DISTRIBUTION AND FREQUENCIES OF MEDICALY RELEVANT MUTATIONS IN THE PUERTO RICAN POPULATION

by

Angelia Caro Monroig

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
BIOLOGY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2018

Approved by:

_____	_____
Juan C. Martínez-Cruzado, PhD Member, Graduate Committee	Date
_____	_____
Audrey Majeske, PhD Member, Graduate Committee	Date
_____	_____
Taras K. Oleksyk, PhD President, Graduate Committee	Date
_____	_____
Ana Velez PhD Chairperson of the Department	Date
_____	_____
Rocio Zapata Medina Graduate School Representative	Date

ABSTRACT

Detection of medically related mutations is fundamental to both the successful treatment and overall improvement in the quality of life for human populations with increased susceptibility to human disorders. On the population level, the identification of people with higher risk for diseases are necessary for maximizing detection and early intervention. The Puerto Rican population is ethnically rich, consisting of a contribution of three ancestral populations from Africa, Europe and native America, and it is expected for inherited disease prevalence to vary greatly due to the local genetic differences. It is useful to map the genetic component of inherited diseases across the island to provide a baseline for future epidemiological studies. In this study, we used the available genome sequences from the 1000 Genomes Project (PUR) and an additional subset of samples representing the entire island's population (LGDS) that was genotyped with Illimina TruSeq exome panel. Replicate analysis was conducted for each municipality to estimate genotyping errors. In the Puerto Rican population from the 1000 Genomes Project as well as in the LGDS cohort, these markers show frequencies close to what is expected in the European population. However, some genes show population frequencies that deviate significantly (>2 s^2 away) from the combined distribution. Most of these exceptions can be explained by the African admixture altering the frequencies of the alleles in Puerto Ricans when the frequencies of those alleles in Africa are highly different to those in Europe, but there are notable exceptions indicating that other evolutionary processes such as drift and selection may also be determining frequencies of the medically relevant markers in the island population of Puerto Rico. This knowledge will be used to develop recommendations for the use of next generation sequencing technology in health service practices, specifically in neonatal testing for the presence of inherited diseases.

RESUMEN

La detección de mutaciones relacionadas a la medicina es fundamental en el tratamiento y la calidad de vida de las poblaciones humanas con una mayor susceptibilidad a estos trastornos. A nivel de la población, la identificación de grupos con mayor riesgo de estas enfermedades es necesaria para maximizar la detección y la intervención temprana de ellas. La población puertorriqueña es étnicamente rica, y consiste en la contribución de las tres poblaciones ancestrales África, Europa y América, y se espera que la prevalencia de la enfermedad hereditaria varíe mucho debido a las diferencias genéticas locales. Es útil mapear el componente genético de las enfermedades hereditarias en toda la isla para proporcionar una base para los futuros estudios epidemiológicos. En este estudio, utilizamos los genomas disponibles del estudio “1000 Genomes Project” (PUR) y un subconjunto adicional de muestras que representan la población de toda la isla (LGDS) que se genotipó con el panel del exoma Illimina TruSeq. Se realizó un análisis replicado para cada municipio para estimar los errores de genotipo. En la población puertorriqueña “1000 Genomes Project y en el conjunto de LGDS, estos marcadores muestran frecuencias cercanas a las esperadas en la población europea. Sin embargo, algunos genes muestran frecuencias de población que se desvían significativamente (> 2 s2 de distancia) de la distribución combinada. La mayoría de estas excepciones pueden explicarse por la mezcla con africanos que altera las frecuencias de los alelos en los puertorriqueños cuando las frecuencias de los alelos en África son marcadamente diferente a Europa, pero hay excepciones notables que indican que otros procesos evolutivos como la deriva y la selección también pueden determinar las frecuencias de los marcadores médicamente relevantes en la población de la isla de Puerto Rico. Este conocimiento se utilizará para desarrollar recomendaciones para el uso de la tecnología de secuenciación de próxima generación en los servicios de salud, específicamente en las pruebas neonatales para la presencia de enfermedades hereditarias.

ACKNOWLEDGMENTS:

I would like to demonstrate sincere gratitude to the influential people in my life. Their continuous motivation served as an inspiration to complete this thesis dissertation.
First off, I would like to thank my family. My parents, Luis Caro and Awilda Monroig, which have always been a tremendous influence in my life. These hard-working individuals taught me to be determined to reach all my dreams. Throughout my graduate studies, they were incredibly supportive, encouraging and were always there to listen when I needed them. Also, I would like to express my sincere gratitude to my younger sister and best friend, Angeliz Caro. Her enthusiasm, determination and life accomplishments were truly inspiring. I could not have reached my goals without her support. Finally, I would like to thank my brother Luis F. Caro for always being there for me when I needed him throughout this process

Second, I would like to acknowledge all those hard-working individuals in my lab. Yashira Afanador, Israel Rivera, Wilfred, Walter Wolfsberger, Kirill Grigorev, Priscila Rodriguez, Frances Marin, Edmundo Torres and Nikole Ayala fellow graduate students that were always there for me throughout the process. Their friendship and laboratory support were hugely influential throughout the years. Also, would like to thank all the students in the LGDS and Biology of Genomes Classes. Their contribution was tremendously valuable for the sampling process.

Third, I would like to thank my advising committee. I would like to demonstrate my most sincere gratitude to my advisor Dr. Taras Oleksyk for taking a leap of faith and accepting me as his graduate student. His guidance has been key in my academic growth. Thank you for being the kind of advisor that turned negatives to positives and always gave me exceptional advice and pep talks whenever needed. Next, I would like to thank his lovely wife and committee member, Dr. Audrey Majeske. She took me under her wing and guided me with the different laboratory techniques. Her help and expertise were essential for my project. Lastly, I would like to thank my third committee member Dr. Juan Carlos Martinez Cruzado for all his help throughout this process. It has truly been an honor collaborating with this committee.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1- INTRODUCTION

The human genome consists of approximately three billion DNA base pairs and is estimated to carry around 20,000 protein-coding genes (Hattori, 2005). Heritable diseases are often associated with mutations in gene sequences, and a change in one of those loci can be their trigger. Human genetics is a branch of science that studies biological inheritance and assesses genomic variation, enabling the description of genetic contributions that cause human diseases. Acknowledging genomic diversity among populations allows the creation of early options for onset detection and prevention of inheritable conditions. A large proportion of population studies concentrate on populations of European and/or African descent, and there is an unbalanced distribution of baseline information for other ethnic groups such as Hispanics (Medicine & Manolio, 2013). Thus, to efficiently correlate European and African populations with Hispanics is unknown (Excoffier, L., Smouse, P. E., & Quattro, 1992). The unique value of studying admixed populations is their heterogeneity that derives from their diverse ancestral origins (Learning, Mckenzie, & Robert, 2014). Admixed populations allow studies of health inequalities that occur through exposure to different environmental and genetic factors in populations. Expanding studies to incorporate diverse admixed populations is needed to identify disease loci that are otherwise difficult to find in the single origin populations. The data generated from this study will help establish baseline information that could be applicable to the Hispanic populations with diverse ancestries formed over the last 400 years on the Caribbean islands, such as Puerto Ricans.

Hispanic Populations:

Puerto Ricans and other Hispanic groups make up a large segment of the U.S. population and therefore must to be considered in health policy decisions. Population estimates published by the U.S. Census Bureau in 2013 established that Hispanic ethnicity constitutes approximately 17% of the total U.S. population making Hispanic origin the nation's most substantial ethnic minority (Week, America, & Salvador, 2014). In that estimation, Mexicans, Puerto Ricans, and Central Americans sum up to 82.4% of all Hispanics living in the U.S. (Dominguez et al., 2015). The large proportion of these minorities justifies the worth of generating reference data to use as a guideline for statistically robust studies. For this reason, four Hispanic populations: MXL- Mexican Ancestry from Los Angeles USA; PUR- Puerto Ricans from Puerto Rico, CLM - Colombians from Medellin, Colombia, PEL - Peruvians from Lima, Peru were amongst the populations selected to create the largest catalog of the human population in a project called 1,000 Genomes (Fig et al., 2015). Through describing the geographic and functional range of human genetic variation, 1000 Genomes project built a global resource to help to understand the genetic contributions to disease and resulted in the identification of variants that occur at a frequency of at least one in 2,508 people. The project focused on 26 different populations and demonstrated that individuals from these locations carry different profiles of rare and common variants (Auton et al., 2015). The results were used to create a public online reference database, which has enabled further genomic studies.

The purpose of this study is to report rare disease mutations not previously published in 1000 Genome project (Consortium, 2012). The Puerto Rican population originated due to historic

admixture between Native Amerindians, whose ancestors had migrated earlier from the Caribbean coast of South America and the Amazon basin, European emigrants from mostly from the Iberian Peninsulawestern Mediterranean, as well as a diversity of sub-Saharan Africans brought by the slave trade (Gravel et al., 2013). As a very distinctive admixed population with European (63.7%), African (21.2%) and Amerindian (15.2%) genetic components creating a unique distribution of heritable diseases across the island (Via et al. 2011) (Pettersson, Lundeberg, & Ahmadian, 2009). Further population studies are necessary to document this uniqueness and diversity and to pave the way to a more comprehensive analysis that would include socioeconomic, historical, evolutionary, environmental and other factors that may contribute to higher incidence of inheritable diseases in Puerto Ricans.

It is estimated that there are 7,300 rare diseases which are inherited through a single gene (Benowitz, 2015). These diseases are transmitted according to a Mendelian pattern of inheritance and are relatively straightforward to assess in case/control studies. However, since these are distributed unevenly across populations, even with new emerging genomic advances, challenges to access these genes are still present. One of the major obstacles is unveiling their presence in populations considering that there are relatively fewer patients that exhibit these conditions. Their absence in the population is due mutation/selection balance: usually purifying selection acts upon populations and eradicates those rare disease alleles (John H. Gillespie, 1998). When these reach low frequencies, it becomes increasingly difficult since most of them would exist in heterozygotes, and therefore would be invisible to the force of selection. Many rare alleles remain hidden but could potentially lead to uncovering functional effects of the disease genes.certain diseases. Their

assessment is crucial and could lead to new insights on biological mechanisms of complex diseases (Beckmann, 2006). Previous studies have demonstrated that the Alzheimer's Disease, maturity-onset diabetes of the young, and familial combined hypertension have also been essential in comprehending and developing themaking a hypothesis for complex diseases such as dementia, diabetes, and dyslipidemias (Peltonen, Perola, Naukkarinen, & Palotie, 2006). Therefore, studies directed towards uncovering rare alleles in admixed populations should be conducted to eliminate the gap of statistical data in the available databases.

Objectives:

This study aims to evaluate the frequencies of the disease alleles in the Puerto Rican population and to make this information accessible to the public for more statistically robust studies. The specific objectives of this study are: 1) assess markers significantly associated with infectious disease phenotypes 2). analisis of single-gene variants associated with increased susceptibility/resistance to pharmacogenomics phenotypes as well as selected markers for variability in drug response. 3) evaluate recent human adaptations that have occurred in the Puerto Rican population. 4). understand the major determinants of the frequencies of the medically related mutations in the Puerto Rican population.

Chapter two is a literature review discussing key concepts of exome sequencing and previous work in populations regarding adaption to environment, susceptibility to infectious disease and pharmacogenomic biomarkers. Chapter three will discuss methods used to prepare the DNA, sequencing platform and assembly and annotation applications. The fourth chapter demonstrates the results. Finally, the fifth chapter provides conclusions of the project.

CHAPTER 2 –LITERATURE REVIEW

Genomics is a relatively new scientific practice. The first DNA sequences were obtained in the 1970s by academic investigators using arduous methods based onincluding two-dimensional chromatography (Koboldt, Steinberg, Larson, Wilson, & Mardis, 2013). Next was the unidimensional thin-gel electrophoresis of radiolabelled fragments generated through chemical or enzymatic synthesis, then the development of fluorescence-based enzymatic sequencing methods combined with capillary electrophoresis, which enabled DNA sequencing to become easier and faster (Wahlberg, Erson, & Uhlen, 1993). Further innovations led to Next Generation Sequencing (NGS) technologies which have revolutionized genetics. The production of an enormous number of low-cost reads makes this approach useful for many applications (Pettersson et al., 2009), and to make research, even more, cost effective, instead of sequencing the entire genomes, researchers can opt to sequence only the coding sequences. Currently there is a dive to make genomes cheaper than $100, but even the study of human genetics in the public databases, notably the public databases, notably the 1000 Genomes Project (Consortium, 2012) and others.

In addition to the whole genome sequencing which is the ultimate source for study individual genome variation, there is a smaller scale alternative, the exome sequencing. Exome sequencing is Exomes are a useful tool when uncovering monogenetic disorders because exomes contain many rare protein-altering variants (Bamshad et al., 2011). Exome sequences cover around 2% of the gene and is a less expensive approach that can nevertheless yield a lot of useful information on functional variation. They may be detected in the form of missense, non-sense, single base

substitutions or small insertions and deletions (indels) and usually present deleterious effects. Through a population genetics approach, the variants assessed through exome sequencing are used to trace heritable, historical, evolutionary and environmental changes that occur. Exome sequences provide a cheaper and more easily interpretable resource for study of human variation that most laboratories can currently access.

Studying Human Genetic Variation

Understanding heritable, historical, evolutionary and environmental changes that occur entails determining genetic variation. *Homo sapiens* is relatively a young species and we have not had enough time to accumulate considerable genetic variation when compared to other species, even to our closest cousins, the chimpanzees (Varki, Geschwind, & Eichler, 2008). Still, significant genetic variation exists amongst humans (Varki et al., 2008). No two humans apart from twins will ever be genetically identical. Humans have been found to have a 1% difference amid their genomes. On a larger scale, however, when comparing genetic variation within and between populations of humans, studies have shown that about 85% of all human genetic variation occurs within human populations, whereas about only 15% of variation exists between populations (Varki et al., 2008).

To understand human variation two significant components are examined, SNPs (Single Nucleotide Polymorphism), and copy number variants (or CNVs). SNPs are changes in a single DNA nucleotide base, that can differ in their frequencies within and among human populations. Studies like 1000 Genomes (Auton et al., 2015), have demonstrated that most common SNPs are shared amongst continents. Studying SNP variation exposes the constant migration and gene flow

of human populations throughout history (Rotimi & Jorde, 2010). Rare mutations, on the other hand, have mutation at levels that are so low, that a given mutation at a position is unlikely to have reoccurred in the small populations of early humans (Draper, 2008). In other words, rare variants are of extreme interest because they are more likely to be population specific because of recent migrations.

To date, millions of SNPs have been found in humans and they have been made available on public online databases such as dbSNP (https://www.ncbi.nlm.nih.gov/SNP/) and the 1000 Genomes browser (http://grch37.ensembl.org/Homo_sapiens/Info/Index). The data generated for the databases through SNP analysis facilitates physical mapping, functional analysis, pharmacogenetic, association studies and evolutionary studies (Sherry, 2002). Apart from SNPs, copy number variants are also used in measuring variation.

A CNV in contrast is a segment of DNA that is a kilobase or larger in length and presents at a variation in comparison with a reference genome. These types of variants can cover millions of bases of DNA, containing entire genes and their regulatory regions. Some of these variants have no phenotypic effects in some regions. Others can afflict gene dosage which could be the root of genetic disease. These genetic disease could be caused alone or in combination with environmental factors (Sebat et al., 2004). CNVs are generally less common the SNPs (Feuk, Carson, & Scherer, 2006).

Evolutionary forces in human populations

The five major forces of evolution are mutations, gene flow, genetic drift, nonrandom mating and natural selection. Together, the interplay of these forces affects in allele frequencies in populations., but the ultimate source of all variation is mutation. A mutation is a random heritable change in a gene or chromosome, resulting from additions, deletions, or substitutions of bases in the DNA sequence. They may create advantageous, deleterious, or neutral traits for the organism (Jobling, Mark, Hurles, 2013).

The mutations could be either hereditary, passed down from the ancestors, or acquired by the organism. Somatic mutations, occur in a person's life in some cells but through environmental factors or through errors in cell division (Jobling, Mark, Hurles, 2013). On the other hand, when referring to germline mutations, we mean that it has been inherited from the parents (sperm/egg) and it is persistent in almost every cell of the body through the person's lifetime. One of the major evolutionary roles of mutations is increasing genetic variation within and amongst populations and could be visualized as the author of genetic variation (**Table 1**).

Rather than increasing genetic variation within populations and reducing divergence between populations, genetic drift decreases genetic variation within population and increases divergence (**Table 1**). Genetic drift causes stochastic changes in allele frequencies in a population due to random fluctuations. When the populations have a limited size, this is the most important force because the lower the size the greater the chance for fixation (Andrews, 2010). The most common form of nonrandom mating (positive assortative mating) tends to result in population substructure, introducing barriers to mating and decreasing population sizes by subdividing larger populations

into smaller compartments. On the other hand, natural selection is the process by which some organisms have a greater chance of surviving and reproducing, and advantageous features are passed on at a higher frequency than less advantageous traits (Maienschein, 2004). In other words, the process of natural selection is going to edit variation and choose the alleles that are going to stay because they are of benefit. Selection may act in three ways, stabilizing, directional or diversifying. Through stabilizing selection, there is a tendency to decrease a population's genetic variance. This occurs when selection favors the average phenotype and selects against extreme variations. In directional selection, a populations genetic variance shifts towards a new phenotype when exposed to environmental changes. Finally, diversifying selection increases genetic variance when natural selection favors two or more extreme phenotypes that have each specific advantage. In diversifying selection, average or intermediate phenotypes are often less fit than either extreme phenotype and are unlikely to feature prominently in the population (Gillespie, 1998). The next evolutionary force is migration, a process that occurs when there is an exchange of genetic material between two populations. Gene flow between two populations tends to decrease genetic variation causing homogenization (Donalson, Daly, Ermini, & Bevitt, 2015).

We can conclude that mutation is the major source of variation. Natural selection works by eliminating deleterious alleles and select for the advantageous ones. Nonrandom mating acts to compartmentalize, while migration acts to homogenize populations. Finally, drift acts a strong force in the smaller populations leading to the random loss of genetic variation and population differentiation. Admixed populations have distinctive patterns in their genome left by history of

migration, random and non-random demographic processes, as well as selection. Thus, a study evolutionary processes in these populations must involve the historical background.

Through a population genetics approach, we can recognize the origin and the amount of genetic variation in a population, and how this variation changes through time. Comprehending how the major evolutionary forces act on genetic diversity enables us to understand modern population histories and composition. With the knowledge of evolutionary forces, we can understand and describe history and diversity in any given population. This work will focus on the history and current genetic diversity in the modern Puerto Rican population in Puerto Rico.

Genetic diversity in Puerto Rico

Like other Caribbean populations, the Puerto Rican population has been formed in the admixture of three ancestral populations: African, European and the Native American. The trihybrid populations from the Americas surfaced because of the admixing between genomes from those of the Old World and the New World. The relocation brought genomes of some ancestral populations from Europe and Africa to external environments in the New World. Simultaneously, this introduced lifestyle changes as well as pathogens, such as smallpox that afflicted people from the Old World (Tang et al., 2007).

In Puerto Ricans, the story of the major admixture began upon the discovery of the island by Christopher Columbus in 1493. The island's settlers at the time were the Taínos that were closely related to the indigenous populations in northwestern Venezuela. Spanish colonization began in 1506 and this process in Puerto Rico caused rapid decline of Taínos due to diseases such as influenza, bubonic plague, smallpox, and pneumonia. Further actions that caused the Taíno decline

were enforced labor, emigration, and war (Via et al., 2011). To strengthen slave labor, captured Africans were imported to the island, which were brought until the middle of 19th century.

The admixture patterns in the genome were not equivalently distributed within the genome of Puerto Ricans. Previous studies demonstrated possible sex-biased gene flow in Caribbean Populations (Moreno-Estrada et al., 2013). They found this by comparing the ancestry proportions of the X chromosome with the autosomes in each population. This highlighted an inclination towards a higher proportion of Native American ancestry on the X chromosome than on the autosomes, where 61 % of the population has Native American ancestry in the maternal line (Martínez-Cruzado et al., 2005). The current genomic admixture in Puerto Rican population is estimated to be on average 64% European, 21% African and 15% native American (Via et al., 2011).

Markers Conferring Susceptibility to Infectious Diseases

It is estimated that infectious diseases take the lives of around 10 million people annually (Hill, 2012), and must constitute one of the most important selective force in the human population. Therefore, scientific studies are needed to contribute to the fabrication of more efficient drugs to prevent or diminish effects of infectious diseases. In many cases, SNPs are a useful tool to underlie differences in our susceptibility to infectious diseases. Two of the major techniques used to asses SNPs associations are candidate gene studies and Genome-wide Association studies (GWAS). In candidate gene studies scientists select relevant genes that have prior implications on functional or biological pathways for a specific disease. In contrast, GWAS studies they scan the whole genome for common genetic variation. Both techniques allow the identification of common and rare

variation underlying susceptibility to infectious disease (Chapman & Hill, 2012). **Tables 2 and 3** demonstrate a list of genome-wide and candidate gene data known to affect infectious disease phenotype. **Table 2** summarizes markers significantly associated with infectious disease phenotypes in genome-wide studies, while **Table 3** lists examples of the single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes.

One example of a successful GWAS study not included in this review is a study of dengue, a global health issue responsible for 100 million infections annually in tropical and subtropical areas. The dengue virus belongs to the flavivirus family and is transmitted to humans mostly by infected *Aedes aegypti* mosquitoes. Patients profiles fluctuate from an asymptomatic infection to rapidly lethal disease. The most recurrent complication is Dengue Shock Syndrome (DSS) that produces an intensification of vascular permeability that causes hypovolaemic shock and often death (Chapman & Hill, 2012). GWAS of Vietnamese children with DSS recognized *MICB* and *PLCE1* susceptibility loci, and proposed that the association between the *MICB* genotype and susceptibility to severe dengue might reflect altered or dysfunctional NK and/or CD8+ T cell activation early in infection that results in a higher viral burden in vivo, a recognized factor in clinical outcome (Khor, Tran, Bich, Pang, & Davila, 2012). The genes mentioned here may be examples of positive selection due to advantage surviving diseases in a local environment. However, in an admixture population like in Puerto Rico, if the local selection has not had sufficient time to act, these frequencies should be altered in accordance to the contributions from the different ancestral populations. In the case of the presence of recent selection, these frequencies should be altered to reflect increase in a protective variant spreading in a population.

Finding a local genetic variant protecting individuals in that population from an infectious disease is one of the most important goals of medical population genetics today.

Pharmacogenetic Markers

In a medical setting, clinicians use immediate information such as basic diagnostic tests, family history, and lifestyle to prescribe patients and alleviate clinical problems. Physicians usually are led by previous clinical trials that have been successful in the general population. In spite of this, not every individual responds in the same manner and at times adverse drug reactions may occur due to different factors (Meyer, Zanger, & Schwab, 2013) (Figure 1). In the United States, adverse side effects from pharmaceutical drugs occur in 2 million people each year and may cause as many as 100,000 deaths, according to the Food and Drug Administration. Costs associated with adverse drug reactions (ADRs) are estimated at $136 billion annually (Adams, 2008). Pharmacogenomics aims to reduce these casualties and analyzes an individual's responses to drugs. Experts in this field, evaluate gene variants disturbing an individual's drug response by pinpointing genetic loci that are associated with known drug responses.

These known loci serve as a reference to assess cases in which individual's drug response is unknown (Adams, 2008). There are two major focus points in this field, first being pharmacokinetics, which studies the dosage of drugs to use to reach optimum efficiency in the body. The second is pharmacodynamics which analyzes how cells such as heart tissue or neurons assimilate a drug (Goldstein, Tate, & Sisodiya, 2003). To study and find associations for drug response SNPs are used. Each combination of SNP is called a haplotype. We obtain one haplotype from the mother, and one from the father and the pair is referred to as a SNP profile. This SNP

profile is used to evaluate the individual's drug response (Jobling, Mark, Hurles, 2013). For Pharmacogenomics, GWAS and candidate gene approaches have been crucial to scientists for distinguishing drug response (Table 5). There is a small proportion of genetic risks that are uncovered in GWAS for common diseases that differs from the results obtained from that of GWAS designed to detect genetic predisposition for ADRs or lack of efficacy of a drug treatment. Resent research has demonstrated that the current genotyping platforms, even those covering > 5 million SNPs, do not completely cover pharmacogenomic variation (Gamazon, Eric R. Skol , Andrew D. and Perera, 2013). However, it is still a useful method and measures should be taken to overcome such limitations. A more powerful approach to discover biomarkers of drug efficacy is focusing on interindividual drug responses. This method has been successful for uncovering chemotherapy biomarkers for cancer patients. Such findings have led to better targeted therapies and a better overall quality of life for these afflicted individuals (Meyer et al., 2013).

GWAS, candidate-gene assessment or individual focused sequencing approaches are useful tools to uncover biomarkers of drug efficacy. Populations such as Puerto Ricans should be examined due to their unique profile caused by admixture. This trihybrid population is likely to contribute significantly to an extensive variability in response to drug therapies, a component that will be missed by traditional studies in more homogeneous populations (Sukumar, Sundar, & Sivarajan, 2010). We have reviewed recent genes that have contributed to the pharmacogenomics and drug response. **Table 4** shows the examples of pharmacogenomic biomarkers with clinical translation while **Table 5** lists genetic markers that have been discovered in the genome-wide association studies on variability in drug response compiled after (Meyer et al., 2013).

An example of marker reviewed in Table 4 are markers within the two genes *CYP2C9* and *VKORC1* involved in warfarin resistance (Johnson et al., 2011). Warfarin is an anticoagulant that helps the blood flow smoothly throughout the body and it is prescribed to prevent thromboembolisms in patients with atrial valve replacement, pulmonary embolism, deep vein thrombosis, myocardial infarction and atrial fibrillation (Claudio-Campos et al., 2017). A recent study by (Claudio-Campos et al., 2017) demonstrated the *CYP2C9* variant *rs2860905* in all the major haplotypes occurring in the Puerto Rican population. This variant showed stronger association with warfarin sensitivity when compared to common variants *CYP2C9∗2* and *CYP2C9∗3*. Such research not only led to the finding of a novel variant, but it also contributed to the description of the structure and LD patterns of *CYP2C9* found in Puerto Ricans. Contributions such as this one allows the establishment of baseline information to fill the gap of knowledge of admixed populations such as the one mentioned. Ours will include pharmacogenomic biomarkers with clinical translation such as the one employed in this research initiative to contribute to the knowledge in this population.

Drug response genes are not usually targets of natural selection, since pharmacological agents are not usually present in natural environments. However, newly admixed populations like that in Puerto Rico are likely to harbor a wide array of genetic variants that have functional effects. Therefore, focusing on Hispanic populations could result in finding novel variants underlying pharmacokinetic and drug response phenotypes

.

Markers Involved in the Adaption to the Environment

Phenotypically, humans demonstrate a wide array of diversity and most of it could be attributed to genetic adaptions caused by environmental pressures. Positive selection is one of the major drivers of evolutionary adaptions. For positive selection to work in organisms they must increase the probability of survival and reproduction. Further, that trait must have the capacity to be passed to their offspring (Schaffner, 2008).

Human populations migrating out of Africa around 100 thousand years ago have encountered several new selective pressures arising from new environments. For example, around 10,000 years ago most human populations changed from a hunter-gather lifestyle to practicing agriculture and pastoralism (Fan, Hansen, Lo, & Tishkoff, 2017). The selective pressures of adapting to an environment and the new nutrition have led to region-specific genetic variants that influence variable phenotypes. In Tables 2 and 3 we can observe a list of candidate genes and GWAS studies that have identified genes of recent human adaptions that have been found in populations studies of Africans, Mexicans, Europeans, East Asians and Puerto Ricans. Identifying these adaptions, allows researchers to learn more about how our species have transformed over time, the obstacles the species have faced and how they overcame them, and about past and present sources of disease. In the next several sections, several well-known examples of adaptations were represented

1.1.1.1 Adaption to diet:

One major example of adaption is the capability to digest lactose, a sugar found in milk (Schaffner, 2008). This ability usually fades before adulthood in most human populations. Nevertheless, for individuals of European ancestry, the ability to break down lactose continues

because of a mutation in the lactase gene (*LCT*). For this reason, it was suggested that the allele became common in Europe due to increased cow milk consumption (Hancock, Alkorta-Aranburu, Witonsky, & Di Rienzo, 2010). Further studies showed that the lactase persistence allele *MCM6* is present in nearly 80% of people of European descent giving evidence of a recent selective sweep in the Northern and Central European population (Schaffner, 2008). This advantageous adaption is an example of convergent evolution denoting that the variant arose independently in geographically diverse populations due to the strong selective pressure (Fan et al., 2017).

1.1.1.2 Adaptions to endemic pathogens:

A further case of selection is the sickle cell allele conferring the resistance to malaria trait in tropical environments such as Sub-Saharan Africa. This mutation causes an error in the fabrication of hemoglobin, the oxygen carrying molecule. The homozygotes with this condition, experience a decrease of oxygen and during long periods of activity, when their erythrocytes take up a sickle shape (Schaffner, 2008). However, the individual must inherit two mutant copies to express the disease, while a heterozygote does not show harmful effects. The mutation is often observed in Sub-Saharan Africa, an area with an increased prevalence of the mosquito-borne parasitic infection, malaria. In this area, the HbA/HbS heterozygote is in fact advantageous to the individual, as it confers resistance to malaria (Fan et al., 2017). The evolution of the sickle cell allele is a textbook example of selection for heterozygote, a type of balancing selection.

1.1.1.3 Adaption to the ultraviolet exposure

Unlike most primates, humans are not shielded by body hair and therefore are not as protected from ultraviolet radiation. Exposure to this radiation was a critical driver in skin pigmentation evolution in humans (Jablonski & Chaplin, 2000), as the adaption to ultraviolet exposure con be

found among human populations living in high exposure environments. Skin color is mainly the result of a pigment called melanin which is present in light and dark-complexioned people. People with light complexioned skin mostly produce pheomelanin, whereas those with dark colored skin mostly produce eumelanin which acts as a protective barrier for the skin preventing DNA damage. Initial studies of human pigmentation were based on candidate genes obtained through observation in model organisms and highly penetrant variants of monogenetic disorders. Through genome-wide association studies scientists uncovered additional candidate loci associated with light complexion such *as TYR, SLC24A5, SLC45A2, OCA2 and TYRP1*. that are found to be affected by selection. (Fan et al., 2017). The alleles in these genes show extremely disjoined distribution, indicating selection pressures at different latitudes.

Studying Populations

Population analysis of the combined evolutionary forces and their role in the change in frequency allows the prediction of long term evolutionary trends. The selection of candidate and GWAS genes in this study will facilitate the way to predict rare disease alleles present in the Puerto Rican population. Focusing on such single-nucleotide variants that occur in populations are significant because they arose relatively recently and are more likely to be population specific. Throughout chapter 2 we highlighted the importance of examining adaption to environment alleles, susceptibility to infectious diseases and pharmacogenetic biomarkers. For our work, we will select genes that were previously studied in other populations and asses them in Puerto Ricans, evaluate their frequencies in the light of their history and the evolutionary forces at play today.

CHAPTER 3 – METHODS

Origin of samples

To address the lack of relevant data available regarding admixed populations such as Hispanics, we decided to assess the Puerto Rican population. A total of ten municipalities were chosen to represent the genetic admixture diversity present across the island (**Figure 2**) based the previous reports of admixture components in these municipalities (Via et al., 2011) (**Figure 2**). Four cities were selected around the blue shaded area that demonstrated strong African descent; four were chosen from the dark orange area that exhibited Native American origin and only two towns were chosen from the purple area to represent European ancestry (see **Figure 2**).

The samples were obtained in the course of a research and educational project supported by a grant from the National Science Foundation. During the course of the implementation of the Local Genomic Diversity Studies (LGDS) project, hundreds of undergraduates were trained and supervised by graduate students to adequately collect saliva samples and extract DNA. The project aimed to extract and genotype DNA from 96 individuals from each of the 78 municipalities of Puerto Rico. However, for this study we used 12 individuals per municipality from 10 different municipalities summing up a total of 120 samples.

In addition, we used publicly available genome sequences (https://www.coriell.org/1/NHGRI/Collections/1000-Genomes-Collections/Puerto-Rican-in-Puerto-Rico-PUR) in a set of samples called "**Puerto Ricans in Puerto Rico**", which were designated as PUR. The PUR samples were collected from mother-father-adult child trios for

which at least six of the eight great-grandparents of the child in the trio were Puerto Rican, and represent an average set of Puerto Rican people in Puerto Rico (https://www.coriell.org/1/NHGRI/Collections/1000-Genomes-Collections/Puerto-Rican-in-Puerto-Rico-PUR). living on the island of Puerto Rico.

1.1.1.4 DNA extraction

During the implementation of the LGDS project, the 10 mL saliva samples (from 8mL dH2O + 2mL 100% EtOH) were spun for 10 minutes at 10,000rpm until the debris was concentrated at the bottom. Then, the supernatant was poured and discarded. Then the pellet was resuspended in 200 µL of DNAzol and 10 µL of Proteinase K (20 mg/ml) and incubated the samples overnight in the water bath at 37◦ C. The next day the samples were spun for 10 minutes at maximum speed (13,200-14,000 rpm). Once finished, 200 µL of 100%ethanol was added and the tubes were inverted 5-8 times and placed at -20C for at least 6 hours. After the given time, the samples were then spun at a maximum speed (13,200-14,000 rpm) for 15 minutes. The supernatant was poured off by decanting and 250 µL of ethanol 70% were added to wash the pellet. Each sample was then transferred to the centrifuge and was set for 5 minutes at maximum speed. Once the round finished, the samples were washed once again using 250 µL of ethanol 70% and were centrifuged at maximum speed for 5 minutes. The supernatant was discarded by decanting and all the leftover ethanol was removed with a pipette without disturbing the pellet. The microtubes were left on the hood to dry for a minimum of one hour and were then resuspended in 200 µL TE. The samples were stored at -20˚C.

To assure the quality of samples for genotyping, only those samples that had a concentration of 50 ng/µl and a 260/280 ratio between 1.6-1.9 were used for this project. Upon finding the

optimal quality samples, they were each diluted to 50 ng/µl. For each township, the twelve samples from each town were pooled together and genotyped.

1.1.1.5 Illumina Sequencing

Exome Sequencing of the LGDS samples was performed at the Laboratory of Translational Genomics, National Cancer Institute in Gaithersburg, MD. The adapter-ligated genomic DNA libraries were prepared with the TruSeq DNA Preparation Kit (Illumina, San Diego, CA, USA) as suggested in the manufacturer's protocol. Then they were amplified by ligation-mediated PCR, purified with the QIAquick PCR Purification kit (Qiagen, Valencia, CA, USA), and evaluated using electrophoresis. Next, exome enrichment was performed with NimbleGen's SeqCap EZ Human Exome Library v2.0 which targeted 44.1 Mb exonic sequences (Roche NimbleGen, Inc., Madison, WI, USA). Libraries were hybridized with the EZ Exome Probe Library. Then DNA was washed and recovered as explained in the NimbleGen SeqCap EZ Library SR protocol. Exome-enriched libraries were amplified by ligation-mediated PCR, purified, and assessed as above. The remaining post capture enriched multiplexed sequencing libraries were used in cluster formation on an Illumina cBOT. Paired-end sequencing was performed using an Illumina HiSeq following Illumina-provided protocols for 2×100-cycle sequencing. The exomes were sequenced to adequate depth to accomplish a minimum threshold of 80% of coding sequence covered with at least 15 reads.

Exome Analysis

To begin annotation the Human Reference genome and transcript annotation were downloaded from the UCSC database (http://genome.ucsc.edu/). Reads were then aligned to the reference genome using Burrows-Wheeler Aligner (Li & Durbin, 2009). The duplicates were marked, and base quality scores were recalibrated. We used GATK (DePristo et al., 2011) and Lofreq (Wilm et al., 2012) programs filtered to filter out common variants that were represented at not more than 10% in 1,000 Genomes data (Auton et al., 2015). GATK was used for variant discovery and genotype calling of substitutions in multiple alleles, insertions and deletions were achieved on all individuals simultaneously using GATK (GATK, http://www.broadinstitute.org/gatk/) (DePristo et al., 2011) with the minimum call quality parameter set to 50. LoFreq (Wilm et al., 2012) was used to predict Single Nucleotide Variants. This assessed base-call qualities and other sources of errors in sequencing such as mapping or base and indel alignment uncertainty.). Reads were then aligned to the reference genome using BWA. The duplicates were marked, and base quality scores were recalibrated. We used GATK (DePristo et al., 2011) and Lofreq (Wilm et al., 2012) programs filtered to filter out common variants that were (represented at no more than 10% in 1,000 Genomes data (Auton et al., 2015). Variant discovery and genotype calling of substitutions in multiple alleles, insertions and deletions were achieved on all individuals simultaneously using GATK (GATK, http://www.broadinstitute.org/gatk/) (DePristo et al., 2011) with the minimum call quality parameter set to 50. LoFreq (Wilm et al., 2012) was used to predict Single Nucleotide Variants (SNVs). This assessed base-call qualities and other sources of errors in sequencing such as mapping or base and indel alignment uncertainty.

Predicting Amino Acid sequences that affect function

We used Sorting Intolerant from Tolerant (SIFT) algorithm (http:// sift.jcvi.org)(Kumar, Henikoff, & Ng, 2009) to predict substitutions that affect protein function. Provean predicted the functional impact for all classes of protein sequence variations for not only single amino acid substitutions but also insertions, deletions, and multiple substitutions (Reference). Polyphen-2 (Adzhubei et al., 2010) (http://genetics.bwh.harvard.edu/pph2) was used to predict the possible impact of an amino acid substitution on the structure and function of a human protein. Automated predictions of this kind are essential for interpreting large datasets of rare genetic variants, which have many applications in modern human genetics research. The database for nonsynonymous SNPs functional predictions (dbNSFP)(Liu, Jian, & Boerwinkle, 2011) was used to identify mutations that cause human diseases, especially rare Mendelian diseases. The results of different methods aimed at predicting the outcome of missense SNVs was combined into an integrated output (González-Pérez & López-Bigas, 2011). SIFT searched the protein databases using PSI-BLAST algorithm to find related sequences. After, SIFT calculated the conservation value and scaled probability for each position.

Candidate marker population frequencies comparisons

We compiled lists of markers in genes significantly associated with infectious disease phenotypes obtained from genome-wide studies, as well as single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes (Chapman and Hill, 2012). Along with these we also included a list of pharmacogenomic biomarkers with clinical

translation as well as the markers associated by GWAS with the variability in drug response (Meyer et al., 2013). Among these we identified the medically related alleles that were significantly different in our study populations (LGDS and PUR) compared to the world reference populations Europe (EUR), East Asia (EAS), and Mexico (MEX).

CHAPTER – 4: RESULTS

1.1.1.6 Frequency of markers suspected for associations with infectious diseases in the Puerto Rican population

There was no difference between the frequency of markers suspected for associations with infectious diseases for the PUR sample data set compared to the combined LGDS dataset (r^2=0.94, Figure 4). Generally, the frequencies in Puerto Rican populations (both PUR and LGDS) also agreed with the European reference (EUR, Figures 4 and 5), but there were certain exceptions, which are outlined below. Some of these differences can be due to the local differences the reference population in Europe from the 1000 Genome Project and the ancestral European component in Puerto Ricans. While it is possible, we do not have conclusive evidence that a fraction of these may be unique to the Puerto Rican population. However, some of these examples do warrant further investigation in a larger population context.

We have compiled the population frequencies among the 32 markers in 17 genes suspected for the association with infectious disease phenotypes in genome-wide studies based on the review in Chapman and Hill (2012). **Table 7** shows minor allele frequency data for Puerto Ricans (PUR), Mexicans (MXL), Europeans (EUR), East Asians (EAS) and Africans (AFR). The table only displays those markers available in the genomes of PUR samples from the 100Genomes (Consortium, 2012). Furthermore, 10 of these markers have been verified in the LGDS pools (see **Verification in LGDS).**

Among **markers suspected to be involved in the infectious diseases** by GWAS (Table 3), the largest differences in Minor Allele Frequency (MAF) is found in *rs3771317* in the ***STAT1*** containing an allele responsible for impaired *IFNγ* response (van de Vosse, van Dissel, & Ottenhoff, 2009) (S.-Y. Zhang et al., 2008) and may play a role in susceptibility to primary biliary cirrhosis (Mells et al., 2011). The frequency of the C allele is 26%, which is higher than that in the European (EUR) population – 14%. In turn, Asian (EAS) and Mexican (MEX) population both have higher proportions of this allele 26% and 31% respectively. Another marker displaying high differences with the European population is *rs6107516* located in the ***PNRP*** gene which is involved in the post-translational modification of host cellular prion protein, and has been shown to be a risk factor for variant Creutzfeldt-Jakob disease (Mead et al., 2009). The minor allele in this locus shows the highest frequency in PUR among all the populations compared (34%, **Table 7**).

rs3771317 marker in ***STAT1*** is also mentioned **among the single-gene variants associated with increased susceptibility or resistance to specific infectious disease** phenotypes presented in **Table 8**, but there are several other loci displaying high level of divergence between populations. In another notable example, we saw high differences with the European populations are observed in the *rs2224234*, and rs573617232 in ***TRAF3*** gene which has been associated with the susceptibility to HSV encephalitis (Pérez de Diego et al., 2010). However, in this case, the population that carries the highest frequency of the minor allele is in Mexico and Africa. Likewise, *rs4251424* locus in ***IRAK4***, a gene implicated in the susceptibility to invasive disease is polymorphic in Puerto Ricans (MAF= 14%) but completely homozygous in EUR, EAS and MEX.

The only other population containing the minor allele is Africa where it is at high frequency (67%). A similar situation is with the rs2814778 in ***DARC***, a gene implicated in the resistance to Plasmodium vivax, and also displaying very high frequency in Africa, during to the local adaptation there (Miller, Mason, Clyde, & McGinniss, 1976). It is likely that these elevated frequency difference is due to the African admixture present in the Puerto Rican population (Via et al., 2011). However, the admixture alone may not be able to explain a very high difference in allele frequency in *rs2546890* located in ***IL12B***, a gene implicated in many deficiencies of human immune signaling during infectious disease (van de Vosse et al., 2009)(S. Zhang et al., 2008). The minor allele in this locus is found at 42% frequency in PUR compared to only 14% in EUR and 31% in MEX, where the frequency in AFR is also comparatively low (26%). Unfortunately, we could not confirm frequencies of any of the markers reported by the whole genome sequencing (WGS) in the LGDS panel.

1.1.1.7 Pharmacogenetic marker frequencies in Puerto Ricans.

Among the **pharmacogenomic biomarkers with clinical translation** (**Table 9**), the largest deviation from the European frequencies (36% in PUR vs 13% in EUR) can be found in *rs887829* located in ***UGT1A1*** gene and associated with neutropenia while administering Irinotecan (Innocenti et al., 2009). Among other, rs3757322 in the estrogen receptor that has been associated with the efficacy of Fulvestrant Tamoxifen in breast cancer patients (Croxtall & McKeage, 2011) is also showing a lower frequency in PUR (41% in EUR vs 30% in PUR). The rs723527 in the epidermal growth factor receptor (*EGFR*), a gene that plays a key role in tumor evolution, proliferation and immune evasion, and is one of the most important targets for biological therapy,

especially for non-small-cell lung cancer (NSCLC) and colorectal cancer (CRC) (Troiani et al., 2016) has been associated with the efficacies of Cetuximab, Erlotinib, Gefitinib, and Panitumumab (Garrett & Eng, 2011) (Troiani et al., 2016), and also is among the genes showing high differences between Europeans and Puerto Ricans (57% in EUR vs 40% in PUR). Finally, *rs4802101* in *CYP2D6*, a gene implicated in the efficacy of Efavirenz (EFV) a nonnucleoside reverse transcriptase inhibitor (NNRTI) used in the treatment of patients with human immunodeficiency virus (HIV) infection(Tanner & Tyndale, 2017). It is likely that all these elevated frequencies are due to the African admixture present in the Puerto Rican population, since in each case the frequency in the African population is high (**Table 9**).

Among the **markers identified by the genome-wide association studies on variability in drug response (Table 10),** the largest difference can be observed in *rs582297* located in ATM serine/threonine kinase. This gene is associated with the glycemic response to metformin in type 2 diabetes (Zhou et al., 2011). In addition, the *rs11001819* in C10orf11 associated with the individual variability in clinical outcomes in breast cancer treatment with the drug tamoxifen shows lower frequency in PUR than in EUR. Both changes in frequency are consistent with the contribution of African admixture.

Finally, we have surveyed 39 loci previously associated with human adaptations based on a recent comprehensive review (Fan et al., 2017), and verified frequencies for 11 of these in LGDS cohort. Among these there are genes involved in diet, lifestyle, altitude and solar radiation. Among these, Puerto Ricans carry alleles that differ in frequency from the European population in adult lactose persistence (*MCM6*: rs4988235, and rs182549), fatty acid biosynthesis,

decreased cholesterol and increased triglycerides (*FAD*: rs74771917, rs3168072, rs12577276, rs7115739, rs174602, rs174570), dark eyes (*OCA2*: rs12913832), skin pigmentation *(MATP (SLC45A2*): rs16891982 and *SLC24A5*: rs1426654), as well as thick hair and showed shaped teeth (*EDAR*: rs365060, rs3827760) (**Table 11**). Most of these changes in frequencies are consistent with the contribution of African admixture to the European background.

Verification in the pools of the LGDS cohort

Markers analyzed in **Tables 7-11** have been independently reassessed in the LGDS pools – combined samples from selected Puerto Rican Municipalities (**Figure 2**). We were only able to confirm a small proportion of the markers: some of these were ablest from the exome because PUR samples included full genome sequences, while LGDS samples contained variation present exclusively in the exons. Nevertheless, among the markers associated with susceptibility to heritable diseases in Puerto Ricans we identified genotypes and estimated population frequencies in 10 loci (**Table 12**). Unfortunately, none of the genotypes for the single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes that were already genotyped in PUR, were present in the Puerto Rican samples (**Table 13**), and for the pharmacogenetic marker frequencies in Puerto Ricans we were only able to confirm frequencies for two markers **(Table 14).** Finally, we were able to estimate genotype frequencies for the 11 markers associated with adaptive traits in the Puerto Rican population **(Table 15).** There is a strong correlation between the two independent estimates of genotype frequencies at these loci in Puerto Ricans from Puerto Rico (PUR) estimated in the 1000Genomes project (Auton et al., 2015)and the Local Genome Diversity Studies (LGDS) (**Figure 4**)

CHAPTER 5 - DISCUSSION

In this study we analyzed frequencies markers involved in infectious disease, pharmacogenomics association, drug response and natural selection to illustrate how admixture has shaped the landscape of the medically related inherited alleles in Puerto Rico, and to search for any indication of other demographic or evolutionary forces that may be at play on the island. We compiled lists of markers in genes significantly associated with infectious disease phenotypes in genome-wide studies, as well as single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes (Chapman and Hill, 2012), along with the list of pharmacogenomic biomarkers with clinical translation as well as the markers associated by GWAS with the variability in drug response (Meyer et al., 2013). To characterize the frequencies of these loci in the population of Puerto Rico, we used two datasets: 1) the PUR samples from 1000 Genome project (Auton et al., 2015) as well as the LGDS cohort collected in this study and sequenced with exome panel TruSeq (Illumina, San Diego, CA, USA). In the Puerto Rican population from the 1000 Genomes Project as well as in the LGDS cohort, these markers show frequencies close to what is expected in the European population. However, some genes show population frequencies that deviate significantly (>2 s^2 away) from the combined distribution. Most of these exceptions can be explained by the African admixture altering the frequencies of the alleles of ancestry-informative markers in Puerto Ricans. However, there are common trends and some exceptions we would like to address for each group of the loci reviewed.

Response to the infectious disease should have a structured additive genetic component and is expected to vary across the populations due to variations in the history of migrations, epidemics

and local environments. In this study, we compiled lists of genetic markers in genes significantly associated with infectious disease phenotypes in genome-wide studies, as well as single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes (Chapman and Hill, 2012). These markers must be under a strong selection pressure as the diseases are one of the most important evolutionary agents in human populations today. Consistent with this expectation, the allele frequencies in Puerto Rican population were most similar with the European population, as this population contributed the most to the modern genetic mix on the island (64%, Via et al. 2011). The deviations from the European frequencies in other genes like *STAT1*, *TRAF3*, *IRAK4*, and *DARC* could be explained by the addition of the African admixture, given that the African populations have very different allele frequencies at thesethe(Pérez de Diego et al., 2010) is loci (Pérez de Diego et al., 2010). However, there was one notable outlier: 42% frequency in PUR compared to only 14% in EUR in the allele frequency in *rs2546890* located in *IL12B*, a gene implicated in many deficiencies of human immune signaling during infectious disease (van de Vosse et al., 2009) (S. Zhang et al., 2008).

The drug response can also be moderated by genetic markers. We compiled a list of pharmacogenomic biomarkers with clinical translation as well as the markers associated by GWAS with the variability in drug response (Meyer et al., 2013). These markers in this reviewed set(Meyer et al., 2013) are not expected to be influenced by natural selection directly as much as those loci in the two previous categories, as recently developed drugs are not natural substances acting persistently on populations, but could rise and fall in frequencies due to the random mutations and drift. This may be the reason why the differences frequencies of pharmacogenomic

biomarkers in the three populations are distributed more evenly, and if there any changes in frequency of the alleles reported they are always consistent with the contribution of African admixture to the European background.

The frequencies of genes under selection are expected to vary greatly between populations due to the variable selective pressures applied by differences in environments encountered by people. We selected genes that have been previously reported by other studies reported as targets of natural selection (reviewed in Fan et al., 2016). These adaptations come in three categories: adaptations to diet, climate, and UV radiation, which are summarized in Table 3. Overall, we observed the same trend in this dataset as in the previous two, indicating that the frequencies in Puerto Rico depend on the percentages of admixture between the ancestral populations. The biggest deviations from the European frequencies were in the categories in which Africans and Europeans deviated as well, indicating that the differences can be explained purely by the admixture proportions.

Frequencies of the medically related SNPs and the Ancestry estimates.

Overall, our results indicate that the frequencies of most of our medically related and evolutionary significant SNPs are a product of mixed ancestry. Among all the markers we have interrogated using genome wide and exome wide data, there is only three possible SNP whose frequency may not be reasonably explained by the effect of admixture between African, European and the Native American ancestral components. In fact, we conclude that the frequencies of all medically related markers can be reasonably predicted using average ancestral proportions estimated earlier for the overall Puerto Rican population (**Figure 6**). However, this relationship changes in specific Puerto Rican municipalities where the ancestral contributions deviate from that

of the average estimates for this island population. This is the main practical and useful contribution to the understating of distribution of the medically and evolutionary relevant markers in the modern Puerto Rican population.

Conclusions

A growing contingent of healthcare researchers are advocating for the personalized approach to medicine using genomics data for tailoring treatment of diseases in which genetic variants confer either the susceptibility to the disease or response to the treatment. The search is on for tailored treatment that will work better in particular populations. For the lack of genome wide data, the Puerto Rican genomes have never been interrogated for the presence of medically significant variants. This project has made the first step in making the personalized approach in genomic medicine for this part of the world.

Table 1. Effects of evolutionary forces in populations. from Donalson, P., Daly, A., Ermini, L., & Bevitt, D. (2015). Genetic Diversity. *Genetic of Complex Disease*, 1–34

		Within Populations	**Between Populations**
Increase variation	**genetic**	• Migration • Mutation • Natural (diversifying) selection	• Genetic drift • Natural selection • Mutation
Decrease variation	**genetic**	• Genetic drift • Natural (directional) selection	• Natural selection • Migration

Table 2. Markers significantly associated with infectious disease phenotypes in genome-wide studies. From Chapman and Hill. 2012. Human genetic susceptibility to infectious disease. *Nature Reviews Genetics* 13, 175–188 (2012)

Disease	Phenotype	Population	Sample size	Most significant marker	Gene(s)	P-value	Odds Ratio	Reference # in review	Reference
HIV-1 and AIDS	Viral load at set point	European	2,554	*rs9264942*	*HLA-C*	5.9×10–32	NA	33,34	*Fellay, J. et al. Common genetic variation and the control of HIV-1 in humans. PLoS Genet. 5, e1000791 (2009).*
				rs2395029	*HLA-B, HCP5*	4.5×10–35	NA	33,34	*Fellay, J. et al. A whole-genome association study of major determinants for host control of HIV-1. Science 317, 944–947 (2007).*
	Viral load at set point‡	African American	515	*rs2523608*	*HLA-B*	5.6×10–10	NA	38	*Pelak, K. et al. Host determinants of HIV-1 control in African Americans. J. Infect. Dis. 201, 1141–1149 (2010).*
	HIV-1 control‡	European	1,712	*rs9264942*	*HLA-C*	2.8×10−35	2.9	35	*The International HIV Controllers Study. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330, 1551–1557 (2010).*
				rs4418214	*MICA*	1.4×10−34	4.4		
				rs2395029	*HLA-B, HCP5*	9.7×10−26	5.3		
				rs3131018	*PSORS1C3*	4.2×10−16	2.1		
		African American	1233	*rs2523608*	*HLA-B*	8.9×10−20	2.6		
				rs2255221	*Intergenic*	3.5×10−14	2.7		
				rs2523590	*HLA-B*	1.7×10−13	2.4		
				rs9262632	*Intergenic*	1.0×10−8	3.1	33,34	*Fellay, J. et al. Common genetic variation and the control of HIV-1 in humans. PLoS Genet. 5, e1000791 (2009).*
	Disease progression‡	European	1071	*rs9261174*	*ZNRD1, RNF39*	1.8×10–8	NA		
	Progression to AIDS 1987‡	European American	755	*rs2395029*	*HLA-B, HCP5*	6.8×10–10	3.47	42	*Limou, S. et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS genomewide association study 02). J. Infect. Dis. 199, 419–426 (2009).*

Disease	Phenotype	Population	Sample size	Most significant marker	Gene(s)	P-value	Odds Ratio	Reference # in review	Reference
	Long-term no progression‡	European	1,627	*rs11884476*	*PARD3B*	3.4×10–9	NA	41	*Troyer, J. L. et al. Genome-wide association study implicates PARD3B-based AIDS restriction. J. Infect. Dis. 203, 1491–1502 (2011).*
	Long-term no progression‡	European	1911	*rs2234358*	*CXCR6*	9.7×10–10	1.85	43	*Limou, S. et al. Multiple-cohort genetic association study reveals CXCR6 as a new chemokine receptor involved in long-term no progression to AIDS. J. Infect. Dis. 202, 908–915 (2010).*
Hepatitis C	Spontaneous clearance	European	1,362	*rs8099917*	*IL28B*	6.1×10–9	2.31	53	*Rauch, A. et al. Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. Gastroenterology 138, 1338–1345 (2010).*
Hepatitis B	Chronic infection	Japanese, Taiwanese	6,387	*rs3077*	*HLA-DPA1*	2.3×10–38	0.56	60	*Kamatani, Y. et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nature Genet. 41, 591–595 (2009).*
				rs9277535	*HLA-DPB1*	6.3×10–39	0.57		
Dengue	Dengue shock syndrome	Vietnamese	8,697	*rs3132468*	*MICB*	4.4×10–11	1.34	65	*Khor, CC. et al. Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. Nature Genet. 43, 1139–1141 (2011).*
				rs3765524	*PLCE1*	3.1×10–10	0.8		
Severe malaria	Susceptibility	African (Gambian)	5,900	*rs11036238*	*HBB*	3.7×10–11	0.63	70	*Jallow, M. et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nature Genet. 41, 657–665 (2009).*
Tuberculosis	Susceptibility	African (Ghana, Gambia, Malawi)	11,425	*rs4334126*	*18q11.2 (GATA6, CTAGE1, RBBP8, CABLES1)*	6.8×10–9	1.19	72	*Thye, T. et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nature Genet. 42, 739–741 (2010).*
Leprosy	Susceptibility	Chinese	11,140	*rs3764147*	*LACC1*	3.7×10–54	1.68	76	*Zhang, F. R. et al. Genomewide association study of leprosy. N. Engl. J. Med. 361, 2609–2618 (2009).*
				rs9302752	*NOD2*	3.8×10–40	1.59		
				rs3088362	*CCDC122*	1.4×10–31	1.52		
				rs602875	*HLA-DR-DQ*	5.4×10–27	0.67		
				rs6478108	*TNFSF15*	3.4×10–21	1.37		
				rs42490	*RIPK2*	1.4×10–16	0.76		

Disease	Phenotype	Population	Sample size	Most significant marker	Gene(s)	P-value	Odds Ratio	Reference # in review	Reference
Meningococcal disease	Protection	European	7,522	*rs1065489*	*CFH*	2.2×10–11	0.64	85	*Davila, S. et al. Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. Nature Genet. 42, 772–776 (2010).*
Variant Creutzfeldt – Jakob disease	Susceptibility	European, Papua New Guinea	5,183	*rs426736* *rs1799990*	*CFHR3* *PRNP*	4.6×10–13 2.0×10–27	0.63 NA	91	*Mead, S. et al. Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. Lancet Neurol. 8, 57–66 (2009).*

Abbreviations: CABLES1, CDK5 and ABL enzyme substrate 1; CCDC122, coiled-coil domain-containing 122; CFH, complement factor H; CFHR3, CFH-related 3; CTAGE1, cutaneous T cell lymphoma-associated antigen 1; CXCR6, chemokine (C-X-C motif) receptor 6; GATA6, GATA-binding protein 6; HBB, haemoglobin beta; HCP5, HLA complex P5 (non-protein-coding); HLA, human leukocyte antigen; IL28B, interleukin-28B; LACC1, laccase (multicopper oxidoreductase) domain-containing 1; MIC, MHC class I polypeptide-related sequence; NA, not applicable or not provided in publication; NOD2, nucleotide-binding oligomerization domain- containing 2; PARD3B, par-3 partitioning-defective 3 homologue B; PLCE1, phospholipase C, epsilon 1; PRMT6, protein arginine methyltransferase 6; PRNP, prion protein; PSORS1C3, psoriasis susceptibility 1 candidate 3 (non-protein coding); RBBP8, retinoblastoma-binding protein 8; RIPK2, receptor-interacting serine–threonine kinase 2; RNF39, ring finger protein 39; TNFSF15, tumour necrosis factor [ligand] superfamily member 15; ZNRD1, zinc ribbon domain-containing 1. *Results are those reported by the study authors. ‡Phenotype definition is provided in the study reference.

Table 3. Single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes. From From Chapman and Hill. 2012. Human genetic susceptibility to infectious disease. *Nature Reviews Genetics* 13, 175–188 (2012)

Disease agent	Phenotype	Mechanism	SNP location	Reference # in review	Disease associated SNPs from NHGRI-EBI Catalog	Reference (s)
Streptococcus pneumoniae	Susceptibility to invasive disease	Impaired TLR–IL-1R signalling	*IRAK4*	94, 100, 103	rs4251424, rs1816854	Picard, C. *et al.* Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. *Medicine* 89, 403–425 (2010). Picard, C. et al. Pyogenic bacterial infections in humans with IRAK-4 deficiency. Science 299, 2076–2079 (2003). Medvedev, A. E. et al. Distinct mutations in IRAK-4 confer hyporesponsiveness to lipopolysaccharide and interleukin-1 in a patient with recurrent bacterial infections. J. Exp. Med. 198, 521–531 (2003).
			MYD88	94, 97	rs3925158	Picard, C. et al. Clinical features and outcome of patients with IRAK-4 and MyD88 deficiency. Medicine 89, 403–425 (2010). von Bernuth, H. et al. Pyogenic bacterial infections in humans with MyD88 deficiency. Science 321, 691–696 (2008).
Neisseria meningitidis	Susceptibility to invasive disease	Membrane attack complex deficiency	*C5-C9*	15, 87	rs429017, rs13157656, rs7713972, rs1901167, rs1013579	Brouwer, M. C. et al. Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. Lancet Infect. Dis. 9, 31–44 (2009). Degn, S. E., Jensenius, J. C. & Thiel, S. Disease-causing mutations in genes of the complement system. Degn, S. E., Jensenius, J. C. & Thiel, S. Disease-causing mutations in genes of the complement system. Skattum, L., van Deuren, M., van der Poll, T. & Truedsson, L. Complement deficiency states and associated infections. Mol. Immunol. 48, 1643–1655 (2011). Am. J. Hum. Genet. 88, 689–705 (2011). Jonsson, G. et al. Hereditary C2 deficiency in Sweden: frequent occurrence of invasive infection, atherosclerosis, and rheumatic disease. Medicine 84, 23–34 (2005).
		Properdin deficiency	*CFP*	167-169		Sprong, T. et al. Deficient alternative complement pathway activation due to factor D deficiency by 2 novel mutations in the complement factor D gene in a family with meningococcal infections. Blood 107, 4865–4870 (2006). Fijen, C. A., Kuijper, E. J., te Bulte, M. T., Daha, M. R. & Dankert, J. Assessment of complement deficiency in patients with meningococcal disease in The Netherlands. Clin. Infect. Dis. 28, 98–105 (1999). Skattum, L., van Deuren, M., van der Poll, T. & Truedsson, L. Complement deficiency states and associated infections. Mol. Immunol. 48, 1643–1655 (2011). Skattum, L., van Deuren, M., van der Poll, T. & Truedsson, L. Complement deficiency states and associated infections. Mol. Immunol. 48, 1643–1655 (2011).

Disease agent	Phenotype	Mechanism	SNP location	Reference # in review	Disease associated SNPs from NHGRI-EBI Catalog	Reference (s)
		Factor D deficiency	*CFD*			
Encapsulated bacteria (for example, S. pneumoniae, N. meningitidis and Haemophilus influenzae)	Susceptibility to invasive disease	Classical complement pathway deficiency	*C2, C1Q, C1R, C1S, C4*	15, 87, 169, 170		Brouwer, M. C. et al. Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. Lancet Infect. Dis. 9, 31–44 (2009). Skattum, L., van Deuren, M., van der Poll, T. & Truedsson, L. Complement deficiency states and associated infections. Mol. Immunol. 48, 1643–1655 (2011).
Mycobacteria	Susceptibility (MSMD)	Impaired IFNγ response	*IFNGR1, IFNGR2, STAT1*	95, 138	rs117070989, rs13201877, rs2284553, rs3771317, rs7582694, rs10168266	van de Vosse, E., van Dissel, J. T. & Ottenhoff, T. H. Genetic deficiencies of innate immune signalling in human infectious disease. *Lancet Infect. Dis.* 9, 688–698 (2009). Zhang, S.-Y. et al. Inborn errors of interferon (IFN)- mediated immunity in humans: insights into the respective roles of IFN-α/β, IFN-γ, and IFN-λ in host defense. Immunol. Rev. 226, 29–40 (2008).
		Impaired IFNγ production	*IL12B, IL12RB1, NEMO*		rs12188300, rs2546890, rs376008, rs12984174	
		Impaired macrophage respiratory burst	*CYBB*	144		Bustamante, J. et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. Nature Immunol. 12, 213–221 (2011).
		Impaired differentiation of dendritic cell subgroups	*IRF8*	145	rs113899791, rs8064111	Hambleton, S. et al. IRF8 mutations and human dendritic-cell immunodeficiency. N. Engl. J. Med. 365, 127–138 (2011).

Disease agent	Phenotype	Mechanism	SNP location	Reference # in review	Disease associated SNPs from NHGRI-EBI Catalog	Reference (s)
HIV-1	Resistance	Absence of coreceptor for pathogen	*CCR5*	26, 46, 47	rs333	Dean, M. et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Science 273, 1856–1862 (1996). Liu, R. et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86, 367–377 (1996). Samson, M. et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382, 722–725 (1996).
HSV-1	Susceptibility to HSV encephalitis	Impaired production of IFNα, IFNβ and/or IFNλ	*UNC93B1*	118		Casrouge, A. *et al.* Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* 314, 308–312 (2006).
			TLR3	119	chr4:187239569	Zhang, S. Y. et al. TLR3 deficiency in patients with herpes simplex encephalitis. Science 317, 1522–1527 (2007).
			TRAF3	120	rs2224234, rs573617232, rs9989163	Perez de Diego, R. *et al.* Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity* 33, 400–411 (2010).
Human herpesvirus-8	Classic Kaposi's sarcoma	T cell deficiency	*STIM1*	171	rs11030122	Byun, M. *et al.* Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J. Exp. Med.* 207, 2307–2312 (2010).
Human papillomaviruses	Epidermodysplasia verruciformis	Unknown	*TMC6, TMC8*	172	rs2748424, rs2748425, chr17:76124810	Ramoz, N. *et al.* Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. *Nature Genet.* 32, 579–581 (2002).
Norovirus	Resistance	Absence of receptor for pathogen	*FUT2*	173,174	rs1047781, rs516316, rs2287921	Lindesmith, L. et al. Human susceptibility and resistance to Norwalk virus infection. Nature Med. 9, 548–553 (2003). Thorven, M. et al. A homozygous nonsense mutation (428G-->A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. J. Virol. 79, 15351–15355 (2005).
Plasmodium vivax	Resistance	Absence of coreceptor for pathogen	*DARC*	175	rs2518564, rs16827466, rs2814778	Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. N. Engl. J. Med. 295, 302–304 (1976).
Candida albicans	Chronic mucocutaneous candidiasis	Impaired IL-17 immunity	*CARD9*	176	rs10781499, rs4077515	Glocker, E. O. *et al.* A homozygous *CARD9* mutation in a family with susceptibility to fungal infections. *N. Engl. J. Med.* 361, 1727–1735 (2009).

Disease agent	Phenotype	Mechanism	SNP location	Reference # in review	Disease associated SNPs from NHGRI-EBI Catalog	Reference (s)
			IL17RA	177	rs140221307, rs2241047, rs3827278, rs140221307	Puel, A. *et al.* Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science* 332, 65–68 (2011).
			IL17F	177	rs763780	Puel, A. *et al.* Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science* 332, 65–68 (2011).
			STAT1	126, 127	rs3771317, rs10168266, rs1517352, rs60976990, rs12468579	Liu, L. *et al.* Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *J. Exp. Med.* 208, 1635–1648 (2011). van de Veerdonk, F. L. et al. STAT1 mutations in autosomal dominant chronic mucocutaneous candidiasis. N. Engl. J. Med. 365, 54–61 (2011).
Bovine spongiform encephalopathy prion	Susceptibility to variant Creutzfeldt–Jakob disease	Post-translational modification of host cellular prion protein	*PRNP*	88, 91	rs1799990, rs6107516, rs6116492	Zeidler, M., Stewart, G., Cousens, S. N., Estibeiro, K. & Will, R. G. Codon 129 genotype and new variant CJD. *Lancet* 350, 668 (1997). Mead, S. et al. Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. Lancet Neurol. 8, 57–66 (2009).

Abbreviations: C1, C2, C4, C5–C9, complement components; CARD9, caspase recruitment domain family member 9; CCR5, chemokine (C-C motif) receptor 5; CFD, complement factor D; CFP, complement factor properdin; CYBB, cytochrome b-245, beta polypeptide (encodes GP91PHOX); DARC, Duffy blood group chemokine receptor; FUT2, fucosyltransferase 2; HSV, herpes simplex virus; IFNγ, interferon-γ; IFNGR, IFNγ receptor; IL, interleukin; IL12RB1, IL-12 receptor β1; IRAK4, interleukin-1 receptor-associated kinase 4; IRF8, interferon regulatory factor 8; MSMD, Mendelian susceptibility to mycobacterial disease; MYD88, myeloid differentiation primary response gene 88; NEMO, nuclear factor-κB essential modulator; PRNP, prion protein; STAT1, signal transducer and activator of transcription 1; STIM1, stromal-interaction molecule 1; TLR, Toll-like receptor; TMC, transmembrane channel-like (also known as EVER genes); TRAF3, TNF receptor-associated factor 3; UNC93B1, unc-93 homologue B1

Table 4. Pharmacogenomic biomarkers with clinical translation. Compiled after Meyer et al., 2013. Omics and Drug Response. Annu. Rev. Pharmacol. Toxicol. 53:475–502.

Drug(s)	Indication	Drug-response phenotype	Disease associated SNPs from NHGRI-EBI Catalog	Gene (s) reported in review	Reference(s) in review	Reference(s)
Imatinib Dasatinib Nilotinib	CML	Efficacy	rs9620247, rs77049423, rs117252219	*BCR-ABL*	146–148	Blay J-Y, von Mehren M. 2011. Nilotinib: a novel, selective tyrosine kinase inhibitor. Semin. Oncol. 38(Suppl. 1):S3–9 147. Druker BJ. 2006. Circumventing resistance to kinase-inhibitor therapy. N. Engl. J. Med. 354:2594–96 148. McCormack PL, Keam SJ. 2011. Dasatinib: a review of its use in the treatment of chronic myeloid leukaemia and Philadelphia chromosome-positive acute lymphoblastic leukaemia. Drugs 71:1771–95
Vemurafenib	Malignant melanoma	Efficacy	rs9648716, rs35407685, rs79811809, rs9648716, rs17623382	*BRAF*	149	SosmanJA, KimKB, SchuchterL, GonzalezR, PavlickAC, et al. 2012. Survival in BRAFV600 – mutant advanced melanoma treated with vemurafenib. *N. Engl. J. Med.* 366:707–14
Maraviroc	HIV-1	Efficacy	rs333, rs1800024	*CCR5 (tropism)*	150	Perry CM. 2010. Maraviroc: a review of its use in the management of CCR5-tropic HIV-1 infection. *Drugs* 70:1189–213
Warfarin	Venous thrombosis, stent thrombosis	Efficacy, adverse reactions	rs9923231, rs1057910, rs1934963, rs1799853, rs9923231, rs10871454, rs749671	*CYP2C9, VKORC1*	151	Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, et al. 2011. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin. Pharmacol. Ther.* 90:625–29
Clopidogrel	Antiplatelet therapy	Efficacy	rs7915414	*CYP2C19*	152, 153	Scott SA, Sangkuhl K, Gardner EE, Stein CM, Hulot J-S, et al. 2011. Clinical Pharmacogenetics Im- plementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin. Pharmacol. Ther.* 90:328–32 Johnson JA, Roden DM, Lesko LJ, Ashley E, Klein TE, et al. 2012. Clopidogrel: a case for indication- specific pharmacogenetics. Clin. Pharmacol. Ther. 91:774–76

Drug(s)	Indication	Drug-response phenotype	Disease associated SNPs from NHGRI-EBI Catalog	Gene (s) reported in review	Reference(s) in review	Reference(s)
Codeine	Pain	Efficacy (i.e., no analgesic effect in carriers of low- or no-activity alleles). Adverse reactions (i.e., respiratory depression in carriers of gene duplications)	rs1065852, rs1058172, rs1065852,	*CYP2D6*	154	Crews KR, Gaedigk A, Dunnenberger HM, Klein TE, Shen DD, et al. 2012. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. *Clin. Pharmacol. Ther.* 91:321–26
Cetuximab Erlotinib Gefitinib Panitumumab	CRC NSCLC NSCLC CRC	Efficacy	rs75061358, rs2736100, rs75061358, rs723527,	*EGFR*	155—157	Argiles G, Dienstmann R, Elez E, Tabernero J .2012. Panitumumab: a summary of clinical development in colorectal cancer and future directions. Future Oncol. 8:373–89 Garrett CR, Eng C. 2011. Cetuximab in the treatment of patients with colorectal cancer. Expert Opin. Biol. Ther. 11:937–49 157. Pal SK, Figlin RA, Reckamp K. 2010. Targeted therapies for non–small cell lung cancer: an evolving landscape. Mol. Cancer Ther. 9:1931–44
Crizotinib	NSCLC	Efficacy	rs11903143, rs2339469, rs13411840, rs13029602, rs7578465	*EML4-ALK*	158	Curran MP. 2012. Crizotinib: in locally advanced or metastatic non-small cell lung cancer. Drugs 72:99– 107
Fulvestrant Tamoxifen	Breast cancer	Efficacy	rs9397437, rs60954078, rs3757322	*Estrogen receptor*	159, 160	Croxtall JD, McKeage K.2011.Fulvestrant: are view of its use in the management o fhormone receptor - positive metastatic breast cancer in postmenopausal women. Drugs 71:363–80 Jordan VC. 2008. Tamoxifen: catalyst for the change to targeted therapy. Eur. J. Cancer 44:30–38

Drug(s)	Indication	Drug-response phenotype	Disease associated SNPs from NHGRI-EBI Catalog	Gene (s) reported in review	Reference(s) in review	Reference(s)
Exemestane Letrozole	Breast cancer	Efficacy	rs1256061, rs58262369, rs10895140	*Estrogen and progesterone receptor*	161, 162	Deeks ED, Scott LJ. 2009. Exemestane: are view of its use in postmenopausal women with breast cancer. Drugs 69:889–918 Barnadas A, Este´vez LG, Lluch-Herna´ndez A, Rodriguez-Lescure A, Rodriguez-Sanchez C, Sanchez-Rovira P. 2011. An overview of letrozole in postmenopausal women with hormone-responsive breast cancer. Adv. Ther. 28:1045–58
Nitrofurantoin Rasburicase	Urinary infection Hyperuricemia	Hemolysis	rs1050828, rs762516	*G6PD*	163	McDonagh EM, Thorn CF, Bautista JM, Youngster I, Altman RB, Klein TE. 2012. PharmGKB sum-mary: very important pharmacogene information for G6PD. Pharmacogenet. Genomics 22:219–28
Trastuzumab Lapatinib	Breast cancer	Efficacy	rs2517959, rs3135718, rs3135724, rs55756123, rs166870, rs10825036	*HER2/ErbB2*	103	Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, Gianni L. 2012. Treatment of HER2- positive breast cancer: current status and future perspectives. Nat. Rev. Clin. Oncol. 9:16–32
Abacavir	HIV-1 infection	Immunologic hypersensitivity reaction, including SJS/TEN	rs2395029, rs10937275, rs1497546, rs6582630, rs10812428	*HLA-B∗5701*	164	Martin MA, Klein TE, Dong BJ, Pirmohamed M, Haas DW, et al. 2012. Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. Clin. Pharmacol. Ther. 91:734–38
Carbamazepine	Epilepsy, bipolar affective disorder, neuralgias	Immunologic hypersensitivity reaction, including SJS/TEN in some Asian populations		*HLA-B∗1502*	165	Chen P, Lin J-J, Lu C-S, Ong C-T, Hsieh PF, et al. 2011. Carbamazepine-induced toxic effects and HLA-B∗1502 screening in Taiwan. N. Engl. J. Med. 364:1126–33
Carbamazepine	Epilepsy, bipolar affective disorder, neuralgias	Immunologic hypersensitivity reaction, including SJS/TEN in		*HLA-B∗3101*	35, 36	Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, et al. 2011. Genome-wide association study identifies HLA-A∗3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. Hum. Mol. Genet. 20:1034–41 36. McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperavic˘iu¯te˙ D, et al. 2011. HLA-A∗3101 and

		Europeans, Japanese				carbamazepine-induced hypersensitivity reactions in Europeans. N. Engl. J. Med. 364:1134–43
Peginterferon-α	HCV infection	Efficacy	rs8099917, rs12979860, rs12980275	*IL28B*	166	Ghany MG, Nelson DR, Strader DB, Thomas DL, SeeffL B .2011. Anupdateontreatment of genotype 1 chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the Study of Liver Diseases. Hepatology 54:1433–44
Imatinib	GIST	Efficacy	rs218238, rs218237, rs172629, rs725344	*KIT (C-KIT)*	167	Corless CL, Barnett CM, Heinrich MC. 2011. Gastrointestinal stromal tumours: origin and molecular oncology. Nat. Rev. Cancer 11:865–78
Simvastatin	High cholesterol	Myotoxicity	rs4149056, rs1871395, rs12317268	*SLCO1B1*	39	Niemi M, Pasanen MK, Neuvonen PJ. 2011. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol. Rev. 63:157–81
Mercaptopurin Thioguanine	ALL, AML, IBD	Myelotoxicity	rs73726531, rs1142345	*TPMT*	169	Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui C-H, et al. 2011. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine
Irinotecan	CRC	Neuropenia	rs6742078, rs887829	*UGT1A1*	170	Innocenti F, Kroetz DL Schuetz E, Dolan ME, Ram´ırez J, et al. 2009. Comprehensive pharmacogenetic analysis of irinotecan neutropenia and pharmacokinetics. J. Clin. Oncol. 27:2604–14

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CML, chronic myeloid leukemia; CRC, colorectal cancer; CYP, cytochrome P450; EGFR, epidermal growth factor receptor; GIST, gastrointestinal stromal tumor; HCV, hepatitis C virus; HIV-1, human immunodeficiency virus type 1; HLA, human leukocyte antigen; IBD, inflammatory bowel diseases; NSCLC, non-small-cell lung carcinoma; SJS, Stevens-Johnson syndrome; TEN, toxic epidermal necrolysis.

Table 5. Genome-wide association studies on variability in drug response. Compiled after Meyer et al., 2013. Omics and Drug Response. Annu. Rev. Pharmacol. Toxicol. 53:475–502

Clinical Problem	Drugs involved	Disease associated SNPs from NHGRI-EBI Catalog (if available)	Gene variant(s) tagged by SNP	Reference(s) in review	Number of published studies	Reference(s)
DILI	Ximelagatran Flucloxacillin Lumiracoxib Amoxicillin-clavulanate		HLA class I and II variants	34	4	Daly AK, Day CP. 2012. Genetic association studies in drug-induced liver injury. Drug Metab. Rev. 44:116–26
Hypersensitivity reaction (SJS, TEN)	Carbamazepine		HLA-B∗ 1502, HLA-B∗ 3101	35–37	3	Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, et al. 2011. Genome-wide association study identifies HLA-A∗3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. Hum. Mol. Genet. 20:1034–41 McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperavic˘iu¯te˙D, et al. 2011. HLA-A∗3101 and carbamazepine-induced hypersensitivity reactions in Europeans. N. Engl. J. Med. 364:1134–43 Shen Y, Nicoletti P, Floratos A, Pirmohamed M, Molokhia M, et al. 2012. Genome-wide association study of serious blistering skin rash caused by drugs. Pharmacogenomics J. 12:96–104
Myotoxicity	Simvastatin	rs4149056, rs1871395, rs12317268	*SLCO1B1*	38, 39	1	Link E, Parish S, Armitage J, Bowman L, Heath S, et al. 2008. SLCO1B1 variants and statin-induced myopathy—a genomewide study. N. Engl. J. Med. 359:789–99 Niemi M, Pasanen MK, Neuvonen PJ. 2011. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol. Rev. 63:157–81
Lack of efficacy	Clopidogrel	rs1074145, rs137891020, rs7915414, rs12777823	*CYP2C19*	40	1	ShuldinerAR,O'ConnellJR,BlidenKP,GandhiA,RyanK,etal.2009.Association ofcytochromeP450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. JAMA 302:849–57

Clinical Problem	Drugs involved	Disease associated SNPs from NHGRI-EBI Catalog (if available)	Gene variant(s) tagged by SNP	Reference(s) in review	Number of published studies	Reference(s)
Efficacy of treatment of HCV infection (viral RNA in serum)	Peginterferon-α	rs8099917, rs12979860, rs12980275	*IL28B*	41–43	3	Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, et al. 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature 461:399–401 Suppiah V, Moldovan M, Ahlenstiel G, Berg T, Weltman M, et al. 2009. IL28B is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. Nat. Genet. 41:1100–4 Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, et al. 2009. Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. Nat. Genet. 41:1105–9
Variable individual dose requirement of coumarin anticoagulants	Warfarin Acenocoumarol Phenprocoumon	rs9923231, rs1057910, rs10871454, rs749671, rs1799853, rs2108622	*VKORC1, CYP2C9, CYP4F2*	44–47	3	Cooper GM, Johnson JA, Langaee TY, Feng H, Stanaway IB, et al. 2008. A genome-wide scan forcommon genetic variants with a large influence on warfarin maintenance dose. Blood 112:1022–27 Takeuchi F, McGinnis R, Bourgeois S, Barnes C, Eriksson N, et al. 2009. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. PLoS Genet. 5:e1000433 Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RHN, et al. 2009. A genome-wide association study of acenocoumarol maintenance dosage. Hum. Mol. Genet. 18:3758–68 Teichert M, Eijgelsheim M, Uitterlinden AG, Buhre PN, Hofman A, et al. 2011. Dependency of phen- procoumon dosage on polymorphisms in the VKORC1, CYP2C9, and CYP4F2 genes. Pharmacogenet. Genomics 21:26–34
Individual variability in glycemic response in type 2 diabetes	Metformin	rs582297, rs1801516	*ATM (?)*	48	1	Zhou K, Bellenguez C, Spencer CCA, Bennett AJ, Coleman RL, et al. 2011. Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. Nat. Genet. 43:117–20
Individual variability in clinical outcomes in breast cancer	Tamoxifen	rs11593840, rs11001819, rs11001765, rs10509373	*C10orf11*	49	1	Kiyotani K, Mushiroda T, Tsunoda T, Morizono T, Hosono N, et al. 2012. Agenome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. Hum. Mol. Genet. 21:1665–72
Individual variability in clearance and toxicity	Methotrexate	rs4149056, rs1871395, rs12317268	*SLCO1B1*	50	1	Trevin˜o LR, Shimasaki N, Yang W, Panetta JC, Cheng C, et al. 2009. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. J. Clin. Oncol. 27:5972–78

Only studies with highly significant association and with replication are listed. Abbreviations: ATM, ataxia telangiectasia mutated; CYP, cytochrome P450; DILI, drug-induced liver injury; HCV, hepatitis C virus; HLA, human leukocyte antigen; SJS, Stevens-Johnson syndrome; TEN, toxic epidermal necrolysis.

Table 6. Recent human adaptations. Compiled after Fan et al., 2016. Going global by adapting local: A review of recent human adaptation. Science Vol. 354, Issue 6308, pp. 54-59

Gene	SNP	Genotype	Main geographical area of repartition	Phenotype	Populations	Reference link to the evidence of phenotype
Adaptations to diet						
References to evidence of selection: *Enattah et al., Nat. Genet. 30, 233–237 (2002); Tishkoff et al., Nat. Genet. 39, 31–40 (2007); Voight, S. Kudaravalli, X. Wen, J. K. Pritchard, PLOS Biol.; Williamson et al., PLOS Genet. 3, e90 (2007); Ranciaro et al., Am. J. Hum. Genet. 94, 496–510 (2014); Enattah et al., Am. J. Hum. Genet. 82, 57–72 (2008); Plantinga et al., Eur. J. Hum. Genet. 20, 778–782 (2012): Mathieson et al., Nature 528, 499–503 (2015).*						
MCM6	rs4988235	CT	-13.910:T	Lactase persistance (TT can digest milk (combined with −22.018:A) −13.915:G)	Eurasia, North Africa, and Central Africa Middle East	*https://www.nature.com/articles/ng826*
	rs41525747	CG	-13.907:G	Lactase persistance	East Africa (Ethiopia and Sudan), Afro-Asiatic Beja	*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3641376/*
	rs869051967	GT	-14.009:G	Lactase persistance	East Africa (Ethiopia and Sudan), Somali camel herders	*https://link.springer.com/article/10.1007%2Fs00239-009-9301-y*
	rs145946881	GC	-14.010:C	Lactase persistance	East Africa (Kenya and Tanzania) and South Africa	*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3641376/*
	rs41380347	TG	-13915:G	Lactase persistance	Sub-Saharan African populations, Arab Bedouin populations	*http://www.bioone.org/doi/full/10.3378/027.086.0101*
	rs41525747	CG	-13907:C	Lactase persistance	Sub-Saharan African populations, Maasai	*http://journals.plos.org/plosone/article/citation?id=10.1371/journal.pone.0044751*
	rs182549	GA	-22018:A	Lactase persistance	European Caucasian populations, Northern China	*http://www.tandfonline.com/doi/full/10.3109/00365520903414176*
		GA	-13838:G	Lactase persistance	Tibetan populations	*https://www.nature.com/articles/jhg201241*
Adaptations to cold climate:						
References to evidence of selection: *Cardona, A., et al. (2014). Genome-Wide Analysis of Cold Adaptation in Indigenous Siberian Populations. PLoS ONE, 9(5), e98076. http://doi.org/10.1371/journal.pone.0098076*						
LRP5	rs4988321		667:A	Reduced bone density, osteoporosis	Siberia	*https://jamanetwork.com/journals/jama/fullarticle/181639*

Gene	SNP	Genotype	Main geographical area of repartition	Phenotype	Populations	Reference link to the evidence of phenotype
	rs3736228		1330:T	Reduced bone density, osteoporosis	Siberia	*https://jamanetwork.com/journals/jama/fullarticle/181639*
CPT1A	rs2924679	GA	57618:A	Energy/Metabolim	Siberia	*https://www.ncbi.nlm.nih.gov/pubmed/23251661?dopt=Abstract*
THADA	rs6732426	CT	T	straighter hair in Europeans	Siberia	*http://www.cell.com/ajhg/fulltext/S0002-9297(09)00464-9*
PRKG1	rs397515330	GA	530:A	Aneurism	Siberia	*http://europepmc.org/articles/PMC3738837*
References to evidence of selection: *Fumagalli et al., Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science 349, 1343–1347 (2015).*						
FAD 2-3	**rs7115739**	GT		fatty acid biosynthesis	Greenland	*http://www.cell.com/ajhg/fulltext/S0002-9297(12)00158-9*
	rs174570	CT		decreased cholesterol and increased triglycerides	Greenland	*http://www.sciencemag.org/content/349/6254/1343.full*
Adaptations to UV radiation						
MATP (SLC45A2)	rs16891982	CG	1122:G	Light skinned Europeans	Europeans	*https://link.springer.com/article/10.1007%2Fs00414-006-0112-z*
	rs6867641	GA	-1169:G	Light skinned Europeans	Europeans	*http://onlinelibrary.wiley.com/resolve/doi?DOI=10.1002/humu.20504*
	rs13289	CG	-1721:C	Light skinned Europeans	Europeans	*http://www.jidonline.org/article/S0022-202X(15)34487-0/fulltext*
HERC2	rs12913832	AG	+874:G	Blue Eyed Europeans		*https://link.springer.com/article/10.1007%2Fs00439-007-0460-x*
OCA	rs1800407	AG	419:A	Green eyed	Europeans	*http://cebp.aacrjournals.org/content/11/8/782.long*
TYR	**rs1126809**	GA	205:A	Red heared Europeans, sun sensitivity	Europeans	*https://www.nature.com/articles/ng.161*

Table 7. Compilation of markers associated with infectious disease phenotypes in genome-wide studies. This table shows minor allele frequency data for Puerto Ricans (PUR), Mexicans (MXL), Europeans (EUR), East Asians (EAS) and Africans (AFR). See Table 2 for the references. From Chapman and Hill. 2012. Human genetic susceptibility to infectious disease. *Nature Reviews Genetics* 13, 175–188 (2012)

Gene	Disease associated SNPs from NHGRI-EBI Catalog	SNP%	PUR	MXL	EUR	EAS	AFR
C5-C9	rs429017	G/A	0.28	0.27	0.25	0.34	0.58
	rs13157656	A/C	0.15	0.22	0.22	0.21	0.03
	rs7713972	T/A	0.17	0.31	0.18	0.50	0.19
	rs1901167	A/G	0.24	0.18	0.25	0.41	0.35
	rs1013579	T/C	0.03	0.05	0.03	0.00	0.001
IFNGR1, IFNGR2, STAT1	rs13201877	A/G	0.11	0.12	0.13	0.05	0.02
	rs2284553	G/A	0.32	0.41	0.38	0.38	0.05
	rs3771317	T/C	0.26	0.31	0.14	0.39	0.26
	rs7582694	G/C	0.26	0.38	0.23	0.35	0.18
	rs10168266	C/T	0.24	0.37	0.18	0.33	0.14
IL12B, IL12RB1, NEMO	rs2546890	G/A	0.42	0.31	0.49	0.41	0.35
	rs376008	C/T	0.24	0.16	0.31	0.37	0.18
	rs12984174	T/C	0.13	0.06	0.13	0.09	0.09
	rs9989163	G/A	0.41	0.48	0.46	0.56	0.41
STIM1	rs11030122	C/G	0.25	0.23	0.33	0.35	0.12
TMC6, TMC8	rs2748424	C/G	0.29	0.17	0.19	0.22	0.46
	rs2748425	G/C	0.29	0.17	0.19	0.22	0.48
FUT2	rs516316	G/C	0.50	0.32	0.44	0.00	0.49
	rs2287921	T/C	0.43	0.26	0.48	0.02	0.12
DARC	rs2518564	A/G	0.28	0.19	0.19	0.43	0.96
CARD9	rs10781499	G/A	0.44	0.54	0.40	0.32	0.25
	rs4077515	C/T	0.44	0.54	0.40	0.32	0.25
IL17RA	rs140221307						
	rs2241047	G/C	0.53	0.50	0.53	0.22	0.54
	rs3827278	C/A	0.23	0.17	0.18	0.29	0.05
IL17F	rs763780	T/C	0.09	0.08	0.06	0.16	0.09
STAT1	rs3771317	T/C	0.26	0.48	0.14	0.39	0.26
	rs10168266	C/T	0.24	0.37	0.18	0.33	0.14
	rs1517352	C/A	0.48	0.50	0.40	0.48	0.04
	rs12468579	A/G	0.49	0.36	0.45	0.66	0.52
PRNP	rs1799990	A/G	0.43	0.38	0.33	0.02	0.35
	rs6107516	G/A	0.34	0.27	0.23	0.02	0.22

Table 8. Frequencies of selected single-gene variants associated with increased susceptibility or resistance to specific infectious disease phenotypes. This table shows minor allele frequency data for Puerto Ricans (PUR), Mexicans (MXL), Europeans (EUR), East Asians (EAS) and Africans (AFR). After Chapman and Hill. 2012.

Phenotype	Gene	SNPs from NHGRI-EBI Catalog	PUR	MXL	EUR	EAS	AFR
Susceptibility to invasive disease	*IRAK4*	rs4251424	0.14	0.00	0.00	0.00	0.67
		rs1816854	0.21	0.07	0.19	0.09	0.38
	MYD88	rs3925158	0.10	0.09	0.12	0.08	0.48
	C5-C9	rs429017	0.28	0.27	0.25	0.34	0.58
		rs13157656	0.15	0.22	0.22	0.21	0.03
		rs7713972	0.17	0.31	0.18	0.50	0.19
		rs1901167	0.24	0.18	0.25	0.41	0.35
		rs1013579	0.03	0.05	0.04	0.00	0.001
Susceptibility (MSMD)	*IFNGR1,*	rs117070989,	0.005	0.00	0.01	0.00	0.001
	IFNGR2,	rs2284553	0.32	0.41	0.38	0.38	0.05
	STAT1	rs3771317	0.26	0.31	0.14	0.39	0.26
		rs7582694	0.26	0.38	0.23	0.35	0.18
		rs10168266	0.24	0.37	0.18	0.33	0.14
	IL12B,	rs12188300					
		rs2546890,	0.42	0.31	0.14	0.39	0.26
	IL12RB1,	rs376008	0.24	0.16	0.31	0.37	0.18
	NEMO	rs12984174	0.14	0.06	0.13	0.09	0.09
	IRF8	rs8064111	0.09	0.09	0.13	0.001	0.003
Susceptibility to HSV encephalitis	*TRAF3*	rs2224234,	0.36	0.46	0.23	0.38	0.56
		rs573617232,	0.38	0.45	0.24	0.37	0.62
		rs9989163	0.41	0.48	0.46	0.56	0.41
Classic Kaposi's sarcoma	*STIM1*	rs11030122	0.25	0.23	0.33	0.35	0.12
Epidermodysplasia verruciformis	*TMC6,*	rs2748424,	0.29	0.17	0.19	0.22	0.46
	TMC8	rs2748425	0.29	0.17	0.19	0.22	0.48
Resistance to Norovirus	*FUT2*	rs1047781,	0.00	0.008	0.00	0.44	0.00
		rs516316,	0.50	0.32	0.44	0.00	0.49
		rs2287921	0.43	0.26	0.48	0.02	0.12
Resistance to Plasmodium vivax	*DARC*	rs2518564,	0.28	0.19	0.19	0.43	0.96
		rs16827466,					
		rs2814778	0.14	0.03	0.006	0.00	0.96
Chronic mucocutaneous candidiasis	*CARD9*	rs10781499	0.44	0.54	0.40	0.32	0.25
		rs4077515	0.44	0.54	0.40	0.32	0.25
	IL17RA	rs2241047,	0.53	0.50	0.53	0.22	0.54
		rs3827278,	0.23	0.17	0.18	0.29	0.05
		rs140221307	0.00	0.00	0.004	0.00	0.00
	IL17F	rs763780	0.09	0.09	0.06	0.15	0.09
	STAT1	rs3771317	0.26	0.48	0.14	0.329	0.26
		rs60976990	0.17	0.27	0.08	0.26	0.33
		rs12468579	0.49	0.36	0.45	0.66	0.52
Susceptibility to variant Creutzfeldt–Jakob disease	*PRNP*	rs1799990	0.43	0.38	0.33	0.02	0.35
		rs6107516,	0.34	0.27	0.23	0.02	0.22
		rs6116492	0.04	0.02	0.02	0.001	0.23

Table 9. Frequencies of selected pharmacogenomic biomarkers with clinical translation in Puerto Ricans From Puerto Rico (PUR) and other populations. Compiled after Meyer et al., 2013. Omics and Drug Response. *Annu. Rev. Pharmacol. Toxicol.* 53:475–502.

Drug(s)	Drug-response phenotype	SNPs from NHGRI-EBI Catalog	Genes/ biomarkers	SNP	PUR	MXL	EUR	EAS	AFR
Vemurafenib	Efficacy	rs9648716	*BRAF*	A/T	0.23	0.09	0.15	0.18	0.74
		rs79811809		A/G	0.08	0.05	0.09	0.00	0.003
		rs9648716		A/T	0.23	0.09	0.15	0.18	0.74
		rs17623382		T/G	0.10	0.07	0.12	0.00	0.009
Maraviroc	Efficacy	rs333	*CCR5*						
		rs1800024							
Warfarin	Efficacy, adverse reactions	rs9923231	*CYP2C9, VKORC1*	C/T	0.39	0.47	0.39	0.88	0.05
		rs1934963		T/C	0.22	0.13	0.20	0.09	0.21
		rs10871454		C/T	0.39	0.48	0.39	0.89	0.06
		rs749671		G/A	0.38	0.46	0.38	0.88	0.05
Codeine	increased drug clearance	rs34223104		T/C		0.14	0.01	0.00	0.04
Efavirenz	Extensive apparent clearance	rs4802101	*CYP2D6*	C/T	0.30	0.23	0.41	0.34	0.06
Cetuximab Erlotinib Gefitinib Panitumumab	Efficacy	rs2736100	*EGFR*	A/C	0.51	0.44	0.50	0.41	0.47
		rs723527		G/A	0.40	0.31	0.57	0.03	0.26
Crizotinib	Efficacy	rs11903143	*EML4-ALK*	A/G	0.22	0.14	0.26	0.16	0.35
		rs2339469		G/A	0.34	0.53	0.34	0.61	0.06
		rs13029602		T/C	0.36	0.41	0.42	0.22	0.39
		rs7578465		C/T	0.40	0.35	0.43	0.68	0.14
Fulvestrant Tamoxifen Exemestane Letrozole	Efficacy	rs3757322	*Estrogen receptor*	T/G	0.41	0.18	0.30	0.34	0.50
		rs1256061	*Estrogen and progesterone receptor*	G/T	0.44	0.37	0.47	0.53	0.31
		rs10895140		G/A	0.40	0.46	0.36	0.54	0,40
Trastuzumab Lapatinib	Efficacy	rs3135718	*HER2/ErbB2*	T/C	0.43	0.48	0.43	0.40	0.63
		rs3135724		C/T	0.21	0.38	0.24	0.29	0.14
		rs166870		C/T	0.32	0.33	0.29	0.17	0.28
		rs10825036		T/G	0.19	0.11	0.29	0.26	0.02
Abacavir	Immunologic hypersensitivity, including SJS/TEN	rs10937275	*HLA-B∗5701*	G/A	0.11	0.09	0.11	0.00	0.14
		rs10812428	*HLA-B∗5701*	C/T	0.39	0.46	0.40	0.33	0.47
Peginterferon-α	Efficacy	rs8099917	*IL28B*	T/G	0.19	0.35	0.17	0.08	0.04
		rs12980275		A/G	0.35	0.46	0.30	0.09	0.56
Imatinib	Efficacy	rs218238	*KIT (C-KIT)*	A/T	0.27	0.35	0.21	0.36	0.77
		rs218237		C/T	0.16	0.28	0.11	0.35	026
		rs172629		C/G	0.15	0.27	0.11	0.35	0.21
		rs725344		C/T	0.15	0.12	0.11	0.04	0.63
Simvastatin	Myotoxicity	rs4149056	*SLCO1B1*	T/C	0.12	0.08	0.16	0.12	0.01
		rs1871395		A/G	0.15	0.11	0.18	0.45	0.22
		rs12317268		A/G	0.15	0.11	0.18	0.45	0.21
Irinotecan	Neuropenia	rs6742078	*UGT1A1*	G/T	0.36	0.37	0.30	0.13	0.47
		rs887829		C/T	0.36	0.37	0.13	0.30	0.49

Table 10. Frequencies of markers from the genome-wide association studies on variability in drug response in Puerto Ricans From Puerto Rico (PUR) and other populations. Compiled after Meyer et al., 2013. Omics and Drug Response. *Annu. Rev. Pharmacol. Toxicol.* 53:475–502

Drug(s)	Clinical Problem	Most significant marker	Gene variant(s) tagged by SNP	SNP	PUR	MXL	EUR	EAS	AFR
Simvastatin	Myotoxicity	rs4149056	*SLCO1B1*	T/C	0.12	0.08	0.16	0.12	0.01
		rs1871395		A/G	0.15	0.11	0.18	0.45	0.22
		rs12317268		A/G	0.15	0.11	0.18	0.45	0.21
Clopidogrel	Lack of efficacy	rs12777823	*CYP2C19*	G/A	0.13	0.13	0.15	0.31	0.25
Peginterferon-α	Efficacy of treatment of HCV infection (viral RNA in serum)	rs8099917	*IL28B*	T/G	0.19	0.35	0.17	0.08	0.04
		rs12980275		A/G	0.35	0.45	0.30	0.09	0.56
Warfarin Acenocoumarol Phenprocoumon	Variable individual dose requirement of coumarin anticoagulants	rs9923231	*VKORC1, CYP2C9, CYP4F2*	C/T	0.39	0.47	0.39	0.88	0.05
		rs10871454		C/T	0.39	0.48	0.39	0.89	0.06
		rs749671		G/A	0.37	0.27	0.38	0.88	0.44
Metformin	Individual variability in glycemic response in type 2 diabetes	rs582297	*ATM*	C/G	0.38	0.42	0.62	0.39	0.67
Tamoxifen	Individual variability in clinical outcomes in breast cancer	rs11593840	*C10orf11*	A/G	0.37	0.27	0.41	0.18	0.44
		rs11001819		G/A	0.35	0.30	0.47	0.16	0.31
		rs10509373		T/C	0.39	0.36	0.43	0.04	0.02
Methotrexate	Individual variability in clearance and toxicity	rs4149056	*SLCO1B1*	T/C	0.12	0.08	0.16	0.12	0.01
		rs1871395		A/G	0.15	0.11	0.18	0.45	0.22
		rs12317268		A/G	0.15	0.11	0.18	0.45	0.21

Table 11. Frequencies of markers associated with the recent human adaptations in Puerto Ricans From Puerto Rico (PUR) and other populations Compiled after Fan et al., 2016.

Function	***Gene***	***SNP***	***AFR***	***MXL***	***EUR***	***EAS***	***PUR***
Ability diet (Lactose)	*MCM6*	rs4988235	0.03	0.24	0.51	0.00	0.21
		rs182549	0.03	0.31	0.51	0.00	0.21
Adaption to artic environment	*FADS*	rs74771917	0.07	0.46	0.03	0.24	0.20
		rs3168072	0.01	0.46	0.03	0.25	0.20
		rs12577276	0.18	0.46	0.03	0.25	0.22
		rs7115739	0.31	0.47	0.04	0.28	0.21
		rs174602	0.74	0.61	0.21	0.36	0.42
		rs174570	0.009	0.59	0.16	0.56	0.32
Short Stature	*DOCK3*	rs13088462	0.00	0.02	0.04	0.001	0.03
		rs4443210	0.38	0.45	0.14	0.40	0.29
		rs7638732	0.29	0.16	0.12	0.05	0.18
	CISH	rs6796769	0.3800	0.44	0.12	0.40	0.23
Adaption to endemic pathogens	*HESX1*	rs9878928	0.27	0.03	0.001	0.001	0.04
	APOL1	rs73885319	0.26	0.00	0.00	0.00	0.03
		rs60910145	0.26	0.00	0.00	0.00	0.03
Adaption to artic environment	*FADS*	rs73885319	0.26	0.00	0.00	0.00	0.03
		rs60910145	0.26	0.00	0.00	0.00	0.03
Adaption to altitudes	*VAV3*	rs10494083	0.42	0.25	0.26	0.23	0.25
		rs345271	0.79	0.26	0.27	0.23	0.29
		rs345261	0.60	0.23	0.24	0.18	0.25
		rs73143	0.82	0.26	0.28	0.23	0.32
		rs2336645	0.79	0.27	0.31	0.36	0.35
	ARNT2	rs7403706	0.73	0.19	0.18	0.09	0.26
	EGLN1	rs186996510	0.00	0.008	0.00	0.03	0.00
		rs12097901	0.34	0.12	0.06	0.46	0.11
	ESPAS1	rs149594770	0.03	0.00	0.00	0.006	0.00
Eye color	*OCA2*	rs12913832	0.03	0.18	0.64	0.002	0.23
		rs1800414	0.001	0.00	0.00	0.60	0.00
		rs74653330	0.00	0.00	0.01	0.00	0.00
	TRYP1	rs1408799	0.19	0.36	0.65	0.02	0.51
Skin pigmentation	*TYR*	rs1126809	0.009	0.14	0.25	0.001	0.18
	SLC24A5	rs1426654	0.07	0.51	0.99	0.01	0.77
	SLC45A2	rs16891982	0.04	0.41	0.94	0.006	0.59
	MC1R	rs885479	0.007	0.39	0.07	0.62	0.10
Thick hair	*EDAR*	rs365060	0.60	0.53	0.09	0.95	0.32
		rs3827760	0.003	0.48	0.01	0.87	0.17

Table 12. Markers associated with susceptibility to inheritable diseases in Puerto Ricans**.** Only the markers present in the pools have been reported.

Phenotype	Mechanism	Gene	Disease associated SNPs from NHGRI-EBI Catalog	Chr	Position	SNP %	LGDS	PUR	Reference
Susceptibility to invasive disease	Membrane attack complex deficiency	*C5-C9*	rs13157656	5	40964750	A/C	0.13	0.15	(Degn, Jensenius, & Thiel, 2011) (Brouwer et al., 2009)
			rs1013579	1	56956811	T/C	0.02	0.03	
Susceptibility (MSMD)	Impaired IFNγ response	*IFNGR1*	rs13201877	6	137354404	A/G	0.12	0.11	(van de Vosse et al., 2009) (S. Zhang et al., 2008)
		STAT1	rs7582694	2	191105394	G/C	0.25	0.26	(van de Vosse et al., 2009) (S. Zhang et al., 2008)
Epidermodysplasia verruciformis									
Resistance	Absence of receptor for pathogen	*FUT2*	rs516316	19	48702888	G/C	0.44	0.50	(Indesmith et al., 2003)(Claus Vogelmeier, M.D., Bettina Hederer, M.D., Thomas Glaab, M.D., Hendrik Schmidt, Maureen P.M.H. Rutten-van Mölken, Ph.D., Kai M. Beeh, M.D., Klaus F. Rabe, M.D., and Leonardo M. Fabbri, & Investigators, 2010)
Chronic mucocutaneous candidiasis	Impaired IL-17 immunity	*CARD9*	rs10781499	9	136371953	G/A	0.45	0.44	(Claus Vogelmeier, M.D., Bettina Hederer, M.D., Thomas Glaab, M.D., Hendrik Schmidt et al., 2010)
			rs4077515	9	136372044	C/T	0.51	0.44	
		IL17RA	rs2241047	22	17105693	G/C	0.48	0.53	
		IL17F	rs763780	6	52236941	T/C	0.11	0.09	(Puel et al., 2011)
Susceptibility to variant Creutzfeldt–Jakob disease	Post-translational modification of host cellular prion protein	*PRNP*	rs1799990	20	4699605	A/G	0.37	0.43	(Zeidler, Stewart, Cousens, Estibeiro, & Will, 1997)(Mead et al., 2009)

Table 13. Pharmacogenetic marker frequencies in Puerto Ricans. Only the markers present in the pools have been reported.

Drug(s)	Indication	Drug-response phenotype	Disease associated SNPs from NHGRI-EBI Catalog	Genes/ biomarkers	Chr	Position	SNP	LGDS	PUR	References
Selected Pharmacogenetic Biomarkers:										
Warfarin	Venous thrombosis, stent thrombosis	Efficacy, adverse reactions	rs749671	*VKORC1*	16	31077026	G/A	0.28	0.38	(Johnson et al., 2011)
Simvastatin	High cholesterol	Myotoxicity	rs4149056	*SLCO1B1*	12	21178615	T/C	0.05	0.12	(Niemi, Pasanen, & Neuvonen, 2011)

Table 14. Markers associated with adaptive traits in the Puerto Rican population.

Adaptation	Phenotype	Populations	Gene	SNP	Chr	Position	SNP	LGDS	PUR	Reference
Cold climate	Reduced bone density, osteoporosis	Siberia	*LRP5*	*rs3736228*	11	68433827	C/T	0.10	0.16	
	Wet earwax. Variation in body odor	Europeans, Asians	*ABCC11*	*rs17822931*	16	48224287	C/T	0.06	0.07	
UV radiation	Light skinned Europeans	Europeans	*MATP (SLC45A2)*	*rs16891982*	5	33951588	C/G	0.61	0.59	(Jablonski & Chaplin, 2000)
Short stature			*HESX1*	*rs9878928*	3	57232504		0.5	0.4	(Scheinfeldt & Tishkoff, 2016)
Adaption to endemic pathogens		Africans	*APOL1*	*rs73885319*	22	36661906		0.03	0.03	(Genovese et al., 2010)
				rs60910145	22	36662034		0.02	0.03	(Genovese et al., 2010)
Adaption to altitude		Asians, west-central South American, Africans	*EGLN1*	*rs12097901*	1	231557623		0.08	0.11	(Scheinfeldt & Tishkoff, 2016)
Skin pigmentation			*TYR*	*rs1126809*	11	89017961		0.18	0.18	(Sturm et al., 2012)
			SLC24A5	*rs1426654*	15	48426484		0.27	0.77	(Scheinfeldt & Tishkoff, 2016)
			SLC45A2	*rs16891982*	5	33951693		0.61	0.59	(Sturm et al., 2012)
			MC1R	*rs885479*	16	89986154		0.06	0.10	(Sturm et al., 2012)

Figure 1: Fluctuation in drug response depend on several factors from: Meyer, U. A., Zanger, U. M., & Schwab, M. (2013). Omics and drug response. Annual review of pharmacology and toxicology, 53, 475-502.

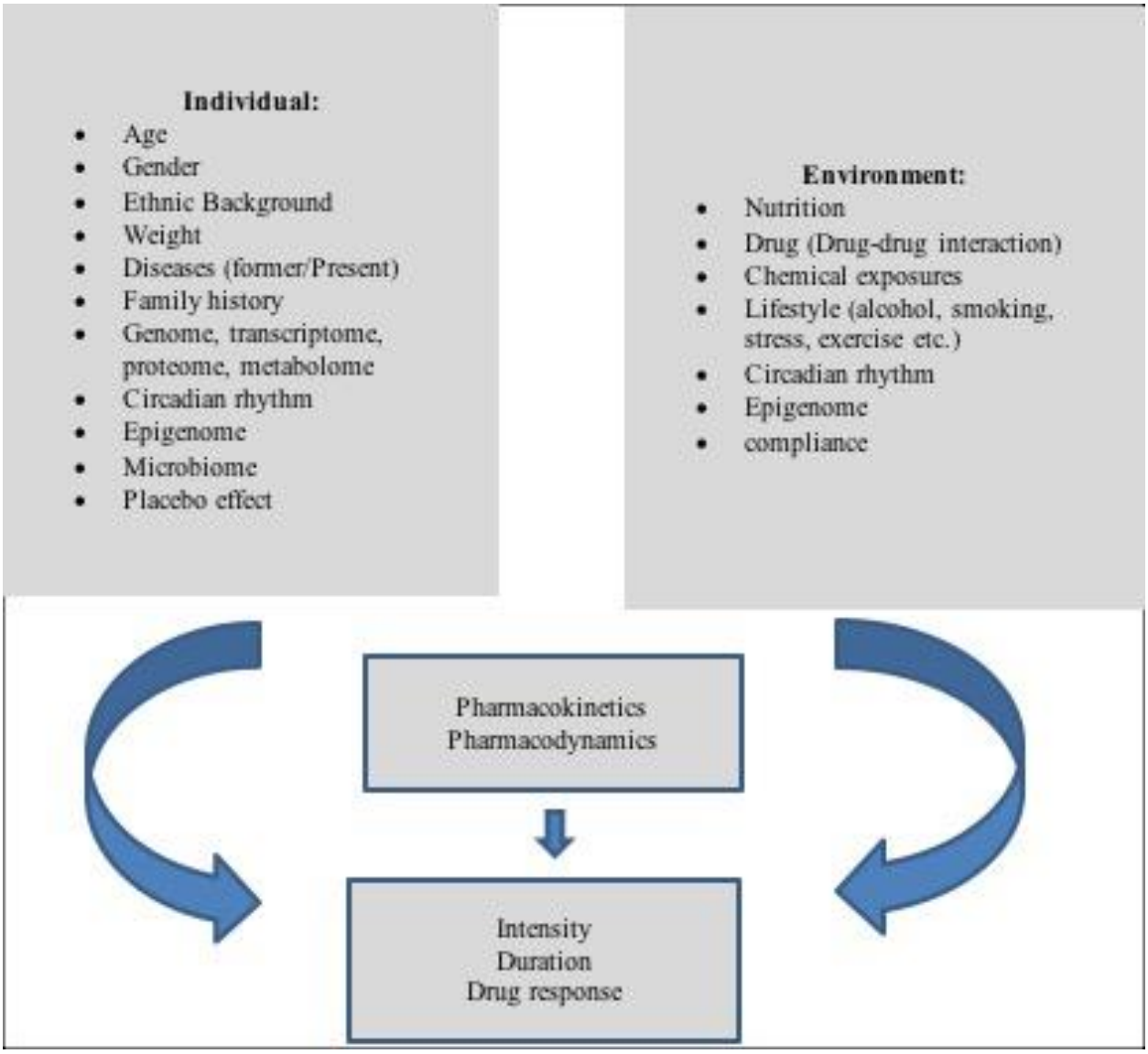


Figure 2. Municipalities selected for Assessment based on previous studies on ancestry distributions. The green represents municipalities with African ancestry (from left to right) Loiza, Rio Grande, Luquillo and Fajardo. Municipalities in blue represent high European ancestry, Mayaguez and Aguadilla. Orange Municipalities represent Native American ancestry, Cabo Rojo, Guanica, sabana Grande and Arecibo. choices have been made using the map of ancestral contributions form Via et al. (2011) (**Figure 3**)

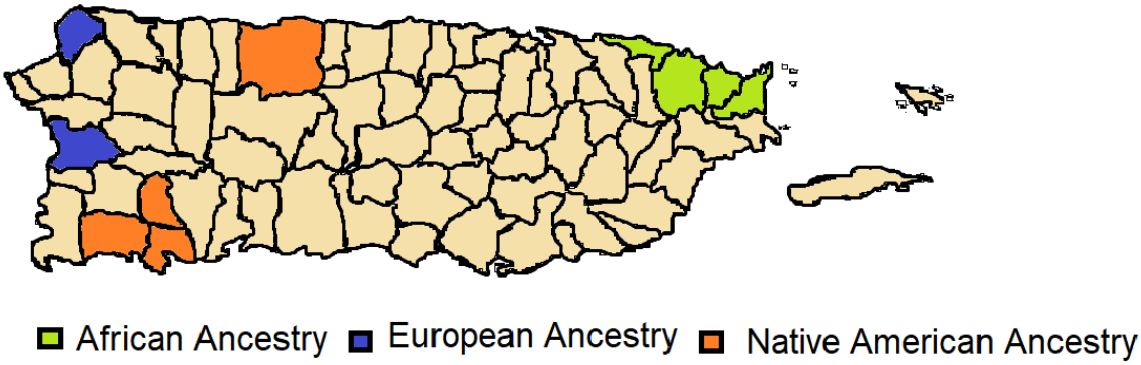


Figure 3. Interpolation plots showing the geographical distribution of ancestry in Puerto Rico. From Via, et al (2011)

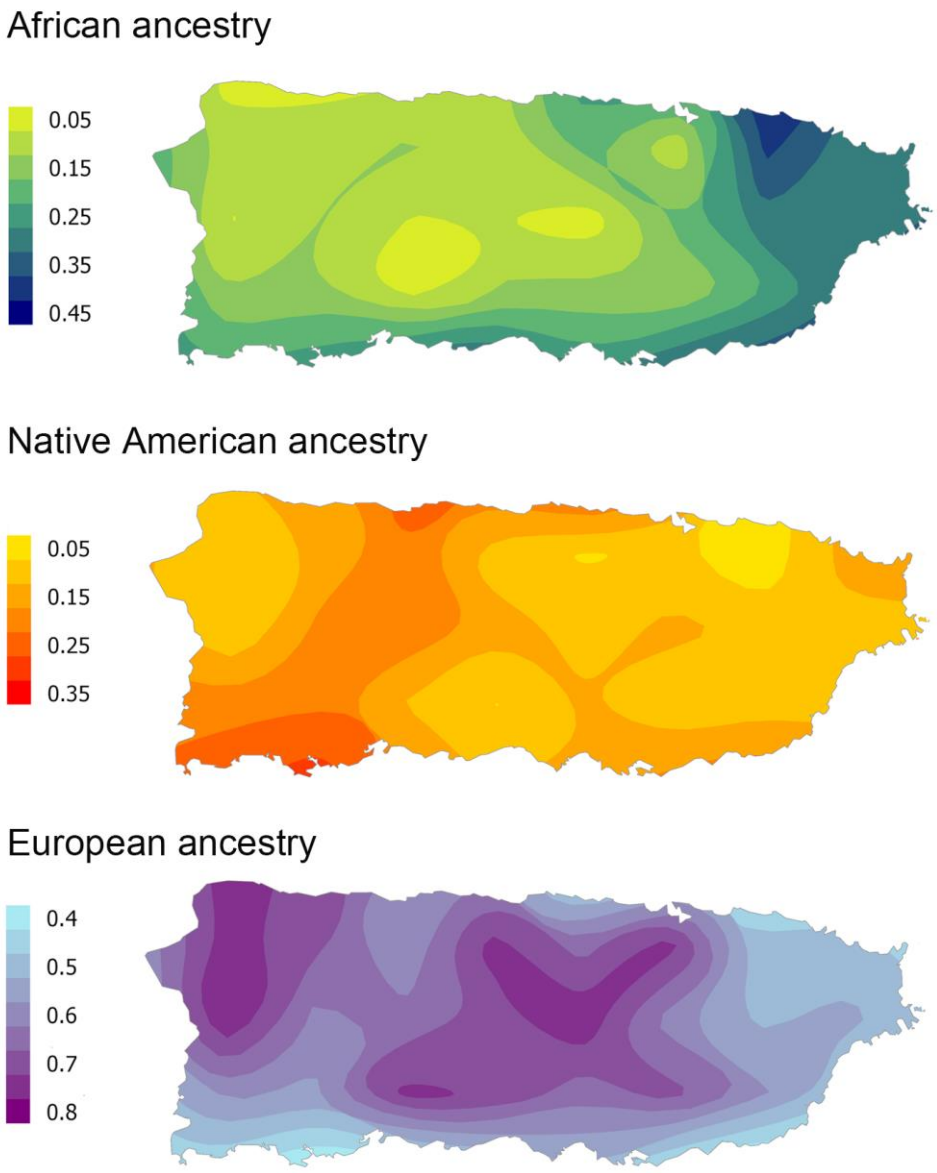


Figure 4. Strong correlation exists between the two independent estimates of genotype frequencies: from the genomes of the Puerto Ricans from Puerto Rico (PUR)(x-axis) in the 1000Genomes project (Auton et al., 2015)and from the exomes in the Local Genome Diversity Studies (LGDS) (y-axis).

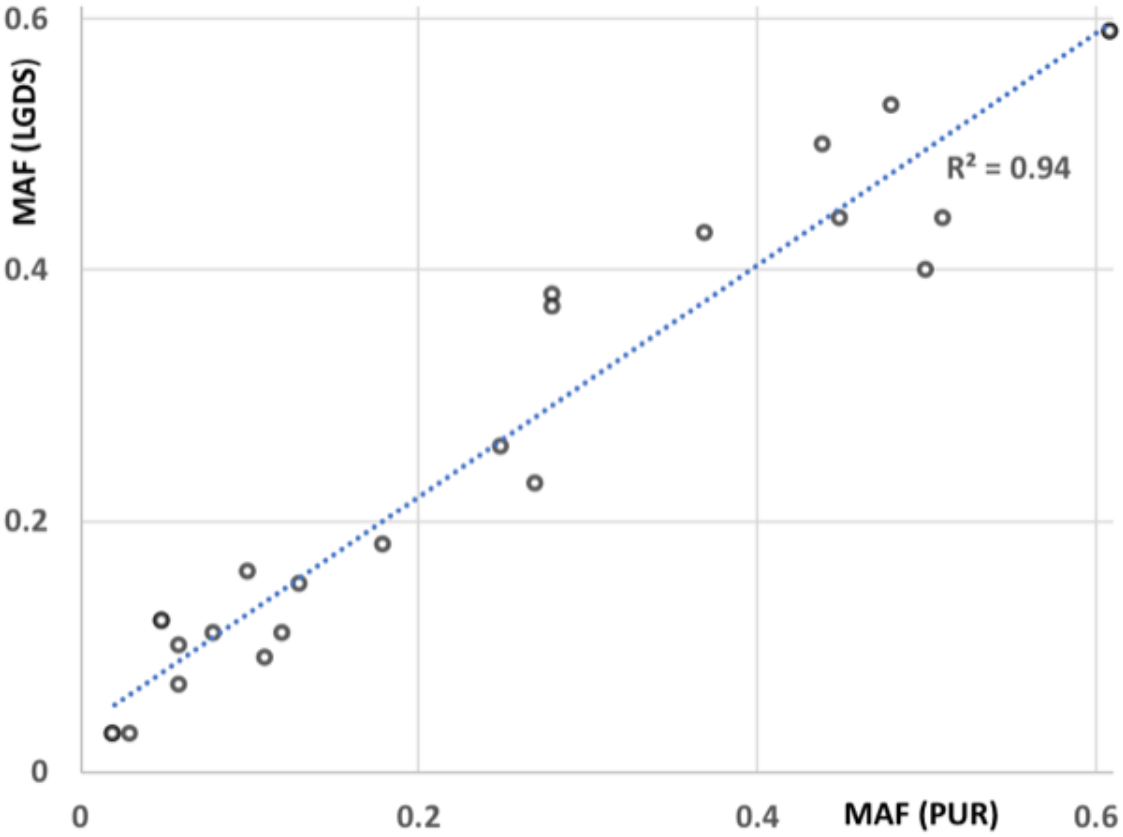
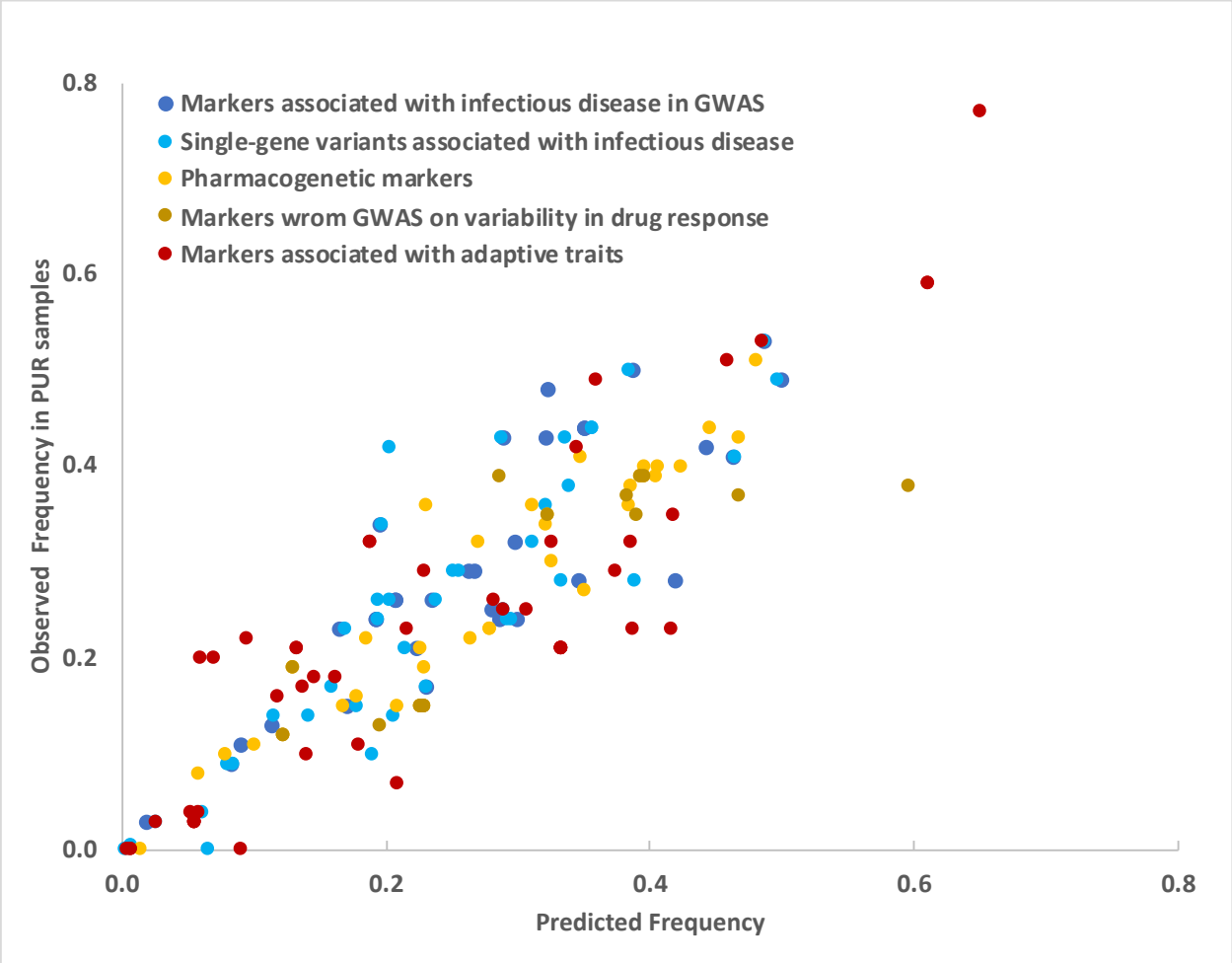


Figure 5. Frequencies of all medically related markers can be reasonably predicted using average ancestral proportions estimated earlier for the Puerto Rican population overall. Observed frequencies (y-axis) are from the PUR samples in the 1000Genomes Project. The Predicted Frequencies (x-axis) have been estimated from multiplying surrogate ancestral populations from the 1000Genomes Project (EUR, AFR, and EAS) by the ancestry proportions estimated for the Puerto Rican population (Via eta l. 2011(64%, 21% and 15% respectively)



APPENDIX:

Towns representing European decent:

Town: Mayaguez	DNA concentration	260/280 ratio
MG 20_A	146.116	1.65385
MG 21_A	851.489	1.686
MG 26_B	33.2	1.79
MG 30_A	74.2	1.8
MG 35_A	91.7449	1.625
MG 37_A	51.5703	1.66
MG 61_A	172.86	1.912
MG 65_A	59	1.74
MG 77_A	134.044	1.894
MG 78_A	416	1.8
MG 79_A	25	1.85
MG 84_B	99.7605	1.721

Town: Aguadilla	DNA concentration	260/280 ratio
AL-70A	35.05	1.88
AL-89B	121.4	1.802
AL- 94A	535.2	1.76
AL-18A	740.6	1.76
AL- 5A	32.2	1.606
AL- 06A	89.75	1.639
AL-17A	24.3	1.60
AL-04B	77.95	1.847
AL-02A	75.4	1.778
AL-03A	34	1.7
AL-16B	70.7	1.81
AL-101 A	52.2	1.77

Towns Representing Native American Descent:

Town: Cabo Rojo	DNA concentration	260/280 ratio
CR-86B	47.5	1.84
CR-27B	454.8	1.83
CR-14B	735	1.86
CR- 09B	23.9	1.69
CR-10A	53.5	1.75
CR-16A	97.73	1.74
CR-17A	66.2	1.66
CR-18A	374.60	1.81
CR-26B	120	1.68
CR-32A	90.39	1.76
CR-45B	118.55	1.775
CR-64B	57.55	1.74

Town: Sabana Grande	DNA concentration	260/280 ratio
SB-50B	49.1	1.72
SB-47B	209.8	1.602
SB-56B	2480.9	1.734
SB-19B	461.75	1.82
SB-34B	168.95	1.71
SB-35B	1360.1	1.763
SB-44B	692.6	1/657
SB-43B	196.05	1.636
SB-53B	67.7	1.686
SB-37B	93	1.635
SB-38B	632.5	1.61
SB-40B	69	1.719

Towns Representing Native American Descent:

Town: Guánica	DNA concentration	260/280 ratio
GN- 12B	110.88	1.7
GN-29B	265.19	1.67
GN34-B	172.15	1.7
GN-43B	31.28	1.68
GN-73B	59.9	1.775
GN-71B	232	1.64
GN-68B	55.1	1.68
GN-94B	63.35	1.668
GN- 93B	91.4	1.726
GN-28 B	216	1.681
GN-92	121.4	1.719
GN-72B	69.75	1.73

Town: Arecibo	DNA concentration	260/280 ratio
AR-79A	90	1.67
AR-43B	51.4	1.91
AR-23B	40.45	1.94
AR-77A	94.65	1.77
AR-41A	143.02	1.77
AR-74A	60	1.74
AR-76A	122	1.749
AR-88A	78.8	1.635
AR-78A	88.7	1.7
AR- 29A	132.75	1.77
AR-101A	109.4	1.83
AR-37B	231	1.67

Towns Representing African Descent:

Town: Loiza	DNA concentration	260/280 ratio
LO- 06A	689.43	1.928
LO-08A	137.23	1.797
LO-18A	166.01	1.781
LO-21A	663.07	1.609
LO-22A	506.8	1.831
LO-24A	344.58	1.610
LO-37A	175	1.74
LO-39A	262	1.75
LO-66A	283	1.78
LO-68A	101	1.67
LO-82B	58.81	1.643
LO-83A	52.72	1.648

Town: Rio Grande	DNA concentration	260/280 ratio
RG-29A	753.8	1.81
RG-36B	107	1.75
RG-52A	224.5	1.764
RG-51A	325.55	1.614
RG-56A	127.9	1.612
RG-74B	161.85	1.825
RG-59B	74.74	1.63
RG-64A	125.9	1.77
RG-71B	121.7	1.77
RG-73A	92.99	1.7
RG-61A	457.2	1.743
RG-50A	86.8	1.82

Town: Fajardo	DNA concentration	260/280 ratio
FA-02B	196.23	1.6351
FA-03A	221.49	1.687
FA-07A	114.9	1.79
FA-09B	73.10	1.969
FA-17A	126.60	1.711
FA-24B	57.4	1.671
FA-36B	394	1.650
FA-37B	113	1.75
FA-20B	377	1.692
FA-18B	405.5	1.776
FA-47B	46.7	1.60
FA-23A	271	1.823

REFERENCES:

Town: Luquillo	DNA concentration	260/280 ratio
LQ-6A	105	1.82
LQ-5A	51.3	1.74
LQ-3A	191.2	1.73
LQ-11B	460.8	1.77
LQ-36A	40.85	1.76
LQ- 21B	35.9	1.60
LQ-34A	66.9	1.77
LQ-20A	137.3	1.71
LQ-19A	184.35	1.70
LQ-17B	110.45	1.73
LQ-10B	26.15	1.60
LQ-7B	139.4	1.75

Adams, J. (2008). Pharmacogenomics and personalized medicine. *Nature Education*, *1*(1), 194.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. http://doi.org/10.1038/nmeth0410-248

Andrews, C. A. (2010). Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education*, *3*(10), 5. Retrieved from https://www.nature.com/scitable/knowledge/library/natural-selection-genetic-drift-and-gene-flow-15186648

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. http://doi.org/10.1038/nature15393

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. a., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, *12*(11), 745–755. http://doi.org/10.1038/nrg3031

Beckmann, S. E. A. and J. S. (2006). mendelian disorders diserve more attention. *Nature*, *7*(4), 277.

Benowitz, S. (2015). Centers for Mendelian Genomics uncovering the genomic basis of hundreds of rare conditions. Retrieved from https://www.genome.gov/27562201/2015-news-feature-centers-for-mendelian-genomics-uncovering-the-genomic-basis-of-hundreds-of-rare-conditions/

Brouwer, M. C., de Gans, J., Heckenberg, S. G., Zwinderman, A. H., van der Poll, T., & van de Beek, D. (2009). Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *9*(1), 31–44. http://doi.org/10.1016/S1473-3099(08)70261-5

Chapman, S. J., & Hill, A. V. S. (2012). Human genetic susceptibility to infectious disease. *Nature Reviews Genetics*, *13*(3), 175–188. http://doi.org/10.1038/nrg3114

Claudio-Campos, K., Labastida, A., Ramos, A., Gaedigk, A., Renta-Torres, J., Padilla, D., … Duconge-Soler, J. (2017). Warfarin anticoagulation therapy in caribbean hispanics of Puerto Rico: A candidate gene association study. *Frontiers in Pharmacology*, *8*(JUN), 1–17. http://doi.org/10.3389/fphar.2017.00347

Claus Vogelmeier, M.D., Bettina Hederer, M.D., Thomas Glaab, M.D., Hendrik Schmidt, P. D., Maureen P.M.H. Rutten-van Mölken, Ph.D., Kai M. Beeh, M.D., Klaus F. Rabe, M.D., and Leonardo M. Fabbri, M. D., & Investigators, for the P.-C. (2010). A Homozygous CARD9 Mutation in a Family with Susceptibility to Fungal Infections. *The New England Journal of Medicine*, *361*(18), 411–422. http://doi.org/10.1056/NEJMoa1402121

Consortium, T. 1000 G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. http://doi.org/10.1038/nature11632

Croxtall, J. D., & McKeage, K. (2011). Fulvestrant. *Drugs*, *71*(3), 363–380. http://doi.org/10.2165/11204810-000000000-00000

Degn, S. E., Jensenius, J. C., & Thiel, S. (2011). Disease-causing mutations in genes of the complement system. *American Journal of Human Genetics*, *88*(6), 689–705. http://doi.org/10.1016/j.ajhg.2011.05.011

DePristo, M. a., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-

generation DNA sequencing data. *Nat Genet*, *43*(5), 491–498. http://doi.org/10.1038/ng.806.A

Dominguez, K., Penman-aguilar, A., Chang, M., Moonesinghe, R., Castellanos, T., Rodriguez-lainz, A., & Schieber, R. (2015). Vital Signs : Leading Causes of Death , Prevalence of Diseases and Risk Factors , and Use of Health Services Among Hispanics in the United States — 2009 – 2013. *Morbidity and Mortality Weekly Report*, *64*(17), 469–478.

Donalson, P., Daly, A., Ermini, L., & Bevitt, D. (2015). Genetic Diversity. In *Genetic of Complex Disease* (pp. 1–34).

Draper, N. (2008). Identification of SNPs, or mutations in sequence chromatograms. *Methods in Molecular Biology*, *439*, 35–52. http://doi.org/10.1007/978-1-59745-188-8-3

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *2*(131), 479–491.

Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2017). Going global by adapting local: A review of recent human adaptation Shaohua. *Science*, *354*(6308), 54–59. http://doi.org/10.1126/science.aaf5098.Going

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, *7*(2), 85–97. http://doi.org/10.1038/nrg1767

Fig, E. D., Note, S., Table, S., Project, G., Table, S., Variants, G., … Table, S. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75. http://doi.org/10.1038/nature15394

Gamazon, Eric R. Skol , Andrew D. and Perera, M. A. (2013). Limits of GWAS for pharmacogenomic testing. *Pharmacogenet Genomics*, *22*(4). http://doi.org/10.1097/FPC.0b013e328350ca5f.The

Garrett, C. R., & Eng, C. (2011). Cetuximab in the treatment of patients with colorectal cancer. *Expert Opinion on Biological Therapy*, *11*(7), 937–949. http://doi.org/10.1517/14712598.2011.582464

Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., … Pollak, M. R. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*, *329*(5993), 841–845. http://doi.org/10.1126/science.1193032

Goldstein, D. B., Tate, S. K., & Sisodiya, S. M. (2003). Pharmacogenetics goes genomic. *Nature Reviews Genetics*, *4*(12), 937–947. http://doi.org/10.1038/nrg1229

González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, *88*(4), 440–449. http://doi.org/10.1016/j.ajhg.2011.03.004

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*, *9*(12).

http://doi.org/10.1371/journal.pgen.1004023
Hancock, A. M., Alkorta-Aranburu, G., Witonsky, D. B., & Di Rienzo, A. (2010). Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2459–2468. http://doi.org/10.1098/rstb.2010.0032
Hattori, M. (2005). Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso. Protein, Nucleic Acid, Enzyme*, *50*(2), 162–168. http://doi.org/10.1038/nature03001
Hill, A. V. S. (2012). Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 840–849. http://doi.org/10.1098/rstb.2011.0275
Indesmith, L. I. S. a L., Oe, C. H. M., Arionneau, S. E. M., Uvoen, N. A. R., Iang, X. I. J., Indblad, L. A. L., … Baric, R. (2003). Human susceptibility and resistance to Norwalk virus infection. *Nat Med*, *9*(5), 548–553. http://doi.org/10.1038/nm
Innocenti, F., Kroetz, D. L., Schuetz, E., Dolan, M. E., Ramírez, J., Relling, M., … Ratain, M. J. (2009). Comprehensive Pharmacogenetic Analysis of Irinotecan Neutropenia and Pharmacokinetics. *Journal of Clinical Oncology*, *27*(16), 2604–2614. http://doi.org/10.1200/JCO.2008.20.6300
Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57–106. http://doi.org/10.1006/jhev.2000.0403
Jobling, Mark, Hurles, M. and C. T.-S. (2013). *Human evolutionary genetics: origins, peoples & disease.*, *2013*. Garland Science.
John H. Gillespie. (1998). *Population Genetics a Concise Guide*. Baltimore, Maryland: The Johns Hopkins University Press Baltimore and London.
Johnson, J. A., Gong, L., Whirl-Carrillo, M., Gage, B. F., Scott, S. A., Stein, C. M., … Altman, R. B. (2011). Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clinical Pharmacology and Therapeutics*, *90*(4), 625–629. http://doi.org/10.1038/clpt.2011.185
Khor, C. C., Tran, C., Bich, N., Pang, J., & Davila, S. (2012). Genome-wide association study identifies susceptibility loci for Dengue shock syndrome at MICB and PLCE1. *Nature Genetics*, *43*(11), 1139–1141. http://doi.org/10.1038/ng.960.Genome-wide
Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). XThe next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27–38. http://doi.org/10.1016/j.cell.2013.09.006
Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1082. http://doi.org/10.1038/nprot.2009.86
Learning, B., Mckenzie, J. F., & Robert, R. (2014). Epidemiology: The Study of Disease, Injury, and Death in the Community. In *Community and public health* (9th ed., pp. 57–83). Jones and Bartlet.
Liu, L., Okada, S., Kong, X.-F., Kreins, A. Y., Cypowyj, S., Abhyankar, A., …

Casanova, J.-L. (2011). Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *The Journal of Experimental Medicine*, *208*(8), 1635–48. http://doi.org/10.1084/jem.20110958

Liu, X., Jian, X., & Boerwinkle, E. (2011). dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Human Mutation*, *32*(8), 894–899. http://doi.org/10.1002/humu.21517

Maienschein, J. (2004). Evolution, Science and Society. *Ludus Vitalis*, *XII*(21), 237–245. http://doi.org/10.1016/S0070-4571(08)00214-8

Martínez-Cruzado, J. C., Toro-Labrador, G., Viera-Vera, J., Rivera-Vega, M. Y., Startek, J., Latorre-Esteves, M., … Valencia-Rivera, P. (2005). Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. *American Journal of Physical Anthropology*, *128*(1), 131–155. http://doi.org/10.1002/ajpa.20108

Mead, S., Poulter, M., Uphill, J., Beck, J., Whitfield, J., Webb, T. E., … Collinge, J. (2009). Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *The Lancet Neurology*, *8*(1), 57–66. http://doi.org/10.1016/S1474-4422(08)70265-5

Medicine, G., & Manolio, T. A. (2013). Genomewide Association Studies and Assessment of the Risk of Disease.

Mells, G. F., Floyd, J. A. B., Morley, K. I., Cordell, H. J., Franklin, C. S., Shin, S.-Y., … Anderson, C. A. (2011). Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nature Genetics*, *43*(4), 329–332. http://doi.org/10.1038/ng.789

Meyer, U. A., Zanger, U. M., & Schwab, M. (2013). Omics and Drug Response. *Annual Review of Pharmacology and Toxicology*, *53*, 475–502. http://doi.org/10.1146/annurev-pharmtox-010510-100502

Miller, L. H., Mason, S. J., Clyde, D. F., & McGinniss, M. H. (1976). The Resistance Factor to Plasmodium vivax in Blacks. *New England Journal of Medicine*, *295*(6), 302–304. http://doi.org/10.1056/NEJM197608052950602

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., … Bustamante, C. D. (2013). Reconstructing the Population Genetic History of the Caribbean. *PLoS Genetics*, *9*(11). http://doi.org/10.1371/journal.pgen.1003925

Niemi, M., Pasanen, M. K., & Neuvonen, P. J. (2011). Organic Anion Transporting Polypeptide 1B1: a Genetically Polymorphic Transporter of Major Importance for Hepatic Drug Uptake. *Pharmacological Reviews*, *63*(1), 157–181. http://doi.org/10.1124/pr.110.002857

Peltonen, L., Perola, M., Naukkarinen, J., & Palotie, A. (2006). Lessons from studying monogenic disease for common disease. *Human Molecular Genetics*. http://doi.org/10.1093/hmg/ddl060

Pérez de Diego, R., Sancho-Shimizu, V., Lorenzo, L., Puel, A., Plancoulaine, S., Picard, C., … Casanova, J.-L. (2010). Human TRAF3 Adaptor Molecule Deficiency Leads to Impaired Toll-like Receptor 3 Response and Susceptibility to Herpes Simplex

Encephalitis. *Immunity*, *33*(3), 400–411. http://doi.org/10.1016/j.immuni.2010.08.014
Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics*, *93*(2), 105–111. http://doi.org/10.1016/j.ygeno.2008.10.003
Picard, C., von Bernuth, H., Ghandil, P., Chrabieh, M., Levy, O., Arkwright, P. D., … Casanova, J.-L. (2010). Clinical Features and Outcome of Patients With IRAK-4 and MyD88 Deficiency. *Medicine*, *89*(6), 403–425. http://doi.org/10.1097/MD.0b013e3181fd8ec3
Puel, A., Cypowyj, S., Bustamante, J., Wright, J. F., Liu, L., Lim, K., … Picard, C. (2011). Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science*, *332*(6025), 65–68. http://doi.org/10.1126/science.1200439.Chronic
Rotimi, C. N., & Jorde, L. B. (2010). Ancestry and Disease in the Age of Genomic Medicine. *New England Journal of Medicine*, *363*(16), 1551–1558. http://doi.org/10.1056/NEJMra0911564
Schaffner, S. F. P. C. S. (2008). Evolutionary Adaptation in the Human Lineage. *Nature Education*, *1*(1), 14. Retrieved from https://www.nature.com/scitable/topicpage/evolutionary-adaptation-in-the-human-lineage-12397
Scheinfeldt, L. B., & Tishkoff, S. A. (2016). Recent human adaptation: genomic approaches, interpretation and insights, *14*(10), 692–702. http://doi.org/10.1038/nrg3604.Recent
Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., … Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.)*, *305*(5683), 525–8. http://doi.org/10.1126/science.1098918
Sherry, A. K. and S. (2002). The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. In O. J. McEntyre J (Ed.), *The NCBI Handbook* (2nd ed.). Bethesda: National Center for Biotechnology Information. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK21088/
Sophie Hambleton1,2, Sandra Salem3,2, Jacinta Bustamante4,5, Venetia Bigley1,5, Stéphanie Boisson-Dupuis6,5, Joana Azevedo7,5, Anny Fortin3, Muzlifah Haniffa1, Lourdes Ceron- Gutierrez8, Chris Bacon9,10, Geetha Menon9, Céline Trouillet7, David McDonald1, 19. (2013). Mutations in IRF8 and Human Dendritic Cell Immunodeficiency. *New England Journal of Medicine*, *360*(9346), 1640–1645. http://doi.org/10.1016/S0140-6736(02)11602-3.Association
Sturm, R. A., Duffy, D. L., Sturm, R., Westerhof, W., Sturm, R., Larsson, M., … Wood, J. (2012). Human pigmentation genes under environmental selection. *Genome Biology*, *13*(9), 248. http://doi.org/10.1186/gb-2012-13-9-248
Sukumar, M., Sundar, M., & Sivarajan, M. (2010). The Emerging Role of Admixture in the Pharmacogenetics of Puerto Rican Hispanics. *Journal of Pharmacogenomics & Pharmacoproteomics*, *1*(1), 1–9. http://doi.org/10.4172/2153-0645.1000102
Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E. G., &

Risch, N. J. (2007). Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *The American Journal of Human Genetics*, *81*(3), 626–633. http://doi.org/10.1086/520769

Tanner, J.-A., & Tyndale, R. F. (2017). Variation in CYP2A6 Activity and Personalized Medicine. *Journal of Personalized Medicine*, *7*(4). http://doi.org/10.3390/jpm7040018

Thorven, M., Grahn, A., Hedlund, K.-O., Johansson, H., Wahlfrid, C., Larson, G., & Svensson, L. (2005). A Homozygous Nonsense Mutation (428G->A) in the Human Secretor (FUT2) Gene Provides Resistance to Symptomatic Norovirus (GGII) Infections. *Journal of Virology*, *79*(24), 15351–15355. http://doi.org/10.1128/JVI.79.24.15351-15355.2005

Troiani, T., Napolitano, S., Della Corte, C. M., Martini, G., Martinelli, E., Morgillo, F., & Ciardiello, F. (2016). Therapeutic value of EGFR inhibition in CRC and NSCLC: 15 years of clinical evidence. *ESMO Open*, *1*(5), e000088. http://doi.org/10.1136/esmoopen-2016-000088

van de Vosse, E., van Dissel, J. T., & Ottenhoff, T. H. (2009). Genetic deficiencies of innate immune signalling in human infectious disease. *The Lancet Infectious Diseases*, *9*(11), 688–698. http://doi.org/10.1016/S1473-3099(09)70255-5

Varki, A., Geschwind, D. H., & Eichler, E. E. (2008). Explaining human uniqueness: genome interactions with environment, behaviour and culture. *Nature Reviews. Genetics*, *9*(10), 749–63. http://doi.org/10.1038/nrg2428

Via, M., Gignoux, C. R., Roth, L. a., Fejerman, L., Galanter, J., Choudhry, S., … Martínez-Cruzado, J. C. (2011). History shaped the geographic distribution of genomic admixture on the island of puerto rico. *PLoS ONE*, *6*(1). http://doi.org/10.1371/journal.pone.0016513

Wahlberg, J., Erson, B. P., & Uhlen, M. (1993). Use of automated sequencing of amplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1 strains . Use of Automated Sequencing of Polymerase Chain Reaction- Generated Amplicons To Identify Three Types of Cholera Toxin Subunit B in, *31*(1), 22–25.

Week, H., America, S., & Salvador, E. (2014). Hispanic Heritage Month 2014 : Sept . 15 – Oct . 15 States and Counties, (September 1968).

Wilm, A., Kim Aw, P. P., Bertrand, Hui Ting Yeo, G., Hoe Ong, S., Chang Hua Wong, … Martin Lloyd Hibberd1 and Niranjan Nagarajan1. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526318/pdf/gks918.pdf

Zeidler, M., Stewart, G., Cousens, S. N., Estibeiro, K., & Will, R. G. (1997). Codon 129 genotype and new variant CJD. *Lancet*, *350*, 668. http://doi.org/10.1016/S0140-6736(05)63366-1

Zhang, S.-Y., Boisson-Dupuis, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., … Casanova, J.-L. (2008). Inborn errors of interferon (IFN)-mediated immunity in

humans: insights into the respective roles of IFN-α/β, IFN-γ, and IFN-λ in host defense. *Immunological Reviews*, *226*(1), 29–40. http://doi.org/10.1111/j.1600-065X.2008.00698.x

Zhang, S.-Y., Jouanguy, E., Ugolini, S., Smahi, A., Elain, G., Romero, P., … Casanova, J.-L. (2007). TLR3 deficiency in patients with herpes simplex encephalitis. *Science (New York, N.Y.)*, *317*(5844), 1522–7. http://doi.org/10.1126/science.1139522

Zhang, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., Picard, C., … Casanova, J. (2008). Inborn errors of interferon (IFN)- mediated immunity in humans: insights into the respective roles of IFN-a/b, IFN-g, and IFN-l in host defense. *Immunological Reviews*, *226*, 29–40. http://doi.org/10.1111/j.1600-065X.2008.00698.x

Zhou, K., Bellenguez, C., Spencer, C. C. A., Bennett, A. J., Coleman, R. L., Tavendale, R., … Pearson, E. R. (2011). Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nature Genetics*, *43*(2), 117–120. http://doi.org/10.1038/ng.735

Adams, J. (2008). Pharmacogenomics and personalized medicine. *Nature Education*, *1*(1), 194.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. http://doi.org/10.1038/nmeth0410-248

Andrews, C. A. (2010). Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education*, *3*(10), 5. Retrieved from https://www.nature.com/scitable/knowledge/library/natural-selection-genetic-drift-and-gene-flow-15186648

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. http://doi.org/10.1038/nature15393

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. a., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, *12*(11), 745–755. http://doi.org/10.1038/nrg3031

Beckmann, S. E. A. and J. S. (2006). mendelian disorders diserve more attention. *Nature*, *7*(4), 277.

Benowitz, S. (2015). Centers for Mendelian Genomics uncovering the genomic basis of hundreds of rare conditions. Retrieved from https://www.genome.gov/27562201/2015-news-feature-centers-for-mendelian-genomics-uncovering-the-genomic-basis-of-hundreds-of-rare-conditions/

Brouwer, M. C., de Gans, J., Heckenberg, S. G., Zwinderman, A. H., van der Poll, T., & van de Beek, D. (2009). Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *9*(1), 31–44. http://doi.org/10.1016/S1473-3099(08)70261-5

Chapman, S. J., & Hill, A. V. S. (2012). Human genetic susceptibility to infectious

disease. *Nature Reviews Genetics*, *13*(3), 175–188. http://doi.org/10.1038/nrg3114
Claudio-Campos, K., Labastida, A., Ramos, A., Gaedigk, A., Renta-Torres, J., Padilla, D., … Duconge-Soler, J. (2017). Warfarin anticoagulation therapy in caribbean hispanics of Puerto Rico: A candidate gene association study. *Frontiers in Pharmacology*, *8*(JUN), 1–17. http://doi.org/10.3389/fphar.2017.00347
Claus Vogelmeier, M.D., Bettina Hederer, M.D., Thomas Glaab, M.D., Hendrik Schmidt, P. D., Maureen P.M.H. Rutten-van Mölken, Ph.D., Kai M. Beeh, M.D., Klaus F. Rabe, M.D., and Leonardo M. Fabbri, M. D., & Investigators, for the P.-C. (2010). A Homozygous CARD9 Mutation in a Family with Susceptibility to Fungal Infections. *The New England Journal of Medicine*, *361*(18), 411–422. http://doi.org/10.1056/NEJMoa1402121
Consortium, T. 1000 G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. http://doi.org/10.1038/nature11632
Croxtall, J. D., & McKeage, K. (2011). Fulvestrant. *Drugs*, *71*(3), 363–380. http://doi.org/10.2165/11204810-000000000-00000
Degn, S. E., Jensenius, J. C., & Thiel, S. (2011). Disease-causing mutations in genes of the complement system. *American Journal of Human Genetics*, *88*(6), 689–705. http://doi.org/10.1016/j.ajhg.2011.05.011
DePristo, M. a., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, *43*(5), 491–498. http://doi.org/10.1038/ng.806.A
Dominguez, K., Penman-aguilar, A., Chang, M., Moonesinghe, R., Castellanos, T., Rodriguez-lainz, A., & Schieber, R. (2015). Vital Signs : Leading Causes of Death , Prevalence of Diseases and Risk Factors , and Use of Health Services Among Hispanics in the United States — 2009 – 2013. *Morbidity and Mortality Weekly Report*, *64*(17), 469–478.
Donalson, P., Daly, A., Ermini, L., & Bevitt, D. (2015). Genetic Diversity. In *Genetic of Complex Disease* (pp. 1–34).
Draper, N. (2008). Identification of SNPs, or mutations in sequence chromatograms. *Methods in Molecular Biology*, *439*, 35–52. http://doi.org/10.1007/978-1-59745-188-8-3
Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *2*(131), 479–491.
Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2017). Going global by adapting local: A review of recent human adaptation Shaohua. *Science*, *354*(6308), 54–59. http://doi.org/10.1126/science.aaf5098.Going
Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, *7*(2), 85–97. http://doi.org/10.1038/nrg1767
Fig, E. D., Note, S., Table, S., Project, G., Table, S., Variants, G., … Table, S. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571),

75. http://doi.org/10.1038/nature15394
Gamazon, Eric R. Skol , Andrew D. and Perera, M. A. (2013). Limits of GWAS for pharmacogenomic testing. *Pharmacogenet Genomics*, *22*(4). http://doi.org/10.1097/FPC.0b013e328350ca5f.The
Garrett, C. R., & Eng, C. (2011). Cetuximab in the treatment of patients with colorectal cancer. *Expert Opinion on Biological Therapy*, *11*(7), 937–949. http://doi.org/10.1517/14712598.2011.582464
Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., … Pollak, M. R. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*, *329*(5993), 841–845. http://doi.org/10.1126/science.1193032
Goldstein, D. B., Tate, S. K., & Sisodiya, S. M. (2003). Pharmacogenetics goes genomic. *Nature Reviews Genetics*, *4*(12), 937–947. http://doi.org/10.1038/nrg1229
González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, *88*(4), 440–449. http://doi.org/10.1016/j.ajhg.2011.03.004
Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*, *9*(12). http://doi.org/10.1371/journal.pgen.1004023
Hancock, A. M., Alkorta-Aranburu, G., Witonsky, D. B., & Di Rienzo, A. (2010). Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2459–2468. http://doi.org/10.1098/rstb.2010.0032
Hattori, M. (2005). Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso. Protein, Nucleic Acid, Enzyme*, *50*(2), 162–168. http://doi.org/10.1038/nature03001
Hill, A. V. S. (2012). Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 840–849. http://doi.org/10.1098/rstb.2011.0275
Indesmith, L. I. S. a L., Oe, C. H. M., Arionneau, S. E. M., Uvoen, N. A. R., Iang, X. I. J., Indblad, L. A. L., … Baric, R. (2003). Human susceptibility and resistance to Norwalk virus infection. *Nat Med*, *9*(5), 548–553. http://doi.org/10.1038/nm
Innocenti, F., Kroetz, D. L., Schuetz, E., Dolan, M. E., Ramírez, J., Relling, M., … Ratain, M. J. (2009). Comprehensive Pharmacogenetic Analysis of Irinotecan Neutropenia and Pharmacokinetics. *Journal of Clinical Oncology*, *27*(16), 2604–2614. http://doi.org/10.1200/JCO.2008.20.6300
Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57–106. http://doi.org/10.1006/jhev.2000.0403
Jobling, Mark, Hurles, M. and C. T.-S. (2013). *Human evolutionary genetics: origins, peoples & disease.*, *2013*. Garland Science.

John H. Gillespie. (1998). *Population Genetics a Concise Guide*. Baltimore, Maryland: The Johns Hopkins University Press Baltimore and London.

Johnson, J. A., Gong, L., Whirl-Carrillo, M., Gage, B. F., Scott, S. A., Stein, C. M., … Altman, R. B. (2011). Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clinical Pharmacology and Therapeutics*, *90*(4), 625–629. http://doi.org/10.1038/clpt.2011.185

Khor, C. C., Tran, C., Bich, N., Pang, J., & Davila, S. (2012). Genome-wide association study identifies susceptibility loci for Dengue shock syndrome at MICB and PLCE1. *Nature Genetics*, *43*(11), 1139–1141. http://doi.org/10.1038/ng.960.Genome-wide

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). XThe next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27–38. http://doi.org/10.1016/j.cell.2013.09.006

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1082. http://doi.org/10.1038/nprot.2009.86

Learning, B., Mckenzie, J. F., & Robert, R. (2014). Epidemiology: The Study of Disease, Injury, and Death in the Community. In *Community and public health* (9th ed., pp. 57–83). Jones and Bartlet.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. http://doi.org/10.1093/bioinformatics/btp324

Liu, X., Jian, X., & Boerwinkle, E. (2011). dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Human Mutation*, *32*(8), 894–899. http://doi.org/10.1002/humu.21517

Maienschein, J. (2004). Evolution, Science and Society. *Ludus Vitalis*, *XII*(21), 237–245. http://doi.org/10.1016/S0070-4571(08)00214-8

Martínez-Cruzado, J. C., Toro-Labrador, G., Viera-Vera, J., Rivera-Vega, M. Y., Startek, J., Latorre-Esteves, M., … Valencia-Rivera, P. (2005). Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. *American Journal of Physical Anthropology*, *128*(1), 131–155. http://doi.org/10.1002/ajpa.20108

Mead, S., Poulter, M., Uphill, J., Beck, J., Whitfield, J., Webb, T. E., … Collinge, J. (2009). Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *The Lancet Neurology*, *8*(1), 57–66. http://doi.org/10.1016/S1474-4422(08)70265-5

Medicine, G., & Manolio, T. A. (2013). Genomewide Association Studies and Assessment of the Risk of Disease.

Mells, G. F., Floyd, J. A. B., Morley, K. I., Cordell, H. J., Franklin, C. S., Shin, S.-Y., … Anderson, C. A. (2011). Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nature Genetics*, *43*(4), 329–332. http://doi.org/10.1038/ng.789

Meyer, U. A., Zanger, U. M., & Schwab, M. (2013). Omics and Drug Response. *Annual Review of Pharmacology and Toxicology*, *53*, 475–502. http://doi.org/10.1146/annurev-pharmtox-010510-100502

Miller, L. H., Mason, S. J., Clyde, D. F., & McGinniss, M. H. (1976). The Resistance Factor to Plasmodium vivax in Blacks. *New England Journal of Medicine*, *295*(6), 302–304. http://doi.org/10.1056/NEJM197608052950602

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., … Bustamante, C. D. (2013). Reconstructing the Population Genetic History of the Caribbean. *PLoS Genetics*, *9*(11). http://doi.org/10.1371/journal.pgen.1003925

Niemi, M., Pasanen, M. K., & Neuvonen, P. J. (2011). Organic Anion Transporting Polypeptide 1B1: a Genetically Polymorphic Transporter of Major Importance for Hepatic Drug Uptake. *Pharmacological Reviews*, *63*(1), 157–181. http://doi.org/10.1124/pr.110.002857

Peltonen, L., Perola, M., Naukkarinen, J., & Palotie, A. (2006). Lessons from studying monogenic disease for common disease. *Human Molecular Genetics*. http://doi.org/10.1093/hmg/ddl060

Pérez de Diego, R., Sancho-Shimizu, V., Lorenzo, L., Puel, A., Plancoulaine, S., Picard, C., … Casanova, J.-L. (2010). Human TRAF3 Adaptor Molecule Deficiency Leads to Impaired Toll-like Receptor 3 Response and Susceptibility to Herpes Simplex Encephalitis. *Immunity*, *33*(3), 400–411. http://doi.org/10.1016/j.immuni.2010.08.014

Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics*, *93*(2), 105–111. http://doi.org/10.1016/j.ygeno.2008.10.003

Puel, A., Cypowyj, S., Bustamante, J., Wright, J. F., Liu, L., Lim, K., … Picard, C. (2011). Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science*, *332*(6025), 65–68. http://doi.org/10.1126/science.1200439.Chronic

Rotimi, C. N., & Jorde, L. B. (2010). Ancestry and Disease in the Age of Genomic Medicine. *New England Journal of Medicine*, *363*(16), 1551–1558. http://doi.org/10.1056/NEJMra0911564

Schaffner, S. F. P. C. S. (2008). Evolutionary Adaptation in the Human Lineage. *Nature Education*, *1*(1), 14. Retrieved from https://www.nature.com/scitable/topicpage/evolutionary-adaptation-in-the-human-lineage-12397

Scheinfeldt, L. B., & Tishkoff, S. A. (2016). Recent human adaptation: genomic approaches, interpretation and insights, *14*(10), 692–702. http://doi.org/10.1038/nrg3604.Recent

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., … Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.)*, *305*(5683), 525–8. http://doi.org/10.1126/science.1098918

Sherry, A. K. and S. (2002). The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. In O. J. McEntyre J (Ed.), *The NCBI Handbook*

(2nd ed.). Bethesda: National Center for Biotechnology Information. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK21088/

Sturm, R. A., Duffy, D. L., Sturm, R., Westerhof, W., Sturm, R., Larsson, M., … Wood, J. (2012). Human pigmentation genes under environmental selection. *Genome Biology*, *13*(9), 248. http://doi.org/10.1186/gb-2012-13-9-248

Sukumar, M., Sundar, M., & Sivarajan, M. (2010). The Emerging Role of Admixture in the Pharmacogenetics of Puerto Rican Hispanics. *Journal of Pharmacogenomics & Pharmacoproteomics*, *1*(1), 1–9. http://doi.org/10.4172/2153-0645.1000102

Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E. G., & Risch, N. J. (2007). Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *The American Journal of Human Genetics*, *81*(3), 626–633. http://doi.org/10.1086/520769

Tanner, J.-A., & Tyndale, R. F. (2017). Variation in CYP2A6 Activity and Personalized Medicine. *Journal of Personalized Medicine*, *7*(4). http://doi.org/10.3390/jpm7040018

Troiani, T., Napolitano, S., Della Corte, C. M., Martini, G., Martinelli, E., Morgillo, F., & Ciardiello, F. (2016). Therapeutic value of EGFR inhibition in CRC and NSCLC: 15 years of clinical evidence. *ESMO Open*, *1*(5), e000088. http://doi.org/10.1136/esmoopen-2016-000088

van de Vosse, E., van Dissel, J. T., & Ottenhoff, T. H. (2009). Genetic deficiencies of innate immune signalling in human infectious disease. *The Lancet Infectious Diseases*, *9*(11), 688–698. http://doi.org/10.1016/S1473-3099(09)70255-5

Varki, A., Geschwind, D. H., & Eichler, E. E. (2008). Explaining human uniqueness: genome interactions with environment, behaviour and culture. *Nature Reviews. Genetics*, *9*(10), 749–63. http://doi.org/10.1038/nrg2428

Via, M., Gignoux, C. R., Roth, L. a., Fejerman, L., Galanter, J., Choudhry, S., … Martínez-Cruzado, J. C. (2011). History shaped the geographic distribution of genomic admixture on the island of puerto rico. *PLoS ONE*, *6*(1). http://doi.org/10.1371/journal.pone.0016513

Wahlberg, J., Erson, B. P., & Uhlen, M. (1993). Use of automated sequencing of amplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1 strains . Use of Automated Sequencing of Polymerase Chain Reaction- Generated Amplicons To Identify Three Types of Cholera Toxin Subunit B in, *31*(1), 22–25.

Week, H., America, S., & Salvador, E. (2014). Hispanic Heritage Month 2014 : Sept . 15 – Oct . 15 States and Counties, (September 1968).

Wilm, A., Kim Aw, P. P., Bertrand, Hui Ting Yeo, G., Hoe Ong, S., Chang Hua Wong, … Martin Lloyd Hibberd1 and Niranjan Nagarajan1. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526318/pdf/gks918.pdf

Zeidler, M., Stewart, G., Cousens, S. N., Estibeiro, K., & Will, R. G. (1997). Codon 129

genotype and new variant CJD. *Lancet*, *350*, 668. http://doi.org/10.1016/S0140-6736(05)63366-1

Zhang, S.-Y., Boisson-Dupuis, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., … Casanova, J.-L. (2008). Inborn errors of interferon (IFN)-mediated immunity in humans: insights into the respective roles of IFN-α/β, IFN-γ, and IFN-λ in host defense. *Immunological Reviews*, *226*(1), 29–40. http://doi.org/10.1111/j.1600-065X.2008.00698.x

Zhang, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., Picard, C., … Casanova, J. (2008). Inborn errors of interferon (IFN)- mediated immunity in humans: insights into the respective roles of IFN-a/b, IFN-g, and IFN-l in host defense. *Immunological Reviews*, *226*, 29–40. http://doi.org/10.1111/j.1600-065X.2008.00698.x

Zhou, K., Bellenguez, C., Spencer, C. C. A., Bennett, A. J., Coleman, R. L., Tavendale, R., … Pearson, E. R. (2011). Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nature Genetics*, *43*(2), 117–120. http://doi.org/10.1038/ng.735

Adams, J. (2008). Pharmacogenomics and personalized medicine. *Nature Education*, *1*(1), 194.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. http://doi.org/10.1038/nmeth0410-248

Andrews, C. A. (2010). Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education*, *3*(10), 5. Retrieved from https://www.nature.com/scitable/knowledge/library/natural-selection-genetic-drift-and-gene-flow-15186648

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. http://doi.org/10.1038/nature15393

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. a., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, *12*(11), 745–755. http://doi.org/10.1038/nrg3031

Beckmann, S. E. A. and J. S. (2006). mendelian disorders diserve more attention. *Nature*, *7*(4), 277.

Benowitz, S. (2015). Centers for Mendelian Genomics uncovering the genomic basis of hundreds of rare conditions. Retrieved from https://www.genome.gov/27562201/2015-news-feature-centers-for-mendelian-genomics-uncovering-the-genomic-basis-of-hundreds-of-rare-conditions/

Brouwer, M. C., de Gans, J., Heckenberg, S. G., Zwinderman, A. H., van der Poll, T., & van de Beek, D. (2009). Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *9*(1), 31–44. http://doi.org/10.1016/S1473-3099(08)70261-5

Chapman, S. J., & Hill, A. V. S. (2012). Human genetic susceptibility to infectious disease. *Nature Reviews Genetics*, *13*(3), 175–188. http://doi.org/10.1038/nrg3114

Claudio-Campos, K., Labastida, A., Ramos, A., Gaedigk, A., Renta-Torres, J., Padilla, D., … Duconge-Soler, J. (2017). Warfarin anticoagulation therapy in caribbean hispanics of Puerto Rico: A candidate gene association study. *Frontiers in Pharmacology*, *8*(JUN), 1–17. http://doi.org/10.3389/fphar.2017.00347

Claus Vogelmeier, M.D., Bettina Hederer, M.D., Thomas Glaab, M.D., Hendrik Schmidt, P. D., Maureen P.M.H. Rutten-van Mölken, Ph.D., Kai M. Beeh, M.D., Klaus F. Rabe, M.D., and Leonardo M. Fabbri, M. D., & Investigators, for the P.-C. (2010). A Homozygous CARD9 Mutation in a Family with Susceptibility to Fungal Infections. *The New England Journal of Medicine*, *361*(18), 411–422. http://doi.org/10.1056/NEJMoa1402121

Consortium, T. 1000 G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. http://doi.org/10.1038/nature11632

Croxtall, J. D., & McKeage, K. (2011). Fulvestrant. *Drugs*, *71*(3), 363–380. http://doi.org/10.2165/11204810-000000000-00000

Degn, S. E., Jensenius, J. C., & Thiel, S. (2011). Disease-causing mutations in genes of the complement system. *American Journal of Human Genetics*, *88*(6), 689–705. http://doi.org/10.1016/j.ajhg.2011.05.011

DePristo, M. a., Banks, E., Poplin, R. E., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, *43*(5), 491–498. http://doi.org/10.1038/ng.806.A

Dominguez, K., Penman-aguilar, A., Chang, M., Moonesinghe, R., Castellanos, T., Rodriguez-lainz, A., & Schieber, R. (2015). Vital Signs : Leading Causes of Death , Prevalence of Diseases and Risk Factors , and Use of Health Services Among Hispanics in the United States — 2009 – 2013. *Morbidity and Mortality Weekly Report*, *64*(17), 469–478.

Donalson, P., Daly, A., Ermini, L., & Bevitt, D. (2015). Genetic Diversity. In *Genetic of Complex Disease* (pp. 1–34).

Draper, N. (2008). Identification of SNPs, or mutations in sequence chromatograms. *Methods in Molecular Biology*, *439*, 35–52. http://doi.org/10.1007/978-1-59745-188-8-3

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *2*(131), 479–491.

Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2017). Going global by adapting local: A review of recent human adaptation Shaohua. *Science*, *354*(6308), 54–59. http://doi.org/10.1126/science.aaf5098.Going

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, *7*(2), 85–97. http://doi.org/10.1038/nrg1767

Fig, E. D., Note, S., Table, S., Project, G., Table, S., Variants, G., … Table, S. (2015). An

integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75. http://doi.org/10.1038/nature15394

Gamazon, Eric R. Skol , Andrew D. and Perera, M. A. (2013). Limits of GWAS for pharmacogenomic testing. *Pharmacogenet Genomics*, *22*(4). http://doi.org/10.1097/FPC.0b013e328350ca5f.The

Garrett, C. R., & Eng, C. (2011). Cetuximab in the treatment of patients with colorectal cancer. *Expert Opinion on Biological Therapy*, *11*(7), 937–949. http://doi.org/10.1517/14712598.2011.582464

Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., … Pollak, M. R. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science*, *329*(5993), 841–845. http://doi.org/10.1126/science.1193032

Goldstein, D. B., Tate, S. K., & Sisodiya, S. M. (2003). Pharmacogenetics goes genomic. *Nature Reviews Genetics*, *4*(12), 937–947. http://doi.org/10.1038/nrg1229

González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, *88*(4), 440–449. http://doi.org/10.1016/j.ajhg.2011.03.004

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*, *9*(12). http://doi.org/10.1371/journal.pgen.1004023

Hancock, A. M., Alkorta-Aranburu, G., Witonsky, D. B., & Di Rienzo, A. (2010). Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2459–2468. http://doi.org/10.1098/rstb.2010.0032

Hattori, M. (2005). Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso. Protein, Nucleic Acid, Enzyme*, *50*(2), 162–168. http://doi.org/10.1038/nature03001

Hill, A. V. S. (2012). Evolution, revolution and heresy in the genetics of infectious disease susceptibility. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 840–849. http://doi.org/10.1098/rstb.2011.0275

Indesmith, L. I. S. a L., Oe, C. H. M., Arionneau, S. E. M., Uvoen, N. A. R., Iang, X. I. J., Indblad, L. A. L., … Baric, R. (2003). Human susceptibility and resistance to Norwalk virus infection. *Nat Med*, *9*(5), 548–553. http://doi.org/10.1038/nm

Innocenti, F., Kroetz, D. L., Schuetz, E., Dolan, M. E., Ramírez, J., Relling, M., … Ratain, M. J. (2009). Comprehensive Pharmacogenetic Analysis of Irinotecan Neutropenia and Pharmacokinetics. *Journal of Clinical Oncology*, *27*(16), 2604–2614. http://doi.org/10.1200/JCO.2008.20.6300

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, *39*(1), 57–106. http://doi.org/10.1006/jhev.2000.0403

Jobling, Mark, Hurles, M. and C. T.-S. (2013). *Human evolutionary genetics: origins,*

peoples & disease., 2013. Garland Science.
John H. Gillespie. (1998). *Population Genetics a Concise Guide*. Baltimore, Maryland: The Johns Hopkins University Press Baltimore and London.
Johnson, J. A., Gong, L., Whirl-Carrillo, M., Gage, B. F., Scott, S. A., Stein, C. M., … Altman, R. B. (2011). Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clinical Pharmacology and Therapeutics*, *90*(4), 625–629. http://doi.org/10.1038/clpt.2011.185
Khor, C. C., Tran, C., Bich, N., Pang, J., & Davila, S. (2012). Genome-wide association study identifies susceptibility loci for Dengue shock syndrome at MICB and PLCE1. *Nature Genetics*, *43*(11), 1139–1141. http://doi.org/10.1038/ng.960.Genome-wide
Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). XThe next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27–38. http://doi.org/10.1016/j.cell.2013.09.006
Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1082. http://doi.org/10.1038/nprot.2009.86
Learning, B., Mckenzie, J. F., & Robert, R. (2014). Epidemiology: The Study of Disease, Injury, and Death in the Community. In *Community and public health* (9th ed., pp. 57–83). Jones and Bartlet.
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. http://doi.org/10.1093/bioinformatics/btp324
Liu, X., Jian, X., & Boerwinkle, E. (2011). dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Human Mutation*, *32*(8), 894–899. http://doi.org/10.1002/humu.21517
Maienschein, J. (2004). Evolution, Science and Society. *Ludus Vitalis*, *XII*(21), 237–245. http://doi.org/10.1016/S0070-4571(08)00214-8
Martínez-Cruzado, J. C., Toro-Labrador, G., Viera-Vera, J., Rivera-Vega, M. Y., Startek, J., Latorre-Esteves, M., … Valencia-Rivera, P. (2005). Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. *American Journal of Physical Anthropology*, *128*(1), 131–155. http://doi.org/10.1002/ajpa.20108
Mead, S., Poulter, M., Uphill, J., Beck, J., Whitfield, J., Webb, T. E., … Collinge, J. (2009). Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *The Lancet Neurology*, *8*(1), 57–66. http://doi.org/10.1016/S1474-4422(08)70265-5
Medicine, G., & Manolio, T. A. (2013). Genomewide Association Studies and Assessment of the Risk of Disease.
Mells, G. F., Floyd, J. A. B., Morley, K. I., Cordell, H. J., Franklin, C. S., Shin, S.-Y., … Anderson, C. A. (2011). Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nature Genetics*, *43*(4), 329–332.

http://doi.org/10.1038/ng.789
Meyer, U. A., Zanger, U. M., & Schwab, M. (2013). Omics and Drug Response. *Annual Review of Pharmacology and Toxicology*, *53*, 475–502. http://doi.org/10.1146/annurev-pharmtox-010510-100502
Miller, L. H., Mason, S. J., Clyde, D. F., & McGinniss, M. H. (1976). The Resistance Factor to Plasmodium vivax in Blacks. *New England Journal of Medicine*, *295*(6), 302–304. http://doi.org/10.1056/NEJM197608052950602
Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., … Bustamante, C. D. (2013). Reconstructing the Population Genetic History of the Caribbean. *PLoS Genetics*, *9*(11). http://doi.org/10.1371/journal.pgen.1003925
Niemi, M., Pasanen, M. K., & Neuvonen, P. J. (2011). Organic Anion Transporting Polypeptide 1B1: a Genetically Polymorphic Transporter of Major Importance for Hepatic Drug Uptake. *Pharmacological Reviews*, *63*(1), 157–181. http://doi.org/10.1124/pr.110.002857
Peltonen, L., Perola, M., Naukkarinen, J., & Palotie, A. (2006). Lessons from studying monogenic disease for common disease. *Human Molecular Genetics*. http://doi.org/10.1093/hmg/ddl060
Pérez de Diego, R., Sancho-Shimizu, V., Lorenzo, L., Puel, A., Plancoulaine, S., Picard, C., … Casanova, J.-L. (2010). Human TRAF3 Adaptor Molecule Deficiency Leads to Impaired Toll-like Receptor 3 Response and Susceptibility to Herpes Simplex Encephalitis. *Immunity*, *33*(3), 400–411. http://doi.org/10.1016/j.immuni.2010.08.014
Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics*, *93*(2), 105–111. http://doi.org/10.1016/j.ygeno.2008.10.003
Puel, A., Cypowyj, S., Bustamante, J., Wright, J. F., Liu, L., Lim, K., … Picard, C. (2011). Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science*, *332*(6025), 65–68. http://doi.org/10.1126/science.1200439.Chronic
Rotimi, C. N., & Jorde, L. B. (2010). Ancestry and Disease in the Age of Genomic Medicine. *New England Journal of Medicine*, *363*(16), 1551–1558. http://doi.org/10.1056/NEJMra0911564
Schaffner, S. F. P. C. S. (2008). Evolutionary Adaptation in the Human Lineage. *Nature Education*, *1*(1), 14. Retrieved from https://www.nature.com/scitable/topicpage/evolutionary-adaptation-in-the-human-lineage-12397
Scheinfeldt, L. B., & Tishkoff, S. A. (2016). Recent human adaptation: genomic approaches, interpretation and insights, *14*(10), 692–702. http://doi.org/10.1038/nrg3604.Recent
Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., … Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science (New York, N.Y.)*, *305*(5683), 525–8. http://doi.org/10.1126/science.1098918
Sherry, A. K. and S. (2002). The Single Nucleotide Polymorphism Database (dbSNP) of

Nucleotide Sequence Variation. In O. J. McEntyre J (Ed.), *The NCBI Handbook* (2nd ed.). Bethesda: National Center for Biotechnology Information. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK21088/

Sturm, R. A., Duffy, D. L., Sturm, R., Westerhof, W., Sturm, R., Larsson, M., … Wood, J. (2012). Human pigmentation genes under environmental selection. *Genome Biology*, *13*(9), 248. http://doi.org/10.1186/gb-2012-13-9-248

Sukumar, M., Sundar, M., & Sivarajan, M. (2010). The Emerging Role of Admixture in the Pharmacogenetics of Puerto Rican Hispanics. *Journal of Pharmacogenomics & Pharmacoproteomics*, *1*(1), 1–9. http://doi.org/10.4172/2153-0645.1000102

Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E. G., & Risch, N. J. (2007). Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. *The American Journal of Human Genetics*, *81*(3), 626–633. http://doi.org/10.1086/520769

Tanner, J.-A., & Tyndale, R. F. (2017). Variation in CYP2A6 Activity and Personalized Medicine. *Journal of Personalized Medicine*, *7*(4). http://doi.org/10.3390/jpm7040018

Troiani, T., Napolitano, S., Della Corte, C. M., Martini, G., Martinelli, E., Morgillo, F., & Ciardiello, F. (2016). Therapeutic value of EGFR inhibition in CRC and NSCLC: 15 years of clinical evidence. *ESMO Open*, *1*(5), e000088. http://doi.org/10.1136/esmoopen-2016-000088

van de Vosse, E., van Dissel, J. T., & Ottenhoff, T. H. (2009). Genetic deficiencies of innate immune signalling in human infectious disease. *The Lancet Infectious Diseases*, *9*(11), 688–698. http://doi.org/10.1016/S1473-3099(09)70255-5

Varki, A., Geschwind, D. H., & Eichler, E. E. (2008). Explaining human uniqueness: genome interactions with environment, behaviour and culture. *Nature Reviews. Genetics*, *9*(10), 749–63. http://doi.org/10.1038/nrg2428

Via, M., Gignoux, C. R., Roth, L. a., Fejerman, L., Galanter, J., Choudhry, S., … Martínez-Cruzado, J. C. (2011). History shaped the geographic distribution of genomic admixture on the island of puerto rico. *PLoS ONE*, *6*(1). http://doi.org/10.1371/journal.pone.0016513

Wahlberg, J., Erson, B. P., & Uhlen, M. (1993). Use of automated sequencing of amplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1 strains . Use of Automated Sequencing of Polymerase Chain Reaction- Generated Amplicons To Identify Three Types of Cholera Toxin Subunit B in, *31*(1), 22–25.

Week, H., America, S., & Salvador, E. (2014). Hispanic Heritage Month 2014 : Sept . 15 – Oct . 15 States and Counties, (September 1968).

Wilm, A., Kim Aw, P. P., Bertrand, Hui Ting Yeo, G., Hoe Ong, S., Chang Hua Wong, … Martin Lloyd Hibberd1 and Niranjan Nagarajan1. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, *40*(22), 11189–11201. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526318/pdf/gks918.pdf

Zeidler, M., Stewart, G., Cousens, S. N., Estibeiro, K., & Will, R. G. (1997). Codon 129 genotype and new variant CJD. *Lancet*, *350*, 668. http://doi.org/10.1016/S0140-6736(05)63366-1

Zhang, S.-Y., Boisson-Dupuis, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., … Casanova, J.-L. (2008). Inborn errors of interferon (IFN)-mediated immunity in humans: insights into the respective roles of IFN-α/β, IFN-γ, and IFN-λ in host defense. *Immunological Reviews*, *226*(1), 29–40. http://doi.org/10.1111/j.1600-065X.2008.00698.x

Zhang, S., Chapgier, A., Yang, K., Bustamante, J., Puel, A., Picard, C., … Casanova, J. (2008). Inborn errors of interferon (IFN)- mediated immunity in humans: insights into the respective roles of IFN-a/b, IFN-g, and IFN-l in host defense. *Immunological Reviews*, *226*, 29–40. http://doi.org/10.1111/j.1600-065X.2008.00698.x

Zhou, K., Bellenguez, C., Spencer, C. C. A., Bennett, A. J., Coleman, R. L., Tavendale, R., … Pearson, E. R. (2011). Common variants near ATM are associated with glycemic response to metformin in type 2 diabetes. *Nature Genetics*, *43*(2), 117–120. http://doi.org/10.1038/ng.735