (0.000)
GLM_IBD1_all	-5.14e-03 (0.000)	-4.06e-03 (0.000)
GLM_IBD1_all	-5.40e-03 (0.000)	-4.29e-03 (0.000)
GLM_IBD1_all	-5.03e-03 (0.000)	-3.78e-03 (0.000)
GLM_IBS1_all	-1.07e-04 (0.183)	-1.58e-04 (0.188)
GLM_IBS1_all	-9.14e-05 (0.272)	-1.64e-04 (0.188)
GLM_IBS1_all	-1.20e-04 (0.163)	-1.61e-04 (0.213)
GLM_IBS1_all	-8.70e-05 (0.343)	-1.91e-04 (0.167)
GLM_IBS1_all	-5.93e-05 (0.538)	-2.13e-04 (0.142)
GLM_IBS1_all	-1.03e-04 (0.332)	-1.39e-04 (0.380)
GLM_IBS1_all	-1.06e-04 (0.341)	-2.67e-04 (0.115)
GLM_IBS1_all	-6.73e-05 (0.561)	-3.98e-04 (0.025)
GLM_IBS1_all	-4.65e-06 (0.970)	-4.05e-04 (0.033)
GLM_IBS1_all	-3.20e-05 (0.813)	-5.55e-04 (0.009)
GLM_IBS1_all	-7.03e-05 (0.625)	-3.81e-04 (0.084)

	geo_dist	time_dist
GLM_IBD2_all	-2.67e-03 (0.000)	-2.18e-03 (0.000)
GLM_IBD2_all	-2.46e-03 (0.000)	-2.52e-03 (0.000)
GLM_IBD2_all	-2.17e-03 (0.000)	-2.89e-03 (0.000)
GLM_IBD2_all	-2.30e-03 (0.000)	-2.80e-03 (0.000)
GLM_IBD2_all	-2.00e-03 (0.000)	-3.08e-03 (0.000)
GLM_IBD2_all	-1.94e-03 (0.000)	-3.31e-03 (0.000)
GLM_IBD2_all	-1.42e-03 (0.000)	-3.29e-03 (0.000)
GLM_IBD2_all	-1.67e-03 (0.000)	-3.21e-03 (0.000)
GLM_IBD2_all	-1.24e-03 (0.000)	-3.67e-03 (0.000)
GLM_IBD2_all	-1.71e-03 (0.000)	-3.64e-03 (0.000)
GLM_IBD2_all	-1.75e-03 (0.000)	-3.54e-03 (0.000)
GLM_IBS2_all	2.97e-05 (0.797)	-1.56e-04 (0.202)
GLM_IBS2_all	6.46e-05 (0.588)	-1.60e-04 (0.207)
GLM_IBS2_all	5.71e-05 (0.645)	-1.51e-04 (0.250)
GLM_IBS2_all	6.58e-05 (0.617)	-1.89e-04 (0.177)
GLM_IBS2_all	4.90e-05 (0.722)	-2.19e-04 (0.137)
GLM_IBS2_all	4.36e-05 (0.774)	-1.55e-04 (0.334)
GLM_IBS2_all	5.51e-05 (0.731)	-2.91e-04 (0.090)
GLM_IBS2_all	9.48e-05 (0.568)	-4.36e-04 (0.015)
GLM_IBS2_all	7.86e-05 (0.656)	-4.43e-04 (0.021)
GLM_IBS2_all	-1.78e-05 (0.927)	-6.02e-04 (0.005)
GLM_IBS2_all	-6.95e-06 (0.973)	-4.11e-04 (0.067)

Conclusion

In summary, there is evidence of a spatial and temporal trends in these data. The spatial trend is driven by within and inter city differences, and so dissapears when within site IBD estimates are removed. Differences are not recovered using IBS in lieu of IBD. However, ChromoPainter, does recover a similar trend?

Future work

Poster