# Correlation between numeric variables

The Pearson’s correlation coefficient was used to determine the correlation between two quantitative variables. The following table shows the small, medium and large correlations between the genetic factors and numeric clinical factors. The genetic factors are ordered from higher to lower correlation values. We can see that large correlations were not detected, except in the case of PCA3exp with RNU11. All scatter plots can be consulted in the folder “Correlations”. The scatter plots did not show good relations…

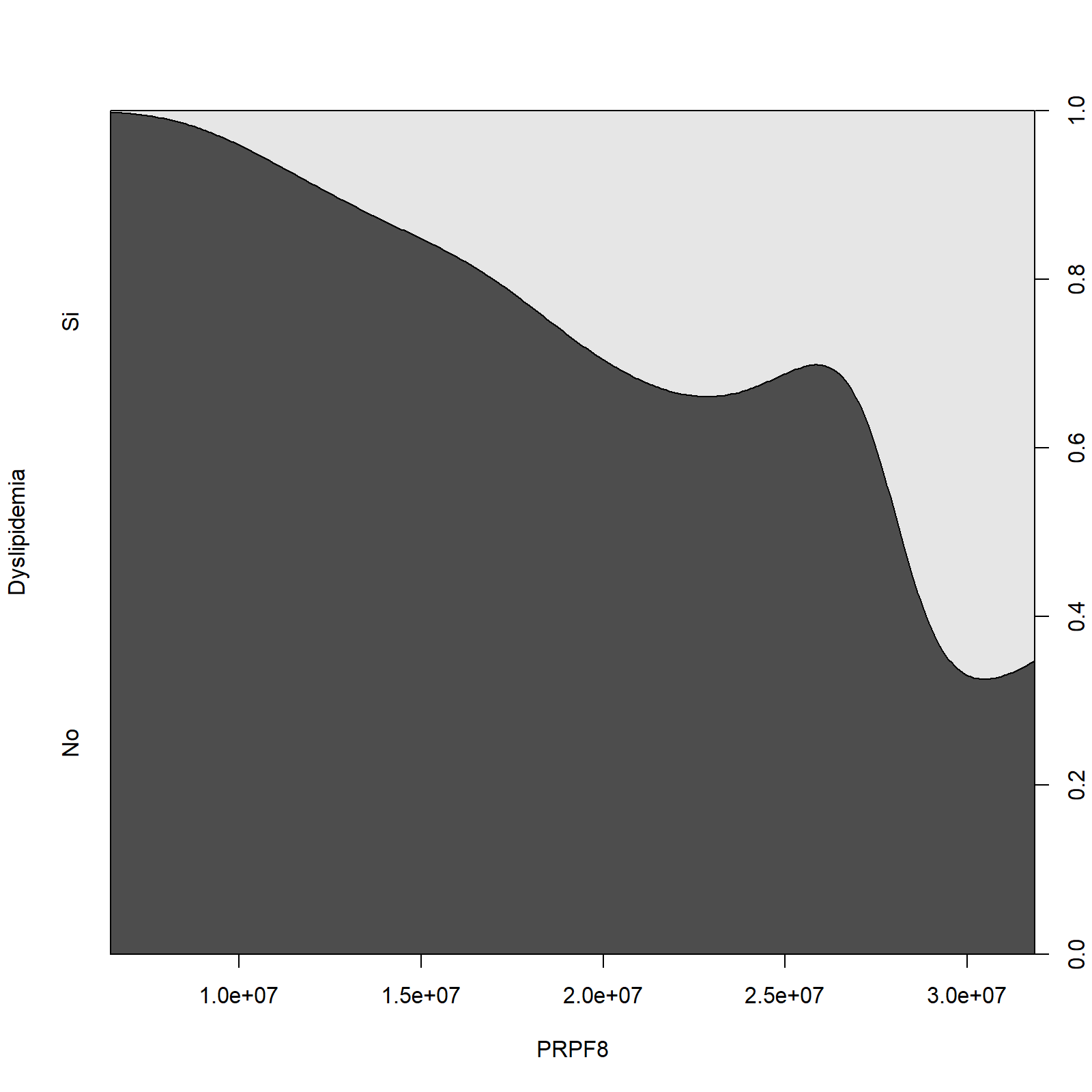
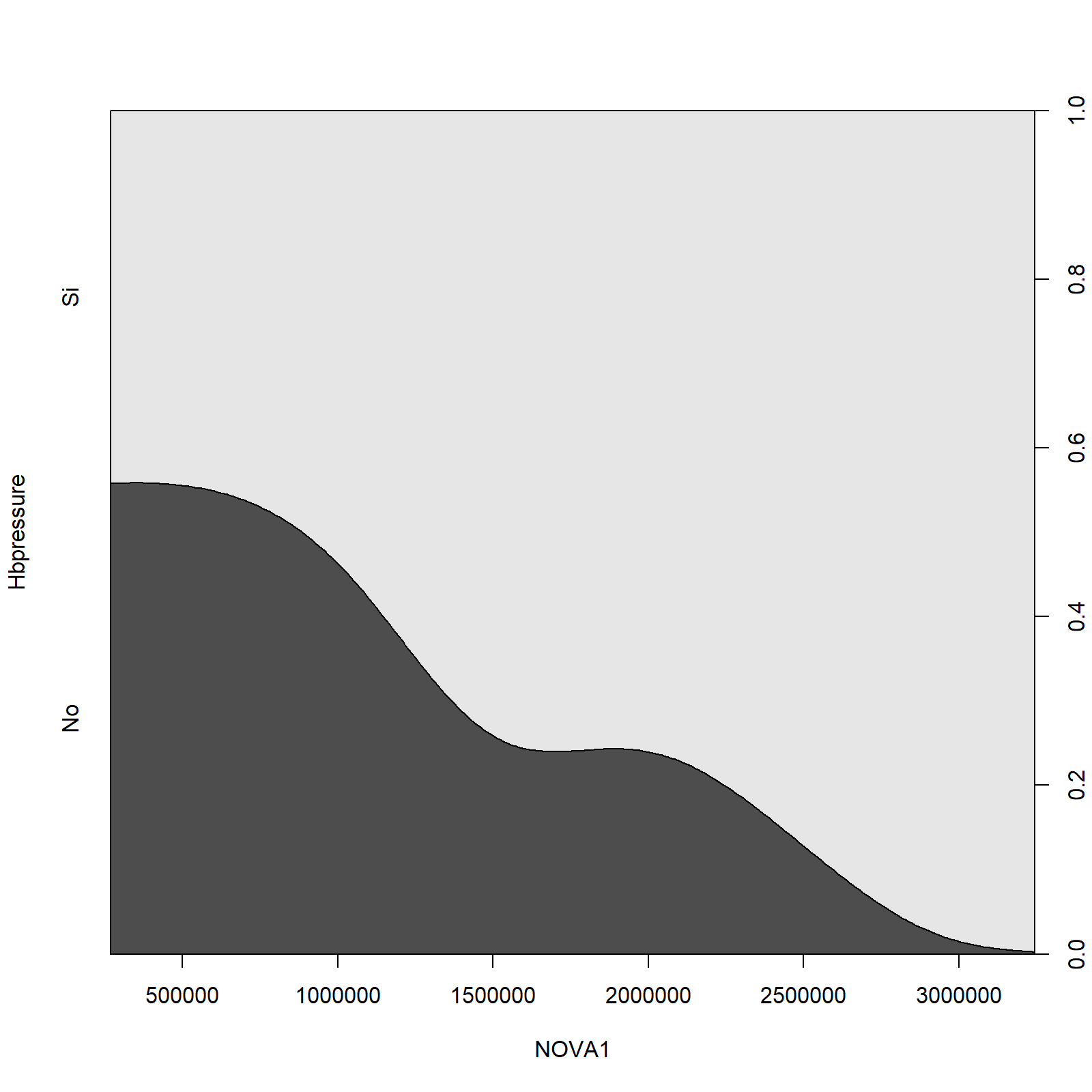
|  |  |  |  |
| --- | --- | --- | --- |
| **Clinical factor** | **Correlations** | | |
| **Small (.1 to .3 , -0.1 to -0.3)** | **Medium (.3 to .5 -0.3 to -0.5)** | **Large (.5 to 1.0 -0.5 to -1.0)** |
| BMI | Positive: RBM22, SRSF3,PRPF8, RBM3, SF3B1,PRPF40A, NOVA1,RAVER1, SRRM1  Negative:U2AF2, snRNP200, MAGOH, U4ATAC, RNU12 | Positive: SF3B1tv1  Negative: SFPQ |  |
| Age | Positive: SRRM1, RNU11, SRSF3, SF3B1tv1  Negative: RAVER1, SRSF6, RBM3 | Positive: RNU12, SF3B1 |  |
| Gleason score | Positive: SRRM4, U4ATAC, RAVER1, snRNP200, MAGOH, KHDRSB1, PRPF40A, RBM3  Negative: SF3B1, SF3B1tv1, RNU12 |  |  |
| PSA | Positive: snRNP200, U2AF2, RBM3, SRSF6  Negative: U4ATAC | Positive: RNU12, |  |
| PSAexp | Positive: SFPQ, RBM3, SF3B1, U4ATAC, KHDRSB1, PRPF8  Negative: PRPF40A , RAVER1 | Positive: NOVA1, SRRM1, SRSF3  Negative: SRRM4 |  |
| PCA3 exp | Positive: RNU12, SRSF6, U2AF2, KHDRSB1, PRPF8  Negative: PRPF40A, SRRM4, SFPQ | Positive: SF3B1tv1, U4ATAC, snRNP200 | Positive: RNU11 |
| sst5TMD4exp | Positive: KHDRSB1, RNU12, SRRM1, NOVA1, RBM22, SRSF3, RBM3  Negative: PRPF40A SFPQ | Positive: RNU11, SRRM4, PRPF8, U2AF2, snRNP200, U4ATAC |  |
| In1Ghrelinexp | Positive: SFPQ, RNU12, SRRM4, PRPF40A, MAGOH, KHDRSB1  Negative: SF3B1tv1, snRNP200, RNU11, U2AF2, SRSF6, RBM22 | Positive: SRSF3, RBM3, |  |
| Arexp | Negative: SRRM4,SRSF3, MAGOH,U2AF2, RNU11, KHDRSB1,SF3B1, snRNP200, PRPF40A | Negative: U4ATAC |  |

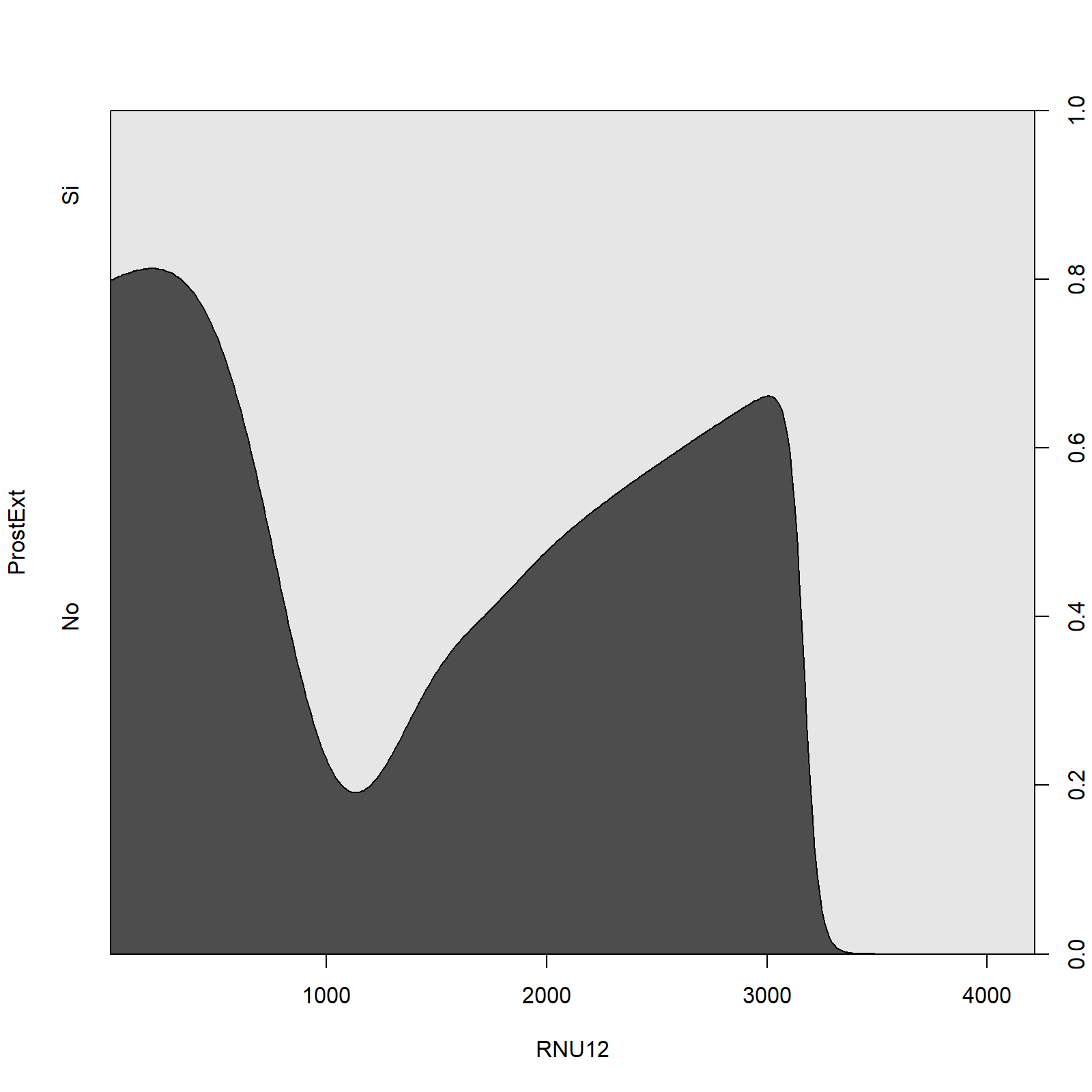
# Association between numeric variables and nominal ones

Eta squared is a measure of effect size, it is analogous to r-squared. It represents the proportion of variance in Y explained by X. It can detect non-linear correlations. The following table shows the small, medium and large correlations between the genetic factors and nominal clinical factors. The genetic factors are ordered from higher to lower correlation values. No large correlations were found.

|  |  |  |  |
| --- | --- | --- | --- |
| **Clinical factor** | **Correlations** | | |
| **Small (0.02-0.13)** | **Medium (0.13-0.26)** | **Large (>0.26)** |
| HB pressure | PRPF40A, SRRM4, KHDRSB1, U4ATAC, RBM3, RBM22, SFPQ, RNU11 | NOVA1 |  |
| Diabetes | U4ATAC, SRSF3, RBM3, SF3B1tv1, PRPF8, SRRM1 |  |  |
| Dyslipidemia | KHDRSB1, NOVA1, RBM3, snRNP200, RNU11, MAGOH, RAVER1, SRRM4, | PRPF8 |  |
| ProstExt | RBM22, SFPQ, SRRM4, RNU11, PRPF40A, SRSF6, RBM3, SRRM1 | RNU12 |  |
| PerineuralInv | PRPF40A, U4ATAC, RBM22, RNU11, RNU12, MAGOH |  |  |

The following conditional density plots show the relationship between the genetic factors and clinical ones. We only show the medium correlations; the rest of graphs are in the folder “Correlations”. The results showed that patients with higher values in the factor NOVA1 are more likely to present an HB pressure. Also, patients with higher values in the factor RNU12 are more likely to have a prost. ext.





# Association between nominal clinical factors

Cramer's V is the most popular of the chi-square-based measures of nominal association because it gives good norming from 0 to 1 regardless of table size. In practice, you may find that a Cramer's V of .10 provides a good minimum threshold for suggesting there is a substantive relationship between two nominal variables.

The following table shows the pairs of variables that have a Cramer's V greater than 0.10. The rest of the pairs were discarded.

|  |  |
| --- | --- |
| **Pair** | **Cramer's V** |
| ProstExt vs PerineuralInv | 0.547 |
| Hbpressure vs PerineuralInv | 0.437 |
| Hbpressure-ProstExt | 0.408 |
| Diabetes vs Dyslipidemia | 0.293 |
| Hbpressure vs Dyslipidemia | 0.267 |
| Diabetes vs ProstExt | 0.158 |

The following mosaic plots show the distributions of the category levels between the nominal variables. As a matter of example, the results show that a major number of patients that don’t have diabetes also don’t have dyslipidemia….