Obesity Exploration

Kushal Kumar Dey

Friday, December 05, 2014

In this report, we perform an exploratory analysis of Obesity and Abdominal Obesity with respect totwo genetic factors (namely the DGAT and the MC4R genes). Note that in the samples availabel, there were two levels of the MC4R gene (G/G and G/A) and the DGAT gene had three levels (TT, CT, CC).

We now form the contingency table of the obese/non-obese individuals against the genotype configuration for the DGAT and the MC4R genes. Here we define a person as obese if his/her BMI is greater than 25, which is the cut off for Asian Indians.

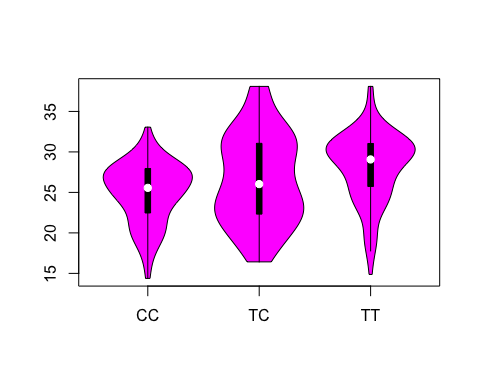
## DGAT  
## Obesity\_index CC TC TT  
## 0 49 55 29  
## 1 66 66 107

|  |  |  |
| --- | --- | --- |
| Levels | Mean BMI | SD BMI |
| TT | 28.1803676470588 | 4.27310632229204 |
| TC | 26.5865289256198 | 5.53656507694434 |
| CC | 24.964347826087 | 3.68932198649847 |

The values of the pairwise comparison t-test between CC/TC, TT/TC and CC/TT are given by **0.0083983**, **0.0111203** and **7.723528310^{-10}**. Also we perform a linear regression of BMI on the genetic effect of DGAT

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 24.9643 | 0.4258 | 58.6239 | 0.0000 |
| DGATTC | 1.6222 | 0.5947 | 2.7277 | 0.0067 |
| DGATTT | 3.2160 | 0.5785 | 5.5591 | 0.0000 |

A violin plot of BMI against the DGAT gene is presented below



Looking at the coefficients, it seems T allele is associated with higher BMI. This is also supported if we fit a GLM model of Obesity (obese/non obese) against DGAT and the output is shown as below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | 0.2978 | 0.1886 | 1.5794 | 0.1142 |
| DGATTC | -0.1155 | 0.2625 | -0.4401 | 0.6599 |
| DGATTT | 1.0077 | 0.2818 | 3.5765 | 0.0003 |

The GLM model shows the TT genotype in the DGAT gene is strongly associated with Obesity. In order compute the odds ratio, we take one group to be **TT** genotype and the other to be **TC+CC**. Then we get the contingency table to be as follows

tab1 =xtabs(~Obesity\_index +DGAT);  
odds1 =((tab1[1,1]+tab1[1,2])\*tab1[2,3])/ ((tab1[2,1]+tab1[2,2])\*tab1[1,3]);

The Odds ratio for TT/ non-TT genotype configuration of DGAT with Obese/Non-Obese is **2.907001** which is indeed significant.

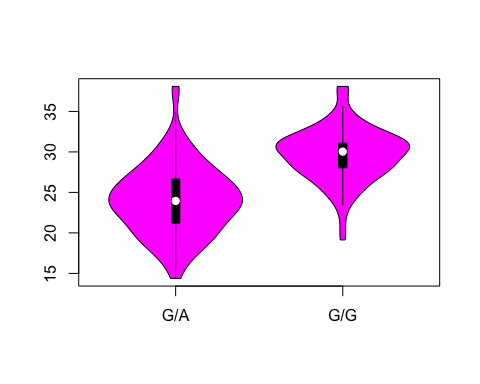
Now we consider the MC4R gene with only two levels given in the sample represented by G/G and G/A. and we compute the summary statistics based on that

|  |  |  |
| --- | --- | --- |
| Levels | Mean BMI | SD BMI |
| G/G | 29.6926315789474 | 3.26087223606477 |
| G/A | 24.094328358209 | 4.26459220461251 |

We use a linear model fit of BMI against the different levels of the MC4R gene and fit obesity against the MC4R gene. We try to see how the genetic factor influences the BMI. The odds atio for the table is given by **20.040625**, which is pretty large and shows a great deal of dependence of obesity on the MC4R levels. We first present a linear model fitting BMI against the MC4R gene

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 24.0943 | 0.2706 | 89.0460 | 0 |
| MC4RG/G | 5.5983 | 0.3991 | 14.0276 | 0 |

The dependence of BMI on the levels of MC4R gene can be depicted graphically as follows



This plot shows enough evidence that the GG genotypic configuration plays a very significant role in increasing the BMI level.

We also fit a generalized linear model fitting obesity (obese/ non obese) against the different levels of the MC4R gene.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 0.3980 | 0.0282 | 14.0930 | 0 |
| MC4RG/G | 0.5318 | 0.0417 | 12.7672 | 0 |

Now we focus on the abdominal obesity, due to WHR and look at how this WHR depends on the levels of the DGAT and the MC4R genes.

|  |  |  |
| --- | --- | --- |
| Levels | Mean BMI | SD BMI |
| G/G | 0.923391812865497 | 0.0836390265133804 |
| G/A | 1.21716417910448 | 5.01818078768695 |

The odds ratio of the Abdominal Obesity against MC4R is **6.7744246**. This again suggests that MC4R is a significant genetic factor for abdominal obesity. We consider fitting a generlaized linear model with abdominal obesity as the response and the levels of the MC4R gene as the predictor.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 0.5721 | 0.0295 | 19.4256 | 0 |
| MC4RG/G | 0.3284 | 0.0434 | 7.5607 | 0 |

This shows that MC4R GG configuration is significantly associated with the abdominal obesity. But then it probably makes sense as abdominal obesity is often correlated with obesity.

We observe how the abdominal obesity varies against the genotypic levels of the DGAT gene namely CC, TC and TT.

|  |  |  |
| --- | --- | --- |
| Levels | Mean BMI | SD BMI |
| TT | 1.42808823529412 | 6.09683469967161 |
| TC | 0.883388429752066 | 0.0950662376013716 |
| CC | 0.882086956521739 | 0.0788902819285193 |

The odds ratio for the TT/non-TT configuration against Abdominal obesity (obese/non obese) was found to be **2.9158936**. Also of interest is how the indicator of obesity changes with th different levels of the DGAT gene (A logistic linear model is fitted).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 0.7043 | 0.0408 | 17.2604 | 0.0000 |
| DGATTC | -0.1010 | 0.0570 | -1.7730 | 0.0771 |
| DGATTT | 0.1412 | 0.0554 | 2.5477 | 0.0112 |

It shows that DGAT levels are indeed associated with abdominal obesity (specially the TT configuration) but it is not as significant as the MC4R gene. This is kind of similar to the patterns we observed in overall obesity study reported above.

Now we present the logistic regression of the Obesity, Overweight data and the Abdominal Obesity data against the genetic factors DGAT gene, MC4R gene, Age and Gender and we see, how these different factors contribute to explaining the probability that a person is obese/ overweight/ abdominally obese. First we look at the abdominally obese individuals.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | -4.6023252 | 0.7813475 | -5.890241 | 0.0000000 |
| DGATTC | -1.0412719 | 0.3749613 | -2.777012 | 0.0054861 |
| DGATTT | 0.5908317 | 0.3940125 | 1.499525 | 0.1337374 |
| MC4RG/G | 2.2293618 | 0.3636883 | 6.129870 | 0.0000000 |
| gender | 3.1294063 | 0.4044674 | 7.737104 | 0.0000000 |
| Age | 0.0265863 | 0.0188477 | 1.410588 | 0.1583660 |

The AIC for this model is given by **280.1552378**.

We now consider the overall obese data (overweight implies BMI>25). We consider an indicator variable that is 1 if a person is overweight (BMI>25) and 0 otherwise.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | -2.2132171 | 0.6537582 | -3.385376 | 0.0007108 |
| DGATTC | -0.5213699 | 0.3273194 | -1.592847 | 0.1111944 |
| DGATTT | 0.6736472 | 0.3337316 | 2.018530 | 0.0435361 |
| MC4RG/G | 2.9237494 | 0.3508340 | 8.333712 | 0.0000000 |
| gender | 0.6854998 | 0.2764627 | 2.479538 | 0.0131553 |
| Age | 0.0243992 | 0.0170006 | 1.435195 | 0.1512316 |

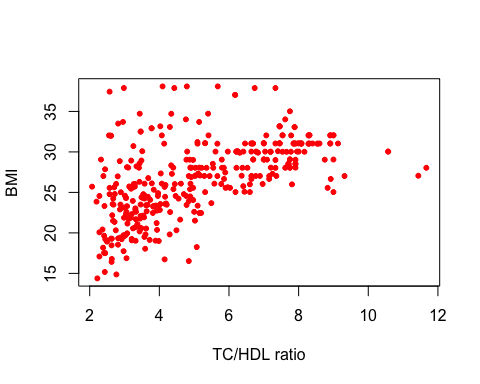
The AIC of the model is **346.6597546**. We consider now the overweight data (overweight implies BMI>23). We consider an indicator variable that is 1 if a person is overweight (BMI>23) and 0 otherwise.

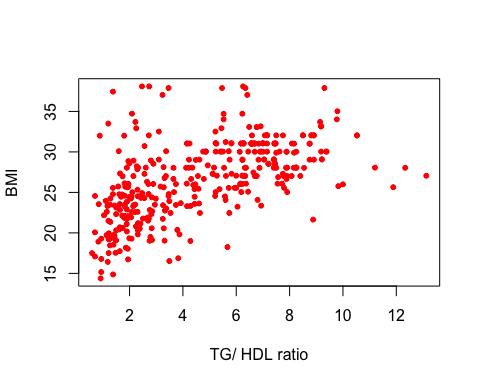
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| (Intercept) | -1.6574622 | 0.6932381 | -2.3908988 | 0.0168072 |
| DGATTC | -0.5815762 | 0.3326993 | -1.7480536 | 0.0804547 |
| DGATTT | 1.1009332 | 0.3980941 | 2.7655098 | 0.0056834 |
| MC4RG/G | 2.7990031 | 0.4897218 | 5.7154963 | 0.0000000 |
| gender | 0.2500216 | 0.3020577 | 0.8277281 | 0.4078245 |
| Age | 0.0551320 | 0.0193698 | 2.8462917 | 0.0044232 |

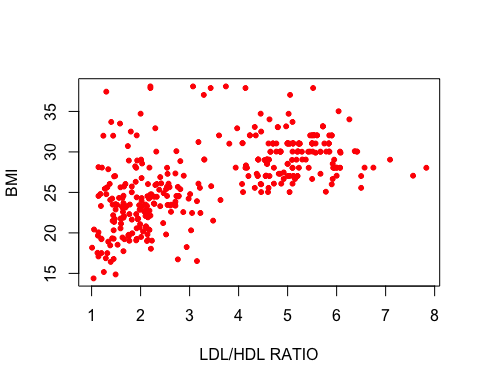
The AIC of the model is **296.4383343**.

Note from the above analysis that we did, it is very clear that the MC4R gene has a very substantial effect on both abdominal and regular obesity, and even for overweight data. The DGAT gene is significant too for obesity, the TC genotype is significant for the abdominal obesity which is slightly strange, however TT and CC levels are not found to be significant for abdominal obesity. The DGAT gene fails to be significant fr the overweight data. Age is found to be a significant impacting factor for obesity but not for abdominal obesity and gender is found to be highly signifiant for abdominal obesity and only midly significant for the general obesity.

Next we perform a linear regression of BMI on the values of the ratios TG/HDL, TC/HDL and LDL/HDL ratios and see if any of these quantities are strongly correlated with BMI. This would be crucial from a medical diagnostic point of view.







For all the three plots above, it seems the lower bound of the BMI tends to linearly increase with increase in the ratios. This tells us that these ratios do indeed raise the minimum value of the BMI.