Analysis of a genetic test in the context of arterial hypertension

João Luís Lopes Amorim
joao.l.amorim@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

September 2018

Abstract

Essential hypertension, characterized by pathological blood pressure elevation, affects globally one billion individuals and constitutes one of the most important risk factors for cardiovascular diseases. In most cases, this condition can be controlled trough pharmacological therapy. Due to advances in DNA analysis technologies and data analysis methodologies the genetics of hypertension can be explored.

The objective of this thesis is the analysis of a HeartGenetics' genetic panel associated with essential hypertension. For this, a genetic risk score model for hypertension and a discriminative model for patients with treatment-resistant hypertension were developed using this panel. This study was supported by a bibliographic revision relative to the association between genetic variants and hypertension.

In a cohort with 417 individuals a logistic regression classifier was used to calculate the risk for hypertension whose performance was evaluated by the area under the curve (AUC = 0.568). Additionally, a univariate analysis was performed from which six genetic variants associated with hypertension were identified. Additionally, using a cohort with 322 individuals it was possible to discriminate patients with treatment-resistant hypertension (p-value = 0.003). The bibliographic revision lead to the suggestion of inclusion of four genetic variants to the HeartGenetics' genetic panel.

The results of this thesis suggest that genetic risk prediction of essential hypertension is complex and requires an extensive genetic panel to obtain results with a clear clinical applicability. Nonetheless, in the context of treatment-resistant hypertension, the results suggest that there is potential for the development of a genetic test for this condition with clinical applicability.

Keywords: hypertension, hypertension genetics, pharmacogenetics, genetic testing, genetic variants

1. Introduction

Cardiovascular diseases (CVDs) are the global leading cause of death. It is estimated that 17 million people die from CVDs each year, corresponding to 30% of all global deaths [1]. Most CVDs can be prevented by addressing risk factors that contribute to their development.

Arterial hypertension is a medical condition that affects almost 1 billion individuals worldwide and constitutes one of the most important risk factors for the development of CVDs such as coronary artery disease, myocardial infarction and heart failure [2].

The development of hypertension involves the complex interplay of both non-genetic and genetic factors. The underlying genetic factors involved in the development of hypertension and in the therapeutic response to antihypertensive drugs have been highlighted in several clinical studies [3, 4, 5, 6]. The most common strategy used for the discovery and validation of these genetic factors is based on association studies. These studies test the association between genetic markers, usually involved in metabolic pathways related with blood pressure control, and the phenotype under study.

Although a plethora of antihypertensive drugs such as diuretics, β-blockers, angiotensin converting enzyme inhibitors, calcium channel blockers, and angiotensin receptor blockers are used for blood pressure control, approximately 40% of all hypertensive patients under treatment do not have hypertension under control [7].

Considering the genetic factors underlying hypertension, the development of a genetic test for this condition with a well-defined clinical utility, using a personalized medicine approach, may contribute to the management and treatment of hypertension which can lead ultimately to a decrease on cardiovascular events and to the sustainability of healthcare systems.

The HeartGenetics project, funded by Portugal2020, in cooperation with the Coimbra Hospital and Universitary Centre (CHUC), and the work developed in this thesis had as main objective the validation of the clinical utility of the HeartGenetics DNArterial genetic test for the identification of the genetic risk for essential hypertension and the

response to hypertensive drugs.

In more detail, the aim of this thesis was to perform a clinical validation study of the DNArterial genetic panel on two components of the test: risk prediction and pharmacogenetic-guided therapy for hypertension. To evaluate the risk prediction component of the test a genetic association study of hypertension was performed. The objectives of this analysis consisted on the identification of genetic variants from the DNArterial genetic panel with statistical significant associations with hypertension and on the construction of a genetic risk score for this condition. To evaluate the pharmacogenetic component of the test an analysis of the genetic influence of pharmacogenetic variants on treatment-resistant hypertension was performed. Additionally, a literature search with the objective of suggesting the addition of additional pharmacogenetic variants to be included in the DNArterial genetic panel was performed.

2. Materials and Methods

2.1. Data characterization

Genotype data used in this thesis was collected from the HeartGenetics biobank. Subjects with cardiovascular disease and/or traditional cardiovascular risk factors, individuals with family history of heart failure and/or traditional cardiovascular risk factors and centenaries with no hereditary diseases or cardiovascular diseases were included in the HeartGenetics biobank. Excluded subjects had either congenital heart disease, myocardiopathy, autoimmune disease, hypothyroidism, diabetes mellitus, renal insufficiency and chronic liver disease or were under therapeutics with either anticonvulsants, antipsychotics, immunosuppressors or antiretrovirals.

Two cohorts were used for this study. The first, from now on referred as HTN cohort, includes 187 hypertensive (cases) and 230 healthy (controls) individuals, from which genotype data of 53 genetic variants involved in blood pressure pathways was gathered. The second, from now on referred as PHARMA cohort, includes 13 individuals with treatment-resistant hypertension (cases) and 309 healthy or hypertensive individuals (controls), from which genotype data of 24 genetic variants known to have pharmacogenetic associations with antihypertensive drugs was gathered. All of the analysed genetic variants, in both cohorts, are currently included in the genetic panel of the DNArterial test.

Two different types of searches were made on PharmGKB to identify genetic variants known to be associated with a drug phenotype: a search per variant in which were considered the 24 pharmacogenetic variants of the DNArterial test genetic panel; and a search per drug or class of drug in which were considered 58 antihypertensive drugs and 5 drug classes identified as relevant drugs used for hypertension treatment in the most recent guidelines from the American Heart Association [8]. From this search, only pharmacogenetic associations between genetic variants and drugs used for hypertension or other cardiovascular conditions treatment were selected. As a result, a total of 396 pharmacogenetic associations involving a total of 203 genetic variants and 82 drugs and 5 classes of drugs were included.

2.2. Genotype data quality control

For the two datasets, HTN and PHARMA, a data quality control was performed. At a first stage, the minor allele frequencies for each variant was computed. Then, it was tested if all the genetic variants were under Hardy-Weinberg equilibrium in the control population. For this, the *HWExact* function from the HardyWeinberg R package which performs an exact Fisher test for HWE was used.

After this analysis, genetic variants with MAF smaller than 5% were removed. Additionally, genetic markers for which there was a deviation from the Hardy-Weinberg equilibrium were flagged for future analyses if proven to have a significant association with the traits studied in this work. This approach was considered, as an alternative of removing these variants, because deviations from the HWE occur frequently mostly due to the fact that this concept is based on multiple assumptions that are not always verified.

Another data quality control procedure is the analysis for population stratification. The approach used in this thesis for this analysis consists of a principal component analysis (PCA) followed by a k-means clustering. For this, the *eigenstrat* function from the AssocTests R package and the function *kmeans* from the Stats R package were used to perform the PCA and the k-means, respectively.

2.3. Univariate association analysis for the identification of genetic variants associated with hypertension

The association with the hypertension trait was studied for each one of the genetic variants, not removed during the quality control procedure, considered in the HTN cohort. For this, three fisher exact tests were performed per variant, one for each genetic model used (dominant, recessive, and additive), using the function *fisher.test* from the package Stats of the R software. A chi-squared test could also have been used, however this method relies on an approximation while the fisher test is exact. In addition, the chi-squared test is not the most appropriate in this case because of the small number of samples used. Therefore, despite being more

computationally expensive, the fisher exact test was chosen.

For each association test an adjusted p-value of 0.001 computed using the Bonferroni correction was considered as statistically significant for rejecting the null hypothesis that cases and controls are independent of the genotype class.

2.4. Construction of a genetic risk score for hypertension

A logistic regression was used to build a genetic risk model and, consequently, compute the genetic risk score for hypertension for each individual. For this, it was used the function *glm* from the package Stats of the R software. To validate this model built a leave-one-out cross-validation was performed. This process is an iterative method in which for each k iterate ($(k\varepsilon\{1, n\})$, where n is the number of samples) is defined a validation dataset with the k sample and a training dataset with the remaining samples. Then, for each iteration the validation dataset is tested against the model build with the training dataset.To evaluate the performance of the model the receiver operating characteristic curve (ROC curve)and its respective area under the curve (AUC) were used.

In this analysis an additive, a dominant, and a recessive model were considered for the genotype data of the HTN dataset. In addition, a mix genetic model in which each genetic variant is encoded by a different genetic model was used to encode the HTN dataset used in the logistic regression. For this, in each iteration of the leave-one-out cross-validation, a different mix model dataset was constructed using the training data in which each genetic variant was encoded with the most adequate genetic model (additive, recessive or dominant). For the identification of the best-fitting genetic model, three Fisher exact tests per variant were performed. The genetic model that achieved the smallest p-value was selected.

2.5. Criteria for the selection of relevant genetic variants with pharmacogenetic associations

After extracting 396 pharmacogenetic associations from the PharmGKB, different criteria were used to select the most relevant genetic variants that should be included in a genetic panel for a pharmacogenetic test for hypertension.

Firstly, genetic variants with at least one pharmacogenetic association with a level of evidence equal or lower than 2 were selected. Those levels of evidence 1 and 2 correspond, respectively, to a high and a moderate level of evidence. In the first level are included variant-drug associations for which the majority of evidence confirms the association. In addition, these associations are replicated in different studies that presented a significant p-value and strong effect size. By comparison, the second level of evidence includes associations that are replicated in different studies although some studies may not show statistical significance or present a small effect size.

Then, from the list of pharmacogenetic associations identified it were selected genetic variants with a FDA drug label containing pharmacogenetic information approved by the FDA [9].

Additionally, a third criterion was considered, it were also selected genetic variants with at least one pharmacogenetic association with a p-value smaller than 0.001. This threshold was used because it was defined that it was preferable to have a genetic panel with genetic variants whose pharmacogenetic associations had a strong statistical significance than including genetic variants with associations with weaker statistical significance. To conclude, given the high risk of complications in patients with treatment-resistant hypertension and the fact that optimization of medical treatment still remains an important concern, it were also selected genetic variants with known pharmacogenetic associations identified in cohorts whose individuals had treatment-resistant hypertension.

2.6. Analysis of the genetic influence on treatment-resistant hypertension

The PHARMA cohort was used to assess the underlying genetic background of treatment-resistant hypertension. Firstly, it was computed a genetic risk score for each individual to compare cases and controls for all the 24 genetic variants with known pharmacogenetic association with antihypertensive drugs. For this analysis three different genetic models were considered (additive, dominant and recessive models). The genetic risk score for each individual was computed using an unweighted approach in which the score is the direct sum of the number of risk alleles. Additionally, it was also used a mix genetic model in which each genetic variant was encoded with its most adequate genetic model to compute the genetic risk score.

The genetic risk score of each cohort subject was then used to compare the controls with the cases. For this, a MannWhitney U test was used to compare the distribution of genetic risks scores between the two groups by testing the null hypotheses that is equally likely that a random observation, in this case the genetic risk score, in one group will be less or greater than a random observation of the second group. A p-value of 0.05 was considered statistical significant.

Aditionally, associations tests for each of the 24 genetic variants considered in the PHARMA cohort to study its association with the treatment-resistant

hypertension trait was also performed. For this, three fisher exact tests were performed per variant, one for each genetic model used (dominant, recessive, and additive).

3. Results

3.1. Genetic variants associated with hypertension

An univariate analysis was performed on the HTN cohort to identify genetic variants associated with the hypertension trait. However, firstly are presented the results obtain from the quality control analyses preformed. Of the 53 genetic variants analysed, 4 had a minor allele frequency less than the considered threshold of 5% (rs75770792, rs111253292, rs5063, and rs5068). As a result, these variants were excluded from the posterior analyses performed on the HTN cohort. In addition, 2 genetic variants were not in Hardy-Weinberg equilibrium (rs11091046 and rs1403543). By doing a PCA and a k-means it was possible to verify that, in both cohorts, only one cluster is present and consequently there is no population stratification.

Of the set of genetic variants considered, a total of 6 variants showed a trend for association with hypertension (p-value $\leq$ 0.05): rs4646994 (ACE), rs3918226 (NOS3), rs5065 (NPPA), rs6693954(REN), rs880054 (WNK1), and rs72811418 (CYBA). However, of the 49 genetic variants analysed, considering as significant a p-value ≤ 0.001 (Bonferroni adjusted p), none showed a statistical significant association with hypertension.

3.2. Genetic risk score model evaluation for essential hypertension

Logistic regression was used to create a genetic risk score to evaluate the impact of the combination of all 49 genetic variants considered in the HTN cohort on hypertension risk. An additive, a dominant, a recessive and mix genetic model were considered for this analysis. The discriminative power of the logistic regression of each model considered was measured by the AUC of the ROC curve. Figure 1 presents the 4 ROC curves obtained and the respective AUCs. Although being higher when using a mix model (AUC = 0.568), the AUCs showed that the logistic regression achieved a modest result using the 49 genetic variants.

3.3. Genetic variants with relevant pharmacogenetic associations

From a total of 396 pharmacogenetic associations involving 203 genetic variants and 87 drugs or classes of drugs, were selected 33 genetic variants with relevant pharmacogenetic associations (detailed information relative to these genetic variants is presented in confidential appendices). Of these 33 genetic variants, 6 variants (rs1799752, rs4961, rs1801253, rs5051, rs13306673) are present in the current genetic panel of the HeartGenetics test for hypertension.

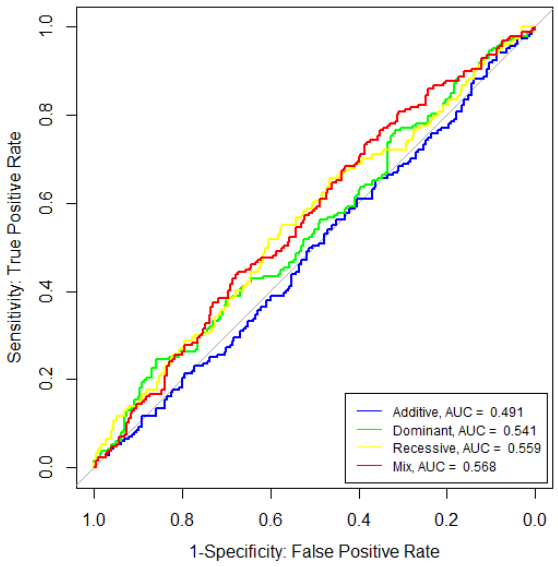


Figure 1: Receiver operator characteristics (ROC) curves of the logistic regressions used for hypertension discrimination using different genetic models.

Considering the first criterion in which are included genetic variants with at least one pharmacogenetic association with level of evidence less thant or equal to 2, 17 genetic variants were selected from a total of 6 different genes. The remaining genetic variants selected had pharmacogenetic associations with weak levels of evidence (level 3), from which 9 were labelled by the FDA as relevant pharmacogenetic variants and 2 had a pharmacogenetic association verified in individuals with treatment-resistant hypertension.

3.4. Genetic risk score model evaluation for treatment-resistant hypertension

The genetic risk score was computed for each subject of the PHARMA cohort. Its distribution between individuals with treatment-resistant hypertension and controls was compared using a Mann-Whitney test. Figure 2 displays, using boxplots, the genetic risk scores, for the additive model, of cases and controls and the respective p-value computed using the Mann-Whitney test.

For the first analysis, in each were considered all 24 pharmacogenetic variants, it was found a statistical difference in the risk score between the controls group and the cases group for all the genetic models used. When an additive model was considered it was achieved a higher statistical difference between groups with a p-value = 0.003. Considering this

genetic model, the upper quartile of the controls group is equal to median and lower quartile of the cases group which means that 75% of the individuals in this group had a higher genetic risk score for treatment-resistant hypertension than 75% of the controls.

An identical second analysis was performed by considering from the HeartGenetics pharmacogenetic panel a total of 6 genetic variants (rs1799752, rs4961, rs1801253, rs5051, rs13306673) identified as relevant pharmacogenetic variants in section 3.3. It was found a statistical difference in the risk score between the controls group and the cases group for the additive, recessive, and mix models. Considering these models, 50% of the cases had a higher genetic risk score than 75% of the controls. When an recessive model model was considered it was achieved a higher statistical difference between groups with a p-value = 0.029. Under a dominant model the non-significant p-value of 0.214 was obtained.

Overall, when all variants were considered versus the 6 selected variants a better discrimination between cases and controls with lower p-values was achieved.

Regarding the univariate analysis performed on this cohort to identify genetic variants associated with the treatment-resistant hypertension trait, of the 24 genetic variants considered, a total of 8 variants showed a trend for association with treatment-resistant hypertension (rs1801252 (ADRB1), rs699 (AGT), rs5186 (AGTR1), rs1048101 (ADRA1A), rs1801253 (ADRB1), rs5051 (AGT), rs1799722 (BDKRB2), and rs2228576 (SCNN1A)). However, of the 24 genetic variants analysed, considered as significant a p-value$\leq$ 0.002 (Bonferroni adjusted p), none showed a statistical significant association with treatment-resistant hypertension.

4. Discussion and Conclusions

The main objective of this thesis was to develop a genetic test for hypertension with a clear clinical value. For this, two different approaches were considered. Firstly, it was evaluated the feasibility of having a genetic test with the aim of predicting hypertension genetic risk. Then, it was considered a second approach in which it was evaluated the clinical validity of a genetic test with the aim of offering a pharmacogenetic-guided treatment.

Using an in-house genetic dataset, from a total of 53 genetic variants, the variants rs4646994 (*ACE*), rs3918226 (*NOS3*), rs5065 (*NPPA*), rs6693954 (*REN*), rs880054 (*WNK1*), and rs72811418 (*CYBA*) showed a trend for association with the hypertension trait. However, none achieved statistical significance.

The *ACE* polymorphism, rs4646994 (also referred as rs1799752), has been previously mentioned in different studies in which a positive correlation of the del allele with hypertension was verified in a Indian and Chinese population, however this finding was not verified in juveline American, Turkish or Sardinian populations [10]. These findings, in line with the results obtained in this thesis, show that the association between this *ACE* polymorphism and hypertension is not robust.

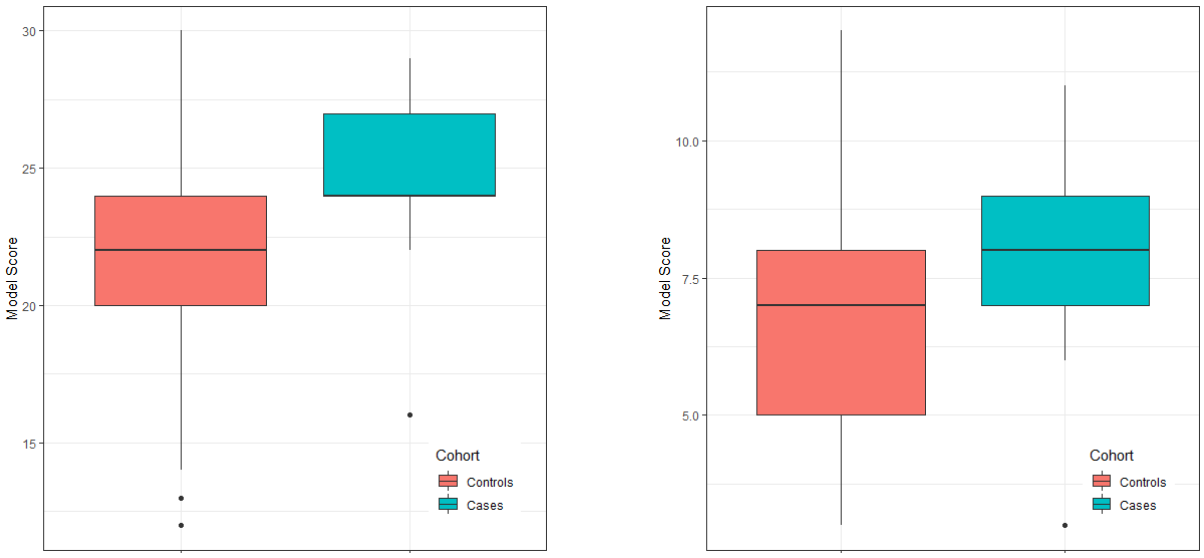
Although, the results of this work show a trend for correlation with hypertension for the locus rs880054, previously reported associations of this SNP with hypertension were not replicated in the British Genetics of Hypertension (BRIGHT) study [A6].

From the remaining six genetic variants mentioned previously, the loci rs3918226 (allele T), rs5065 (allele T), and rs6693954 (allele A) have been associated with hypertension with OR = 1.54 (95% CI: 1.37-1.73), OR = 0.94 (95% CI: 0.88-1.00), and OR = 1.98 (95% CI: 1.40-2.80), respectively [11, 12, 13]. Lastly, the polymorphism rs72811418 of *CYBA* gene was shown to be associated with hypertension by Moreno et al., however the findings were not replicated in a different cohort [14].

The differences observed between the results obtained in this work and current evidence may be the consequence of using a cohort with a small sample size, at least in comparison with the sample sizes currently used in genome-wide association studies.

Regarding the genetic risk score model created to predict hypertension risk, it can be said that it did achieved a modest result which is confirmed by the AUCs of the 4 models created using an additive, a dominant, a recessive and mix genetic model. However, the results obtained seem to be in line with the state of the art. Held et al. compared the performance of the logistic regression method to predict the risk of hypertension using non-genetic factors alone and in combination with genetic data [15]. Reported AUCs were 0.785 and 0.778 for the model considering non-genetic factors alone and the model considering both genetic and non-genetic factors, respectively. Although, were only considered 10 causal genes for the genetic component of the model used in this study, the results reported point that the inclusion of genotype data for risk prediction may not improve substantially the prediction ability of these models. Additionally, these results clearly show, as discussed previously in this work, that non-genetic factors play a crucial role in hypertension development which also limits the accuracy of a genetic test for hypertension risk that uses only genotype data.

As mentioned previously, the analysis of the UK biobank dataset shows that the genetic factors influencing hypertension include a broad number of



(a) 24 genetic variants

(b) 6 genetic variants

Figure 2: Boxplot of the genetic risk scores calculated using an additive model for the cases and controls groups. (a) p-value = 0.003, (b) p-value = 0.035.

genetic variants, each with small effect size. These results in line with the results obtained in this work point that it is not clear that developing a genetic test with the aim of predicting hypertension risk is feasible and therefore its clinical value may not be sufficient to be adopted by clinical practitioners.

A total of 33 genetic variants with relevant pharmacogenetic associations were selected with the aim of evaluating if they could guide hypertension treatment by considering the underlying interactions between the genotype and the response to antihypertensive drugs.

Different criteria were used for the selection of genetic variants with interesting pharmacogenetic associations. As a result, the selected genetic variants have pharmacogenetic associations with different levels of evidence, from which only a few present pharmacogenetic associations with a good level of evidence *i.e.* a level of evidence smaller than or equal to 2.

This is the case of the association between the *CYP2D6* polymorphisms and the β-blocker metoprolol. It has been shown in multiple studies that these polymorphism may lead to a decreased metabolism or clearance of metoprolol [16, 17, 18]. Despite the reported differences in metoprolol pharmacokinetics in individuals carrying these polymorphisms, this may not translate into differences in adverse event rate. In fact, it was shown that there is no evidence for association between the variable pharmacokinetics of metoprolol, a result of the differences in *CYP2D6* metabolic activity, and a higher rate of adverse events or a lower efficacy in patients with hypertension [17].

In opposition, it has been shown that patients carrying an insertion of an Alu element, a short stretch of repetitive DNA, in the *ACE* polymorphism rs1799752 may have a decrease response to captopril treatment. This result have been observed in patients with heart failure, chronic obstructive pulmonary disease or diabetes mellitus type 2 [19, 20, 21, 22].

Regarding the rs4961 *ADD1* polymorphism, it was verified in patients with liver cirrhosis that individuals carrying allele T are more likely to have a bad response (OR = 2.89, 95% CI: 1.28-2.81) to diuretic treatment using furosemide and spironolactone than those carrying allele G [23]. Despite these interesting results, this association was only verified in patients with liver cirrhosis in which diuretics were used for the treatment of ascites and not as an antihypertensive therapy. In other studies, the association between this polymorphism and diuretics has been found in hypertensive patients however with contradictory results. Turner et al. have shown in a cohort of 259 hypertensive adults that allele T of this polymorphism is not related to hydrochlorothiazide response [24]. In opposition, in a study with similar sample size it was observed a higher reduction in SBP and DBP in hypertensive patients carrying allele T of the rs4961 genetic variant [25].

Another polymorphism with a pharmacogenetic association whose level of evidence is at least 2 is the rs4149601 from the gene *NEDD4L*. In white patients carrying the AA genotype it was observed

a poorer response to hydrochlorothiazide therapy [26]. However, in individuals with Asian ancestry, the opposite result was observed, at a lower level of evidence, in which allele A was associated with increased blood pressure reduction [27]. A decreased response to hydrochlorothiazide it was also reported for hypertensive patients carrying allele T of the rs7297610 polymorphism (*YEATS4*) and patients carrying allele G of the rs16960228 polymorphims (*PRKCA*). The association between this diuretic and each of these polymorphisms, rs7297610 and rs16960228, was observed in a European cohort and in a cohort with black individuals, respectively [28, 29, 30, 27, 31].

Concerning the genetic variants with FDA labelled pharmacogenetic associations, *NAT2* polymorphisms have been correlated with adverse effects after hydralazine therapy in patients with treatment-resistant hypertension [32] and the rs1057910 and *CYP2D6* polymorphisms (CYP2D6*1, CYP2D6*10, CYP2D6*4, CYP2D6*5) have been associated with decreased clearance of losartan and carvedilol, respectively [33, 34, 35]. Despite having an FDA label, these associations were observed in a small number of studies with a small study size. Given this, a level of evidence 3 was attributed to these associations by the PharmGKB which indicates that the reported pharmacogenetic associations may need to be replicated in different studies. The same applies to the remaining genetic variants selected in this work with pharmacogenetic associations whose level of evidence is smaller than 2.

The pharmacogenetic associations involving these last genetic variants may indicate that these polymorphisms can have a relevant role in the pharmacokinetics and pharmacodynamics of antihypertensive drugs. Additionally, it could also be hypothesized that the selected genetic variants may be involved in some underlying genetic factors responsible for the development of treatment-resistant hypertension.

Given this, a final analysis was performed in which using the PHARMA cohort it was studied if genetic variants with known pharmacogenetic associations could predict the risk of treatment-resistant hypertension. Using 24 pharmacogenetic variants from the HeartGenetics genetic panel it was possible to discriminate with statistical significance individuals with treatment-resistant hypertension (p-value = 0.003, additive model). A worse result, despite being statistical significant, was achieved using a subset of those initial genetic variants with relevant pharmacogenetic associations identified in section **??** (p-value = 0.035, additive model). These results support the idea that there are underlying genetic factors involved in the development of treatment-resistant hypertension [36]. Of the 24 genetic variants analysed, 8 showed a trend for association with the treatment-resistant hypertension trait. However, none reached statistical significance. Of these polymorphisms the SNPs rs5051 and rs699 have been previously associated with resistant hypertension. In the GenHAT study, a study with 2203 treatment resistant cases and 2354 controls, it was found that carries of the T allele of rs699 and carriers of the G allele of rs5051 were associated with resistant hypertension with OR = 1.27(95% CI: 1.12-1.44, p-value = 0.0001) and OR = 1.36 (95% CI: 1.20-1.53, p-value $<$ 0.0001), respectively [37].

Given this, it can be hypothesized that these results may indicate that an individual carrying risk alleles of these genetic variants may have a worse response to treatment or could have a higher risk for treatment-resistant hypertension. Therefore, this information can assist the clinician in the definition of the most adequate pharmacological approach for hypertension treatment for each individual.

Different limitations were present in the analyses performed in this thesis. Firstly, the lack of positive results from the association tests performed using th data from the HTN and PHARMA cohorts and the logistic regression done using the data from the HTN cohort may be a consequence of the small sample size of the cohorts. Hong et al. mentioned that for testing a single genetic marker it would be required at least 248 samples, under the assumption of a 5% MAF, equal ratio of cases and controls, and a 5% error rate during genotyping, to achieve an adequate statistical power [38]. By comparison with the sample size used in this thesis, it is clear that the small number of samples may have contributed for the differences between the results obtained and current evidence. These same analyses may have also been limited by the fact that it were only considered genetic variants from the HeartGenetics genetic panel. Additionally, in the logistic regression it were considered as variables all genetic markers. An equivalent approach could have been taken if only genetic variants with statistical associations with the hypertension trait were considered. By doing this, it could be evaluated if a genetic risk model constructed using only genetic variants correlated with the given trait would have better results than a genetic risk model constructed using all of the genetic variants under study.

Concerning the selection of relevant pharmacogenetic variants thought the information on the PharmGKB repositiory, it can be said that the criteria used for the selection of genetic variants may have been too strict. After selecting genetic variants with level of evidence smaller than or equal to 2 or variants with a FDA label, polymorphisms with at least one pharmacogenetic association with a p-

value $\leq$0.001 were selected as relevant. By adopting this threshold for the p-value, genetic variants with important pharmacogenetic associations and significant p-values reported in several studies may have been not considered. However, this limitation was mitigated by the fact that for all variant/drug association pairs analysed the level of evidence was taken into consideration. Given this, if these genetic variants were not selected under the first inclusion criteria (level of evidence $\leq$2) it can be hypothesized that the pharmacogenetic associations involving these polymorphisms still lack a strong supporting evidence. Regarding the PHARMA cohort used to discriminate treatment-resistant hypertension patients, the small percentage of cases used was not ideal. It can be hypothesized that a more balanced cohort could have achieved higher discriminative power however further studies are needed to prove this hypothesis.

The results from this thesis are preliminary and would benefit, considering an academic point of view, from further studies involving a larger cohort of patients with hypertension and treatment-resistant hypertension so that it could be performed new association tests with a stronger statistical power. In addition, it would be also interesting to perform a study using a large cohort of individuals with treatment-resistant hypertension under antihypertensive therapy to understand the underlying genetic factors involved in this condition and in the response to antyhypertensive treatment.

Overall, the results obtained point that a genetic test has a modest capacity for predicting hypertension using only genetic data, mostly because hypertension is a complex condition involving a variety of non-genetic and genetic factors. Given the results obtained in the pharmacogenetic analyses performed, it may be interesting to develop a test with the aim of offering a pharmacogenetic-guided treatment. Therefore, despite the modest capacity for predicting the genetic risk for hypertension, the genetic panel used demonstrates the potential for identifying patients with treatment-resistant hypertension.

Acknowledgements

I would like to express my most sincere thanks to everyone that helped during the elaboration of this thesis, particularly to my advisors professor Ana Teresa Freitas and doctor José Caldas for all the guidance provided and important feedback given. I would also like to thank all the HeartGenetics staff for welcoming me and making me feel as member of the team.

References

[1] Haidong Wang, Mohsen Naghavi, Christine Allen, Ryan M Barber, Zulfiqar A Bhutta, Austin Carter, Daniel C Casey, Fiona J Charlson, Alan Zian Chen, Matthew M Coates, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *The lancet*, 388(10053):1459–1544, 2016.

[2] Mohammad H Forouzanfar, Patrick Liu, Gregory A Roth, Marie Ng, Stan Biryukov, Laurie Marczak, Lily Alexander, Kara Estep, Kalkidan Hassen Abate, Tomi F Akinyemiju, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm hg, 1990-2015. *Jama*, 317(2):165–182, 2017.

[3] Derek Klarin, Qiuyu Martin Zhu, Connor A Emdin, Mark Chaffin, Steven Horner, Brian J McMillan, Alison Leed, Michael E Weale, Chris CA Spencer, François Aguet, et al. Genetic analysis in uk biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nature genetics*, 49(9):1392, 2017.

[4] Helen R Warren, Evangelos Evangelou, Claudia P Cabrera, He Gao, Meixia Ren, Borbala Mifsud, Ioanna Ntalla, Praveen Surendran, Chunyu Liu, James P Cook, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature genetics*, 49(3):403, 2017.

[5] Praveen Surendran, Fotios Drenos, Robin Young, Helen Warren, James P Cook, Alisa K Manning, Niels Grarup, Xueling Sim, Daniel R Barnes, Kate Witkowska, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature genetics*, 48(10):1151, 2016.

[6] Aldi T. Kraja et al. New blood pressure–associated loci identified in meta-analyses of 475 000 individualsclinical perspective. *Circulation: Genomic and Precision Medicine*, 10(5), 2017.

[7] George Bakris, Pantelis Sarafidis, Rajiv Agarwal, and Luis Ruilope. Review of blood pressure control rates and outcomes. *Journal of the American Society of Hypertension*, 8(2):127–141, 2014.

[8] PK Whelton, RM Carey, WS Aronow, DE Casey Jr, KJ Collins, C Dennison Himmelfarb, et al. Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the amer-

ican college of cardiology/american heart association task force on clinical practice guidelines. *Hypertension*, 2017.

[9] Center for Drug Evaluation and Research. Science & research (drugs) - table of pharmacogenomic biomarkers in drug labeling.

[10] Paul R Gard. Implications of the angiotensin converting enzyme gene insertion/deletion polymorphism in health and disease: a snapshot review. *International journal of molecular epidemiology and genetics*, 1(2):145, 2010.

[11] Erika Salvi, Zoltán Kutalik, Nicola Glorioso, Paola Benaglio, Francesca Frau, Tatiana Kuznetsova, Hisatomi Arima, Clive Hoggart, Jean Tichet, Yury P Nikitin, et al. Genome-wide association study using a high-density snp-array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of enos. *Hypertension*, 59(2):248, 2012.

[12] David Conen, Robert J Glynn, Julie E Buring, Paul M Ridker, and Robert YL Zee. Natriuretic peptide precursor a gene polymorphisms and risk of blood pressure progression and incident hypertension. *Hypertension*, 50(6):1114–1119, 2007.

[13] Bei Sun, Jonathan S Williams, Luminita Pojoga, Bindu Chamarthi, Jessica Lasky-Su, Benjamin A Raby, Paul N Hopkins, Xavier Jeunemaitre, Nancy J Brown, Claudio Ferri, et al. Renin gene polymorphism: its relationship to hypertension, renin levels and vascular responses. *Journal of the Renin-Angiotensin-Aldosterone System*, 12(4):564–571, 2011.

[14] María U Moreno, Gorka San José, Ana Fortuño, Óscar Beloqui, Javier Díez, and Guillermo Zalba. The c242t cyba polymorphism of nadph oxidase is associated with essential hypertension. *Journal of hypertension*, 24(7):1299–1306, 2006.

[15] Elizabeth Held, Joshua Cape, and Nathan Tintle. Comparing machine learning and logistic regression methods for predicting hypertension using a combination of gene expression and next-generation sequencing data. In *BMC proceedings*, volume 10, page 34. BioMed Central, 2016.

[16] JA Batty, AS Hall, HL White, J Wikstrand, RA de Boer, DJ van Veldhuisen, P van der Harst, F Waagstein, Å Hjalmarson, J Kjekshus, et al. An investigation of cyp2d6 genotype and response to metoprolol cr/xl during dose titration in patients with heart failure: A merit-hf substudy. *Clinical Pharmacology & Therapeutics*, 95(3):321–330, 2014.

[17] Issam Zineh, Amber L Beitelshees, Andrea Gaedigk, Joseph R Walker, Daniel F Pauly, Kathleen Eberst, J Steven Leeder, Michael S Phillips, Craig A Gelfand, and Julie A Johnson. Pharmacokinetics and cyp2d6 genotypes do not predict metoprolol adverse events or efficacy in hypertension. *Clinical Pharmacology & Therapeutics*, 76(6):536–544, 2004.

[18] SK Jin, HJ Chung, MW Chung, J-I Kim, J-H Kang, SW Woo, S Bang, SH Lee, HJ Lee, and J Roh. Influence of cyp2d6* 10 on the pharmacokinetics of metoprolol in healthy korean volunteers. *Journal of clinical pharmacy and therapeutics*, 33(5):567–573, 2008.

[19] Bilge Volkan-Salanci, Selcuk Dagdelen, Mehmet Alikasifoglu, Tomris Erbas, Mutlu Hayran, and Belkis Erbas. Impact of renin-angiotensin system polymorphisms on renal haemodynamic responsiveness to acute angiotensin-converting enzyme inhibition in type 2 diabetes mellitus. *Journal of the Renin-Angiotensin-Aldosterone System*, 10(1):41–50, 2009.

[20] Sonoo Mizuiri, Hiromichi Hemmi, Akira Inoue, Masaaki Takano, Satoshi Kadomatsu, Hiroyuki Tanimoto, Masateru Tanegashima, Ikuko Hayashi, Tatsuo Fushimi, and Akira Hasegawa. Renal hemodynamic changes induced by captopril and angiotensin-converting enzyme gene polymorphism. *Nephron*, 75(3):310–314, 1997.

[21] Peter Jacobsen, Kasper Rossing, Peter Rossing, Lise Tarnow, Christine Mallet, Odette Poirier, Francois Cambien, and Hans-Henrik Parving. Angiotensin converting enzyme gene polymorphism and ace inhibition in diabetic nephropathy. *Kidney international*, 53(4):1002–1006, 1998.

[22] H Kanazawa, K Hirata, and J Yoshikawa. Effects of captopril administration on pulmonary haemodynamics and tissue oxygenation during exercise in ace gene subtypes in patients with copd: a preliminary study. *Thorax*, 58(7):629–631, 2003.

[23] Ying-Ying Yang, Han-Chieh Lin, Ming-Wei Lin, Chi-Jen Chu, Fa-Yauh Lee, Ming-Chih Hou, Shou-Dong Lee, Wei-Ping Lee, Tze-Tze Liu, and Tjin-Shing Jap. Identification of diuretic non-responders with poor long-term clinical outcomes: a 1-year follow-up of 176 non-azotaemic cirrhotic patients with moderate ascites. *Clinical Science*, 121(11):509–521, 2011.

[24] Stephen T Turner, Arlene B Chapman, Gary L Schwartz, and Eric Boerwinkle. Effects of endothelial nitric oxide synthase, α-adducin, and other candidate gene polymorphisms on blood pressure response to hydrochlorothiazide. *American journal of hypertension*,

16(10):834–839, 2003.

[25] Paolo Manunta, Gail Lavery, Chiara Lanzani, Peter S Braund, Marco Simonini, Claire Bodycote, Laura Zagato, Simona Delli Carpini, Cristina Tantardini, Elena Brioni, et al. Physiological interaction between α-adducin and wnk1-nedd4l pathways on sodium-related blood pressure regulation. *Hypertension*, 52(2):366–372, 2008.

[26] Stephen T Turner, Eric Boerwinkle, Jeffrey R Oconnell, Kent R Bailey, Yan Gong, Arlene B Chapman, Caitrin W McDonough, Amber L Beitelshees, Gary L Schwartz, John G Gums, et al. Genomic association analysis of common variants influencing antihypertensive response to hydrochlorothiazide. *Hypertension*, 62(2):391–397, 2013.

[27] Fang Luo, Yibo Wang, Xiaojian Wang, Kai Sun, Xianliang Zhou, and Rutai Hui. A functional variant of nedd4l is associated with hypertension, antihypertensive response, and orthostatic hypotension. *Hypertension*, 54(4):796–801, 2009.

[28] Julio D Duarte, Stephen T Turner, B Tran, Arlene B Chapman, Kent R Bailey, Yan Gong, John G Gums, Taimour Y Langaee, Amber L Beitelshees, Rhonda M Cooper-Dehoff, et al. Association of chromosome 12 locus with antihypertensive response to hydrochlorothiazide may involve differential yeats4 expression. *The pharmacogenomics journal*, 13(3):257, 2013.

[29] Stephen T Turner, Kent R Bailey, Brooke L Fridley, Arlene B Chapman, Gary L Schwartz, High Seng Chai, Hugues Sicotte, Jean-Pierre Kocher, Andréi S Rodin, and Eric Boerwinkle. Genomic association analysis suggests chromosome 12 locus influencing antihypertensive response to thiazide diuretic. *Hypertension*, 52(2):359–365, 2008.

[30] Patrik Svensson-Färbom, Björn Wahlstrand, Peter Almgren, Jonas Dahlberg, Cristiano Fava, Sverre Kjeldsen, Thomas Hedner, and Olle Melander. A functional variant of the nedd4l gene is associated with beneficial treatment response with β-blockers and diuretics in hypertensive patients. *Journal of hypertension*, 29(2):388–395, 2011.

[31] Caitrin W McDonough, Sarah E Burbage, Julio D Duarte, Yan Gong, Taimour Y Langaee, Stephen T Turner, John G Gums, Arlene B Chapman, Kent R Bailey, Amber L Beitelshees, et al. Association of variants in nedd4l with blood pressure response and adverse cardiovascular outcomes in hypertensive patients treated with thiazide diuretics. *Journal of hypertension*, 31(4):698, 2013.

[32] Lizania Borges Spinasse, Adalberto Rezende Santos, Philip Noel Suffys, Elizabeth Silaid Muxfeldt, and Gil Fernando Salles. Different phenotypes of the nat2 gene influences hydralazine antihypertensive response in patients with resistant hypertension. *Pharmacogenomics*, 15(2):169–178, 2014.

[33] Ümit Yasar, Cecilia Forslund-Bergengren, Gunnel Tybring, Pedro Dorado, Adrián LLerena, Folke Sjöqvist, Erik Eliasson, and Marja-Liisa Dahl. Pharmacokinetics of losartan and its metabolite e-3174 in relation to the cyp2c9 genotype. *Clinical Pharmacology & Therapeutics*, 71(1):89–98, 2002.

[34] Yoh Takekuma, Toru Takenaka, Masami Kiyokawa, Koujiro Yamazaki, Hiroshi Okamoto, Akira Kitabatake, Hiroyuki Tsutsui, and Mitsuru Sugawara. Evaluation of effects of polymorphism for metabolic enzymes on pharmacokinetics of carvedilol by population pharmacokinetic analysis. *Biological and Pharmaceutical Bulletin*, 30(3):537–542, 2007.

[35] Mutsuko Honda, Yumi Ogura, Wakako Toyoda, Masato Taguchi, Takashi Nozawa, Hiroshi Inoue, and Yukiya Hashimoto. Multiple regression analysis of pharmacogenetic variability of carvedilol disposition in 54 healthy japanese volunteers. *Biological and Pharmaceutical Bulletin*, 29(4):772–778, 2006.

[36] Nihal El Rouby and Rhonda M Cooper-DeHoff. Genetics of resistant hypertension: a novel pharmacogenomics phenotype. *Current hypertension reports*, 17(9):71, 2015.

[37] Amy I Lynch, Marguerite R Irvin, Barry R Davis, Charles E Ford, John H Eckfeldt, and Donna K Arnett. Genetic and adverse health outcome associations with treatment resistant hypertension in genhat. *International journal of hypertension*, 2013, 2013.

[38] Eun Pyo Hong and Ji Wan Park. Sample size and statistical power calculation in genetic association studies. *Genomics & informatics*, 10(2):117–122, 2012.