UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN MEDICINA MOLECOLARE

Curriculum: Genomics, Proteomics e Tecnologie correlate

XXV ciclo

TESI DI DOTTORATO DI RICERCA

GENETIC BIOMARKERS OF RENAL FUNCTION

Dottoranda: Federica Rizzi

Matricola: R08697

TUTOR Prof. Cristina Barlassina

COORDINATORE del Dottorato: Prof. Mario Clerici

A.A. 2011-2012

ABSTRACT

The present study is part of HYPERGENES Project, focused on the definition of a comprehensive genetic-epidemiological model of complex traits like essential hypertension and of intermediate phenotypes related to hypertension. We specifically focused on the identification of genetic biomarkers related to Glomerular Filtration Rate (GFR) as index of renal function.

Chronic kidney disease (CKD), defined as either kidney damage or decreased kidney function for 3 or more months (GFR < 60 ml/min per $1.73m^2$) affects about 6-11% of the general population. A strong relationship has been reported between renal dysfunction, hypertension and cardiovascular diseases; hypertension is present in more than 80% of patients with CKD and contributes to progression to end stage renal disease and to cardiovascular events as well.

Apart from the risk conferred by traditional cardiovascular risk factors, eGFR has a strong genetic component. We therefore performed a genotype-phenotype association analysis in subjects affected by essential hypertension and in a population based cohort. The analyses have been carried on using: 1) 1,634 hypertensives from the HYPERGENES Discovery phase and 1,198 hypertensives from the HYPERGENES Validation phase; 2) 2,697 subjects from Epogh-Flemengho population based sample 3) 1,952 subjects from Epogh-Flemengho population based sample with GFR follow-up data, using the difference between the last and the first measurement as phenotype of renal function.

Using a Genome wide association (GWA) approach, in the HYPERGENES Discovery sample we identified a potential novel variant associated with hypertension and renal function in STAC gene (rs4678878, p-value of $3.83x10^{-8}$). This gene is likely involved in a neuron-specific signal transduction, although its function is not yet completely clear. This result was not validated in HYPERGENES Validation and in the population based studies. The finding needs further investigation performing a fine mapping of the region.

Through a candidate gene approach, we could identify some variants, mapping in genes already known as associated to renal function (PPARA, CST3-CST9 region, PRKAG2, ABCA1, NEDD4L).

The identification of genetic variants affecting renal function could help to better understand not only the GFR variability in the general population but also the pathophysiology of CKD and progressive kidney function decline. Ultimately, this could lead to novel tools for diagnosis, prevention and therapy of CKD.

Il presente studio fa parte del progetto HYPERGENES, che si pone come obiettivo la definizione di un modello genetico-epidemiologico di patologie a tratto complesso come l'ipertensione e di fenotipi intermedi ad essa correlati. In particolare lo scopo di questo lavoro è l'identificazione di biomarkers genetici correlati al tasso di filtrazione glomerulare (Glomerular Filtration Rate, GFR) come indice di funzionalità renale.

L'insufficienza renale cronica (Chronic kidney disease, CKD), definita come danno renale o come diminuita funzionalità renale per 3 o più mesi (GFR < 60 ml/min per 1.73m^2), colpisce il 6-11% della popolazione. Esiste uno stretto legame tra diminuita funzionalità renale, ipertensione e patologie cardiovascolari; l'ipertensione è presente in più dell'80% dei pazienti con CKD e contribuisce al progredire del danno renale sia in termini di insufficienza renale terminale che in termini di patologie cardiovascolari.

Oltre ai tradizionali fattori di rischio cardiovascolari, l'eGFR ha una forte componente genetica. Abbiamo quindi analizzato l'associazione genotipo-fenotipo in soggetti ipertesi e in uno studio di popolazione. Sono stati analizzati : 1) 1,634 pazienti ipertesi facenti parte della fase di Discovery del progetto HYPERGENES e 1,198 pazienti ipertesi della fase di Validation 2) 2,697 soggetti apparteneti allo studio di popolazione Epogh-Flemengho 3) 1,952 soggetti apparteneti allo studio di popolazione Epogh-Flemengho con dati di follow up per il fenotipo GFR, usando la differenza tra l'ultima e la prima misurazione come fenotipo di funzionalità renale.

Usando un approccio Genome wide (GWA), nel campione di Discovery del progetto HYPERGENES abbiamo identificato una nuova variante potenzialmente associata con l'ipertensione e con la funzionalità renale nel gene STAC (rs4678878, p-value=3.83×10^{-8}). Questo gene è presumibilmente coinvolto nella trasduzione neuronale, nonostante la sua funzione non sia ad oggi completamente chiarita. Questo risultato non è stato confermato nella Validation di HYPERGENES e negli studi di popolazione e sono necessarie ulteriori analisi eseguendo un fine mapping della regione.

Utilizzando l'approccio gene candidato abbiamo identificato alcune varianti localizzate in geni tradizionalmente associati alla funzionalità renale (PPARA, regione CST3-CST9, PRKAG2, ABCA1, NEDD4L).

L'identificazione di varianti genetiche che influenzano la funzionalità renale potrebbe essere utile per comprendere non solo la variabilità del GFR nella popolazione ma anche la patofisiologia della CKD e della diminuzione progressiva della funzionalità renale. Infine potrebbe aiutare a sviluppare nuovi metodi di diagnosi, prevenzione e terapia della CKD.

SYMBOLS

BP: Blood Pressure
CKD: Chronic Kidney Disease
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
CVD: Cardiovascular Disease
D': D prime
DBP: Diastolic Blood Pressure
EH: Essential Hypertension
eGFRcys: Glomerular Filtration rate estimated from cystatin C
eGFRcrea: Glomerular Filtration rate estimated from serum creatinine
eGFR: estimated Glomerular Filtration Rate
FA: Fatty acid
GFR: Glomerular Filtration Rate
GWAs: Genome Wide Association Study
LD: Linkage Disequilibrium
MDRD: Modification of Diet in Renal Disease
PCA: Principal Component Analysis
r^2: r squared
SBP: Systolic Blood Pressure
TOD: Target Organ Damage

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Outline of Renal anatomy and physiology

1.1.1. Renal Morphology

Kidneys are organs that serve essential regulatory role in the urinary system and also serve homeostatic functions such as the regulation of electrolytes, maintenance of acid-base balance, and regulation of blood pressure (via maintaining salt and water balance). They serve the body as a natural filter of the blood, and remove wastes, which are diverted to the urinary bladder. In producing urine, the kidneys excrete wastes such as urea and ammonium, and they are also responsible for the readsorption of water, glucose, and aminoacids. The kidneys also produce hormones including calcitriol, erythropoietin, and the enzyme renin.

The basic unit of the kidney is the *nephron*. Each *nephron* consists of a *glomerulus*, a tuft of capillaries interposed between two arterioles (the afferent and efferent arterioles), and a series of tubules lined by a continuous layer of epithelial cells. The *glomeruli* are located in the outer part of the kidney, the *cortex*, whereas the tubules are presented in both the *cortex* and the *medulla*, the inner part of the kidney.

The initial step in the excretory function of the *nephron* is the formation of an ultrafiltrate of plasma across the glomerulus. This fluid then passes through the tubules and is modified by reabsorption and by secretion. Fluid filtered across the glomerulus enters *Bowman's space* and then the proximal tubule, composed by an initial convoluted segment and a later straight segment, the pars recta. The *loop of Henle* begins at the end of the pars recta and includes a thin *descending limb* and a thin and thick segment of the *ascending limb*. This hairpin configuration plays a major role in the excretion of a hyperosmotic urine.

The thick *ascending limb* has a cortical segment that returns to the region of the parent *glomerulus*. In this area, where the tubule approaches the afferent *glomerular arteriole*, are located the specialized tubular cells of the *macula densa*. The *juxtaglomerular cells* of the afferent arteriole and the *macula densa* compose the *justaglomerular apparatus*, which plays a central role in renin secretion.

After the *macula densa*, there are three cortical segments: the *distal convoluted tubule*, the *connecting segment* and the *cortical connecting tubule*. The *connecting segments* of many nephrons drain into a single *collecting tubule*. Fluid leaving the cortical *collecting tubule* flows into the *medullary collecting tubule* and drains sequentially into the calyces, the renal pelvis, the ureters, and the bladder.

The segmental subdivision of the nephron is based upon different permeability and transport characteristics that translate in important differences in function [1].

1.1.2. Renal circulation and Glomerular Filtration Rate regulation

Kidney has a blood flow of approximately 400 ml/100g of tissue per minute (20 percent of the cardiac output), much greater than that observed in other well perfused vascular beds such as heart, liver and brain.

Bloods enters the kidney through the renal arteries and passes through serial branches (interlobar, arcuate, interlobular) before entering the *glomeruli* via the capillary wall. It then leaves the *glomeruli* via the efferent arterioles and enters the postglomerular capillaries. In the cortex, these capillaries run in apposition to the adjacent tubules, although not necessarily to the tubule segments from the same *glomerulus*. In addition,

branches from the efferent arterioles of the *juxtamedullary glomeruli* enter the medulla and form the *vasa recta capillaries*. Blood returns to the systemic circulation through veins similar to the arteries in name and location.
The renal circulation affects urine formation in the following ways:

1. The rate of glomerular filtration is an important determinant of solute and water excretion
2. The peritubular capillaries in the cortex return reabsorbed solutes and water to the systemic circulation and modulate the degree of proximal reabsorption and secretion
3. The *vasa recta* capillaries in the *medulla* return reabsorbed salt and water to the systemic circulation and participate in the countercurrent mechanism, permitting the conservation of water by the excretion of a hyperosmotic urine

Systemic blood pressure has a crucial role in controlling kidney function. Since the rate of glomerular filtration is directly related to the hydrostatic blood pressure in glomerular capillaries, any increase in blood pressure results in a corresponding increase in filtration and urine output. Since this process ultimately decreases blood volume, the kidneys are a powerful mean of controlling long term blood pressure. Conversely, drops in blood pressure have the opposite effect. This mechanism is important because it illustrates how circulatory and renal systems interact as blood pressure changes affect kidney function and vice versa [2].
Glomerular Filtration Rate (GFR) describes the flow rate of filtered fluid through the kidney. It is normally maintained within relatively narrow limits to prevent inappropriate fluctuations in solute and water excretion. Regulation of the GFR is primarily achieved by alterations in arteriolar tone that influence both the hydraulic pressure in the glomerular capillary and renal blood flow. In normal subjects GFR is maintained by autoregulation, a phenomenon intrinsic to the kidney that allow GFR and plasma filtration rate to remain roughly constant over a wide range of arterial pressures through changes in afferent arteriolar resistance. Autoregulation is mediated by at least three factors: stretch receptors in the afferent arteriole, angiotensin II and tubuloglomerular feedback.

- Strech receptors: myogenic stretch receptors in the wall of the afferent arteriole increase the degree of stretch and promote arteriolar constriction in response to an elevation in renal perfusion pressure.
- Angiotensin II: The afferent arteriolae of each glomerulus contains specialized cells called *juxtaglomerular cells*. These cells synthesize the precursor prorenin, which is cleaved into the active proteolytic enzyme renin. Active renin is then stored in and released from secretory granules. More proximal cells in the interlobulary artery can also be recruited for renal release when the stimulus is prolonged. Renal hypoperfusion, produced by hypotension or volume depletion, and increased sympathetic activity are the major stimuli to renin secretion. Renin initiates a sequence of steps that begins with cleavage of a decapeptide angiotensin I from renin substrate (angiotensinogen), an alfa2-globulin produced in the liver (and other organs including the kidney). Angiotensin I is then converted into angiotensin II through a reaction catalyzed by the angiotensin converting enzyme (ACE), located in the lung, in the luminal membrane of vascular endothelial cells, the glomerulus itself and other organs. Angiotensin II

has two major systemic effects: systemic vasoconstriction (by arteriolar vasoconstriction and consequent elevation of systemic vascular resistance) and sodium and water retention.

- Tubuloglomerular feedback (TGF) refers to the alterations in GFR that can be induced by changes in tubular flow rate. This phenomenon is mediated by specialized cells in the *macula densa* segment at the end of the cortical thick ascending limb of the *loop of Henle*; these cells sense changes in the delivery and subsequent reabsorption of chloride. An elevation in renal perfusion pressure can activate TGF via an initial rise in GFR; the ensuing increase in macula densa chloride delivery will then initiate a response that returns both GFR and *macula densa* flow toward normal.

In hypovolemic states these responses can be overridden by neurohumoral vasoconstriction, in an attempt to maximize coronary and cerebral perfusion, mainly through a marked stimulation of the vasocostrictor sympathetic nervous and renin-angiotensin systems. Norepinephrine directly increases the afferent tone and indirectly enhances efferent resistance via activation of the release of renin and angiotensin II. Angiotensin II and norepinephrine also stimulate glomerular prostaglandin production, which plays an important role in modifying vasocostrictive effects.

A major function of autoregulation and TGF is preventing excessive salt and water losses.

Endothelin, released locally from endothelial cells, is another important renal vasoconstrictor that affects both afferent and efferent glomerular arterioles, leading to reductions in renal blood flow and GFR. This is not an important regulator of renal hemodynamics in normal subjects but plays a major role in subjects with posthischemic acute renal failure. Another vasoactive factor released from the endothelial cells is nitric oxide, lowering renal vascular resistance.

In contrast volume expansion tends to be associated with increased renal perfusion and a mild rise in GFR. Reduced secretion of angiotensin II and norepinephrine and enhanced release of dopamine and atrial natriuretic peptide contribute to this response:

- Dopamine dilates both the afferent and the efferent arterioles, raising renal blood flow while producing a lesser increment or no change in GFR
- Atrial natriuretic peptide determines afferent dilation and efferent constriction, causing a raise in the pressure of the glomerular capillaries and therefore in GFR

These hormonal alterations also facilitate excretion of the excess of sodium.

1.2. Estimation of Glomerular Filtration Rate

GFR is considered the best index of the level of kidney function. The normal level of GFR varies according to age, sex, and body size. Normal GFR in young adults is approximately 120 to 130 ml/min per 1.73 m^2 and declines with age.

The gold standard markers for GFR extimation have the characteristic of being completely filtered by the glomerulus and not reabsorbed along the tubule. Exogenous substances like inulin, iothalamate, diethylenetriaminepentaacetic acid and iohexol fulfill these characteristics. However they aren't commonly used in clinical practice because they require invasive, expensive and time-consuming procedures.

Serum creatinine concentration is currently used to assess kidney function: creatinine is derived from the metabolism of creatine in skeletal muscle and it is released into the plasma at a relatively constant rate. Plasma creatinine concentration is therefore very

stable, varying less than 10 percent per day in serial observations in normal subjects [1]. Like inulin and the exogenous markers mentioned above, it is freely filtered across the glomerulus and is neither reabsorbed nor metabolized by the kidney. However creatinine is secreted at a low extent in the tubule via the organic cation secretory pump in the proximal tubule, resulting in creatinine excretion exceeding the amount filtered by 10 to 20 percent.

The use of serum creatinine per se as an index of kidney function is by the way confounding, as it can be influenced by several factors, including different skeletal muscle masses, individual age, sex and race [3]. At birth it reflects the maternal serum creatinine level, it rapidly decrease during the first month of life to about 0.25 mg/dL and then gradually increases with age until adolescence, where further growth and development of muscle mass results in different constant values for men and women. Finally, at older ages, serum creatinine slightly goes up again (figure 1)[4]. Concerning race, African-American men have higher values, on average, than Caucasian males, mainly because they have more muscle mass.

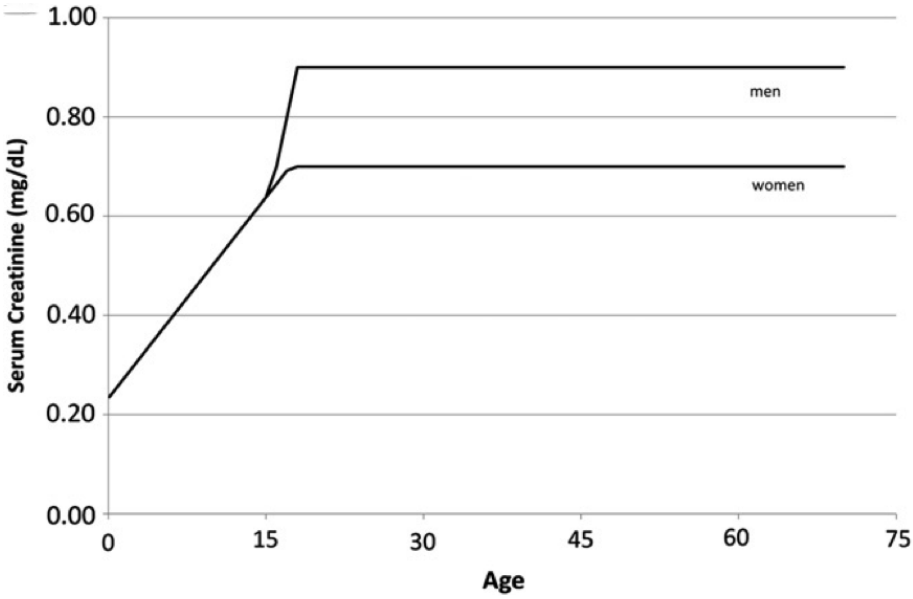


Figure 1: Expected time-evolution of serum creatinine concentration in healthy Caucasian [4].

A further critical point in using serum creatinine data is the calibration of laboratory serum creatinine assay, which can vary across laboratories. The National Health and Nutrition Examination Surveys (NHANES), one of the most important sources of data about the prevalence of chronic diseases and health conditions, including kidney function, in the US population, broadly analyzed this calibration bias in most routine serum creatinine measurement procedures. They could validate a serum creatinine calibration procedure in order to compare serum creatinine measurement in the NHANES III (1988-

1994) and NHANES 1999-2000, 2001-2002, and 2003-2004 surveys to standard creatinine by using an assay traceable to known gold-standard reference methods (National Institute of Standards and Technology) [3,5].

1.2.1. Impact of age on GFR

The ideal all-purpose GFR prediction equation probably does not exist, as GFR changes during life and the major determining factor in all eGFR-equations, serum creatinine, changes also during life, but in a completely different way. With particular reference to children, the normal level of GFR varies with age, and consequently with body size, and increases with maturation from infancy, approaching adult mean values at approximately 2 years of age. It remains constant during growth from child to adult and then gradually starts to decline with age [4]. GFR declines with normal ageing usually begin after 30–40 years of age. The rate of decline may accelerate after 50–60 years of age. This decline appears to be a part of the normal physiologic process of cellular and organ senescence and is associated with structural changes in the kidneys [6].

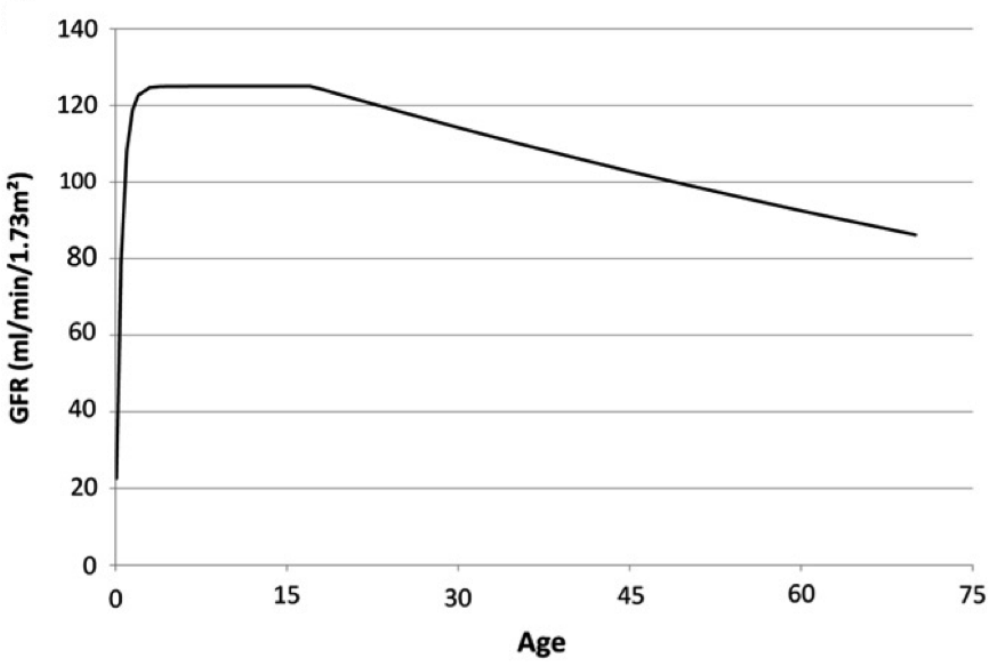


Figure 2: Expected time-evolution of GFR in healthy Caucasian [4].

Verhave et al. [7] reported a study on 8,446 subjects where measured creatinine clearance showed a parabolic curve over age with an acceleration of clearance decline above 50 years in both males and females. Grewal et al. reported their experience in 428 kidney donors evaluated with radioactive chromium EDTA. GFR remained constant until the age of 40 years and then declined at a rate of 9.1 ml/min/1.73m^2 per decade. In agreement with these studies, Douville et al. reported a relative stable GFR up to the age

of 30, followed by a slow decline [8]. In particular they showed a regular decrease of GFR from ~120 ml/min/1.73m^2 in early adulthood down to ~60 ml/min/1.73m^2 in the 80s with a continuous trend over 50 years [9].

1.2.2. eGFR equations in adults

Several equations have therefore been developed to predict creatinine clearance using serum creatinine together with patient's anthropometric data. The predictions given by these equations represent an approximated estimate of true GFR. In adults the most commonly used formulas are:

1. The Cockcroft-Gault equation
2. The Modification of Diet in Renal Disease (MDRD) equation
3. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation

The three equations are reported in details in table 1 [10, 11].

Cockcroft-Gault equation for estimated GFR in ml/min	
Men	(140-age × weight) / (72 × serum creatinine)
Women	(140-age × weight) / (72 × serum creatinine) × (0.85)
Where age is in years, weight in kg and serum creatinine in mg/100 ml	
Correction factor for ethnicity: none	
Four-variable MDRD Study equation for estimated GFR in ml/min per 1.73 m2	
Men	(186* x serum creatinine$^{-1.154}$ x age$^{-0.203}$)
Women	(186* x serum creatinine$^{-1.154}$ x age$^{-0.203}$) x (0.742)
Where serum creatinine is measured in mg/100 ml and age in years.	
Correction factor for ethnicity: if black, GFR estimates must be multiplied by 1.210	
CKD-EPI equation	
Men	
serum creatinine ≤0.9 mg dL	141^{α} x (serum creatinine/0.9)$^{-0.411}$ x (0.993)^{age}
serum creatinine >0.9 mg/dL	141^{α} x (serum creatinine/0.9)$^{-1.209}$ x (0.993)^{age}
Women	
serum creatinine ≤0.7 mg/dL	144^{α} x (serum creatinine/0.7)$^{-0.329}$ x (0.993)^{age}
serum creatinine >0.7 mg/dL	144^{α} x (serum creatinine/0.7)$^{-1.209}$ x (0.993)^{age}
Where serum creatinine is measured in in mg/100 ml and age in years Correction factor for ethnicity: if black, the coefficient α is 163 in men and 166 in women CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; GFR = glomerular filtration rate; MDRD = Modification of Diet in Renal Disease. *186 must be replaced with 175 when creatinine measurements are standardized with reference material from the National Institute of Standards and Technology.	

Table 1: eGFR equations in adults.

The Cockcroft-Gault equation tends to overestimate creatinine clearance in edematous or obese people.
The MDRD equation has gained widespread acceptance and is reported by most clinical laboratories. The MDRD study equation has also been widely used to assess the burden of chronic kidney disease in many epidemiologic studies, including the CKD-Gen consortium study, a large, still ongoing, meta-analysis of genome-wide association data in 67,093 Caucasian individuals from 20 population based studies to identify new susceptibility loci for reduced renal function, estimated by serum creatinine (eGFRcrea), cystatin C (eGFRcys), and Chronic Kidney Disease incidence [12, 13]. As MDRD Study equation was developed in people with chronic kidney disease (CKD), it's not accurate (underestimates) in estimating GFR values greater than 60 ml/min/1.73 m^2.
The CKD_EPI equation was recently developed and validated [12] and showed to have better performance than the MDRD equation, especially at higher GFR values.
We chose to use the MDRD formula, despite it's probably less accurate at higher GFR values, as this formula is currently widely used and allowed us to compare our results with published data.

1.2.3. eGFR equation in children

Accurate estimation of kidney function in children is important from a clinical, research, and public health perspective and given the wide range of maturation of skeletal muscle and growth observed in the pediatric population, recognition of abnormal creatinine values is challenging [14].
The Schwartz formula, devised for children in the mid-1970s, estimates GFR from an equation that uses serum creatinine, height, and an empirical constant. GFR as estimated by the Schwartz formula has been used as one of the enrollment criteria for the Chronic Kidney Disease in Children (CKiD) study, a National Institutes of Health–funded North American cohort study whose goal is to recruit children and adolescents with mild to moderate CKD and characterize progression and effects of CKD on cardiovascular, growth, and behavioral indices [15, 16]. The equation is reported in table 2.

Schwartz equation for estimated GFR in ml/min	
Children and adolescent girls	(0.55 × height) / serum creatinine
Adolescent boys (≥ 13 years)	(0.7 × height) / (serum creatinine)
Where age is in years, height in m and serum creatinine in mg/100 ml	

Table 2: eGFR equation in children and adolescent.

1.3. Chronic Kidney Disease

Chronic kidney disease affects about 6-11 % of the general population. It is defined as either kidney damage or decreased kidney function (decreased GFR) for 3 or more months. Kidney disease can be diagnosed without knowledge of its cause and kidney damages are usually ascertained by markers rather than by kidney biopsy. Proteinuria is the principal marker of kidney damage, together with the albumin-creatinine ratio. An albumin–creatinine ratio greater than 17 mg/g in men and greater than 25 mg/g in

women is considered abnormal. Other markers of damage include abnormalities in urine sediment, abnormalities in blood and urine chemistry measurements, and abnormal findings on imaging studies. Persons with normal GFR but with markers of kidney damage are at increased risk for adverse outcomes of chronic kidney disease.

A GFR level less than 60 ml/min per 1.73 m^2 represents loss of half or more of the adult level of normal kidney function. Below this level, the prevalence of complications of chronic kidney disease increases. Because GFR declines with age, the prevalence of chronic kidney disease increases with age; approximately 17% of persons older than 60 years of age have an estimated GFR less than 60 ml/min per 1.73 m^2.

Nowadays CKD is a common condition, continuously increasing in prevalence. Due to its impact on quality of life, cardiovascular disease (CVD) incidence, and mortality, it is an important public health challenge [17].

Despite its increasing prevalence, our understanding of the underlying risk factors and pathophysiologic mechanisms remains incomplete. Obesity, smoking, hypertension, high cholesterol, and diabetes are key risk factors for CKD. However, additional underlying factors contribute to its etiology. In particular, studies have consistently demonstrated important genetic contributions to the disease onset [13].

Identification of biomarkers that allow early identification of CKD patients at high risk of progression would be useful for early and targeted treatment to improve patient care [18].

1.3.1. CKD Classification

Data from the Third National Health and Nutrition Examination Survey (NHANES III) show the increasing prevalence of complications of chronic kidney disease at lower levels of GFR. This provide a strong basis for using GFR to classify the stage of severity of chronic kidney disease. Table 3 shows the classification of stages of chronic kidney disease estimated using data from NHANES III [19, 20].

Stage	Description	GFR, ml/min per 1.73 m^2
1	Kidney damage with normal or increased GFR	≥ 90
2	Kidney damage with mildly decreased GFR	60 to 89
3	Moderately decreased GFR	30 to 59
4	Severely decreased GFR	15 to 29
5	Kidney failure	< 15 or dialysis

Table 3: Stages of CKD.

1.4. Renal Dysfunction and Hypertension

A strong relationship has been reported between renal dysfunction, hypertension and cardiovascular diseases. Structural and functional cardiac abnormalities have been demonstrated in subjects with mild renal function impairment. Cardiovascular disease is a frequent complication of renal failure and is the most common cause of death in patients with CKD. Kidney dysfunction and loss of blood pressure control have a reciprocal crucial

role in the pathophysiology of hypertension: renal disease could be either a cause or an effect of hypertension. It is epidemiologically more common to see renal failure leading on to hypertension, but the converse is controversial [19, 21].

Glomerular vessels regulate the blood flow by vasoconstriction or vasodilatation depending on the perfusion pressure, to keep the perfusion at the glomerulus constant. Prolonged high perfusion pressures can lead to significant vasoconstriction, which can then cause localized damage to the glomeruli. This can cause necrosis of the glomeruli leading to microalbuminuria, which could lead to significant proteinuria if the disease is not treated. Renal failure could also be an effect of atherosclerosis affecting the renal arteries, leading to under-perfusion [21].

Hypertension also plays a major role in cardiac damage in CKD via left ventricular hypertrophy induction. In addition, a reduction in coronary reserve and capillary density that occurs in CKD patients exposes them to coronary ischemia, which in turn leads to worsening of ventricular dysfunction [19, 22].

Several mechanisms are involved in the strong inter-relationship between hypertension and renal failure, including the renin-angiotensin system, the sympathetic nervous system activation and the endothelial dysfunction.

Renin-angiotensin System

Sodium retention and activation of the renin-angiotensin system have been considered the most important mechanisms involved in the elevation of blood pressure in subjects with kidney disease.

Angiotensin II is a vaso-active peptide that causes blood vessels to constrict, resulting in increased blood pressure. Angiotensin II also stimulates the secretion of the hormone aldosterone from the adrenal cortex. Aldosterone causes the tubules of the kidneys to increase the reabsorption of sodium and water into the blood. This increases the volume of fluid in the body, which also increases blood pressure. If the renin–angiotensin–aldosterone system is too active, blood pressure will be too high [22].

Sympathetic nervous system activation

In kidney failure plasma catecholamine concentration is elevated and increased nerve sympathetic traffic has been demonstrated.

There is clear evidence that Angiotensin II enhances sympathetic activity, both at peripheral and central sites. It stimulates the muscle sympathetic nerve activity (MSNA), which indicates central sympathetic activation and enhances norepinephrine release through a presynaptic effect. Sympathetic activation results in further activation of the renin-angiotensin system.

Furthermore during renal ischemia adenosine is released. Adenosine evokes an increase in afferent renal nerve traffic.

Basal activity of central sympathetic activity is inhibited by central nitric oxide production. Inhibition of nitric oxide synthesis results in an increase of central sympathetic activity. Experimental studies further suggest that stimulation of the central sympathetic nervous system activity by renal afferent impulses may be mediated by local activation of Angiotensin II, which stimulates central sympathetic outflow by inhibition of NOSmRNA abundance [22].

Renalase, a new regulator of cardiac function and blood pressure produced by the kidney has been recently identified. Renalase metabolizes catecholamines (dopamine-

epinephrine-norepinephrine) and is secreted into plasma and urine of healthy persons (not in uremic individuals). To what extent the impairment of renalase production contributes to sympathetic hyperactivity and blood pressure elevation in CKD remains to be established [22].

Endothelial dysfunction

Endothelial dysfunction and remodeling of blood vessels may participate not only in vascular complications in patients with kidney disease but also in the maintenance of elevated blood pressure.

Hypertension, together with dyslipidemia and diabetes, are considered major risk factors for the development of endothelial dysfunction and progression of atherosclerosis. Inflammatory mediators and activation of the renin-angiotensin system contribute, through enhanced production of reactive oxygen species, to atherogenesis in CKD [23].

Endothelins are powerful vasoconstrictor peptides that also play numerous other functions in many different organs. Endothelin-1 (ET-1) is the most abundant and important of this family of peptides in blood vessels. ET-1 is involved in renal and cardiac target organ damage in hypertension. Production of ET-1 is increased in the endothelium and the kidney in salt-dependent models of hypertension. It induces an inflammatory response by increasing oxidant stress in the vascular wall, which leads to vascular remodeling and endothelial dysfunction. ET-1 production is increased by salt loading, which by activation of renal ETB receptors inhibits sodium reabsorption. ET-1 acts by stimulating ETA and ETB receptors. Both ETA and ETB receptors are localized on vascular smooth muscle cells where they induce their vasoconstrictor, proliferative and hypertrophic action. ETA receptors are the predominant ET vasoconstrictor receptors in arteries. Vasoconstrictor ETB receptors are present in the veins and pulmonary vessels in larger numbers than in arteries, although ETA still predominate over ETB receptors in these vessels. ETB receptors are also localized on endothelial cells and act through the production of NO and prostacyclin to exert a vasodilator effect. The endothelin system has been demonstrated to participate patho-physiologically in numerous conditions, including hypertension and renal failure [24].

1.5. Genome Wide Association Study

A genome-wide association study (GWAS) is a powerful hypothesis-free approach to unravel the genetic component by association analyses of CKD with several million genetic variants distributed across the genome. The aim of an association study is to map genetic loci associated to a complex phenotype by comparing the genetic variations between individuals with the disease (cases) and without the disease (controls), in order to identify allelic variants that are enriched among sufferers of the disease. The phenotype used in genome-wide association studies can be either qualitative (disease yes or no) or quantitative [25].

Mapping genetic loci with GWAS is based on Linkage Disequilibrium (LD), which is defined as a condition in which some combinations of alleles or genetic markers occur more or less frequently than can be accounted by chance. LD indicates that the two alleles rely on the same DNA strand and are transmitted together [26].

Linear (for continuous traits, e.g. eGFR) and logistic (for dichotomous traits, e.g. CKD) regression models are used to calculate the mean shift in the distribution (for quantitative traits) or in the disease probability (for dichotomous phenotypes) per risk allele

(compared to the other, the reference allele).
With respect to the candidate genes approach, the genome wide association studies allow to identify new chromosomal regions containing disease-susceptibility loci by detecting differences in allele frequencies between cases and controls. Association analysis is a good choice of methodology to look for genetic factors implicated in complex diseases but we must be careful with the many factors that can lead to false-positive or false-negative results: issues related to sample size, power, genetic effect sizes, replication and population stratification must be taken into account when planning an association study of any complex disease.
GWAS do not provide data on all genetic variability but only low risk-common alleles (i.e those that are represented in more than 1-5% of the population) with large effects are identified: GWAS identify loci and not genes and cannot easily identify loci at which there are many rare allele in any given population. This approach is designed to find loci that fit the common disease-common variant hypothesis of human disease and such variants are not sufficiently frequent to be captured by current GWA genotyping arrays [27].
Since the publication of the first GWAS in 2005, this method has led to the discovery of novel loci for numerous human common diseases and phenotypes. The GWAS approach has been fostered by the publication of the human genome just over ten years ago [28, 29], by the identification of millions of single nucleotide polymorphisms across populations from the "International HapMap Project" (International HapMap Consortium, 2005), by the knowledge that subsets of these SNPs can capture common genetic variations via linkage disequilibrium and by advances in the microarray-based technology. With the high-throughput genotyping technologies offered by companies such as "Illumina" and "Affymetrix", it is now possible to rapidly genotype more than one million SNPs across the whole genome per person in a single analytical process.

1.6. State of the art on genetic loci found associated with eGFR

Multiple studies such as familial studies have provided evidence for a genetic component to kidney disease. Heritability estimates of eGFR are reported between 0.41 and 0.75 in individuals with the major CKD risk factors hypertension or diabetes, and as 0.33 in a population-based sample [30].
The first large GWAS Consortium (CHARGE) [30] identified susceptibility variants for renal function and CKD at the UMOD, SHROOM3, and STC1 loci in nearly 20,000 individuals. By the way, together, single nucleotide polymorphisms (SNPs) at these loci, explain only 0.43% of the variance in eGFR, suggesting that additional loci remained to be identified.
The CKDGen Consortium realized a huge meta-analysis in up to 130,600 European ancestry participants belonging to 25 different cohorts for the discovery study and 17 different cohorts for the replication study. They analyzed renal function using three different phenotypes: eGFR evaluated from serum creatinine, eGFR evaluated from serum cystatin c and CKD, defined as eGFRcrea < 60 ml/min/1.73m^2. Using eGFR estimate from serum cystatin allows discriminating genetic variants truly related to kidney function from those related to serum concentrations of the estimation marker and creatinine metabolism.
Kottgen et al (2010) identified 20 new loci associated with eGFR and CKD. Of these, 13 are likely to be involved in renal function and in susceptibility to CKD (in or near LASS2, GCKR, NAT8/ALMS1, TFDP2, DAB2, SLC34A1, VEGFA, PRKAG2, FAM122A/PIP5K1B, ATXN2,

DACH1, UBE2Q2/FBXO22, and SLC7A9), whereas 7 are likely to be associated with creatinine production or secretion (CPS1, SLC22A2, TMEM60, WDR37, SLC6A13, WDR72, and TBX2/BCAS3) [19].

SNPs in the GATM and CST3 gene regions have been associated repeatedly with GFR. Variants in the GATM gene, involved in creatine synthesis are associated with GFR estimated from serum creatinine [13, 19, 31].

The UMOD risk variant, that shows association with GFR estimated from serum creatinine and with GFR estimated from serum cystatin C, likely is related to true GFR [31].

The CKD-Gen Consortium further investigated association with 16 genetic marker loci previously identified as associated to kidney function and prevalent CKD evaluating the relevance of these SNPs to the initial development of CKD or to ESRD risk. Of the 16 SNPs analyzed, 11 showed association with incident CKD: SNPs in UMOD, PRKAG2, ANXA9, DAB2, SHROOM3, DACH1, STC1, SLC34A1, ALMS1/NAT8, UBE2Q2 and GCKR showed p-values ranging from p = 4.161029 in UMOD to p = 0.03 in GCKR. The odds ratios (OR) for incident CKD of the minor alleles at each of the 11 loci ranged from 0.76 per copy of the T allele (allele frequency 18%) at the UMOD locus to 1.19 per copy of the A allele (allele frequency 22%) at PRKAG2. After additional adjustment for baseline eGFR, 6 SNPs (at the UMOD, PRKAG2, ANXA9, DAB2, DACH1 and STC1 loci) remained significantly associated with incident CKD, with minimal attenuation of effect size. Only two SNPs showed nominally significant associations with ESRD: rs1260326 in GCKR (OR= 0.93; p-value=0.03) and rs12917707 in UMOD (OR = 0.92; p-value = 0.04). From a functional point of view, the genes identified in this study were involved in calcium and phosphate homeostasis (STC1), the production of Tamm-Horsfall protein (UMOD), epithelial cell shape regulation (SHROOM3), nephrogenesis (ALMS1, VEGFA, potentially also DACH1), glomerular filtration barrier integrity and podocyte function DAB2, VEGFA), angiogenesis (VEGFA), metabolic kidney function (PRKAG2, potentially CKR, LASS2), ciliary function (ALMS1, GCKR-IFT172) and solute transport (SLC7A9, SLC34A1).

Pattaro and collegues (2012) then identified 6 new loci in association with eGFR in or near MPPED2 (rs3925584), DDX1 (rs6431731), SLC47A1 (rs2453580), CKD12 (rs11078903), CASP9 (rs12124078) and INO80 (rs2928148). Significant associations between SNPs in SLC22A2 and eGFR were also identified in previous GWAS performed by the CKDGen Consortium [13, 32]. Differential expression of CDK12 (Cyclin-dependent kinase 12) inhibitors have been described in human glomerular disease. The CASP9 (caspase-9) gene encodes the third apoptotic activation factor that is involved in the activation of cell apoptosis, necrosis and inflammation. In the kidney, caspase-9 may play an important role in the medulla response to hyperosmotic stress and in cadmium-induced toxicity. For the remaining variants no previous link with renal function are reported [32].

Böger et al. (2011) analyzed the association of markers with the initiation of CKD in more than 26,000 individuals from the general population using serial estimations of kidney function, and with ESRD in four case-control studies in subjects of European ancestry (3,775 cases, 4,577 controls) [33].

Chambers et al. (2010) identified common variants at 2p12–p13, 6q26, 17q23 and 19q13 associated with serum creatinine, marker of kidney function. Of these, rs10206899 (near NAT8, 2p12–p13) and rs4805834 (near SLC7A9, 19q13) were also associated with CKD [34].

Wang et al. (2001) demonstrated that the incidence of hypertension is influenced by genes encoding the angiotensin-converting enzyme (ACE; insertion/deletion [I/D]) polymorphism, alfa-adducin (Gly460Trp), and aldosterone synthase (344C/T). By interfering with blood pressure or sodium homeostasis, these genetic polymorphisms also may change renal function. They therefore investigated serum creatinine level, calculated and measured creatinine clearances, and 24-hour urinary protein excretion in subjects previously genotyped for these three polymorphisms and found that renal function was slightly but consistently impaired when both the ACE D and alfa-adducin Trp alleles were present, while the aldosterone synthase T allele did not strengthen genetic associations with the ACE D allele considered alone or in combination with the alfa-adducin Trp allele [35].

The Framingham Heart Study investigated possible association with GFR as marker of kidney function testing about 1,000 participants for 70,980 SNPs and identified as best hits rs2839235 (p-value = $1.6x10^{-5}$) using the GEE (generalized estimating equation) method and rs6434804 using the FBAT (family-based association test) method (p-value = $2.4x10^{-5}$) [36].

1.7. Project Background: The HYPERGENES EUROPEAN PROJECT

Data presented in this thesis are part of data generated within the HYPERGENES Project, an EU supported IP, under FP7 (http://www.HYPERGENES.eu), aimed at the definition of a comprehensive genetic- epidemiological model of essential hypertension and of the intermediate phenotypes of hypertension, specifically associated Target Organ Damage (TOD).

The HYPERGENES Project investigated associations between genetic variants and EH pursuing a two- stage study by recruiting cases and controls from extensively characterized cohorts recruited over many years in different European regions Continental Italy, Sardinia and North Europe.

The discovery phase consisted of 1,865 cases and 1,750 controls genotyped with the 1 Million SNPs Illumina array. Best hits were followed up in a validation panel of 1,385 additional cases and 1,595 controls that were genotyped with a custom array of 14,055 SNPs. A new hypertension susceptibility locus was identified, localized in the promoter region of endothelial Nitric Oxide Synthase (*eNOS*) gene (rs3918226) showing strong association with EH (combined p-value = $2.58x10^{-13}$, OR = 1.54, 95% CI 1.37-1.73). This finding was further validated with a meta-analysis, using other *in-silico*/*de novo* genotyping data for a total of 21,714 subjects. Meta-analysis' results showed an overall OR of 1.34 (95% CI 1.25-1.44, p-value of $1.032x10^{-14}$). A quantitative analysis on a sample of 1,820 population-based individuals was then performed and revealed an effect size of 1.91 (95% CI 0.16-3.66) for systolic BP and 1.40 (95% CI 0.25-2.55) for diastolic BP. A potential binding site for transcription-factors of ETS family directly next to rs3918226 was identified *in-silico,* suggesting a potential modulation of eNOS expression. The impact of rs3918226 on hypertension was further validated in a Longitudinal cohort study. The incidence of hypertension was investigated in relation to eNOS rs3918226 genotypes and their association with longitudinal changes in systolic and diastolic blood pressure in EPOGH (European Project on Genes in Hypertension) and FLEMENGHO (Flemish Study on Environment, Genes, and Health Outcomes) cohorts [37].

This last part of the project focuses on the association between hypertension and estimated Glomerular Filtration Rate (eGFR), a target parameter to evaluate kidney function damages and on the identification of biomarkers related to eGFR in a population based study.

2. AIM OF THE STUDY

The aim of the present work is to identify genetic markers of GFR in hypertensive subjects and in a population based study that includes a Cross sectional and a Longitudinal study.
The analyses have been carried on using three different approaches:

- Study 1: A genotype-phenotype association with eGFR as quantitative phenotype in two groups of hypertensives: 1,634 individuals from the Discovery phase and 1,198 subjects from the Validation phase of HYPERGENES.
- Study 2: An association analysis between SNPs and baseline eGFR in the Epogh-Flemengho population based sample (n=2,697).
- Study 3: An association analysis between SNPs and Longitudinal changes in eGFR in subjects described at point 2 that have been followed for up to 23 years (n=1,952).

3. MATERIAL AND METHODS

3.1. Sample Description

3.1.1. HYPERGENES hypertensive samples

This sample included hypertensives recruited during the Discovery and Validation phases of the HYPERGENES project for which eGFR phenotype was available. Subjects belong to three different geographic areas: North Europe, Continental Italy and Sardinia.

Participants were included in the study as hypertensives if they could self report to be of Caucasian origin, were unrelated with other participants, had diastolic blood pressure (DBP)>90 mmHg and systolic blood pressure (SPB)>140 mmHg or were under antihypertensive treatment before the age of 50.

We studied 1,634 hypertensive subjects of HYPERGENES Discovery phase genotyped for 1,111,731 SNPs. In order to validate the results of this study, we performed a meta-analysis with the CKDGen Consortium results. The meta-analysis was performed only for the SNPs that in both dataset showed a p-value $\leq 10^{-3}$. The CKDGen sample included 74,354 subjects analyzed for renal function using eGFR evaluated from serum creatinine, eGFR evaluated from serum cystatin c and also defining chronic kidney disease based on eGFRcrea<60 ml/min. eGFR based on serum cystatin allows to discriminate true susceptibility loci for renal function from those related to creatinine production and secretion [14, 32].

We also studied 1,198 cases from the Validation phase genotyped for 14,055 SNPs (custom chip).

3.1.2. Population Based Sample

The population based sample was recruited during several years for the Epogh-Flemengho study. It included 2,697 individuals, from 10 to 83 years old, with eGFR measurement available. From the overall sample, 1,952 subjects had multiple follow up measurements for eGFR. In the population based sample we performed a genotype-phenotype association in:

1. A Cross sectional analysis on 2,697 individuals using baseline eGFR as quantitative phenotype.
2. A Longitudinal analysis in a subset of the Cross sectional sample composed by 1,952 subjects with multiple eGFR measurements. We calculated delta eGFR as the difference between the last and the first measurement available for every subject.

3.2. Cohorts contributing to the Study

Table 4 summarizes the cohorts used during the different stages of the analysis. Each cohort is then described in detail below.

Cohort	Hypergenes Discovery	Hypergenes Validation	Population based Cross sectional	Population based Longitudinal
EPOGH-FLEMENGHO	X	X	X	X
Milano and Sassari	X	X		
Immidiet	X			
WHSS	X			
Paris		X		
Progress		X		

Table 4: Cohorts involved in the different stages of the analysis.

- The **FLEMENGHO** (Flemish Study on Environment, Genes and Health Outcomes) - **EPOGH** (European Project on Genes in Hypertension) cohort has been the first large-scale study on genetic epidemiology of blood pressure and associated phenotypes in Europe and has already produced 35 scientific papers. It recruited family-based random samples in 5 eastern and 2 western European countries [38, 39]. Subjects were followed up to 23 years. According to the HYPERGENES "macro-region concept", the FLEMENGHO-EPOGH cohort is considered a "North-European" cohort. In the Hypergenes project, from this cohort, 186 cases with eGFR data available were analyzed in the discovery phase while 43 hypertensives were used in the Validation analysis. For the Population Based study we could analyze 2,697 subjects in the Cross sectional analysis on eGFR. As 1,951 subjects also had multiple follow up measurement available, we analyzed them in the Longitudinal study on delta eGFR.
- **Milano** and **Sassari cohorts** (Italian and Sardinian cohorts) have been collected with a focus on the pathophysiology of essential hypertension and TOD. In most hypertensives phenotypes were collected before any pharmacological treatment was started [40, 41].

  Overall the Italian cohort available for HYPERGENES includes 514 hypertensives with eGFR data available in the Discovery phase and 238 in the Validation phase. Sardinian cohort includes 704 hypertensives with eGFR available in the Discovery phase and 372 in the Validation phase.
- The **IMMIDIET** study is a population-based Cross sectional study, funded by the European Union (FP5) [42]. It compares healthy couples from regions of England, Belgium and Italy in order to evaluate the present dietary habits and the risk profile of the three communities at different risk of myocardial infarction. A large body of clinical (mostly cardiovascular) and environmental data is available. In the HYPERGENES discovery phase we analyzed 158 cases with eGFR data

available.

- The cohort of the **Wandsworth Heart & Stroke Study** (**WHSS**) [43] has the characteristic of being a multiethnic community-based study, drawn from the same geographical area of South London. The sample totals 1,577 individuals (~40% hypertensives), evenly distributed for ethnic origin; 33% Caucasians, 33% of African descent, 33% of South Asian descent. The population has been extensively phenotyped. Mortality follow-up through death certificates and cancer registrations are available. It is a unique multiethnic cohort sharing the same urban environment. 72 cases of Caucasian origin and with eGFR phenotype available were included in the Discovery phase.
- **Paris (France):** The French hypertensive patients were selected from the HYPERGENES dataset of hypertensive families consecutively recruited in Paris, since 1990 [44]. The dataset comprises more than 1,000. In this study, 447 hypertensives had eGFR phenotype available for the Validation analysis.
- **PROGRESS cohort.** The "Perindopril protection against recurrent stroke study" (PROGRESS) was designed to determine the effects of a blood-pressure-lowering regimen in hypertensive and non- hypertensive patients with a history of stroke or transient ischemic attack [45]. 113 affected subjects had eGFR measurement available and were used for the Validation analysis.

3.3. Phenotype

The quantitative phenotype was eGFR calculated using the MDRD equation for adults and using the Schwartz formula for children (table 5).

Four-variable MDRD Study equation for estimated GFR in ml/min per 1.73 m^2	
Men	(186 x serum creatinine$^{-1.154}$ x age$^{-0.203}$)
Women	(186 x serum creatinine$^{-1.154}$ x age$^{-0.203}$) x (0.742)
Where serum creatinine is measured in mg/100 mL and age in years.	
Correction factor for ethnicity: if black, GFR estimates must be multiplied by 1.210	
Schwartz equation for estimated GFR in ml/min	
Children and adolescent girls	(0.55 × height) / serum creatinine
Adolescent boys (≥ 13 years)	(0.7 × height) / (serum creatinine)
Where age is in years, height in m and serum creatinine in mg/100 mL	

Table 5: MDRD and Schwartz equation.

In the Longitudinal study we used delta eGFR, calculated as the difference between the first and the last measurement available in each patient.

3.4. Genotyping and data filtering

3.4.1. Genotyping and data filtering in HYPERGENES Discovery

DNA was extracted from peripheral blood with standard procedures.

In the discovery phase all samples were genotyped using the Illumina 1M-duo arrays (Illumina Inc, San Diego, CA, USA). The chip captures 1,199,187 SNPs. In addition to markers necessary for broad genome coverage the chip contains 21,877 non-synonymous

SNPs, 51,207 SNPs in sex chromosomes, 138 in mitochondrial DNA, 35,969 SNPs covering recently reported copy number variant regions, 30,908 SNPs in MHC and ADME regions. Genotyping was performed in two different genotyping centres: Milan University (UNIMI) and Lausanne University (UNIL). UNIMI genotyped 2,064 subjects (1,270 controls and 794 cases) whereas UNIL performed the genotyping of 1,995 subjects (665 controls and 1,330 cases).

All raw intensity data were collected in UNIMI genotyping centre and analyzed with the Illumina Software Genome Studio for genotype calling, using the Illumina reference cluster file. A DNA call rate threshold was set at 0.95 and DNAs with call rate ≤ 0.95 were excluded from the final data set. For each DNA, data from X chromosome were used to check for discordance with ascertained sex.

After association analysis, a final assessment of genotype quality was performed for the significantly associated SNPs (p-value<$1 \cdot 10^{-4}$) with visual inspection of cluster plots. As genotyping was performed in two Laboratories, replications were designed in order to estimate the genotyping error rate between the two Genotyping Centres. DNAs from 20 individuals were genotyped with the 1M-duo Illumina's BeadChips.

Within samples with Call Rates >0.95, 99.624% of the genotype calls were concordant and 0.376% were discordant. A cross-check between the scan performances of the two Illumina iScan platforms was also performed. Ten chips (20 samples) were processed and scanned in UNIMI (average DNA call rate 0.994). The same chips were re-scanned in UNIL. We could demonstrate that 99.189% of the calls were concordant and 0.0053% was discordant between the two platforms. For 0.805% of the calls we could not do the comparison since this percentage accounts for calls available in the chip scanned in one Lab but not in the other and vice-versa.

3.4.2. Genotyping and data filtering in HYPERGENES Validation

Genotyping was performed on 2,869 samples by Milan University (UNIMI) and Lausanne University (UNIL), who genotyped 1,700 and 1,169 samples respectively, for 14,055 SNPs, using an Illumina custom chip (15K) built *ad hoc* on the basis of the Discovery results. All raw intensity data were collected in UNIMI genotyping centre and analyzed with the Illumina Software Genome Studio. For custom chips Illumina doesn't supply the reference cluster file for genotype calling, we therefore ran training samples to generate a reference cluster file. To create this file all SNPs were visually inspected in order to check the accuracy of the cluster plots. Three hundred and forty ambiguous SNPs were excluded.

Genotype calling for the whole sample was performed using the custom reference file. In the Validation phase the DNA call rate threshold was set at 0.98 and this led to the exclusion of 173 DNAs from the final data set that comprised of 2,696 DNAs.

3.4.3. Genotyping and data filtering in Epogh Flemengho Population based study

Genotyping was performed on 3,430 samples by Milan University (UNIMI) and Lausanne University (UNIL), who genotyped 192 and 3,238 samples respectively for 14,055 SNPs, using the 15K Illumina custom chip. All raw intensity data were collected in UNIMI genotyping centre and analyzed with the Illumina Software Genome Studio.

Genotype calling for the whole sample was performed using the custom reference file previously generated. The DNA call rate threshold was set at 0.98 and this led to the exclusion of 533 DNAs from the final data set that comprised of 2,897 DNAs.

3.5. Statistical softwares

3.5.1. PLINK

PLINK [46] is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner. The focus of PLINK is purely on analysis of genotype/phenotype data. PLINK is being developed by Shaun Purcell at the Center for Human Genetic Research (CHGR), Massachusetts General Hospital (MGH), and the Broad Institute of Harvard & MIT, with the support of others.

3.5.2. Eigensoft

The Eigensoft package (version 3.0 for Linux platform, Department of Genetics, Harvard Medical School, Boston, USA) combines functionality from population genetics methods and EIGENSTRAT stratification correction method [47, 48]. The EIGENSTRAT method uses principal components analysis to explicitly model ancestry differences between cases and controls along continuous axes of variation; the resulting correction is specific to a candidate marker's variation in frequency across ancestral populations, minimizing spurious associations while maximizing power to detect true associations. The EIGENSOFT package has a built-in plotting script and supports multiple file formats and quantitative phenotypes.

3.5.3. STATA SE

STATA SE (Version 11) is a commercial data analysis and Statistical software. The software offers different tools to data analysis, management, graphics, matrix language, linear models, survey statistics, multivariate methods, multinomial etc.

3.5.4. R

R (version 2.14.1 Copyright 2011 The R Foundation for Statistical Computing) is a free language and environment for statistical computing and graphics. It is a GNU project similar to the S language and environment, which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, it is highly extensible and presents some packages implementing statistical methods and algorithms for the analysis of genetic data and for related population genetics studies.

Particularly we used GenABEL package, a widely used GWAS R package aiming to facilitate statistical analyses of GWAS data using fast procedures for specific genetic tests such as data quality control, testing of association of single nucleotide polymorphisms with binary or quantitative traits, visualization of results and easy interfaces to standard statistical and graphical procedures [49].

3.6. Quality Control

3.6.1. Quality Control on the HYPERGENES Discovery sample

All QC steps were performed in accordance with the protocol written by C.A Anderson et al [50]. The HYPERGENES Discovery sample included 4,059 individuals genotyped for 1M SNPs. 143 Samples having Call Rate <0.95 were excluded. 56 subjects with genotypic sex mismatch (difference between the gender reported in clinical data and the one estimated with sex SNPs genotyped) were identified and removed from the analysis. Using genome-wide IBD estimation (PLINK version 1.7 [46]) we identified and removed from the analysis

64 duplicated and 156 related subjects (44 family components, 63 siblings, 23 parent/offspring, 26 second degree).
In this study, we selected a sample composed by hypertensives with eGFR data available (1,634 individuals). We performed a heterozigosity rate test (PLINK version 1.7) and based on this results (threshold: ± 3 sd from average) 22 subjects were excluded. Principal Component Analysis (PCA) was therefore performed on this sample, using the EIGENSOFT package (version 2.0) [47, 48]: 8 outliers were removed as individuals that exceed a default number of standard deviations (6.0) from the whole sample. After this quality control the final hypertensive sample comprised 1,604 individuals (1,083 males and 521 females).
173,624 SNPs with a minor allele frequency (MAF) <1% and 74,501 with call rate <99% were removed leaving 881,299 SNPs for analysis.

3.6.2. Quality Control on the HYPERGENES Validation sample

The HYPERGENES Validation sample was composed by 2,696 individuals genotyped using the 15K custom chip. After data quality control for relatedness/duplicates (32 duplicated and 33 related subjects) the final case-control sample was composed by 2,631 subjects, 1,385 hypertensives and 1,246 controls (1,417 males and 1,214 females).
We selected a hypertensive sample of 1,198 individuals with eGFR data. A heterozigosity rate test (PLINK version 1.7) was performed and 12 individuals were excluded (threshold: ± 3 sd from average). Principal Component Analysis (PCA) was then performed using the EIGENSOFT package (version 2.0). 3 outliers were removed as individuals that exceed a default number of standard deviations (6.0) from the whole sample.
200 SNPs were filtered out on the basis of SNP genotype call rates (<99%) and 2,657 SNPs for MAF (MAF<0.01). After quality control the final hypertensive sample comprised 1,183 individuals and 10,800 SNPs.

3.6.3. Quality Control on the Epogh Flemengho Population based study

The Epogh Flemengho Population based sample was composed by 2,897, genotyped using the 15K custom chip. After data quality control for duplicates (59 duplicated subjects) the final case-control sample was composed by 2,838 subjects.
We analyzed 2,697 subjects in the Cross sectional analysis and 1,951 individuals for the Longitudinal study. In the Cross sectional analysis 462 SNPs were excluded for low call rate (< 99%) while 2,749 SNPs for Minor Allele Frequency (< 0.01): 10,527 SNPs were therefore available for the analysis. In the Longitudinal study 491 SNPs were excluded for low call rate (< 99%) and 2,764 SNPs for Minor Allele Frequency (< 0.01): 10,483 SNPs were finally available for the analysis.

3.6.4. Quality control summary

The following table (6) resumes the number of SNPs and subjects used in each analysis.

Study	Samples	SNPs
HYPERGENES Discovery analysis	1,604	881,299
HYPERGENES Validation analysis	1,183	10,800
Population based: Cross sectional analysis	2,697	10,527
Population based: Longitudinal analysis	1,951	10,483

Table 6: Summary of number of subjects and SNPs analyzed in each analysis.

3.7. Statistical Analysis

3.7.1. HYPERGENES hypertensive sample

The basic statistical analyses were performed in accordance with the protocol written by G.M. Clarke et al [51]. In the HYPERGENES Discovery sample, each SNP was tested for association with eGFR using a linear regression under an additive model with adjustment for sex, age, body mass index, systolic blood pressure, diastolic blood pressure, antihypertensive treatment (Y/N), ethnic group and for the first 10 PCs, as implemented in PLINK.

In the Validation phase, linear regression analysis was carried out using an additive genetic model adjusted for sex, age, body mass index, systolic blood pressure, diastolic blood pressure, antihypertensive treatment, ethnic group and for the first 10 PCs.

Combined analysis for HYPERGENES discovery and CKDGen Meta-analysis results was conducted using both Z-score and inverse variance weighting meta-analysis as implemented in METAL [52].

3.7.2. Population Based study

In the Cross sectional study each SNP was tested for association with eGFR using a polygenic model with mmscore as implemented in R, GenABEL package [53, 54, 55, 56, 49]. The model was adjusted for pedigree (random effect), sex, age, body mass index, systolic and diastolic blood pressure, antihypertensive treatment (Y/N) and recruitment centre (fixed effects).

In the Longitudinal study each SNP was tested for association with delta eGFR using a polygenic model with mmscore as implemented in R, GenABEL package [53, 54, 55, 56, 49]. Delta eGFR was calculated as the last available eGFR minus the baseline eGFR. The model was adjusted for pedigree (random effect), age at baseline, eGFR at baseline, gender, BMI at baseline, time to follow up, centre (fixed effects).

4. RESULTS

4.1. Sample Characteristics

4.1.1. HYPERGENES Discovery and Validation Sample

We studied 1,634 hypertensive subjects from the HYPERGENES Discovery phase and 1,198 hypertensives from the HYPERGENES Validation phase. Characteristics of the study samples, detailed by ethnic groups, are reported in table 7 and 8. The mean age of the Discovery participants varied from 45 years in south Europe subjects to 50 years in Sardinia subjects and from 42 to 50 years in the south Europe and Sardinia Validation groups respectively. The mean observed eGFR across the three ethnic groups ranged from 80.78 ml/min/1.73m^2 (north Europe) to 95.49 ml/min/1.73m^2 (south Europe) in the Discovery sample and from 83.49 ml/min/1.73m^2 (north Europe) to 97.33 ml/min/1.73m^2 (south Europe) in the Validation sample. eGFR distribution in the two study groups is presented in figure 3 and 4. The mean eGFR in males and females is significantly different in both groups, with a p-value $\leq 10^{-5}$: eGFR distribution according to sex for the HYPERGENES Discovery and Validation samples is shown in table 9 and 10. Figure 5 and 6 show the distribution of eGFR by decade in the two groups; as previously mentioned, GFR declines with normal ageing usually begin after 30–40 years of age and the rate of decline may accelerate after age of 50-60 years: in the Discovery sample it is possible to notice the expected eGFR plateau until 40 years old followed by a progressive decrease (figure 5).

	North Europe		**Sardinia**		**South Europe**	
Number of subjects (%)	410		684		510	
Number (%) with characteristic						
Women	45.61		32.46		21.96	
Treated for hypertension	56.83		0		1.18	
Mean (SD) characteristic						
SBP (mm Hg)	141.90	(20.26)	158.46	(14.47)	148.36	(9.04)
DBP (mmHg)	89.43	(12.19)	102.44	(8.97)	96.35	(5.80)
Age (yrs)	46.74	(9.29)	50.77	(10.60)	45.39	(8.05)
Body mass index (kg/m^2)	27.90	(5.23)	27.48	(3.86)	26.50	(3.11)
eGFR	80.78	(16.46)	84.54	(18.20)	95.49	(18.90)
Serum Creatinine	0.97	(0.19)	0.95	(0.19)	0.89	(0.15)

Table 7: Discovery sample description by ethnic group.

	North Europe		Sardinia		South Europe	
Number of subjects (%)	588		372		238	
Number (%) with characteristic						
Percent women (%)	47.45		48.66		18.91	
SBP (mm Hg)	152.43	(20.62)	158.96	(13.83)	150.22	(14.47)
DBP (mmHg)	95.76	(12.76)	101.83	(7.96)	97.01	(9.75)
age (yrs)	46.9	(8.2)	50.18	(10.14)	42.9	(8.25)
body mass index (kg/m^2)	26.2	(4.29)	27.41	(4.29)	26.36	(3.10)
Treated for hypertension (%)	72.45		0		13.87	
eGFR	83.49	(18.71)	90.83	(20.70)	97.33	(19.29)
Serum Creatinine	0.95	(0.25)	0.86	(0.15)	0.8925	(0.16)

Table 8: Validation sample description by ethnic group.

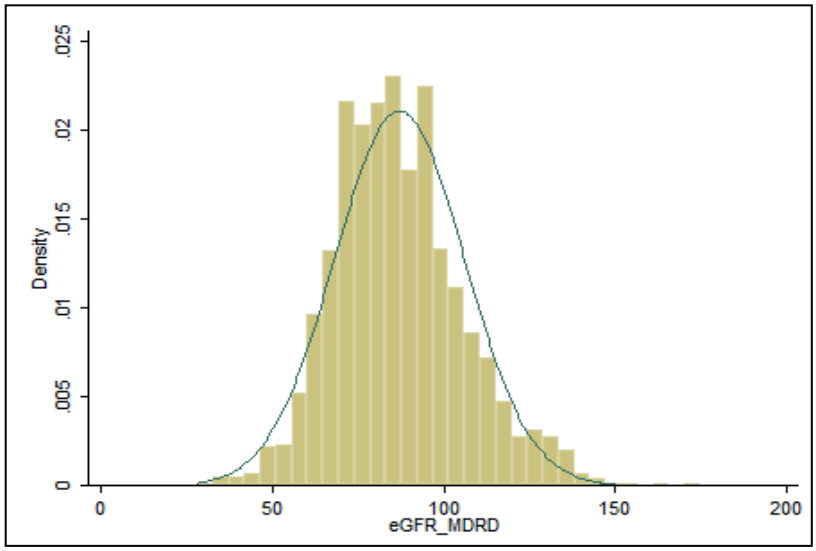


Figure 3: Histogram of eGFR distribution in the Discovery sample.

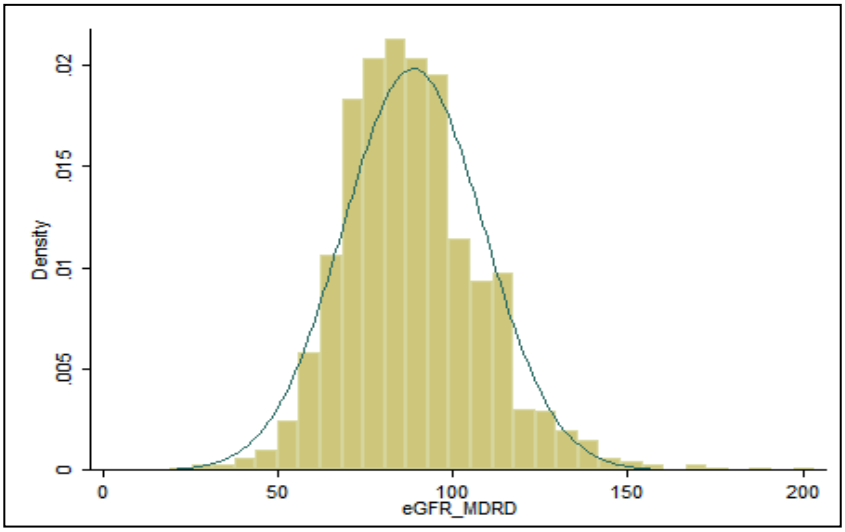


Figure 4: Histogram of eGFR distribution in the Validation sample.

Variable	Sample	N	Mean	Std. Dev.	Min	Max
eGFR	All	1,604	87.06	18.95	28.18	174.81
	Males	1,083	90.71	18.27	35.24	165.18
	Females	521	79.48	18.08	28.18	174.81

Table 9: eGFR distribution according to gender in the Discovery sample.

Variable	Sample	Obs	Mean	Std. Dev.	Min	Max
eGFR	All	1,183	88.63	20.12	19.08	203.48
	Males	686	91.82	18.71	27.54	167.22
	Females	497	84.22	21.16	19.08	203.48

Table 10: eGFR distribution according to sex in the Validation sample.

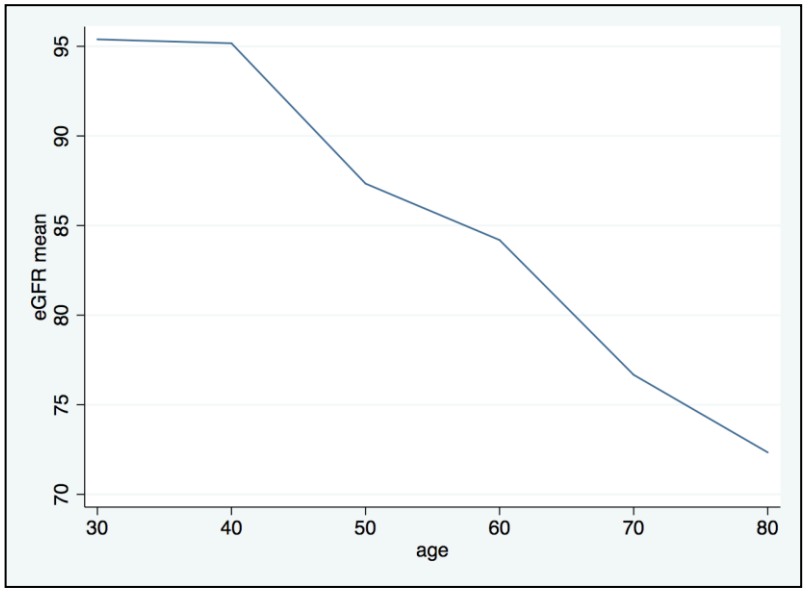


Figure 5: Distribution of eGFR by decade in the Discovery sample.

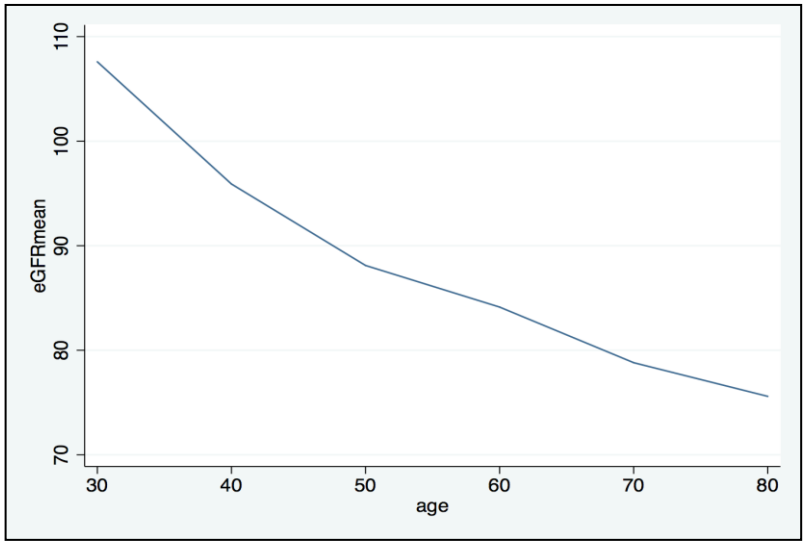


Figure 6: Distribution of eGFR by decade in the Validation sample.

4.1.2. Population Based Sample

For the Cross sectional study we analyzed 2,697 subjects from the EPOGH-FLEMENGHO cohort used in the Cross sectional study. A subset of 1,952 subjects from the same cohort was used in the Longitudinal study. Characteristics of the study samples are reported in table 11. The mean age of the Cross sectional sample was 39 years while it ranged from 40 years at the baseline measurement to 48 years at the follow-up measurement in the Longitudinal study. The mean observed eGFR was 83.49 ml/min/1.73m^2 in the Cross sectional sample and ranged from 84.86 at baseline to 84.26 at follow-up in the Longitudinal group. The mean eGFR value is significantly different between baseline and follow-up measurement (p-value=0.03). The distribution of eGFR in the Cross sectional sample and the distribution of delta eGFR in the Longitudinal sample are presented in figures 7 and 8. Finally, figure 9 reports the distribution of eGFR by decade in the Cross sectional sample.

	Cross Sectional		**Longitudinal**	
Percent women (%)	51.28		52.49	
SBP (mm Hg) baseline	124.31	(17.04)		
DBP (mmHg) baseline	76.39	(11.26)		
Age (yrs) baseline	39.31	(15.99)	39.96	(15.16)
Age (yrs) follow up			47.83	(16.14)
Body Mass Index (kg/m^2) baseline	25.1	(4.7)	25.28	(4.64)
Treated for hypertension (%)	11.38			
eGFR baseline	83.49	(20.31)	84.86	(19.9)
eGFR follow up			84.26	(20.18)
Serum Creatinine baseline	0.98	(0.18)	0.97	(0.19)
Serum Creatinine follow up			0.93	(0.26)

Table 11: Population based sample description.

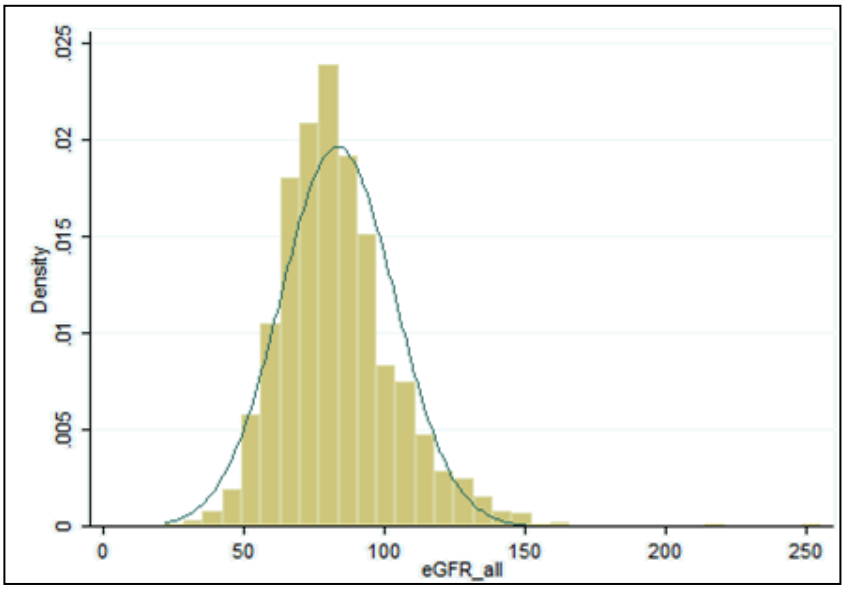


Figure 7: eGFR distribution in the Cross sectional sample.

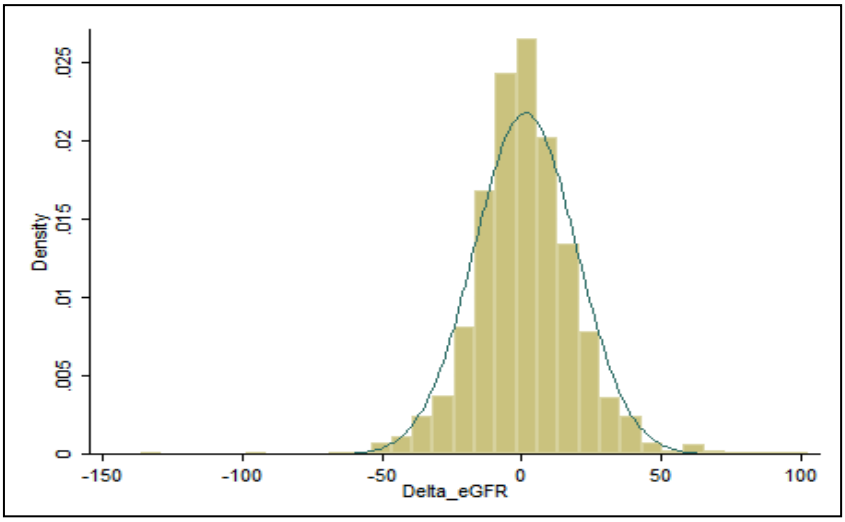


Figure 8: Delta eGFR distribution in the Longitudinal sample.

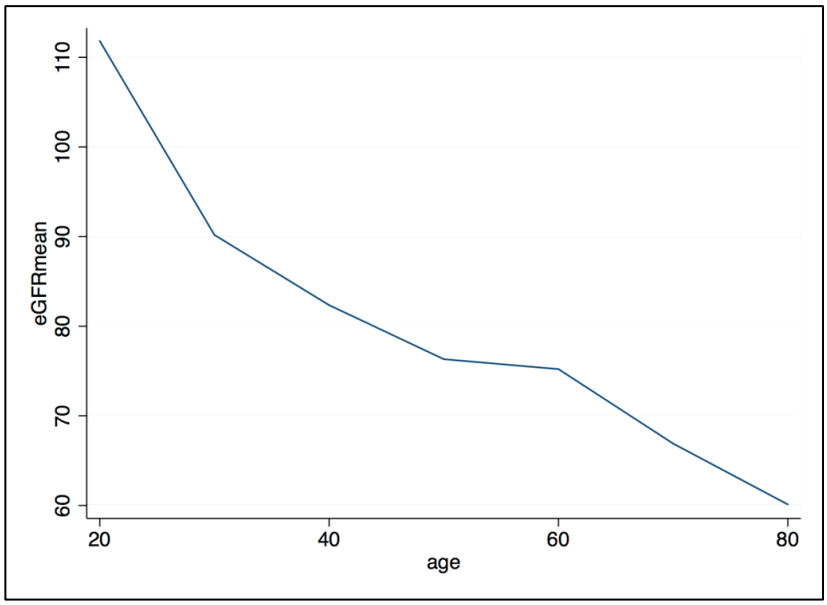


Figure 9: Distribution of eGFR by decade in the Cross sectional sample.

4.2. Results of the Genome-wide study

4.2.1. HYPERGENES Discovery and Validation Analysis

Best results of the single SNP analysis are reported in table 12 for the Discovery sample and table 13 for the Validation sample. Rs4678878 reaches, in the Discover sample, the genome-wide significance ($3.83x10^{-8}$). The first two findings of the Discovery sample, rs4678878 and rs11129714, map in STAC gene, which encodes for a SH3 and cysteine rich domain protein involved in a neuron-specific signal transduction pathway. In the Validation analysis none of the SNPs reach the Bonferroni significance threshold at 10^{-6} (calculated on the basis of the SNPs present on 15K custom chip). In this study, five markers (rs7666226, rs7666425, rs13150531, rs13114026, rs10014602), having a p-value 10^{-4}, lie in PIGG gene, coding for an ethanolamine phosphate transferase involved in glycosyl phosphatidyl inositol anchor biosynthesis. Rs802374, rs17142666, rs17142724 and rs17142737 (p-value =10^{-4}) are localized in KCND2 gene coding for a potassium voltage-gated channel; mutations in this gene are implicated in cardiac conduction disorders.

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
3	rs4678878	36445311	G	-3.421	0.6191	3.83×10^{-8}	STAC	intron	-14549
3	rs11129714	36448844	G	-3.257	0.6381	3.74×10^{-7}	STAC	intron	-11016
9	rs10809319	11054428	A	-3.006	0.6343	2.34×10^{-6}	LOC646087	flanking_5UTR	-234120
3	rs9311801	6149815	A	3.913	0.8469	4.14×10^{-6}	GRM7	flanking_5UTR	-728112
13	rs9557072	98238372	G	-2.696	0.6057	9.14×10^{-6}	DOCK9	flanking_3UTR	-5373
12	rs4765044	124547089	A	2.825	0.6375	1.00×10^{-5}	TMEM132B	intron	-22849
22	rs16989636	30638680	G	-12.57	2.846	1.08×10^{-5}	DEPDC5	flanking_3UTR	-5679
4	rs2046770	169534575	G	2.856	0.6484	1.13×10^{-5}	DDX60L	intron	-1325
17	rs2659030	76792569	A	-2.645	0.602	1.19×10^{-5}	AZI1	intron	-567
8	rs11784679	22887095	C	2.98	0.6816	1.31×10^{-5}	RHOBTB2	flanking_5UTR	-26312
5	rs216124	149442202	A	3.263	0.7494	1.42×10^{-5}	CSF1R	intron	-1422
13	rs6490699	21614323	A	3.413	0.7848	1.45×10^{-5}	FGF9	flanking_3UTR	-440136
16	rs9930152	48389067	A	-2.694	0.6228	1.61×10^{-5}	ZNF423	intron	-8109
9	rs10733219	11068006	G	-2.656	0.6197	1.92×10^{-5}	LOC646087	flanking_5UTR	-247698
13	rs772311	98266480	C	2.652	0.6197	1.99×10^{-5}	DOCK9	intron	-5934
7	rs889870	137302692	A	4.524	1.059	2.06×10^{-5}	CREB3L2	intron	-34198
4	rs12499919	115391056	A	3.197	0.7506	2.17×10^{-5}	ARSJ	flanking_5UTR	-270530
21	rs6517073	32351155	G	2.773	0.6519	2.23×10^{-5}	HUNK	flanking_3UTR	-52907
19	rs467439	6867493	G	2.579	0.607	2.28×10^{-5}	EMR1	intron	-114
3	rs728082	36425383	A	-2.741	0.6469	2.40×10^{-5}	STAC	intron	-28133
1	rs570762	176156776	A	2.636	0.6242	2.56×10^{-5}	SEC16B	flanking_3UTR	-8128
7	rs6973843	105773180	A	-4.213	1.003	2.82×10^{-5}	PBEF1	flanking_5UTR	-60587

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
5	rs1215655	3223620	A	2.522	0.6014	$2.89x10^{-5}$	LOC729099	flanking_5UTR	-275739
9	rs6474626	11069752	A	-2.598	0.6205	$2.98x10^{-5}$	LOC646087	flanking_5UTR	-249444
1	rs11210537	42118310	A	-2.682	0.6434	$3.23x10^{-5}$	HIVEP3	intron	-38361
9	rs10809326	11065596	A	-2.665	0.6423	$3.52x10^{-5}$	LOC646087	flanking_5UTR	-245288
5	rs778447	66314470	A	3.929	0.9471	$3.53x10^{-5}$	MAST4	intron	-3558
6	rs9295289	158387494	G	-5.527	1.335	$3.63x10^{-5}$	SYNJ2	intron	-3075
3	rs4505641	36407085	G	2.531	0.6129	$3.82x10^{-5}$	STAC	intron	-9835
1	rs12024772	42125931	G	-2.639	0.6402	$3.95x10^{-5}$	HIVEP3	intron	-30740
15	rs2387110	89672208	G	2.624	0.6369	$3.99x10^{-5}$	SV2B	flanking_3UTR	-32554
2	rs6546443	68952977	A	-2.601	0.6334	$4.23x10^{-5}$	BMP10	flanking_5UTR	-824
16	rs17752673	47034594	A	3.189	0.7798	$4.53x10^{-5}$	LOC643114	flanking_5UTR	-12362
15	rs8024792	47881243	G	-2.832	0.6928	$4.58x10^{-5}$	LOC645453	flanking_3UTR	-5081
2	rs11676320	232890310	G	4.697	1.149	$4.60x10^{-5}$	DIS3L2	intron	-12530
1	rs12030998	57957630	A	3.911	0.9588	$4.76x10^{-5}$	DAB1	intron	-69705
2	rs3817368	68951366	C	-2.573	0.632	$4.90x10^{-5}$	BMP10	intron	-295
11	rs12363667	43657386	A	5.015	1.234	$5.10x10^{-5}$	HSD17B12	flanking_5UTR	-1495
21	rs2835144	36121367	G	-2.458	0.6052	$5.13x10^{-5}$	SETD4	flanking_3UTR	-207342
11	rs487518	131283728	A	-2.396	0.5926	$5.54x10^{-5}$	HNT	flanking_5UTR	-2194
18	rs275875	20105278	G	-2.435	0.603	$5.64x10^{-5}$	OSBPL1A	intron	-306
2	rs1370650	140113214	A	2.697	0.6679	$5.66x10^{-5}$	LRP1B	flanking_3UTR	-592252
16	rs7197521	47031763	G	3.408	0.8453	$5.79x10^{-5}$	LOC643114	flanking_5UTR	-15193
2	rs2312078	68948656	G	-2.537	0.631	$6.07x10^{-5}$	BMP10	intron	-1449

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
11	rs11037553	43653467	A	4.948	1.231	$6.08x10^{-5}$	HSD17B12	flanking_5UTR	-5414
9	rs10738194	10921959	A	-2.385	0.5933	$6.10x10^{-5}$	LOC646087	flanking_5UTR	-101651
6	rs1075379	33845778	G	2.916	0.7254	$6.10x10^{-5}$	LEMD2	flanking_3UTR	-1190
11	rs522856	131283569	C	-2.387	0.5939	$6.12x10^{-5}$	HNT	flanking_5UTR	-2353
15	rs4551980	47910160	A	-2.749	0.684	$6.13x10^{-5}$	ATP8B4	flanking_3UTR	-27567
19	rs2276470	50666508	A	2.458	0.6129	$6.37x10^{-5}$	FOSB	intron	-109
15	rs7182178	47911691	A	-2.743	0.684	$6.37x10^{-5}$	ATP8B4	flanking_3UTR	-26036
9	rs10815718	7925389	G	-3.2	0.7982	$6.37x10^{-5}$	C9orf123	flanking_5UTR	-135590
2	rs1438026	140057490	G	2.641	0.66	$6.56x10^{-5}$	LRP1B	flanking_3UTR	-647976
10	rs2758965	77749539	A	6.183	1.545	$6.57x10^{-5}$	C10orf11	intron	-4626
3	rs6798664	6184095	G	2.637	0.6596	$6.67x10^{-5}$	GRM7	flanking_5UTR	-693832
19	rs2276469	50666433	G	2.449	0.6133	$6.83x10^{-5}$	FOSB	intron	-34
3	rs9811274	144830174	A	-2.469	0.6185	$6.86x10^{-5}$	SLC9A9	intron	-23613
10	rs967948	77748165	A	6.135	1.54	$7.06x10^{-5}$	C10orf11	intron	-6000
18	rs1893468	53769609	A	3.451	0.8667	$7.16x10^{-5}$	NEDD4L	flanking_5UTR	-93169
2	rs3762613	68952476	G	-2.515	0.6328	$7.35x10^{-5}$	BMP10	flanking_5UTR	-323
2	rs12994269	115566962	A	4.208	1.059	$7.44x10^{-5}$	DPP10	flanking_5UTR	-69240
18	rs4245295	53741452	A	2.847	0.7168	$7.44x10^{-5}$	LOC284288	flanking_3UTR	-84183
3	rs7617700	36392255	G	-2.66	0.6708	$7.67x10^{-5}$	STAC	flanking_5UTR	-4846
4	rs1946860	178404692	C	-6.16	1.556	$7.83x10^{-5}$	NEIL3	flanking_5UTR	-63330
6	rs7755323	33869137	G	2.892	0.7322	$8.15x10^{-5}$	MLN	flanking_3UTR	-1290
5	rs7702151	141096026	G	-2.415	0.6115	$8.21x10^{-5}$	CENTD3	flanking_5UTR	-54042

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
5	rs2359425	76193792	A	5.172	1.311	8.29×10^{-5}	S100Z	intron	-7955
9	rs4421446	11083422	A	-2.526	0.6406	8.42×10^{-5}	LOC646087	flanking_5UTR	-263114
6	rs9346649	168233081	A	2.418	0.6152	8.86×10^{-5}	FRMD1	flanking_5UTR	-10393
1	rs10797912	182051797	G	-5.454	1.388	8.92×10^{-5}	RGL1	intron	-9555
5	rs4380707	41441161	A	3.763	0.9579	8.92×10^{-5}	PLCXD3	intron	-22768
13	rs7997540	20548501	C	3.11	0.7921	8.98×10^{-5}	LATS2	flanking_5UTR	-14847
5	rs11956840	76191426	A	5.251	1.338	9.03×10^{-5}	S100Z	intron	-9690
13	rs9571607	66102132	A	2.521	0.6424	9.09×10^{-5}	PCDH9	intron	-1211
10	rs1408405	9398374	A	-5.603	1.429	9.17×10^{-5}	LOC389936	flanking_5UTR	-90656
2	rs4954786	140077226	G	2.611	0.6667	9.38×10^{-5}	LRP1B	flanking_3UTR	-628240
16	rs6498964	63502473	G	7.564	1.933	9.48×10^{-5}	CDH11	flanking_3UTR	-35711
18	rs1539932	53761811	G	2.799	0.7153	9.52×10^{-5}	NEDD4L	flanking_5UTR	-100967
7	rs454066	51998407	A	-2.548	0.6514	9.54×10^{-5}	LOC642663	flanking_3UTR	-580450
4	rs13137544	190229767	A	-2.453	0.6271	9.57×10^{-5}	LOC728856	flanking_3UTR	-130677
3	rs4678874	36432376	G	-2.688	0.6881	9.74×10^{-5}	STAC	intron	-27484
1	rs10493043	30418573	G	-3.117	0.7984	9.85×10^{-5}	MATN1	flanking_3UTR	-540010
4	rs10520439	181002632	A	-2.621	0.6717	9.93×10^{-5}	LOC643179	flanking_3UTR	-729074
2	rs1347729	140039735	A	2.601	0.6669	9.98×10^{-5}	LOC129560	flanking_5UTR	-662948

Table 12: Best Discovery Analysis Results (p-value $\leq 10^{-5}$). The results are ranked in a descending order of p-value. For each SNP, chromosome, position, effect allele, beta and related standard error, p-value, gene, position relative to gene and distance to gene are reported.

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
18	rs1905641	61714830	A	6.569	1.716	$1.36x10^{-4}$	CDH7	flanking_3UTR	-15676
4	rs7666226	507376	G	4.039	1.084	$2.04x10^{-4}$	PIGG	coding	[128/326]
4	rs7666425	507461	G	4.039	1.084	$2.04x10^{-4}$	PIGG	coding	[213/241]
6	rs9462639	40727471	A	11.24	3.053	$2.42x10^{-4}$	LRFN2	flanking_5UTR	-64367
4	rs13150531	507622	A	3.905	1.086	$3.36x10^{-4}$	PIGG	coding	[374/80]
4	rs13114026	510853	A	3.883	1.088	$3.75x10^{-4}$	PIGG	coding	[25/166]
7	rs802374	119960866	A	-9.812	2.787	$4.47x10^{-4}$	KCND2	intron	-199327
7	rs17142666	119777393	G	-9.695	2.869	$7.51x10^{-4}$	KCND2	intron	-74356
9	rs7034422	118667250	A	-4.22	1.251	$7.69x10^{-4}$	ASTN2	intron	-1417
5	rs891921	168145529	A	4.424	1.316	$8.00x10^{-4}$	SLIT3	coding	[39/32]
19	rs8104211	11591545	A	-2.75	0.8242	$8.74x10^{-4}$	ZNF627	flanking_3UTR	-571
4	rs10014602	524965	G	3.666	1.102	$9.11x10^{-4}$	PIGG	flanking_3UTR	-1653
7	rs17142724	119853752	A	-8.951	2.704	$9.60x10^{-4}$	KCND2	intron	-150715
7	rs17142737	119861886	A	-8.951	2.704	$9.60x10^{-4}$	KCND2	intron	-158849
6	rs9475417	55796320	G	3.254	0.9851	$9.83x10^{-4}$	BMP5	intron	-3716

Table 13: Best Validation Analysis Results (p-value ≤ 10^{-4}). For each SNP chromosome, position, effect allele, beta and related standard error, p-value, gene, position relative to gene and distance to gene are reported.

4.2.2. Population-based study Analysis

Best hits from the Cross sectional and Longitudinal analyses are reported in table 14 and 15 respectively. In these analyses no SNPs reach the Bonferroni significance threshold of 10^{-6} (based on the number of SNPs present in the custom chip). In the Cross sectional study, rs5767743 and rs5766743 map in PPARA gene, that encodes for a peroxisome proliferator-activated receptor alpha; this protein has two functions: 1) it is a transcription factor activated by oleoylethanolamide, a naturally occurring lipid that regulates satiety, 2) it is a receptor for peroxisome proliferators such as hypolipidemic drugs and fatty acids involved in cardiovascular and metabolic pathways. The 3 best findings of the Longitudinal study (rs7730041, rs10067052, rs4867490), having p-value 10^{-5}, are located in LOC340113 which function is unknown. Rs3890733, p-value = 5.94×10^{-4} maps in VDR gene, the vitamin D receptor. Vitamin D deficiency has been reported as an underestimated risk factor for cardiovascular disease in patients with chronic kidney disease [57].

4.2.3. Comparative Results

Table 16 shows the cross-check analysis between the different studies (Discovery, Validation, Cross-sectional and Longitudinal study). In particular, the table reports the results in the different studies for markers having a significance p-value less or equal to 10^{-4} in the Discovery sample. No SNP has been validated with significant p-values.

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
22	rs5767743	45000657	G	2.26312	0.573589	7.96×10^{-5}	PPARA	intron	-3634
16	rs4781967	17529520	A	-1.9564	0.548865	3.65×10^{-4}	XYLT1	flanking_5UTR	-57282
13	rs17056501	37192478	A	2.811752	0.790778	3.77×10^{-4}	TRPC4	intron	-25595
22	rs5766743	44986041	G	1.777124	0.533238	8.60×10^{-4}	PPARA	intron	-3692

Table 14: Best Cross sectional analysis results (p-value ≤ 10^{-4}). For each SNP, chromosome, position, effect allele, beta and related standard error, p-value, gene, position relative to gene and distance to gene are reported.

CHR	SNP	BP	EFFECT ALLELE	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
5	rs7730041	32986121	G	2.35109	0.535683	$1.25x10^{-5}$	LOC340113	coding	[143/24]
5	rs10067052	32976917	A	2.274925	0.537495	$2.52x10^{-5}$	LOC340113	flanking_5UTR	-6388
5	rs4867490	32955758	A	2.174796	0.537612	$5.66x10^{-5}$	LOC340113	flanking_5UTR	-27547
8	rs2059652	18901785	A	2.146122	0.552253	$1.10x10^{-4}$	PSD3	intron	-13567
18	rs16960057	32120092	A	-4.74426	1.222113	$1.12x10^{-4}$	FHOD3	flanking_5UTR	-11607
10	rs9732211	128666930	G	-1.88228	0.503303	$1.97x10^{-4}$	DOCK1	intron	-689
3	rs925819	2847153	G	3.527368	0.951663	$2.25x10^{-4}$	CNTN4	intron	-10889
16	rs2897205	57038642	A	-4.42194	1.196842	$2.35x10^{-4}$	NDRG4	flanking_5UTR	-16475
11	rs11020063	92182121	A	-1.94493	0.559552	$5.41x10^{-4}$	FAT3	intron	-535
12	rs3890733	46575639	A	-1.84173	0.533771	$5.94x10^{-4}$	VDR	intron	-4247
4	rs7661978	100503221	A	-2.05162	0.598138	$6.40x10^{-4}$	ADH1C	flanking_5UTR	-10282
3	rs2616570	2602816	A	1.698215	0.501791	$7.55x10^{-4}$	CNTN4	intron	-14575

Table 15: Best Longitudinal analysis results (p-value ≤ 10^{-4}). For each SNP, chromosome, position, effect allele, beta and related standard error, p-value, gene, position relative to gene and distance to gene are reported.

Chr	SNP	BP	Effect Allele	BETA Disc	SE Disc	P Disc	BETA Val	SE Val	P Val	BETA Cross	SE Cross	P Cross	BETA Long	SE Long	P Long
3	rs4678878	36445311	G	-3.421	0.619	$3.83x10^{-8}$	0.259	0.755	$7.32x10^{-1}$	-0.319	0.466	$4.94x10^{-1}$	-0.333	0.519	$5.24x10^{-1}$
22	rs16989636	30638680	G	-12.570	2.846	$1.08x10^{-5}$	6.111	3.552	$8.56x10^{-2}$						
15	rs2387110	89672208	G	2.624	0.637	$3.99x10^{-5}$	0.387	0.809	$6.32x10^{-1}$	0.258	0.485	$5.95x10^{-1}$	-0.200	0.539	$7.12x10^{-1}$
2	rs334162	127980901	A	-4.943	1.292	$1.36x10^{-4}$	-0.320	1.718	$8.52x10^{-1}$	-0.449	1.094	$6.81x10^{-1}$	-1.255	1.164	$2.83x10^{-1}$
2	rs334142	127953892	G	-4.956	1.298	$1.39x10^{-4}$	-1.331	1.753	$4.48x10^{-1}$	-0.305	1.103	$7.82x10^{-1}$	-1.108	1.177	$3.49x10^{-1}$
2	rs777559	127955490	A	-4.954	1.298	$1.41x10^{-4}$	-1.331	1.753	$4.48x10^{-1}$	-0.305	1.103	$7.82x10^{-1}$	-1.108	1.177	$3.49x10^{-1}$
2	rs2078028	127980716	G	-4.943	1.298	$1.45x10^{-4}$	-1.331	1.753	$4.48x10^{-1}$	-0.298	1.103	$7.87x10^{-1}$	-1.112	1.177	$3.47x10^{-1}$
1	rs3813649	176169348	A	-2.592	0.687	$1.67x10^{-4}$	-0.489	0.889	$5.82x10^{-1}$	-0.083	0.543	$8.79x10^{-1}$	0.186	0.602	$7.58x10^{-1}$
2	rs334149	127971354	G	-4.845	1.289	$1.78x10^{-4}$	-1.333	1.734	$4.42x10^{-1}$	-0.305	1.103	$7.82x10^{-1}$	-1.108	1.177	$3.49x10^{-1}$
2	rs6731270	127978345	G	-4.845	1.289	$1.78x10^{-4}$	-1.333	1.734	$4.42x10^{-1}$	-0.305	1.103	$7.82x10^{-1}$	-1.108	1.177	$3.49x10^{-1}$
15	rs17364665	46672121	C	-4.560	1.274	$3.54x10^{-4}$	-0.523	1.493	$7.26x10^{-1}$	0.779	0.864	$3.67x10^{-1}$	-0.448	0.978	$6.49x10^{-1}$
14	rs7153295	98836375	A	-2.346	0.663	$4.17x10^{-4}$	1.256	0.839	$1.34x10^{-1}$	0.415	0.509	$4.14x10^{-1}$	-0.104	0.563	$8.55x10^{-1}$
1	rs7522194	176172092	A	-2.415	0.697	$5.47x10^{-4}$	-0.404	0.894	$6.51x10^{-1}$	0.026	0.549	$9.62x10^{-1}$	0.030	0.606	$9.61x10^{-1}$
1	rs11119014	206457721	G	-4.812	1.393	$5.66x10^{-4}$	-0.681	1.797	$7.05x10^{-1}$	0.281	0.867	$7.46x10^{-1}$	-0.983	0.978	$3.17x10^{-1}$
16	rs3803667	48227891	G	2.633	0.765	$5.91x10^{-4}$	-0.900	0.999	$3.68x10^{-1}$	0.145	0.558	$7.94x10^{-1}$	-0.892	0.629	$1.58x10^{-1}$
8	rs10105026	85313533	A	-2.111	0.614	$6.03x10^{-4}$	-0.858	0.759	$2.59x10^{-1}$	0.037	0.456	$9.35x10^{-1}$	-0.137	0.500	$7.85x10^{-1}$
6	rs1043784	7826930	G	3.061	0.901	$6.94x10^{-4}$	-0.053	1.151	$9.63x10^{-1}$	-0.470	0.753	$5.32x10^{-1}$	-0.421	0.828	$6.13x10^{-1}$
1	rs3738081	37097270	C	-2.264	0.679	$8.80x10^{-4}$	-0.196	0.835	$8.15x10^{-1}$	-0.028	0.535	$9.58x10^{-1}$			
1	rs12067006	37097522	G	-2.240	0.674	$9.15x10^{-4}$	-0.296	0.840	$7.25x10^{-1}$	-0.017	0.534	$9.74x10^{-1}$	-0.394	0.569	$4.90x10^{-1}$
1	rs2334915	115802203	A	-2.088	0.632	$9.76x10^{-4}$	-0.565	0.775	$4.66x10^{-1}$	0.026	0.464	$9.56x10^{-1}$	0.299	0.513	$5.62x10^{-1}$

Chr	SNP	BP	Effect Allele	BETA Disc	SE Disc	P Disc	BETA Val	SE Val	P Val	BETA Cross	SE Cross	P Cross	BETA Long	SE Long	P Long
8	rs13263453	18773776	A	3.113	0.943	9.85×10^{-4}	1.517	1.147	1.86×10^{-1}	1.231	0.733	9.33×10^{-2}	-1.344	0.813	1.00×10^{-1}
19	rs8105305	43094939	A	-1.979	0.600	9.93×10^{-4}	-0.535	0.775	4.90×10^{-1}	0.743	0.483	1.24×10^{-1}	-0.434	0.533	4.18×10^{-1}

Table 16: Discovery SNPs with a p-value $\leq 10^{-4}$, that were on the 15K chip.
Association results for the Discovery, Validation, Cross-sectional and Longitudinal studies are reported.

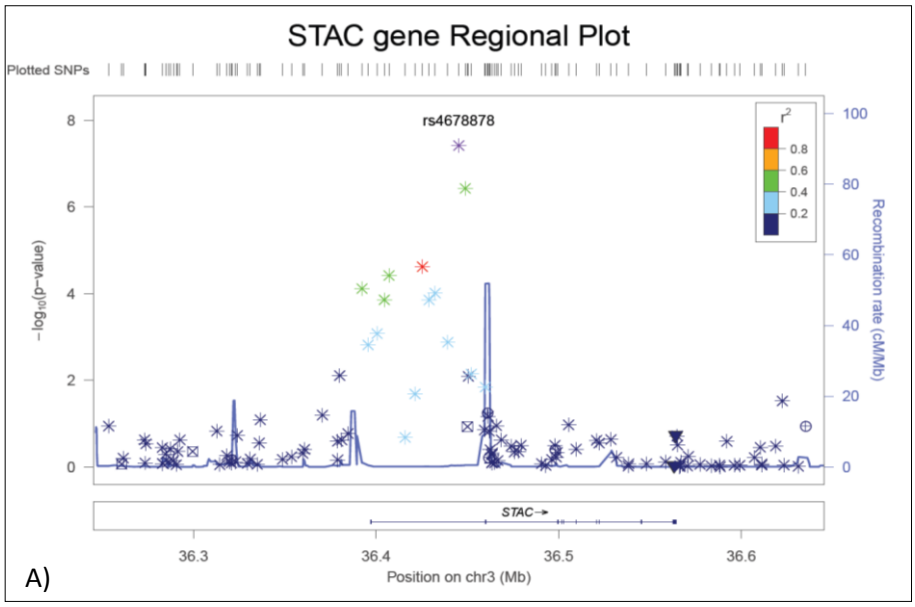
4.2.4. Principal Findings: STAC Gene

The STAC gene was covered by 159 SNPs in the 1M chip used for the Discovery phase and just one SNP, rs4678878 (P_{disc}=3.83x10^{-8}), was present on the custom chip used for the Validation and Population based study. In the Discovery phase 15 SNPs in this gene reach a significance ranging from 3.83x10^{-8} (rs4678878) to 8.14x10^{-3} (rs11921700). They are presented in table 17. These markers are all located just before a hot-spot of recombination and map in the same linkage disequilibrium block; in fact the D' between rs4678878, our best SNP, and each of the other 14 SNPs ranged from 1 to 0.638. Nevertheless, r^2 ranged from 0.783 to 0.041 (figure 10 and table 18).

In Validation and Population based study we do not replicate our best Discovery's finding rs4678878: $beta_{val}$=-3.421 and p_{val}=7.32x10^{-1} in validation, $beta_{Cross}$=0.259 and p_{Cross}=4.94x10^{-1} in Cross-sectional and a $beta_{long}$=-0.319 and p_{long}=5.24x10^{-1} in Longitudinal study, respectively (Table 16).

SNP	BP	Minor Allele	BETA	SE	P
rs4678878	36445311	G	-3.421	0.6191	3.83x10^{-8}
rs11129714	36448844	G	-3.257	0.6381	3.74x10^{-7}
rs728082	36425383	A	-2.741	0.6469	2.40x10^{-5}
rs4505641	36407085	G	2.531	0.6129	3.82x10^{-5}
rs7617700	36392255	G	-2.66	0.6708	7.67x10^{-5}
rs4678874	36432376	G	-2.688	0.6881	9.74x10^{-5}
rs2361037	36404488	G	-2.684	0.7028	1.39x10^{-4}
rs2885140	36429011	A	-2.631	0.6893	1.40x10^{-4}
rs1531134	36400540	G	-2.428	0.7257	8.39x10^{-4}
rs10514698	36439279	G	-2.286	0.7102	1.32x10^{-3}
rs6550422	36395632	G	-2.814	0.8865	1.53x10^{-3}
rs1598529	36452120	G	-1.641	0.6093	7.17x10^{-3}
rs2127232	36380078	C	-3.42	1.284	7.80x10^{-3}
rs11921700	36450546	A	-3.277	1.237	8.14x10^{-3}

Table 17: Best SNPs localized in gene STAC analyzed in the Discovery sample.



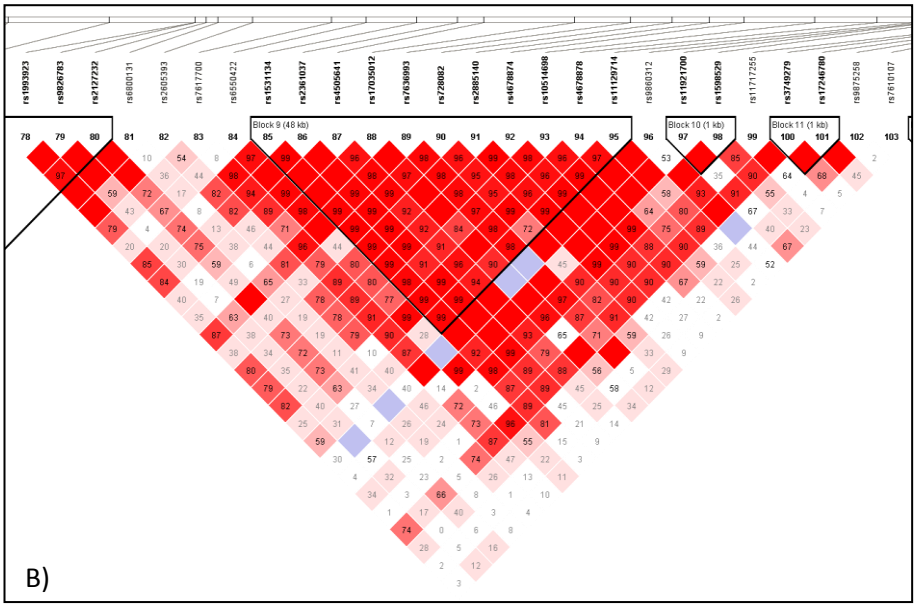


Figure 10: Genetic association and LD distribution of the STAC gene:
A) Regional association plot in the Hypergenes Discovery sample. Degree of association is expressed on the left. Recombination rate is expressed on the right.
B) LD structure. The figure was made with Haploview software.

Best SNP	other SNPs	D'	r^2
rs4678878	rs11129714	0.975	0.733
rs4678878	rs11921700	1	0.107
rs4678878	rs1598529	0.646	0.276
rs4678878	rs10514698	0.965	0.475
rs4678878	rs1531134	0.995	0.469
rs4678878	rs2127232	0.638	0.041
rs4678878	rs2361037	0.995	0.521
rs4678878	rs2885140	0.968	0.525
rs4678878	rs4505641	0.966	0.492
rs4678878	rs4678874	0.968	0.525
rs4678878	rs6550422	0.99	0.245
rs4678878	rs728082	0.997	0.783
rs4678878	rs7617700	0.908	0.531

Table 18: D' and r^2 between our best SNP and the other best findings in STAC gene.

4.2.5. Meta-analysis with CKDGen consortium data

In order to validate our Discovery results, we performed a meta-analysis between these findings and the CKDGen Consortium meta-analysis results [13, 32]. The meta-analysis was performed only for the SNPs that in both dataset presented a p-value lower than 10^{-3}. The CKDGen Consortium analyzes renal function using three different phenotypes: 1) eGFR evaluated from serum creatinine; 2) eGFR evaluated from serum cystatin c and 3) CKD, defined as eGFRcrea < 60 ml/min/1.73m^2. We compared our best results with each of the three different phenotypes analyzed by the CKDGen consortium. Results are reported in table 19. Due to the different sample sizes, the meta-analysis results are strongly steered by CKDGen meta-analysis.

Our best meta-analysis results map in the CST3_CST9 region (rs17751897, rs12625716, rs1158167), and in gene PRKAG2 (rs2536067). CST3_CST9 is a region that has been repeatedly associated with eGFR estimated from cystatin c [31, 32, 34] and PRKGA2 encodes for the enzyme 5'-AMP-activated protein kinase subunit gamma-2. Mutations in PRKAG2 gene have been associated with ventricular pre-excitation (Wolff-Parkinson-White syndrome), progressive conduction system disease and cardiac hypertrophy [13, 32, 58].

Marker Name	Chr	Position	Gene	BETA_disc (SE)	P_disc	BETA_CKDGen (SE)	P_CKDGen	BETA combined	SE combined	Inverse variance weighted P combined	Z-score P combined
eGFR Creatinine											
rs2536067	7	151056382	PRKAG2	2.121 (0.595)	$3.75x10^{-4}$	-0.0063 (0.0013)	$5.90x10^{-7}$	-0.0063	0.0013	$1.21x10^{-6}$	$4.81x10^{-8}$
rs1444039	1	107750779	NTNG1	-2.576 (0.725)	$3.94x10^{-4}$	-0.0039 (0.0013)	$2.98x10^{-3}$	-0.0039	0.0013	$2.64x10^{-3}$	$5.56x10^{-4}$
rs17115661	12	74454267	LOC390345	-2.218 (0.627)	$4.18x10^{-4}$	-0.0038 (0.0013)	$3.68x10^{-3}$	-0.0038	0.0013	$3.55x10^{-3}$	$1.82x10^{-2}$
rs7815850	8	18782841	PSD3	2.922 (0.827)	$4.22x10^{-4}$	-0.0044 (0.0016)	$6.82x10^{-3}$	-0.0044	0.0016	$6.09x10^{-3}$	$3.04x10^{-2}$
rs1041434	21	36561559	DOPEY2	-2.852 (0.809)	$4.34x10^{-4}$	-0.0046 (0.0016)	$3.84x10^{-3}$	-0.0046	0.0016	$4.13x10^{-3}$	$1.88x10^{-2}$
rs11614414	12	74452811	LOC390345	-2.205 (0.628)	$4.53x10^{-4}$	0.0038 (0.0013)	$3.69x10^{-3}$	0.0038	0.0013	$3.55x10^{-3}$	$1.81x10^{-2}$
rs17809619	3	20715647	LOC728976	-3.449 (0.989)	$4.99x10^{-4}$	0.0079 (0.0021)	$1.92x10^{-4}$	0.0079	0.0021	$1.74x10^{-4}$	$1.45x10^{-3}$
rs4529500	8	18784118	PSD3	2.821 (0.812)	$5.22x10^{-4}$	-0.0043 (0.0016)	$9.08x10^{-3}$	-0.0043	0.0016	$7.35x10^{-3}$	$3.77x10^{-2}$
rs7312284	12	74403903	LOC390345	-2.274 (0.655)	$5.33x10^{-4}$	-0.0054 (0.0014)	$1.07x10^{-4}$	-0.0054	0.0014	$1.18x10^{-4}$	$8.68x10^{-4}$
rs920835	1	107748325	NTNG1	-2.498 (0.726)	$5.98x10^{-4}$	-0.0039 (0.0013)	$3.10x10^{-3}$	-0.0039	0.0013	$2.65x10^{-3}$	$6.17x10^{-4}$
rs4256249	4	77677603	SHROOM3	-1.935 (0.617)	$1.74x10^{-3}$	0.0047 (0.0013)	$1.60x10^{-4}$	0.0047	0.0013	$3.08x10^{-4}$	$1.04x10^{-3}$
eGFR Cystatin											
rs686053	1	46987372	KIAA0494	3.342 (0.9337)	$3.55x10^{-4}$	-0.0094 (0.0032)	$3.36x10^{-3}$	-0.0094	0.0032	$3.18x10^{-3}$	$6.26x10^{-4}$
rs1566538	15	68142926	TLE3	2.617 (0.7576)	$5.66x10^{-4}$	-0.0084 (0.0027)	$1.88x10^{-3}$	-0.0084	0.0027	$1.79x10^{-3}$	$3.49x10^{-4}$
rs6475827	9	2489345	VLDLR	-2.261 (0.6544)	$5.66x10^{-4}$	0.0088 (0.0025)	$3.58x10^{-4}$	0.0088	0.0025	$4.11x10^{-4}$	$5.53x10^{-5}$
rs17068808	3	63450334	SYNPR	-2.115 (0.616)	$6.10x10^{-4}$	0.0063 (0.0022)	$4.32x10^{-3}$	0.0063	0.0022	$4.03x10^{-3}$	$8.96x10^{-4}$
rs17751897	20	23540714	CST9	-2.455 (0.7162)	$6.24x10^{-4}$	-0.0701 (0.0027)	$1.69x10^{-152}$	-0.0701	0.0027	$1.83x10^{-148}$	$9.37x10^{-144}$
rs10510168	3	1290929	CNTN6	3.481 (1.033)	$7.73x10^{-4}$	-0.0097 (0.0036)	$6.34x10^{-3}$	-0.0097	0.0036	$7.30x10^{-3}$	$2.69x10^{-2}$

Marker Name	Chr	Position	Gene	BETA_disc (SE)	P_disc	BETA_CKDGen (SE)	P_CKDGen	BETA combined	SE combined	Inverse variance weighted P combined	Z-score P combined
rs12625716	20	23554845	CST3	-2.399 (0.7168)	$8.36x10^{-4}$	0.0702 (0.0027)	$1.88x10^{-152}$	0.0702	0.0027	$6.91x10^{-149}$	$7.70x10^{-144}$
rs10966539	9	2488434	VLDLR	-2.31 (0.6925)	$8.71x10^{-4}$	0.0082 (0.0026)	$1.98x10^{-3}$	0.0082	0.0026	$1.54x10^{-3}$	$3.95x10^{-4}$
rs7692604	4	163462185	FSTL5	-2.07 (0.6257)	$9.59x10^{-4}$	-0.0071 (0.0024)	$2.69x10^{-3}$	-0.0071	0.0024	$2.97x10^{-3}$	$5.64x10^{-4}$
rs6986444	8	121695275	SNTB1	-3.321 (1.005)	$9.77x10^{-4}$	0.0097 (0.0037)	$8.27x10^{-3}$	0.0097	0.0037	$9.07x10^{-3}$	$3.28x10^{-2}$
rs1158167	20	23526189	CST3_CST9 regions	-2.059 (0.709)	$3.74x10^{-3}$	-0.0687 (0.0026)	$3.43x10^{-150}$	-0.0687	0.0026	$9.91x10^{-154}$	$2.23x10^{-142}$
CKD											
rs775291	18	8775897	KIAA0802	-3.16 (0.8682)	$2.81x10^{-4}$	0.1283 (0.0489)	$8.65x10^{-3}$	0.1179	0.0488	$1.57x10^{-2}$	$3.84x10^{-2}$
rs2536067	7	151056382	PRKAG2	2.121 (0.5949)	$3.75x10^{-4}$	0.0684 (0.0206)	$9.20x10^{-4}$	0.0658	0.0206	$1.40x10^{-3}$	$5.74x10^{-3}$

Table 19: Meta-analysis between SNPs with a p-value ≤ 10^{-3} both in the Hypergenes Discovery Analysis and in CKDGen Consortium analysis (Pattaro C. et al. Genome-wide association and functional follow-up reveals new loci for kidney function. PLoS Genet. 2012; 8(3):e1002584).

4.3. Analysis Results: Candidate Gene approach

We also performed a candidate gene approach looking for genes traditionally studied in cardiovascular diseases and renal function studies. To follow this approach we set a cut off of 10^{-5} in the Discovery analysis and 10^{-3} in the 15K analysis and identified several SNPs that mapped in genes already reported as associated to cardiovascular and renal pathways (see table 20). Two of them, NEDD4L and ABCA1, are suggestive of significance in all four studies even if the SNPs are different among the studies. The associated markers in NEDD4L and ABCA1 had a p-value < 10^{-4} in the Discovery dataset and p-value < 10^{-2} in the Validation and Population based study (table 21). According to the number of SNPs analyzed in NEDD4L gene, Bonferroni's threshold was $2x10^{-4}$ in the Discovery and $1x10^{-3}$ in the Validation, Cross sectional and Longitudinal analyses. On the contrary, for ABCA1 gene, the significance threshold was $1.93x10^{-4}$ in the Discovery and $2.17x10^{-3}$ in the Validation, Cross sectional and Longitudinal analyses.

NEDD4L is a regulator of the amiloride-sensitive epithelial sodium channel (ENaC) and is a candidate gene for salt sensitivity. ABCA1 is involved in lipid secretory pathways and especially in the co-transport of cholesterol and phospholipids and has been associated to cardiac diseases, hypertension and renal failure. Figures 11 and 12 show the regional plot for the two genes, considering all the SNPS analyzed in the Discovery, Validation, Cross sectional and Longitudinal studies, highlighted with different forms. For markers identified in more than one study, best p-value is reported.

SNP	GENE	EFFECT ALLELE	BETA	SE	P	CHR	BP	POSITION	DISTANCE TO GENE	RELATED PATHWAY
Discovery Hypergenes										
rs6973843	PBEF1	A	-4.213	1.003	$2.82X10^{-5}$	7	105773180	flanking_5UTR	-60587	Atherosclerosis, Renal, Cardiovascular and Coronary artery diseases, Metabolic disorder
rs1893468	NEDD4L	A	3.451	0.8667	$7.16X10^{-5}$	18	53769609	flanking_5UTR	-93169	Essential Hypertension, Salt sensitivity, Renin activity
rs1539932	NEDD4L	G	2.799	0.7153	$9.52X10^{-5}$	18	53761811	flanking_5UTR	-100967	
Validation Hypergenes										
rs10991411	ABCA1	G	4.749	1.619	$3.43X10^{-3}$	9	106724097	intron	-5940	Atherosclerosis, Ischemia, Heart and Coronary artery disease, Hyperlipidemia
rs6136	SELP	C	-3.666	1.397	$8.80X10^{-3}$	1	167830575	coding	[21/164]	Nephropathy, Heart disease, Ischemic stroke, Essential hypertension
rs9649847	EGFR	A	-3.011	1.158	$9.42X10^{-3}$	7	55154632	intron	-22841	Polycystic kidney disease
rs10439056	NEDD4L	A	4.742	1.835	$9.88X10^{-3}$	18	53875962	intron	-13024	Essential Hypertension, Salt sensitivity, Renin activity
Cross-sectional										
rs5767743	PPARA	G	2.263	0.574	$7.96X10^{-5}$	22	45000658	intron	-3634	Coronary Heart Disease, Type II diabetes, Plasma Lipid Levels, Hypertension, Atherosclerosis, Metabolic syndrome, Cardiac growth
rs5766743	PPARA	G	1.777	0.533	$8.60X10^{-4}$	22	44986042	intron	-3692	

SNP	GENE	EFFECT ALLELE	BETA	SE	P	CHR	BP	POSITION	DISTANCE TO GENE	RELATED PATHWAY
rs5742912	SCNN1A	G	5.247	1.666	$1.63X10^{-3}$	12	6328611	coding	[20/37]	Blood pressure
rs11888123	SLC8A1	A	1.547	0.505	$2.17X10^{-3}$	2	40408981	intron	-100136	Hypertension
rs3794764	NOS2A	A	-1.634	0.551	$2.98X10^{-3}$	17	23135555	intron	-1296	Atherosclerosis, Nephropathy, Microalbuminuria
rs11854914	FBN1	A	-2.434	0.849	$4.16X10^{-3}$	15	46590506	intron	-848	Coronary artery disease, Aortic stiffness
rs4253711	PPARA	A	1.579	0.571	$5.69x10^{-3}$	22	44973697	intron	-545	Coronary Heart Disease, Type II diabetes, Plasma Lipid Levels, Hypertension, Atherosclerosis, Metabolic syndrome, Cardiac growth
rs10478694	EDN1	I	1.378	0.516	$7.58x10^{-3}$	6	12398718	5UTR	[73/130]	Kidney dysfunction, Left ventricular hypertrophy, Blood pressure, Cardiac plasma endothelin-1 levels, Cholesterol plasma levels
rs3771426	ADD2	G	1.751	0.660	$7.99x10^{-3}$	2	70811011	intron	-17187	Hypertension
rs708498	PTGER2	A	-1.751	0.671	$9.07x10^{-3}$	14	51862250	intron	-1439	Myocardial infarct, Stroke ischemic
rs6566955	NEDD4L	G	-1.338	0.515	$9.35x10^{-3}$	18	54073769	intron	-3503	Essential Hypertension, Salt sensitivity, Renin activity
Longitudinal										
rs3890733	VDR	A	-1.842	0.534	$5.94x10^{-4}$	12	46575640	intron	-4247	Chronic renal failure related to vitamine D deficiency

SNP	GENE	EFFECT ALLELE	BETA	SE	P	CHR	BP	POSITION	DISTANCE TO GENE	RELATED PATHWAY
rs12309274	WNK1	C	2.478	0.752	$1.04x10^{-3}$	12	846209	intron	-4483	Essential Hypertension
rs6028	F5	G	-1.816	0.594	$2.33x10^{-3}$	1	167818306	coding	[13/78]	Cerebrovascular disease
rs3917750	SELP	A	-1.740	0.569	$2.34x10^{-3}$	1	167843377	intron	-381	Nephropathy, Heart disease, Ischemic stroke, Essential hypertension
rs1800779	NOS3	G	-1.721	0.563	$2.34x10^{-3}$	7	150320876	intron	-898	Coronary Heart Disease, Chronic renal failure, Hypertension
rs5517	KLK1	G	-1.736	0.578	$2.80x10^{-3}$	19	56015044	coding	[77/59]	Hypertension, Chronic renal failure
rs1800783	NOS3	T	-1.684	0.561	$2.81x10^{-3}$	7	150320330	intron	-1009	Coronary Heart Disease, Chronic renal failure, Hypertension
rs6951150	NOS3	A	-1.648	0.563	$3.60x10^{-3}$	7	150312847	flanking_5UTR	-6233	
rs9804992	WNK1	A	2.160	0.754	$4.34x10^{-3}$	12	859155	coding	[155/303]	Essential Hypertension
rs760694	SELP	C	-1.576	0.555	$4.70x10^{-3}$	1	167835322	intron	-2284	Nephropathy, Heart disease, Ischemic stroke, Essential hypertension
rs2066715	ABCA1	A	-2.985	1.092	$6.52x10^{-3}$	9	106627854	coding	[69/135]	Atherosclerosis, Ischemia, Heart and Coronary artery disease, Hyperlipidemia

SNP	GENE	EFFECT ALLELE	BETA	SE	P	CHR	BP	POSITION	DISTANCE TO GENE	RELATED PATHWAY
rs12445698	SLC12A3	A	-1.794	0.670	7.69×10^{-3}	16	55485717	intron	-227	Diabetic nephropathies, Renal disease, Essential hypertension, Pseudohypoaldosteronism, Hyperaldosteronism
rs3212833	KLK1	A	-1.592	0.595	7.75×10^{-3}	19	56014691	intron	-274	Hypertension, Chronic renal failure
rs2241054	EGFR	A	-1.665	0.589	4.91×10^{-3}	7	55207473	intron	-697	Polycystic kidney disease
rs3818109	SYNE1	G	1.445	0.540	7.77×10^{-3}	6	152498168	intron	-108	Renal Failure

Table 20: SNPs with a p-value ≤ 10^{-5} in the Hypergenes Discovery and with a p-value ≤ 10^{-3} in the Validation, Cross Sectional and Longitudinal studies found related to Renal or Cardiovascular pathways.

CHR	SNP	BETA	SE	P	GENE	POSITION	DISTANCE TO GENE
Discovery							
18	rs1893468	3.451	0.867	$7.16x10^{-5}$	NEDD4L	flanking_5UTR	-93169
18	rs1539932	2.799	0.715	$9.52x10^{-5}$	NEDD4L	flanking_5UTR	-100967
9	rs4149286	4.650	1.404	$9.44x10^{-4}$	ABCA1	intron	-1722
Validation							
18	rs10439056	4.742	1.835	$9.879x10^{-3}$	NEDD4L	intron	-13024
9	rs10991411	4.749	1.619	$3.427x10^{-3}$	ABCA1	intron	-5940
Cross sectional							
18	rs6566955	-1.338	0.515	$9.350x10^{-3}$	NEDD4L	intron	-3503
9	rs2230808	-1.102	0.549	$4.468x10^{-2}$	ABCA1	coding	[13/61]
Longitudinal							
18	rs17064378	1.242	0.531	$1.987x10^{-2}$	NEDD4L	intron	-46661
9	rs2066715	-2.985	1.092	$6.518x10^{-3}$	ABCA1	coding	[69/135]

Table 21: Analysis results in the four studies for different SNPs localized in NEDD4L and ABCA1.

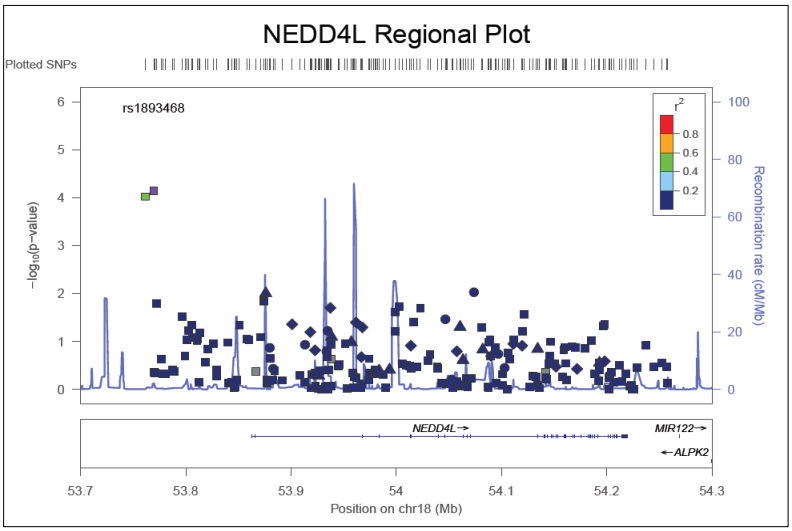


Figure 11: Regional association plot for NEDD4L gene in the four studies. Degree of association is expressed on the left. Recombination rate is expressed on the right.

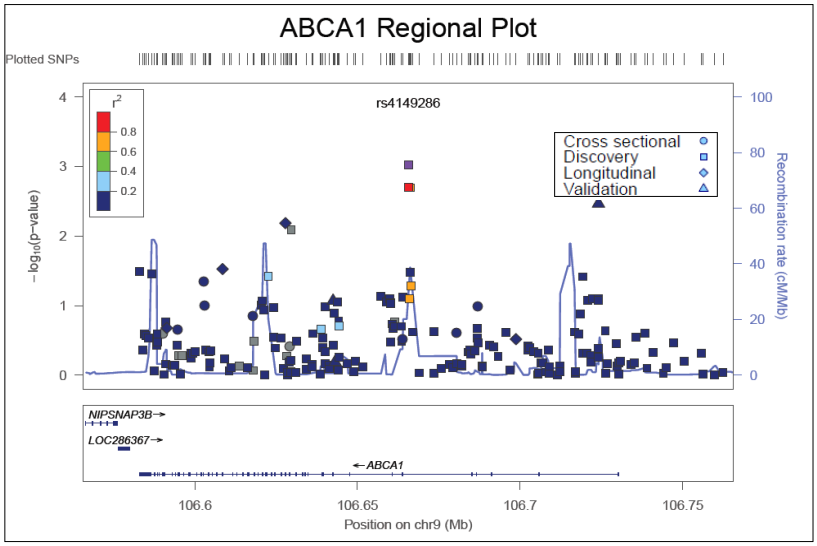


Figure 12: Regional association plot for ABCA1 gene in the four studies. Degree of association is expressed on the left. Recombination rate is expressed on the right.

5. DISCUSSION

The present study is part of HYPERGENES Project (European Network for Genetic-Epidemiological Studies: building a method to dissect complex genetic traits using essential hypertension as a disease model) focused on the definition of a comprehensive genetic-epidemiological model of complex traits like essential hypertension and of intermediate phenotypes related to hypertension. Among the intermediate phenotypes available, in the present study we focused on eGFR as index of renal function.

A strong relationship has been reported between renal dysfunction, hypertension and cardiovascular diseases; hypertension is present in more than 80% of patients with chronic kidney disease and contributes to progression of renal disease toward end stage as well as to cardiovascular events. Being renal function strongly related to blood pressure control and physiopathology of hypertension, the aim of this work was to identify genetic loci associated with renal function, in subjects affected by essential hypertension and in a population based study.

CKD affects about 10% of the general population and causes significant morbidity and mortality. Apart from the risk conferred by traditional cardiovascular risk factors, there is a strong genetic component. Moreover heritability of GFR has been estimated to range from 36 to 75% [27]. Being the genetic component of GFR quite consistent, we performed a GWA study. In GWAs a large number of SNPs distributed in the genome are characterized in large numbers of individuals, with the aim to identify variants with a low or moderate effect on the phenotype of interest. To address this issue the precise characterization of the phenotype is crucial and inaccurate results can be the consequence of a poorly defined phenotype.

The strength of our GWA study is based on sample selection: the hypertensive individuals have been very carefully selected for different phenotypes related to blood pressure control and renal function; the population based sample belongs to an extensively characterized and followed up cohort. This accurate selection lead the sample size to be small compared to other published studies, but decreases the bias of combining cohorts recruited with different criteria. On the other hand, the limited sample size can be the cause of the lack of validation of the findings in the different association studies performed.

In the Discovery hypertensive dataset, two new loci were identified, rs4678878 and rs11129714, reaching a p-value of 3.83×10^{-8}, above the genome-wide significance, and 3.74×10^{-7} respectively. In both SNPs, the homozygosity for the risk allele G causes a mean eGFR value decrease of 8 ml/min/1.73 m^2 (p-value < 10^{-7}) compared to the AA genotype.

The two SNPs mapped in STAC gene. These SNPs were not replicated in our further studies as associated to renal function or hypertension. This can be probably due to the small sample size and to the lack of markers mapping in STAC genes in the custom chip used to analyze the HYPERGENES Validation sample and the Population based study. The custom chip (including about 15,000 SNPs) has been created on the basis of the results obtained in the case-control Discovery of HYPERGENES. Moreover this chip has been built to interrogate candidate SNPs historically studied in hypertension field and SNPs involved in relevant pathways related to blood pressure control. Rs4678878 was the only SNP in STAC present on the chip and was not validated in any of the three studies performed. All best STAC SNPs identified in the discovery analysis were in linkage disequilibrium, with a D' ranging from 1 to 0.638. Nevertheless the associated p-values, ranging from 3.83×10^{-8}

(rs4678878) to $8.14x10^{-3}$ (rs11921700), were not consistent with this finding. This can be explained by the R^2 between rs4678878 and the other 15 best STAC SNPs, ranging from 0.783 to 0.041. These markers are all located just before a hot spot of recombination and they could tag a causal SNP towards the 5' of the gene.

STAC gene encodes for a protein that consists of two well characterized domains: a cysteine-rich domain (CRD) of the protein kinase C (PKC) family and a Src homology 3 (SH3) domain, both of them are frequently found in proteins involved in signal transduction. In tissue-expression patterns, the STAC mRNA was expressed predominantly in brain, although weak signals were detected in heart, liver, lung and kidney. The function of STAC is now under investigation. However, the primary structure consisting of domains important in signal transduction and its neuron specific expression pattern suggest that STAC plays a crucial role in neuron specific signal transduction pathways [59].

Legha and colleagues (2010) identify STAC1 and STAC2 gene to define discrete and distinct subsets of dorsal root ganglia (DRG) neurons. They found that STAC1 mainly marks peptidergic nociceptive neurons, while STAC2 is principally expressed in a subset of nonpeptidergic nociceptors, in all trkB+ neurons and in a subpopulation of proprioceptive neurons [60].

In the analysis performed using the 15K chip none of the SNPs analyzed reached the significance Bonferroni threshold of 10^{-6}. However, in the Population based Cross sectional study, two findings, rs5767743 and rs5766743, with p-values of $7.96x10^{-5}$ and $8.60x10^{-4}$ respectively, mapped in PPARA gene that has been described as involved in hypertension onset and progression. Peroxisome proliferator-activated receptors (PPARs) are members of a large family of ligand-inducible transcription factors that include receptors for retinoids and vitamin D, thyroid, and steroid hormones. PPARs regulate the expression of target genes by binding to DNA sequence elements, named PPAR response elements (PPRE). PPREs have been identified in the regulatory regions of a variety of genes involved in lipid and energy balance.

Three different PPAR genes have been identified (α, δ, and γ). PPARA, maps in chromosome 22, position 22q13.31. It is mainly expressed in tissues that have a high ratio of fatty acid (FA) oxidation such as liver, kidney, heart and muscle. Numerous studies have demonstrated that several genes encoding enzymes involved in FA oxidation are modulated by PPARA. In addition to its role in FA oxidation, PPARA mediates the action of hypolipidemic drugs of the fibrate class on lipid and lipoprotein metabolism, it alters the synthesis of lipoprotein lipase and regulates the transcription of apolipoprotein C-III and other lipid-related genes. Fatty acids are essential biological components, used as metabolic fuels, covalent regulators of signaling molecules and essential components of cellular membranes. Altered levels of FA are linked to a variety of metabolic diseases including atherosclerosis, hyperlipidemia, obesity, insulin resistance, and type 2 diabetes. The finding that PPARA modulates gene expression of enzymes controlling FA oxidation suggests that it may be a candidate gene for type 2 diabetes and dyslipidemia. Moreover, Vohl and colleagues (2000) demonstrated that the heterozygosity for the L162V mutation in PPARA is unlikely to be a major risk factor for the development of type 2 diabetes but is associated with increased plasma apolipoprotein B levels [61]. Jamshidi et al (2002) showed how PPARA regulates left ventricular growth in response to exercise and hypertension stimuli and illustrated the important role of cardiac fatty acid metabolism in cardiac growth using the previously described functional L162V variant and a novel G/C

polymorphism in intron 7 of the human PPARA gene. These data may have therapeutic implications for patients with left ventricular hypertrophy (LVH) [62].

We also performed a meta-analysis between our Discovery findings and the CKDGen Consortium meta-analysis, setting a p-value cut off of 10^{-3} in both dataset. We used the CKDGen results obtained from three different phenotypes of renal function: eGFR calculated from serum creatinine, eGFR calculated from cystatin C and incidence of chronic kidney disease. Since the three phenotypes describe renal function, a finding associated to more than one phenotype would be more relevant and informative from a physiological point of view. CKDGen meta-analysis provides association data from 67,093 individuals, with a further replication in an independent sample of 22,982, reaching very high p-values. We could therefore confirm some of our findings. However, due to the different sample sizes, the results are strongly steered by CKDGen meta-analysis results.

Rs17751897 (risk allele G), rs12625716 (risk allele A) and rs1158167 (risk allele G), having in our sample a p-value of $6.24x10^{-4}$, $8.36x10^{-4}$ and $3.74x10^{-3}$ respectively, reached an inverse variant weighted p-value of $1.83x10^{-148}$, $6.91x10^{-149}$ and $9.91x10^{-154}$ respectively in the meta-analysis with CKDGen data. The three markers map in the CST3-CST9 region, repeatedly associated with eGFR estimated from cystatin c; in all of them the homozygous for the risk allele cause a significant decrease in mean eGFR of around 4 ml/min/1.73 m^2 (pvalue < 0.05) compared to the no-risk genotype. The cystatin superfamily consists of a large group of cystatin domain-containing proteins, most of which are reversible and tight-binding inhibitors of cysteine proteases. On the basis of the sequence similarity, presence or lack of disulfide bonds and physiological localization, this superfamily has been divided into three families: family I or stefins, family II or cystatins and family III or kininogens [63]. All of the characterized cystatins exhibit sequence homologies [64]. The cystatin locus on chromosome 20 contains the majority of the type 2 cystatin genes and pseudogenes.

CST3 gene is located in chromosome 20, locus 20p11.21 and encodes for cystatin C protein, the most abundant extracellular inhibitor of cysteine proteases; the protein is found in high concentrations in all biological fluids, such as the cerebrospinal fluid and plasma and it is highest expressed in the epididymis, vas deferens, brain, thymus, and ovary. Cystatin C is mainly used as a biomarker of kidney function [31, 32, 36]. It is a cationic, low-molecular-weight protein (13 kD) freely filtered by the glomerulus and almost completely reabsorbed and catabolized by proximal tubular epithelial cells. It is not secreted by the tubule, it's completely absent in urine and it is therefore measured only in plasma. It is produced at a constant rate by all cells and is not related to weight, age, sex and race. Several studies have reported significantly higher performance of cystatin C in estimating glomerular filtration rate [65].

Defects in CST3 are the cause of amyloidosis type 6 (AMYL6), a hereditary generalized amyloidosis due to cystatin C amyloid deposition. Amyloid deposition in the cerebral vessels results in cerebral amyloid angiopathy, cerebral hemorrhage and premature stroke. Cystatin C levels in the cerebrospinal fluid are abnormally low. Genetic variations in CST3 are also associated with age-related macular degeneration type 11 (ARMD11), a multifactorial eye disease representing the most common cause of irreversible vision loss in industrialized countries. It also seems to play a role in Alzheimer's disease. Recently, CST3 gene has been studied for its role in predicting new-onset or deteriorating

cardiovascular disease, as its expression is severely reduced in both atherosclerotic and aneurysmal aortic lesions [66, 67, 63].

CST9 gene (testatin gene) is located in chromosome 20, locus 20p11.21. This gene belongs to a subgroup within the family 2 cystatins of cysteine protease inhibitors, together with cystatin T gene and cystatin-related epididymal spermatogenic (CRES) gene. Töhönen (1998) described testatin as expressed during testis cord formation in pre-Sertoli cells. Therefore, this gene may be involved in early testis development pathways [68]. Furthermore it may play a role in hematopoietic differentiation or inflammation [66].

In the meta-analysis Rs2536067, with a p-value of $3.75x10^{-4}$ in the discovery sample, reached an inverse variant weighted p-value of $1.21x10^{-6}$ and a z-score p-value of $4.81x10^{-8}$ in the meta-analysis on eGFR measured from serum creatinine. It maps to gene PRKAG2, encoding for the enzyme 5'-AMP-activated protein kinase (AMPK) subunit gamma-2. AMPKs are heterotrimeric proteins composed of a catalytic alpha subunit and two beta and gamma regulatory subunits. They are widely expressed in human tissues. They act as major regulators of cellular ATP levels and protect cells against stresses that cause ATP depletion. Although AMPKs are abundantly expressed in the kidney, their role in renal physiology is less clear than in other organs. In recent years various studies described roles for AMPKs in multiple aspects of renal physiology and disease including ion transport, podocyte function, renal hypertrophy, ischemia, inflammation, diabetes and polycystic kidney disease.

PRKAG2 human gene maps to chromosome 7q36.1 and is composed of 16 exons. It encodes for a protein with four cystathionine beta-synthase domains.

PRKAG2 protein plays a key role in regulating cellular energy metabolism in response to reduction of intracellular ATP levels. Mutations in this gene have been associated with ventricular pre-excitation (Wolff-Parkinson-White syndrome), a progressive conduction system disease, associated to cardiac hypertrophy. Defects in PRKAG2 are the cause of two congenital heart diseases: cardiomyopathy familial hypertrophic type 6 (CMH6), a hereditary heart disorder characterized by ventricular hypertrophy, which is usually asymmetric and often involves the interventricular septum, and glycogen storage disease of heart lethal congenital (GSDH), a glycogen storage diseases with clinically prominent cardiac involvement. Kottgen and colleagues (2010) identified rs7805747 in PRKAG2 as associated with renal function in CKDGen discovery analyses for both eGFRcrea and CKD phenotypes. The A allele of the SNP is associated with reduced glomerular filtration rate and thus a slightly increased risk for CKD [13, 32, 58].

Through a candidate gene approach we could identify two genes, NEDD4L and ABCA1, suggestive of significance in all four analyses performed in our datasets (HYPERGENES Discovery and Validation studies and the two population based studies).

NEDD4L reached the candidate gene's Bonferroni threshold of $2.53x10^{-4}$ in the Discovery association study, while none of the SNPs analyzed using the 15K chip could reach a p-value < $1.09x10^{-3}$. As showed in the regional plot (fig. 11), the best SNPs, rs1893468 and rs1539932, (p-value of $7.16x10^{-5}$ and $9.52x10^{-5}$ respectively) are located in the 5'-UTR flanking region of the gene. NEDD4L gene is located in chromosome 18, locus 18q21.31 and encodes for the NEDD4L enzyme (neural precursor cell expressed developmentally down-regulated 4-like). The majority of rare monogenic forms of human hypertension identified are caused by mutations that directly or indirectly increase activity or expression of the amiloride-sensitive epithelial Na^+ channel (ENaC) in the distal nephron.

This results in enhanced renal sodium reabsorption and salt-sensitive hypertension with compensatory low levels of plasma renin. NEDD4L protein acts as a channel remover of ENaC. Given the central role for ENaC in monogenic hypertension, the search for more common genetic variation in pathways controlling ENaC expression is an attractive approach for identification of essential hypertension susceptibility genes, as a complement to genome-wide association studies (GWAS) [69]. Manunta et al (2008) showed a consistent combined effect on renal sodium handling and blood pressure response of the NEDD4L (rs4149601 G/A), ADD1 (Gly460Trp) and WNK1 (rs880054 A/G) polymorphisms. NEDD4L rs4149601 A allele, at the first nucleotide of exon 1, leads to an alternative splice site which generates a isoform of NEDD4L downregulating ENaC more potently than the wild form of NEDD4L (rs4149601-G), suggesting that carriers of the G allele have higher ENaC expression and higher renal sodium reabsorption through ENaC than carriers of the A allele [70]. Jin and colleagues (2010) further identified several SNPs in NEDD4L gene that were linked to blood pressure regulation and hypertension. None of these was significantly validated in our analysis [71].

The other gene identified through a candidate gene approach, ABCA1, did not exceed the candidate gene's Bonferroni threshold of $1.93x10^{-4}$ in the Discovery phase and of $2.17x10^{-3}$ in the analysis performed using the 15K chip. The top SNPs identified in this gene show high linkage disequilibrium and are located in a high recombination region (fig. 12). ABCA1 map to chromosome 9q31.1 and it is composed of 50 exons. It encodes for ABCA1 membrane protein, belonging to the adenosine triphosphate-binding cassette family, a group of proteins located in different tissues where they translocate different substrates to various compartment. It is implicated in co-transporting of cholesterol and phospholipids across plasma membrane, in intracellular lipids transport and transfer to apoA-1 e apoE, in the process of HDL particles formation, in electroneutral anion transport, in secretion of peptides like IL-1ß, and in engulfment of apoptotic cells by macrophages. It is expressed on cell surface, Golgi membranes and endolysosomal compartments of various tissues involved in the turnover of lipids such as the liver, the small intestine and adipose tissue, and in other tissue such as macrophages, lung, testis, pregnant uterus and placenta [72]. Among members of ABC family, several have been linked to inherited diseases, such as severe familial HDL deficiency, atherosclerotic complications and increased risk of myocardial infarction independent of HDL cholesterol levels [73].

A correlation has been demonstrated between ABCA1 and proteins coded by genes SNTB1 and VLDLR. Beta 1-syntrophin (SNTB1) was identified in our study in the meta-analysis with CKDGen data (rs6986444, p-value $9.77x10^{-4}$); it regulates the cellular distribution and activity of ABCA1 [74]. Very low-density lipoprotein receptor, VLDLR, one of our meta-analysis findings (rs6475827, p-value $5.66x10^{-4}$), regulates expression of ABCA1 and cholesterol efflux [75].

In conclusion our study identified a potential novel variant associated with hypertension and renal function in STAC gene (rs4678878, p-value of $3.83x10^{-8}$); this gene is likely involved in a neuron-specific signal transduction, although its function is not yet completely clear. The finding we obtained in STAC gene needs further investigation for different reasons: 1. We could validate only one SNP with the custom chip. Therefore it is necessary to fine map the region with a denser number of markers. This step could be followed by a validation in a larger sample. 2. STAC gene is implicated in neuronal

transduction and from a functional point of view this finding is not easy to explain in relation to the renal phenotype we are studying.
Finally we could also identify some variants, mapping in loci already known as associated to renal function, in an accurately selected population (PPARA, CST3-CST9 region, PRKAG2, ABCA1, NEDD4L).
The identification of genetic variants that can affect renal function may lead to a better understanding not only of GFR variability in the general population but also of the pathophysiology of CKD and progressive kidney function decline. Ultimately, this could lead to novel tools for diagnosis, prevention and therapy of CKD.

6. REFERENCES

1. Burton D. Rose, Theodore W. Post. **Clinical Physiology of Acid-Base and Electrolyte Disorders**. 2001. Fifth Edition. Mc Graw Hill edition.
2. http://science.kennesaw.edu/~bodavis/
3. Selvin E, Manzi J, Stevens LA, Van Lente F, Lacher DA, Levey AS, Coresh J. **Calibration of serum creatinine in the National Health and Nutrition Examination Surveys (NHANES) 1988-1994, 1999-2004**. Am J Kidney Dis. 2007 Dec; 50(6):918-26. PubMed PMID: 18037092.
4. Pottel H, Hoste L, Delanaye P, Cavalier E, Martens F. **Demystifying ethnic/sex differences in kidney function: Is the difference in (estimating) glomerular filtration rate or in serum creatinine concentration?** Clin Chim Acta. 2012 Oct 9; 413(19-20):1612-7. Epub 2012 May 3. PubMed PMID: 22584028.
5. Coresh J, Astor BC, McQuillan G, Kusek J, Greene T, Van Lente F, Levey AS. **Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate.** Am J Kidney Dis. 2002 May; 39(5):920-9. PubMed PMID: 11979335.
6. Glassock RJ, Winearls C. **Ageing and the glomerular filtration rate: truths and consequences**. Trans Am Clin Climatol Assoc. 2009; 120:419-28. Review. PubMed PMID: 19768194; PubMed Central PMCID: PMC2744545.
7. Verhave JC, Balj´e-Volkers CP, Hillege HL et al. **The reliability of different formulae to predict creatinine clearance**. J Intern Med 2003; 253: 563–73.
8. Grewal GS, Blake GM. **Reference data for 51Cr-EDTA measurements of the glomerular filtration rate derived from live kidney donors**. NuclMed Commun 2005; 26: 61–65.
9. Douville P, Martel AR, Talbot J, Desmeules S, Langlois S, Agharazii M. **Impact of age on glomerular filtration estimates**. Nephrol Dial Transplant. 2009 Jan; 24(1):97-103. Epub 2008 Sep 4. PubMed PMID: 18772177.
10. Cirillo M. **Evaluation of glomerular filtration rate and of albuminuria/proteinuria.** J Nephrol. 2010 Mar-Apr; 23(2):125-32. Review. PubMed PMID: 20213606.
11. Ratto E, Leoncini G, Viazzi F, Pontremoli R. **Glomerular filtration rate and cardiovascular risk: prognostic and therapeutic implications**. G Ital Nefrol. 2008 Jan-Feb; 25(1):21-31. Review. Italian. PubMed PMID: 18264915.
12. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J; **CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate.** Ann Intern Med. 2009 May 5; 150(9):604-12. Erratum in: Ann Intern Med. 2011 Sep 20;155(6):408. PubMed PMID: 19414839; PubMed Central PMCID: PMC2763564.
13. Köttgen A, Pattaro C, Böger CA, Fuchsberger C, Olden M, Glazer NL, Parsa A, Gao X, Yang Q, Smith AV, O'Connell JR, Li M, Schmidt H, Tanaka T, Isaacs A, Ketkar S, Hwang SJ, Johnson AD, Dehghan A, Teumer A, Paré G, Atkinson EJ, Zeller T, Lohman K, Cornelis MC, Probst-Hensch NM, Kronenberg F, Tönjes A, Hayward C, Aspelund T, Eiriksdottir G, Launer LJ, Harris TB, Rampersaud E, Mitchell BD, Arking DE, Boerwinkle E, Struchalin M, Cavalieri M, Singleton A, Giallauria F,

Metter J, de Boer IH, Haritunians T, Lumley T, Siscovick D, Psaty BM, Zillikens MC, Oostra BA, Feitosa M, Province M, de Andrade M, Turner ST, Schillert A, Ziegler A, Wild PS, Schnabel RB, Wilde S, Munzel TF, Leak TS, Illig T, Klopp N, Meisinger C, Wichmann HE, Koenig W, Zgaga L, Zemunik T, Kolcic I, Minelli C, Hu FB, Johansson A, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Schreiber S, Aulchenko YS, Felix JF, Rivadeneira F, Uitterlinden AG, Hofman A, Imboden M, Nitsch D, Brandstätter A, Kollerits B, Kedenko L, Mägi R, Stumvoll M, Kovacs P, Boban M, Campbell S, Endlich K, Völzke H, Kroemer HK, Nauck M, Völker U, Polasek O, Vitart V, Badola S, Parker AN, Ridker PM, Kardia SL, Blankenberg S, Liu Y, Curhan GC, Franke A, Rochat T, Paulweber B, Prokopenko I, Wang W, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Shlipak MG, van Duijn CM, Borecki I, Krämer BK, Rudan I, Gyllensten U, Wilson JF, Witteman JC, Pramstaller PP, Rettig R, Hastie N, Chasman DI, Kao WH, Heid IM, Fox CS. **New loci associated with kidney function and chronic kidney disease.** Nat Genet. 2010 May; 42(5):376-84. Epub 2010 Apr 11. PubMed PMID: 20383146; PubMed Central PMCID: PMC2997674.

14. Fadrowski JJ, Neu AM, Schwartz GJ, Furth SL. **Pediatric GFR estimating equations applied to adolescents in the general population**. Clin J Am Soc Nephrol. 2011 Jun; 6(6):1427-35. Epub 2011 May 12. PubMed PMID: 21566103; PubMed Central PMCID: PMC3109941.
15. Schwartz GJ, Muñoz A, Schneider MF, Mak RH, Kaskel F, Warady BA, Furth SL. **New equations to estimate GFR in children with CKD**. J Am Soc Nephrol. 2009 Mar; 20(3):629-37. Epub 2009 Jan 21. PubMed PMID: 19158356; PubMed Central PMCID: PMC2653687.
16. Schwartz GJ, Brion LP, Spitzer A. **The use of plasma creatinine concentration for estimating glomerular filtration rate in infants, children, and adolescents**. Pediatr Clin North Am. 1987 Jun; 34(3):571-90. PubMed PMID: 3588043.
17. Fox CS, Muntner P. **Trends in diabetes, high cholesterol, and hypertension in chronic kidney disease among U.S. adults: 1988-1994 to 1999-2004**. Diabetes Care. 2008 Jul; 31(7):1337-42. Epub 2008 Apr 24. PubMed PMID: 18436617; PubMed Central PMCID: PMC2453673.
18. Ju W, Smith S, Kretzler M. **Genomic biomarkers for chronic kidney disease**. Transl Res. 2012 Apr; 159(4):290-302. Epub 2012 Feb 9. Review. PubMed PMID: 22424432; PubMed Central PMCID: PMC3329158.
19. Schiffrin EL, Lipman ML, Mann JF. **Chronic kidney disease: effects on the cardiovascular system**. Circulation. 2007 Jul 3; 116(1):85-97. Review. PubMed PMID: 17606856.
20. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G; **National Kidney Foundation. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification**. Ann Intern Med. 2003 Jul 15; 139(2):137-47. Erratum in: Ann Intern Med. 2003 Oct 7; 139(7):605. PubMed PMID: 12859163.
21. Nadar SK, Tayebjee MH, Messerli F, Lip GY. **Target organ damage in hypertension: pathophysiology and implications for drug therapy**. Curr Pharm Des. 2006; 12(13):1581-92. Review. PubMed PMID: 16729871.

22. Neumann J, Ligtenberg G, Klein II, Koomans HA, Blankestijn PJ. **Sympathetic hyperactivity in chronic kidney disease: pathogenesis, clinical relevance, and treatment**. Kidney Int. 2004 May; 65(5):1568-76. Review. PubMed PMID: 15086894.
23. Jara A, Mezzano S. **Vascular damage in chronic kidney disease.** Rev Med Chil. 2008 Nov; 136(11):1476-84. Review. PubMed PMID: 19301781.
24. Schiffrin EL. **Vascular endothelin in hypertension**. Vascul Pharmacol. 2005 Jun; 43(1):19-29. Review. PubMed PMID: 15955745.
25. Balding DJ, Bishop M, Cannings C. **Handbook of statistical genetics**, third edition (2007). Wiley edition.
26. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. **Linkage disequilibrium in the human genome**. Nature. 2001 May 10; 411(6834):199-204. PubMed PMID: 11346797.
27. Böger CA, Heid IM. **Chronic kidney disease: novel insights from genome-wide association studies**. Kidney Blood Press Res. 2011; 34(4):225-34. Epub 2011 Jun 21. Review. PubMed PMID: 21691125.
28. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K,Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W,Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M,Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M,Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly

M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. **The sequence of the human genome. Science.** 2001 Feb 16; 291(5507):1304-51. Erratum in: Science 2001 Jun 5; 292(5523):1838. PubMed PMID: 11181995.

29. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing

Consortium. **Initial sequencing and analysis of the human genome.** Nature. 2001 Feb 15; 409(6822):860-921. Erratum in: Nature 2001 Aug 2;412(6846):565. Nature 2001 Jun 7; 411(6838):720. Szustakowki, J [corrected to Szustakowski, J]. PubMed PMID: 11237011.

30. Köttgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, Yang Q, Gudnason V, Launer LJ, Harris TB, Smith AV, Arking DE, Astor BC, Boerwinkle E, Ehret GB, Ruczinski I, Scharpf RB, Chen YD, de Boer IH, Haritunians T, Lumley T, Sarnak M, Siscovick D, Benjamin EJ, Levy D, Upadhyay A, Aulchenko YS, Hofman A, Rivadeneira F, Uitterlinden AG, van Duijn CM, Chasman DI, Paré G, Ridker PM, Kao WH, Witteman JC, Coresh J, Shlipak MG, Fox CS. **Multiple loci associated with indices of renal function and chronic kidney disease**. Nat Genet. 2009 Jun; 41(6):712-7. Epub 2009 May 10. PubMed PMID: 19430482; PubMed Central PMCID: PMC3039280.
31. Köttgen A. **Genome-wide association studies in nephrology research**. Am J Kidney Dis. 2010 Oct; 56(4):743-58. Epub 2010 Aug 21. Review. PubMed PMID: 20728256.
32. Pattaro C, Köttgen A, Teumer A, Garnaas M, Böger CA, Fuchsberger C, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gierman HJ, Feitosa M, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson Å, Tönjes A, Dehghan A, Chouraki V, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Kollerits B, Pistis G; CARDIoGRAM Consortium; ICBP Consortium; CARe Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2), Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu FB, Demirkan A, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Koenig W, Illig T, Döring A, Wichmann HE, Kolcic I, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Endlich K, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Giulianini F, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, Thun GA, Paulweber B, Haun M, Sala C, Metzger M, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki I, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman JC, Hayward C, Ridker P, Parsa A, Bochud M, Heid IM, Goessling W, Chasman DI, Kao WH, Fox CS. **Genome-wide association and functional follow-up reveals new loci for kidney function**. PLoS Genet. 2012; 8(3):e1002584. Epub 2012 Mar 29. PubMed PMID: 22479191; PubMed Central PMCID: PMC3315455.

33. Böger CA, Gorski M, Li M, Hoffmann MM, Huang C, Yang Q, Teumer A, Krane V, O'Seaghdha CM, Kutalik Z, Wichmann HE, Haak T, Boes E, Coassin S, Coresh J, Kollerits B, Haun M, Paulweber B, Köttgen A, Li G, Shlipak MG, Powe N, Hwang SJ, Dehghan A, Rivadeneira F, Uitterlinden A, Hofman A, Beckmann JS, Krämer BK, Witteman J, Bochud M, Siscovick D, Rettig R, Kronenberg F, Wanner C, Thadhani RI, Heid IM, Fox CS, Kao WH; CKDGen Consortium. **Association of eGFR-Related Loci Identified by GWAS with Incident CKD and ESRD**. PLoS Genet. 2011 Sep; 7(9):e1002292. Epub 2011 Sep 29. PubMed PMID: 21980298; PubMed Central PMCID: PMC3183079.
34. Chambers JC, Zhang W, Lord GM, van der Harst P, Lawlor DA, Sehmi JS, Gale DP, Wass MN, Ahmadi KR, Bakker SJ, Beckmann J, Bilo HJ, Bochud M, Brown MJ, Caulfield MJ, Connell JM, Cook HT, Cotlarciuc I, Davey Smith G, de Silva R, Deng G, Devuyst O, Dikkeschei LD, Dimkovic N, Dockrell M, Dominiczak A, Ebrahim S, Eggermann T, Farrall M, Ferrucci L, Floege J, Forouhi NG, Gansevoort RT, Han X, Hedblad B, Homan van der Heide JJ, Hepkema BG, Hernandez-Fuentes M, Hypponen E, Johnson T, de Jong PE, Kleefstra N, Lagou V, Lapsley M, Li Y, Loos RJ, Luan J, Luttropp K, Maréchal C, Melander O, Munroe PB, Nordfors L, Parsa A, Peltonen L, Penninx BW, Perucha E, Pouta A, Prokopenko I, Roderick PJ, Ruokonen A, Samani NJ, Sanna S, Schalling M, Schlessinger D, Schlieper G, Seelen MA, Shuldiner AR, Sjögren M, Smit JH, Snieder H, Soranzo N, Spector TD, Stenvinkel P, Sternberg MJ, Swaminathan R, Tanaka T, Ubink-Veltmaat LJ, Uda M, Vollenweider P, Wallace C, Waterworth D, Zerres K, Waeber G, Wareham NJ, Maxwell PH, McCarthy MI, Jarvelin MR, Mooser V, Abecasis GR, Lightstone L, Scott J, Navis G, Elliott P, Kooner JS. **Genetic loci influencing kidney function and chronic kidney disease**. Nat Genet. 2010 May; 42(5):373-5. Epub 2010 Apr 11. PubMed PMID: 20383145.
35. Wang JG, Staessen JA, Tizzoni L, Brand E, Birkenhäger WH, Fagard R, Herrmann SM, Bianchi G. **Renal function in relation to three candidate genes.** Am J Kidney Dis. 2001 Dec; 38(6):1158-68. PubMed PMID: 11728946.
36. Hwang SJ, Yang Q, Meigs JB, Pearce EN, Fox CS. **A genome-wide association for kidney function and endocrine-related traits in the NHLBI's Framingham Heart Study**. BMC Med Genet. 2007 Sep 19; 8 Suppl 1:S10. PubMed PMID: 17903292; PubMed Central PMCID: PMC1995611.
37. Salvi E, Kutalik Z, Glorioso N, Benaglio P, Frau F, Kuznetsova T, Arima H, Hoggart C, Tichet J, Nikitin YP, Conti C, Seidlerova J, Tikhonoff V, Stolarz-Skrzypek K, Johnson T, Devos N, Zagato L, Guarrera S, Zaninello R, Calabria A, Stancanelli B, Troffa C, Thijs L, Rizzi F, Simonova G, Lupoli S, Argiolas G, Braga D, D'Alessio MC, Ortu MF, Ricceri F, Mercurio M, Descombes P, Marconi M, Chalmers J, Harrap S, Filipovsky J, Bochud M, Iacoviello L, Ellis J, Stanton AV, Laan M, Padmanabhan S, Dominiczak AF, Samani NJ, Melander O, Jeunemaitre X, Manunta P, Shabo A, Vineis P, Cappuccio FP, Caulfield MJ, Matullo G, Rivolta C, Munroe PB, Barlassina C, Staessen JA, Beckmann JS, Cusi D. **Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter**

region of endothelial NO synthase. Hypertension. 2012 Feb; 59(2):248-55. Epub 2011 Dec 19. PubMed PMID: 22184326; PubMed Central PMCID: PMC3272453.

38. Staessen JA, Roels HA, Emelianov D, Kuznetsova T, Thijs L, Vangronsveld J, Fagard R. **Environmental exposure to cadmium, forearm bone density, and risk of fractures: prospective population study**. Public Health and Environmental Exposure to Cadmium (PheeCad) Study Group. Lancet. 1999; 353:1140-1144.
39. Staessen JA, Wang JG, Brand E, Barlassina C, Birkenhäger WH, Herrmann SM, Fagard R, Tizzoni L, Bianchi G. **Effects of three candidate genes on prevalence and incidence of hypertension in a Caucasian population**. J Hypertens. 2001; 19:1349-1358.
40. Cusi D, Barlassina C, Azzani T, Casari G, Citterio L, Devoto M, Glorioso N, Lanzani C, Manunta P, Righetti M, Rivera R, Stella P, Troffa C, Zagato L, Bianchi G. **Polymorphisms of alpha-adducin and salt sensitivity in patients with essential hypertension**. Lancet. 1997; 349:1353-1357.
41. Piazza A, Cappello N, Olivetti E, Rendine S. **A genetic history of Italy**. Ann Hum Genet. 1988; 52:203-213.
42. Iacoviello L, Arnout J, Buntinx F, Cappuccio FP, Dagnelie PC, de Lorgeril M, Dirckx C, Donati MB, Krogh V, Siani A. **Dietary habit profile in European communities with different risk of myocardial infarction: the impact of migration as a model of gene-environment interaction. The IMMIDIET Study.** Nutr Metab Cardiovasc Dis. 2001; 11:122-126.
43. Cappuccio FP, Cook DG, Atkinson RW, Strazzullo P. **Prevalence, detection, and management of cardiovascular risk factors in different ethnic groups in south London**. Heart. 1997; 78:555-563.
44. Charru A, Jeunemaitre X, Soubrier F, Corvol P, Chatellier G. **Short report: HYPERGENE: a clinical and genetic database for genetic analysis of human hypertension**. J Hypertens. 1994; 12:981-985.
45. PROGRESS Collaborative Group. **Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack**. Lancet.2001; 358:1033-1041.
46. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. **PLINK: a toolset for whole-genome association and population-based linkage analysis**. AJHG. 2007; 81:559-575.
47. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. **Principal components analysis corrects for stratification in genome-wide association studies**. Nat Genet. 2006; 38: 904–909.
48. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. PLoS Genet. 2006; 2: 2074-2093.
49. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. **GenABEL: an R library for genome-wide association analysis. Bioinformatics**. 2007 May 15; 23(10):1294-6. Epub 2007 Mar 23. PubMed PMID: 17384015.
50. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. **Data quality control in genetic case-control association studies**. Nat Protoc. 2010; 5: 1564-1573.

51. Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. **Basic statistical analysis in genetic case-control studies**. Nat Protoc. 2011 Feb; 6:121-33.
52. Willer CJ, Li Y, Abecasis GR. **METAL: fast and efficient meta-analysis of genomewide association scans**. Bioinformatics. 2010; 26(17):2190-1.
53. Chen WM, Abecasis GR. **Family-based association tests for genome-wide association scans**. Am J Hum Genet. 2007 Nov; 81(5):913-26.
54. Thompson EA, Shaw RG (1990) **Pedigree analysis for quantitative traits: variance components without matrix inversion**. Biometrics 46, 399-413.
55. Aulchenko YS, de Koning DJ, Haley C. **Genomewide rapid association using mixed model and regression: a fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis**. Genetics. 2007 177(1): 577-85.
56. Amin N, van Duijn CM, Aulchenko YS. **A genomic background based method for association analysis in related individuals**. PLoS ONE. 2007 Dec 5; 2(12):e1274.
57. Gravellone L, Rizzo MA, Martina V, Mezzina N, Regalia A, Gallieni M. **Vitamin d receptor activators and clinical outcomes in chronic kidney disease**. Int J Nephrol. 2011; 2011: 419524. Epub 2011 May 15. PubMed PMID: 21647319; PubMed Central PMCID: PMC3106992.
58. Hallows KR, Mount PF, Pastor-Soler NM, Power DA. **Role of the energy sensor AMP-activated protein kinase in renal physiology and disease**. American Journal of Renal Physiology, 2010; 298(5): F1067–F1077.
59. Suzuki H, Kawai J, Taga C, Yaoi T, Hara A, Hirose K, Hayashizaki Y, Watanabe S. **Stac, a novel neuron-specific protein with cysteine-rich and SH3 domains**. Biochem Biophys Res Commun. 1996 Dec 24; 229(3):902-9. PubMed PMID: 8954993.
60. Legha W, Gaillard S, Gascon E, Malapert P, Hocine M, Alonso S, Moqrich A. **stac1 and stac2 genes define discrete and distinct subsets of dorsal root ganglia neurons**. Gene Expr Patterns. 2010 Oct-Dec;10(7-8):368-75. Epub 2010 Aug 22. PubMed PMID: 20736085.
61. Vohl MC, Lepage P, Gaudet D, Brewer CG, Bétard C, Perron P, Houde G, Cellier C, Faith JM, Després JP, Morgan K, Hudson TJ. **Molecular scanning of the human PPARa gene: association of the L162v mutation with hyperapobetalipoproteinemia**. J Lipid Res. 2000 Jun;41(6):945-52. PubMed PMID: 10828087.
62. Jamshidi Y, Montgomery HE, Hense HW, Myerson SG, Torra IP, Staels B, World MJ, Doering A, Erdmann J, Hengstenberg C, Humphries SE, Schunkert H, Flavell DM. **Peroxisome proliferator-activated receptor alpha gene regulates left ventricular growth in response to exercise and hypertension**. Circulation. 2002 Feb 26;105(8):950-5. PubMed PMID: 11864924.
63. Kordis D, Turk V. **Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes**. BMC Evol Biol. 2009 Nov 18;9:266. PubMed PMID: 19919722; PubMed Central PMCID: PMC2784779.

64. Hsia N, Cornwall GA. **Cres2 and Cres3: new members of the cystatin-related epididymal spermatogenic subgroup of family 2 cystatins**. Endocrinology. 2003 Mar; 144(3):909-15. PubMed PMID: 12586767.
65. Fesler P, Mimran A. **Estimation of glomerular filtration rate: what are the pitfalls?** Curr Hypertens Rep. 2011 Apr; 13(2):116-21. Review. PubMed PMID: 21207252.
66. Sun H, Li N, Wang X, Liu S, Chen T, Zhang L, Wan T, Cao X. **Molecular cloning and characterization of a novel cystatin-like molecule, CLM, from human bone marrow stromal cells**. Biochem Biophys Res Commun. 2003 Jan 31; 301(1):176-82. PubMed PMID: 12535658.
67. Abrahamson M, Olafsson I, Palsdottir A, Ulvsbäck M, Lundwall A, Jensson O,Grubb A. **Structure and expression of the human cystatin C gene**. Biochem J. 1990 Jun 1;268(2):287-94. PubMed PMID: 2363674; PubMed Central PMCID: PMC1131430.
68. Töhönen V, Osterlund C, Nordqvist K. **Testatin: a cystatin-related gene expressed during early testis development.** Proc Natl Acad Sci U S A. 1998 Nov 24;95(24):14208-13. PubMed PMID: 9826679; PubMed Central PMCID: PMC24352.
69. Svensson-Färbom P, Wahlstrand B, Almgren P, Dahlberg J, Fava C, Kjeldsen S, Hedner T, Melander O. **A functional variant of the NEDD4L gene is associated with beneficial treatment response with β-blockers and diuretics in hypertensive patients**. J Hypertens. 2011 Feb; 29(2):388-95. PubMed PMID: 21052022.
70. Manunta P, Lavery G, Lanzani C, Braund PS, Simonini M, Bodycote C, Zagato L, Delli Carpini S, Tantardini C, Brioni E, Bianchi G, Samani NJ. **Physiological interaction between alpha-adducin and WNK1-NEDD4L pathways on sodium-related blood pressure regulation**. Hypertension. 2008 Aug; 52(2):366-72. Epub 2008 Jun 30. PubMed PMID: 18591455.
71. Jin HS, Hong KW, Lim JE, Hwang SY, Lee SH, Shin C, Park HK, Oh B. **Genetic variations in the sodium balance-regulating genes ENaC, NEDD4L, NDFIP2 and USP2 influence blood pressure and hypertension**. Kidney Blood Press Res.2010;33(1):15-23. Epub 2010 Jan 15. PubMed PMID: 20090362.
72. Schmitz G, Langmann T. **Structure, function and regulation of the ABC1 gene product**. Curr Opin Lipidol. 2001 Apr; 12(2):129-40. Review. PubMed PMID: 11264984.
73. Iatan I, Alrasadi K, Ruel I, Alwaili K, Genest J. **Effect of ABCA1 mutations on risk for myocardial infarction**. Curr Atheroscler Rep. 2008 Oct; 10(5):413-26. Review. PubMed PMID: 18706283.
74. Okuhira K, Fitzgerald ML, Sarracino DA, Manning JJ, Bell SA, Goss JL, Freeman MW. **Purification of ATP-binding cassette transporter A1 and associated binding proteins reveals the importance of beta1-syntrophin in cholesterol efflux**. J Biol Chem. 2005 Nov 25;280(47):39653-64. Epub 2005 Sep 28. PubMed PMID: 16192269.
75. Chen X, Guo Z, Okoro EU, Zhang H, Zhou L, Lin X, Rollins AT, Yang H. **Up-regulation of ATP binding cassette transporter A1 expression by very low**

density lipoprotein receptor and apolipoprotein E receptor 2. J Biol Chem. 2012 Feb 3; 287(6):3751-9. Epub 2011 Dec 14. PubMed PMID: 22170052; PubMed Central PMCID: PMC3281687.

ACKNOWLEDGMENTS

I acknowledge the Filarete Foundation Genomics and Bioinformatics unit and KOS Genetic s.r.l. for their collaboration to the present study, and in particular Prof. Cristina Barlassina, Prof. Daniele Cusi, Erika Salvi, Francesca Frau, Sara Lupoli, Martina Chittani and Fabiola Canavesi for their support and precious scientific contribution. I also want to thank Tatyana Kouznetsova, Prof. Caroline Fox, Matthias Olden and the CKDGen Consortium for sharing their data and for the keen interest shown on the project.