Grant agreement no:
279171

Project acronym:

MEDIGENE

Project title:

Genetic and environmental factors of insulin resistance syndrome and its long-term complications in immigrant Mediterranean populations

Final report

Funding Scheme:
Specific International Cooperation Action (SICA)

Name of the scientific representative (co-ordinator) :

Florin Grigorescu MD, PhD (INSERM)
Managing: **Christophe Normand**
University of Montpellier 1, France
Phone (+33) 411 75 98 95
Fax (+33) 467 54 27 31
Email: florin.grigorescu@inserm.fr

Final publishable summary report

Executive summary

MEDIGENE (standing for *Mediterranean Genes*) project in FP7 will study genetic and environmental (G x E) determinants of the *metabolic syndrome* (MetS) in natives and recent immigrants in Europe by a novel approach integrating ancestry of Mediterranean populations in epidemiology, locus refining and *Genome Wide Association Studies* (GWAS). Complex disorders associated with insulin resistance such as MetS have a major contribution to the overall morbidity and mortality in human populations. In modern societies, these diseases have an important economic impact and immigrants received a particular attention. From millennia, West Mediterranean shores were place of pre-historical termini of population expansion from Southern Europe and North Africa. Archaeogenetic studies in Europe indicated that Y chromosome and *mitochondrial* (mt)DNA patterns or *Ancestry Informative Markers* (AIMs) revealed a close relationship between genetic and geographic distances able to locate an individual DNA within few hundreds kilometers. The project will use this information and that on autosomal *single nucleotide polymorphism* (SNP) markers in mapping genes for insulin resistance, cardiovascular or metabolic complications in *Albanians, Romanians, Turkish, Tunisians, Algerians and Moroccans* immigrated in host countries such as *France, Spain, Italy* and *Greece* compared to native populations. Ancestry markers and exploration on ancient (a)DNA from Antique Roman historical migration in *Catalonia* (Spain) will help to give a better picture of the genetic landscape of Europe and North Africa. Populations from East of Europe or in the Volga-Ural region, being genetically more distant, will help the understanding of the genetic background in Europe. Genes for MetS and insulin resistance will be studied in existing samples from host and home countries or from novel recruitment in GWAS, locus refining and haplotype mapping. Informative filtered SNPs will be then used in epidemiology and on novel DNA samples, in revealing G x E interactions and specificities of the pathogenesis of MetS. Genetic findings, including the gene variation, will be tentatively replicated in home countries (*Anatolia* and *North Africa*) in the goal to develop ethnic specific markers more significant at a clinical scale. Major impact is expected from dissemination of our findings to prevent occurrence of MetS and obesity in children and adolescents or in descendants of modern immigration, understand variability of clinical manifestations of MetS in the context of malnutrition and finally – integrating GWAS strategies – ameliorate the association signal and bursting R&D activities of SMEs.

Summary description of context and objectives

MEDIGENE (*Mediterranean Genes*) project proposed to study genetic and environmental (G x E) interaction in the pathogenesis of *metabolic syndrome* (MetS) in recent immigrants in Europe compared to native populations by novel approaches tempting to integrate phylogeny and ancestry of Mediterranean populations, locus refining and *Genome Wide Association Studies* (GWAS). Strategically, the program specificity is to bring together competences of *anthropologists*, *geneticists*, *archaeologists* and *clinicians* in understanding of genetic and nutritional aspects of insulin resistance. The challenging goal is to use information from genetic anthropology to better understand gene variability in population, potential stratification and thus, to ameliorate diagnosis of insulin resistance and obtain more powerful genetic markers.

Mediterranean Immigration. Complex disorders such as obesity, *type 2 diabetes* (T2D) and MetS have a major contribution to the overall morbidity and mortality in human populations. In modern societies these diseases also have an economic impact since current *cardiovascular* (CV) and metabolic complications are very likely to paralyze health expenses in the 21st century. One major concern in complex disorders is the impact on immigrant populations, recently settled among autochthons (*natives*). They represent 3% of the world population (191 millions) and by themselves form another virtual continent. In Europe, there are 41 millions immigrants, equivalent to 8.6 % of the European Community (EC) population. In modern times, immigrants are carrying genetic susceptibility from the home country contributing to the general variability in epidemiological studies for a country. As function of admixture and cultural habits (*aculturation*), genetic predisposition of immigrants enters into collision with Westernized way of life resulting in protective and pathogenic effects. This may be a major source of difficulty in understanding the contribution of genetic background in the etiology of complex disorders. Therefore, one goal for research would be to first decipher ethnic components and eventually different genetic susceptibility among populations.

The interest in *Mediterranean* area is coming form the numerous population movements during millennia and from the wide diversity of life styles. Epidemiological data also indicate disparities in health indices for CV complications. Since the finality is to ameliorate the genetic diagnosis of MetS and insulin resistance we intend to amplify the association signals by considering anthropological *heterogeneity* and population structure. This may be obtained by multiple sampling in populations around the Mediterranean Sea. We have focused on some actual immigrants (*Albanians, Romanians, Turkish, Tunisians, Algerians and Moroccans*) in four major host countries *France, Spain, Italy and Greece,* choice explained by the severe economic impact in healthcare and vulnerability of immigrants.

Anthropology of Europeans. History may help understanding of population movements from one country to another. One challenging question for researchers is how far one should go back in the population history to detect genetic effects that may help stratification in case-control association studies. Indeed, West Mediterranean shores were place of pre-historical termini of population expansion from Southern Europe and North Africa. Understanding of the genetic architecture of Europeans implies a careful analysis of its genetic variation. In pre-historical times, numerous *archaeogenetic* studies have indicated

a progressive colonization of the Europe from East to West by *demic* diffusion and mass migration. North South axis was considered impermeable over the Mediterranean, at lest from some studies with Chr Y markers. Actual immigration from one side to another of the Mediterranean Sea contradicts this idea and the impact on susceptibility for complex disorders remains poorly studied.

Antique Romans and bio-geography. Studies of historical populations such as Antique Romans are a good model to understand population movements and genetic make-up. Created in 218 AD, the *Tarraco* city was one of the most important settlements of Western Mediterranean basin as Capital of *Provincia Hispania Tarraconensis* comprising almost 20.000 inhabitants. The actual city of Tarragona has brought to light numerous remains from cemeteries dating from the 2nd to the 7th century AD. The goal of MEDIGENE was to approach the Roman society from a wide perspective, as an interdisciplinary work where archaeologists, odontologists, paleo-pathologists and epigraphists would combine their work with geneticists to obtain information on gene variability. Although studies of *ancient*(a)DNA were already performed in ancestral populations, there are poor population studies. Therefore we considered that numerous remains in the *Tarragona Necropolis* were excellent opportunities to study hundreds of individuals in the antique Roman society. The admixture in the Mediterranean area was supposed to be high since there were numerous population movements with mass migration. Classical examples are *Phoenicians* and *Greeks* who established colonies all around the Mediterranean Sea. Similarly *Gothic* populations from the North traversed Europe and through Gibraltar strait settled in North Africa. The popular belief among researchers is that all these movements generated a high degree of admixture. Despite this, recent genetic studies with markers from non-recombinant regions of Chr Y and mitochondrial(mt)DNA or with more specific *Ancestry Informative Markers* (AIMs) revealed however a close relationship between genetic and geographic distances in Europe. Autosomal markers, submitted to erosion by genetic recombination through numerous generations are supposed to dissipate this information. One of the most interesting observations over the last several years was that using *Affymetrix* DNA chips for autosomal *Single Nucleotide Polymorphism* (SNP) markers, anthropologists were able to locate an individual DNA in the geographic landscape with an error of several hundreds km. This suggest a very high degree of fixed patterns in Europe from Mesolithic or early Neolithic times up to modern times. In the context of low genetic differentiation among Europeans, this observation is noteworthy. Admixture definitively was operating, but genetic makeup followed perhaps more intricate roads. Therefore, our project had the idea to use genetic information on bio-geography to obtain a better classification (*stratification*) of populations and thus, to ameliorate the association signal. Clustering populations by autosomal SNPs, which are the same as those used in GWAS itself for a disease is very challenging.

GWAS in populations. GWAS are unprecedented opportunities to decipher culprit genes for complex disorders such as MetS but their success was compromised by the fact that genes discovered so far displayed weak association being unable to explain the heritability in population. Researchers are referring to "*missing heritability*". Although, there may be many explanations for this, MEDIGENE consortium considered that one main reason would be the disregard of *rare* SNPs with *Minor Allelic Fr*equency (MAF) < 1% but which may have strong effect. Moreover, there is a lot of heterogeneity and still unknown etiology of complex disorders and in particular, specific G x E interaction. Processes such as "*decanalization*" and *evolutionary adaptation* to ecological niches may contribute to actual epidemics of obesity, diabetes or MetS. These considerations strongly point out towards the

necessity for novel epidemiological strategies with multidisciplinary vocation involving collaborations between anthropologists and medical geneticists.

Locus refining. The project has the objective to refine findings from GWAS or other gene candidates in MetS either by sequence or bioinformatics strategies. Rare variants might be a source for completing the "*missing heritability*" in complex disorders. Private mutations may also contribute to this. In the particular case of MEDIGENE, refining was proposed by resolving haplotype structure of candidate genomic regions, which is a challenging but original point. One of the most efficient strategies is the *imputing* having as source publicly available database from *1000 Genome Project* (1kGP). Thus, MEDIGENE project proposed to use this information for imputing or to create a new customized chip at a much higher density of markers than International HapMap database.

New sampling. In view of actual epidemiological data, consortium estimated that a central part of the project should involve new recruitment of immigrant populations in Europe under *high-quality epidemiological standards*, including *24 h dietary records* (24HDR). Objectives would to evaluate prevalence of MetS in local samples of immigrants, to assess association with age, sex, socio-economic, lifestyle and environment, including nutritional factors. Moreover, new hypothesis were launched as for instance, the involvement *branched chain amino-acids* (BCAA). We estimated that novel sampling should concern 1700 individuals completing previous samples and focusing on the *Romanian* community in Spain, *Turkish or Algerian* communities in France and *Albanian* community in Greece and Italy. In view of epidemiological observations, North African's diaspora in Europe received particular attention. One of the most interesting observations was the existence of lower mortality by CV causes of immigrants from Maghreb in France, demonstrated for men but not for women – in other terms – women from Maghreb lost the advantage of immigration. At a more general level, these examples suggest the importance of gender specificity for insulin resistance and metabolic complications. One of the excellent models in this area is the MetS and underling insulin resistance in *polycystic ovarian disease* (PCOS) in women with infertility.

Replication in home-countries. Potential discoveries and novel GWAS associations impose obviously subsequent replications in home countries. There may be more than 150 candidate genes for MetS from previous GWAS of European descendants, but these genes were not all replicated in ethnic populations, including *North Africans*. Previous studies in other geographical areas (e.g. Canadians) revealed that some genes were indeed [*universally*] replicated but others did not appearing as new and perhaps ethnic specific. Another original aspect of the project is the interest in insulin resistance rather than *dyslipidemia, diabetes or hypertension* as components of MetS. Indeed, previous GWAS for MetS did not address specifically insulin resistance *per se*, perhaps due to the lack of data. Simplistic regard on insulin level is not sufficient. By contrast, how insulin resistance varies as function of cumulative occurrence of components [*criteria*] of MetS may be a new way to investigate pathogenesis of MetS.

Gene environment (G x E) interaction. Another goal of the program was to understand pathogenesis of MetS and interactions with confounding factors. While most of studies correlated insulin resistance uniquely with the *energy allocation mechanisms*, MEDIGENE program proposed to regard this alteration from a much broader perspective, as a component of long term and *geographically delimitated adaptation*. Genes involved in

insulin action (e.g. *Insulin Receptor Substrate* – IRS or *Insulin Recepto*r – INSR) or obesity genes such as *fat mass and obesity associated* (FTO) gene may act on brain regulating the eating behavior. The global approach imposes evaluation of prevalence of MetS components in relation to *comprehensive questionnaires on dietary intake, physical activity, age at migration and social ties*. G x E interactions have a major impact in health, although it remains underappreciated due in principal, to the lack of data in geographically specific populations and still insufficient collaboration between research in anthropology, ecology and sociocultural fields. The gaps are progressively recovered, particularly by the accumulating knowledge in public databases integrating GWAS data and specific observations in evolution, nutrition and ecology.

Description of main S&T results and foregrounds

Anthropologic studies in native populations

To offer new solutions for population stratification in *Genome Wide Association Studies* (GWAS) for *Metabolic Syndrome* (MetS) and/or insulin resistance we intended to better characterize the *genetic landscape of Southern Europe, Balkans and North Africa* based on anthropological data. The Russian partner conducted these studies and was involved genotyping Mediterranean populations with the commercially available EUR chip from *Affymetrix* or with a new customized chip (MEDISCOPE) from the University of Montpellier. We generated GWAS data with EUR chip from a first set of 480 anthropological samples of European populations (*France, Spain, Italy, Greece, Romania, Moldova, Albania and Croatia*) and North Africans (*Morocco, Algeria, Tunisia)* or Anatolia (*Turkey* and *Lebanon*). Russian partner also studied other Turkic-speaking populations (*Karachays, Balkars, Kumyks, Nogays, Chuvashes, Tatars, Bashkirs, Yakuts*) from Russia, Kazakhstan and Uzbekistan, which are genetically more distant but useful in understanding differences among populations. The first GWAS data revealed extraordinary results, indicating that all Mediterranean samples were superimposed on the geographical map, having *Spanish Basques* at the West extreme and *Lebanese* at the East extreme. Interestingly, North Africans were located in the South respecting the natural barrier of the Mediterranean Sea. Turks were close to other Turkic-speakers of *Caucasus* while those from Volga-Ural region and *Yakuts* were situated on the most East extreme. *Principal Component Analysis* (PCA) of autosomal SNP markers performed before GWAS of MetS indicated that most of Europeans (native or immigrants) might be considered together because all were closed to standard CEU populations included in HapMap database. Exceptions were evidently North Africans (*Algerians, Moroccans and Tunisians*) forming a separate group. These PCA clusters allowed exclusion from each population of outliers. Unexpected, a relative high number of outliers were detected in French samples randomly recruited in two regions in Southern France, suggesting a certain level of admixture with North Africans.

Russian partner used uniparental markers on mtDNA and Chr Y to confirm and enrich geographical separation. These markers were absent from commercially EUR chip, but were included in the customized MEDISCOPE chip. Mitochondrial (mt)DNA markers were used in stratification of women with MetS and *hyperandrogenism* (HA) in the frame of PCOS – subproject conducted by the French partner - which included Romanians, North Italians and Greeks. Uniparental Chr Y markers were used for stratification of male populations with MetS from natives or immigrants. PCA revealed a clear separation between *Atlantic* populations in the West of Europe (Spanish or French) from East Europe and *Balkan* region (*Albanians, Greeks, Romanians, Croatian*). For practical reasons these stratifications were designated as *gates*: *Atlantic, Anatolia, North Africa* gates, which apparently deserved separate GWAS. While differences between French (North) and Algerians or Tunisians (South) were easily detected, no difference were found for instance between two regions in France (*Languedoc Roussillon* and *Rhône-Alpes* regions). Similarly no differences were found among Romanians, *Wallachians* from Muntenia versus *Moldovans*. Interestingly, a high proportion of haplogroup E-M81 was found in Spanish, which was also the largest in *Algerians* suggested a gene flow over the Gibraltar barrier. Both French and Spanish males

contained high proportion of R1b-M269 indicating close similarities, at least as strong as found in Italians. In addition, Romanians males displayed more frequent haplogroup I-CTS674 and E-M243 presumably of Near East origin. Immigrant Albanians in Greece and Italy can be easily discriminated because of E-M35 frequency. Another gradient was observed between *French, Spanish and Algerians* for E-M81 haplogroup, which was at higher frequency in Algerians. These studies suggested the necessity to include Spanish population in the so-called "France-Algerian gate" supposed to compare in GWAS one side and another of the Mediterranean Sea. Taking together, these PCA data allowed detection of sensitive differences between studied ethnic populations. However, it should be noted that the number of individuals remained relatively low, which hamper phylogenetic resolution of haplogroups and narrows the interpretation. Definitively, the main separation among *Albanian, French, Spanish and Algerians* can be demonstrated and thus, the main initial goal was entirely achieved.

Genome Wide Association Studies (GWAS) in metabolic syndrome

French partner performed GWAS in 1480 samples with 657,856 SNPs in 553 cases and 876 controls and revealed positive hits on Chr 9, 11, 16, 9, 3, 18 and 12. Some rare SNPs on Chr X had P values < $8.8 \times 10^{-10.}$ The most associated gene was located on Chr 9 encoding for an aminopeptidase, with a potential role in the generation of angiotensin IV and pathogenesis of hypertension. Many other SNPs were strongly associated, but with and inter-genic location. These results should be interpreted with caution because of the imbalance between man and women in these collections. Moreover, GWAS in natives versus immigrants were still insufficient as size. Interestingly, there was a good correlation between insulin resistance and cumulative components of MetS in all populations. Thus, *Homeostasis Model Assessment* (HOMA) index was proportionally increased with level of BMI and cumulative components. Therefore, another GWAS was performed on 479 controls and 524 cases of MetS targeting insulin resistance ($HOMA_{IR}$). This study revealed strong signals on Chr 2, 8, 9 and 16. Most of genes corresponding to these signals were confirmatory, except SYT9, gene. As expected, we found positive signals on TOX and DUSP9 genes, previously studied in French laboratory. To further understand significance of these findings, we were searching for the overlap between these genes and 92 genes recorded in HUGE database for MetS from previous GWAS. Numerous genes were confirmed such as HNF1A, FTO, TCF7L2 and ABCA1. *Correlation trend test* (CTT) with HOMA values revealed 40 positively hits, one of the best being FTO gene, but also ALMS1, SLC22A10 and adrenergic receptor ADRA1D.

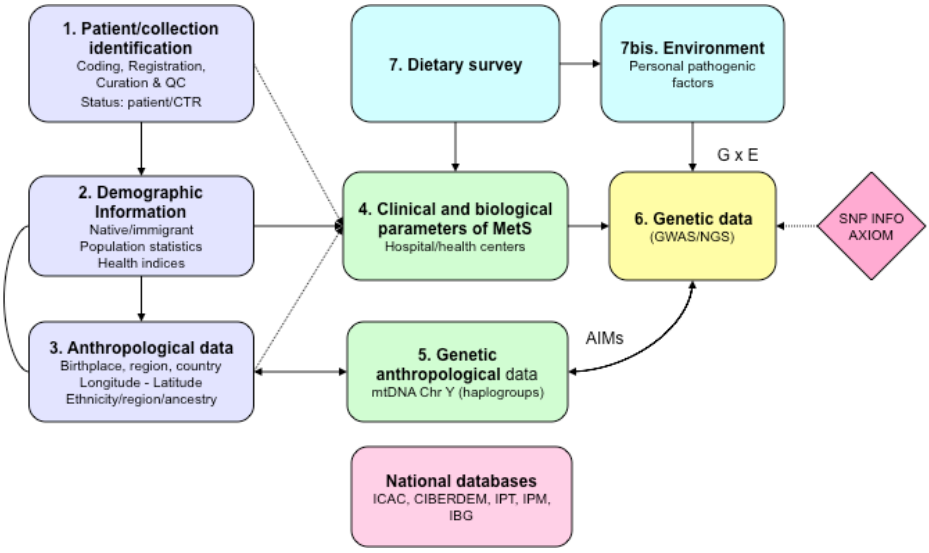
Based on anthropological studies, we initiated GWAS in stratified populations called "gates". GWAS revealed strong signals on Chr 9, Chr 11, Chr 16 and Chr X in Atlantic gate while in *Balkan*, the strongest signals were found on Chr 6 and 20. When *Algerians* were compared to *French* (*French-Algerian* Gate) the strongest signal remains on Chr 20, followed by ChrX (a signal different from DUSP9) and Chr 3.

We concluded that GWAS for MetS based on anthropological study in Mediterranean populations revealed significant signals of association in ethnic groups. Further systematic studies coupled with *meta-analysis* of strata will reveal important signals that may be used for diagnosis of MetS, insulin resistance and components of MetS. Variation in allelic frequency should be regarded not only as trait that should be statistically corrected (e.g. by

genomic control) but as a *diversity impeding diagnosis and personalized medicin*e. It is expected that in-depth studies by *haplotype mapping* and locus refining may reveal new ways to develop ethnic specific tests in genetic epidemiology over the Mediterranean area.

Bioinformatics – MAGDB and MEDIPAD software

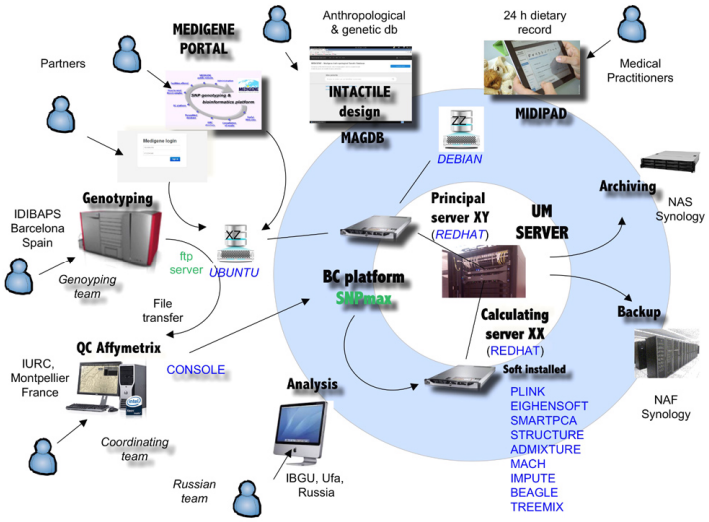
MAGDB. One prerequisite condition for the project was the careful evaluation of dataflow and implementation of bioinformatics tools to assure secure management and data integrity. A LINUX server was implemented at the *University Montpellier* with periodic backup and archiving systems and able to manage huge amount of data. A MEDIGENE *Vortal* (https://xz.univ-montp1.fr) was created to assure administration, links with the official website (www.medigene-fp7.eu), tutorials and exchange of results. The BC-platform (SNP_{max}) of the partner from Finland was implemented on the server able to manage GWAS and *Next Generation Sequencing* (NGS) data (http://bcplatforms.com/). A number of 4621 DNA samples from previous collections were curated, which together with another 681 samples from Mediterranean countries were collected during WP1 displaying information on birthplace. Another 2019 samples from patients with MetS were obtained adding to 1921 samples from women with PCOS with or without MetS. Russian partner contributed with 12 ethnic groups (336 samples) of Turkic-speaking populations while Romania with more than 800 PCOS samples, all containing information on birthplace of individuals two generation upstream. Data were collected in a new database called MAGDB (*Medigene Anthropological and Genetic DataBase*) composed by several modules and intended to integrate anthropological, nutritional, clinical, genetic and environmental data (figure below).



Modules composing the MAGDB database for MEDIGENE program

The information flow in MEDIGENE among partners is depicted in the next figure. Clinical data from partners were collected through new software MADIPAD. The program was developed by the French partner INTACTILE DESIGN SA (France) (http://intactile.com/). DNA samples were submitted to *quality control* (QC) and then genotyped in IDIBAPS (Barcelona) on *Affymetrix* platform of the Genomic Facility. Raw data from the Spanish

partner (CIBER) were then transferred to Montpellier for QC and for further statistical analysis. Virtual machines were also created for the use of partners. The server contained numerous computer programs for analysis including PLINK and imputing programs such as IMPUTE2 and BEAGLE 4.1.



Informatics and bioinformatics structures implemented for MEDIGENE program

MEDIPAD. By far, the most important realization during the program was the MEDIPAD software developed by *Intactile Design SA* (http://www.intactile.info/www-medigene) able to record a dietary data for 24HDR from all countries around the Mediterranean area. Data were transmitted to the server and included in the MAGDB containing clinical profile of patients and GPAQ data on lifestyle. MEDIPAD helps recording dietary intake in a patient (*breakfast, lunch, dinner and snacks*) through *face-to-face* interview and uses publicly available French CIQUAL composition of aliments. The dietician or medical practitioner can have access from an iPAD or a local computer. The French CIQUAL database was completed up to 2193 food items from France, as well as with foods from *Turkish, Greek, Spanish* and *Romanian* partners. *Tunisian* and *Moroccan* partners also added specific dishes. Nutritional assessment parameters were standardized and calibrated based on 5124 dietary recalls obtained from *Saint Pierre Institute* (*Palavas,* France) on a population with 80% obesity at different ages. BMI values were corrected for age and yielded 28.5% 1st degree obesity and 46% 2nd degree obesity. More importantly, a completely new database was compiled in CIQUAL for the content of BCAA by the French partner, which was proposed for investigation in *nutrigenomics*. The hypothesis was that BCAA (*leucine, isoleucine and valine*) - being essential amino-acids and abundant in human diet - may be alimentary factors in the pathogenesis of insulin resistance. Several studies have tempted to define correlations between HOMA index, parameters of intravenous glucose tolerance test (IVGTT) and plasma levels of BCAA. Other studies intended to define correlation between BCAA intake and genetic data.

Ancient (a)DNA studies of Romans in Tarragona Necropolis

Historical populations such as *Antique Romans* represent a good model for understanding genetic make-up of Europe. Remains from Tarragona city were excellent opportunities to approach the Roman society from a *wide perspective, as an interdisciplinary work between archaeologists, odontologists, paleo-pathologists, epigraphists and geneticists.* Studies were

conducted by ICAC and CIBER in Spain with the goal to obtain *ancient* (a)DNA samples. Excavations have been analyzed according to their economic and historical context, skeletons being located at the *National Archeological Museum* of Tarragona where gender and age-at-death were determined from 1982 skeletons followed by odontological study.

National Archeological Museum of Tarragona.

A relational database has been generated to gather information on social environment and contextualize samples. This database will help now researchers to understand ritual and funerary practices and reconstruct social structure, family ties or ancient migrations of the Roman society. Teeth were used to establish a bank of aDNA called *TarraCorpus* managed by CIBER. Dental pathology assessment and paleopathology studies have been carried out in 218 adults and 46 sub-adults from the Roman Imperial age necropolis (1^{st} – 5^{th} century AD). The study included 4366 permanent and 375 deciduous teeth. There were investigated pathological lesions (*caries, abscesses and ante mortem tooth loss*) or non-pathological ones (*calculus*) or *linear enamel hypoplasia*. In fine, a number of 1339 teeth were provided for Biobank (IDIBAPS) from 492 individuals. Presence of *dental hypoplasia* was related with nutritional aspects. It was found high degree of hypoplasia in the lower canine teeth correlated with significant change in childhood nutrition (around age 7). Females have a lower percentage of dental wear in comparison to males and *alofisos* and 17.5% have lower percentage of alveolar bone reabsorption in comparison to 28.8% in males. Not statistically significant differences between man and women for caries and calculus suggested a similar diet in both sexes. Hypoplasia also corroborates to this theory for children. A significant positive correlation between abrasion and alveolar bone height would explain the increase of the distance between *enamel–cementum* junction and alveolar bone crest. These findings were in accordance with those in *Romano-British* and medieval populations. Since teeth wear unsurprisingly increases with age, the percentage of calculus also increases, making it statistically significant for adolescent individuals, youth, adults and senile. ICAC found statistical significance for wear, cavities, calculus and hypoplasia. As centuries pass, the percentage of hypoplasia decreases in a significant manner. Because of the inability of enamel to remodel, these easily observable ring-like defects can provide an *indelible chronological record* of stress occurring during teeth crown formation. We found a 6% of hypoplasia in the total studied sample. Noteworthy the high presence of hypoplasia during 2^{nd} and 3^{rd} century corroborated with data from the literature emphasizing the spread of various diseases, amongst which the *smallpox* is the best known. The crisis of the 3^{rd} century reversed the demographic curb and, in this context, one should recall the "*plague of the Antonines*" (164-180) that it could have killed a third of the Roman Empire population. When development of a population slows down, wear and calculus decreases due to

malnutrition and low protein intake. Coinciding with that, from the 3rd century onwards, end the golden age in *Tarraco* and paralleled demographic and nutritional wealth reduction with influences on dental wear. A change in food preparation methods throughout Roman culture was also recorded, such as the switch from *stone mortars* to *wooden mortars* for cereal processing that could introduce little or no grit contaminants into the food and therefore decrease the dietary abrasiveness of the grain ingested in the latest centuries. Finally, the highest percentage of calculus in the teeth of individuals from low and privileged socioeconomic status, as well as those in the medium-low, proves to be statistically significant. We can explain the significance in the calculus in the privileged group due to a higher protein intake and the low group due to poor hygiene and a higher index of periodotal disease.

A number 489 samples were amplified in CIBER for DNA analysis. The quality control (QC) tests were performed for each sample processed by amplification of *Hypervariable Region* I (HVRI) of mtDNA and of nuclear DNA (nuDNA). No contamination coming from reagents or samples was detected. Almost 80% of samples amplified for mtDNA and 18% for nuclear DNA. The *whole genome amplification* (WGA) was used to increase DNA amount, which was successful for 76 samples with similar proportion in male and female. aDNA after WGA was studied by *Affymetrix* chip MEDISCOPE from which we recovered 11.4% of SNPs (28,859 SNPs). To assess the bio-geographic location we performed analysis on autosomal SNP by PCA after *pruning* with combined HapMap markers from European, Asian and African populations. Roman population was located in a cluster upside of CEU and African populations, suggesting that Romans may be distinguished from other modern populations and their bio-geographic location was compatible with Indo-European origin. By contrast, mtDNA and Chr Y SNPs markers failed in these samples because of the presence mutually exclusive mutations. To study individual genes, we focused on FTO (*fat mass and obesity associated*) gene, which contains markers for obesity and MetS in almost all populations. To increase the signal, we imputed from 1000 Genome Project reference panel 257 SNPs from Romans and generated 355 SNPs with a *linkage disequilibrium* (LD) profile comparable with modern populations. From the total number of SNPs, 9% were identical to modern populations. However, a fraction of 3% of SNPs was specific for Romans. Ancient and modern populations were compared at this locus for which major differences were found in intron 4, but the intron 1 (containing the leader-SNP for obesity) was similar. It is expected that resolution of haplotype-structure and detection of differences among ethnic populations will help the interpretation of data from ancient DNA in Roman population.

Data from Roman samples indicate that major objectives were attained, offering the possibility to obtain genetic information from *ancient* (a)DNA. Although data still require refining, bio-geographic location and genes may be determined in such ancient populations. At a more general level, ICAC studies demonstrated successful collaboration between archaeologists, clinicians and geneticists as an original mark of MEDIGENE program.

Locus refining for candidate loci

Prioritized genes & imputing. The general objective was to refine findings from GWAS or other gene candidates in MetS either by sequence information around the associated SNP, imputing or functional understanding by haplotype-mapping. Development of novel imputing procedures using *hyplotype clustering method* considerably reduced the

computational time and offered new opportunities to refine loci of biological interest. Although at least two imputing methods were tested (IMPUTE2 and BEAGLE) we used essentially BEAGLE v.4.1 for at least 22 regions in the human genome. Further analysis consisted in *determination of haplotypes* as function of *linkage disequilibrium* (DL) blocks, reconstruction of haplotypes in population and assignment of haplotype-pairs to individuals. These studies were preceded by a prioritization of gene candidates, as indicated in the table below.

Prioritized genes for MetS after scoring for significance in GWAS and pathogenic SNP content

Symbol	Gene name	Location	Score[a] GWAS	SNP[b] Pathogenic
CETP	Cholesteryl ester transfer protein	Chr16q21	4,5	-
ADIPOQ	Adiponectin	Chr3q27	4	0,13
G6PC2	Glucose-6-phosphatase	Chr2q24.3	3,7	-
CRP	C-reactive protein	Chr1q23.2	3,6	-
LPL	Lipoprotein lipase	Chr8p22	3,6	3,89
LIPC	Lipase, hepatic	Chr15q21-q23	3,6	0,08
DGKB	Diacylglycerol kinase	Chr7p21.2	3,5	-
FADS1	Fatty acid desaturase 1	Chr11q12.2-q13.1	3,5	-
FADS2	Fatty acid desaturase 2	Chr11q12.2	3,5	-
GCK	Glucokinase (hexokinase 4)	Chr7p15.3-p15.1	3,5	9,00
GCKR	GCK regulator	Chr2p23	3,5	-
HNF1A	HNF1 homeobox A	Chr12q24.2	3,5	2,52
LEPR	Leptin receptor	Chr1p31	3,5	0,03
MTNR1B	Melatonin receptor 1B	Chr11q21-q22	3,5	-
ABCA1	ATP-binding cassette	Chr9q31.1	3,4	0,35
APOB	Apolipoprotein B	Chr2p24-p23	3,4	0,81
FTO	Fat mass and obesity associated	Chr16q12.2	3,4	0,01
IL6R	Interleukin 6 receptor	Chr1q21	3,4	0,38
MLXIPL	MLX interacting protein-like	Chr7q11.23	3,4	-
NR1H3	Nuclear receptor subfamily 1	Chr11p11.2	3,4	-
APOA1	Apolipoprotein A-I	Chr11q23-q24	3,3	11,59
APOE	Apolipoprotein E	Chr19q13.2	3,3	
ABCB11	ATP-binding cassette	Chr2q24	3,2	0,18
TFAP2B	Transcription factor AP-2 beta	Chr6p12	3,2	1,46
TRIB1	Tribbles homolog 1 (Drosophila)	Chr8q24.13	3,2	-
APOA5	Apolipoprotein A-V	Chr11q23	3	0,69
TCF7L2	Transcription factor 7-like 2	Chr10q25.3	2,1	-
CDH13	Cadherin 13, H-cadherin (heart)	Chr16q23.3	1,6	-
EDIL3	EGF-like repeats and discoidin I-like domains 3	Chr5q14	1	-
FER	Fer (fps/fes related) tyrosine kinase	Chr5q21	1	-

a, Score in GWAS represents the sum of the scoring as function of P values and that of the OR of association in HuGe databese; b, Pathogenic score is based on the presence of pathogenic or potentially pathogenic SNPs from dbSBP; Genes in yellow indicate the identity with the list retrieved from GAD database.

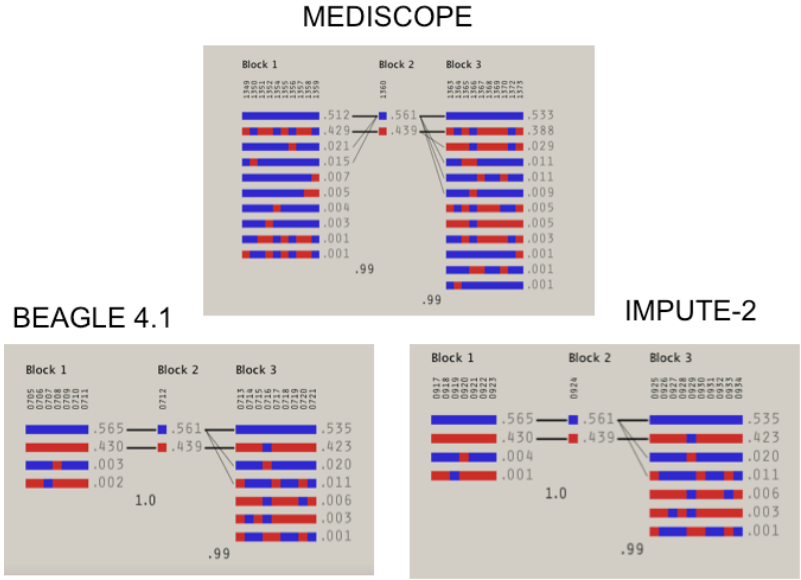
MEDISCOPE chip. Another method to increase density of genotyped SNP was to develop customized "*MyDesign*" chip. Together with experts from *Affymetrix*, the French partner end up with the solution to add to the commercially available EUR chip 100,000 additional SNP for the needs of MEDIGENE. This was possible eliminating the second probeset for validated SNP in *Affymetrix* database. This "intelligent chip" finally contained 758,223 SNPs covering genes for MetS, insulin signaling & action, specific genes in PCOS with insulin resistance, pharmaco-genes and several ethnic markers including some of the Neanderthal genome. Russian partner added markers for mtDNA and Chr Y for anthropological needs. The composition of this chip from the University of Montpellier is indicated in the table below.

Composition of the customized chip for MEDIGENE called MEDISCOPE (Montpellier)

Category	# of Probesets**
Classical AXIOME_EUR SNPs	650,963
Additional SNP for anthropological studies	
Ancestral Informative Markers (AIMs)	3,608
Mitochondrial (mt) SNPs	86
Chr Y SNPs	204
Genes at maximum density (1KGP)	
MetS loci at 1000 GP density	26,605
Other genes at HapMap density	
MetS loci (intra & extragenic)	40,070
MetS in HA women	14,603
Genes for insulin signaling	9,652
MetS and complications	12,332
Grand Total	**758,223**

**, Number of SNPs is cumulative

The use of MEDISCOPE corroborated with imputing allowed extension of fine-scale mapping on several targets such as FTO, AVPR1A, INSR, IRS, TRIB3, APOA5 and SIK1 genes. At the FTO locus (Chr16q12.2) we were interested to understand variability of leader SNP rs9939609 in ethnic populations. By studying 701 individuals (377 lean controls and 324 obese patients with MetS) with 460 SNP from EUR chip and then with SNPs from MEDISCOPE chip, the imputing generated 2142 SNPs with BEAGLE and 2818 with IMPUTE2, among which 44 SNPs were common and positively associated with obesity. Strong associations were also found for several rare SNPs with OR > 4.0. This study demonstrated the successful imputing procedure offering new signals of better significance. Imputation changed the LD block structure at the locus (58 blocks for BEAGLE and 54 for IMPUTE2 in intron 1) reducing the number of haplotypes around the leader SNP rs9939609 as indicated in the figure below. Therefore, a much better signal might be obtain from the locus by combining the increase in density of SNPs and haplotype mapping. The use of such strategies in comparing for instance, French and Algerian populations, revealed several other signals better than the leader SNP, which is used in almost all publications. Moreover, several SNPs indicated involvement of AKTIP or IRX3 genes upstream and downstream from the locus in intron 1. These signals were different among ethnic populations.



Haplotype mapping of the intron 1 of FTO gene before and after imputing of novel SNPs at higher density.

At the locus of IRS-1 gene (2q36) we completed haplotype mapping of a genomic region, including the Gly972Arg variant (rs1801278) generating by imputing at least 10 SNPs associated. Similarly, at the *IRS-2* locus (Chr 15q36-37) we generated better signals. Finally, TRIB3, SIK1 and ADRA1D genes were investigated by *Italian* (IRCCS, Rome), *Turkish* and *Russian* partners respectively, and revealed interesting associations and ethnic specificity. These studies on locus refining strongly indicate that imputing or genotyping at high density might improve by *fine-scale mapping* candidate loci, thus challenging NGS, except perhaps private mutations. Recent available databases containing reference panels at 10K genomes density would considerable improve fine mapping. Imputing, combined with *haplotyping* of the entire genome are definitively efficient strategies in ameliorating diagnosis of MetS. Our studies on locus refining revealed the absence or poor SNP panels for Mediterranean area, particularly for North Africa, Anatolia and Eastern countries.

Epidemiological studies on immigrants – new collections

New recruitments in immigrant populations in Europe under high-quality epidemiological standards were a central part of the MEDIGENE program (WP5) offering opportunity for new GWAS and specific subprojects. Partners beneficiated from implementation of the new MAGDB database and MEDIPAD software for 24HDR all around the Mediterranean area. Major objectives were to evaluate the prevalence of MetS and its components in local samples of immigrants and perform DNA sampling from 1700 individuals from *Romanian* community in Spain, *Turkish and Algerian* communities in France, *Albanian* communities in Italy, Greece and Kosovo. All collected DNA samples were declared to MESR in France (CODECOH DC-2014-222) or Biobank *Barcelona* (IDIBAPS) following ethical approval. Patients were recruited as "*clinical series*" of MetS defined by IDF criteria, although during investigation we used also the ATP-III definition. Nutritional assessment was determined in 987 patients using *face-to-face* interview with a dietician and performing 24HDR. To help the nutritional investigation, CIQUAL alimentary table was completed up to 2193 food items, including items from *Greece*, *Romania, Turkey, Tunisia and Morocco*. Traditional food and dishes in *Greece* were compiled from *Hellenic Health Foundation* while *Turkish* foods and dishes were included from *Turkish Food Composition Database.* Romanian food items were introduced from « *Gourmet European Foods as cooked by European families* » and calculated with an in-house program. As previously indicated, a completely new database was compiled in CIQUAL comprising the content in *Branched-chain amino acids* (BCAA) in foods. All nutritional data were curated, corrected for *energy intake* and validated and calibrated on 5124 dietary-recalls in France.

Albanians immigrants in Greece, Italy, Kosovo and native Albanians. Albanian immigrants encompass Albanians outside of *Albania, Kosovo, Republic of Macedonia, Montenegro, and Serbia.* The greatest concentrations are found in *Italy, Greece, Germany, Switzerland* and *Turkey*. Albanian exodus has been the largest emigration movement in Europe after World War II involving more than 800,000 persons. A population of 440,000 is located in Greece and 1,616,869 in Kosovo while in Italy there are 800,000 Albanians. The Greek partner at *University of Ioanninia* studied 73 Albanian immigrants for whom 46.3% had MetS. While in native Greeks the prevalence of MetS was estimated at 23.6% with little differences between man and women, in native Albanians in Albania the MetS was estimated at 41.4 % in men and 58.6% in women among 401 individuals studied by the Albanian partner in Tirana. There were major differences in the prevalence components MetS. Thus, *High Blood Pressure* (HBP) was the most representative in Albanian

immigrants in Greece (90.9%) while the low HDL (65%) characterized native Albanians. It may be possible that predominance of male immigration would play a role. Similar features of Albanian immigrants were found in Italy where the Italian partner studied at *Medical Center of the University of Rome* "*Tor Vergata*" 74 Albanians among whom 30 received the full diagnosis of MetS. The prevalence of MetS in native Italians was previously estimated only at 17%, again with significant but small differences between men and women. As in the case of Albanians in Greece, we cannot define causes of such prevalence of MetS in immigrants nor potential genetic or environmental factors. Although we have no clinical data, in another Albanian population immigrated in *Kosovo*, the Russian and Croatian partners studied genetic differences of these immigrants compared to those in Greece and Italy. Variation on autosomal SNPs, uniparental markers on Chr Y and mtDNA indicated clear similarity among Albanians who can be distinguished from other European populations. To understand potential factors in the pathogenesis of MetS nutritional assessment was performed in Albanians in Greece and Italy. In Greece, there were no significant differences in the *total energy intake* (less than 2240 kcal/day) in both controls and patients with MetS. No major differences were found in Albanians immigrated in Italy by *Food Frequency Questionnaires* (FFQ), including *junk food score* or physical activity level by IPAQ.

Romanians and North Africans immigrated in Spain and France. The Spanish partner CIBER investigated Romanians and Moroccans immigrated in Spain. Romanians made up 15.6% of Spain's total foreign population of 4,676,022 people. The reasons of migration are multiple, but after reaching 900.000 people in 2011, Romanian immigrants have been decreasing in Spain. Moroccans are also numerous in Spain and compared to Romanians in 2013 (23,594 arrivals), the Moroccans were 21,338 entries in the same year. CIBER and IDIBAPS Bio-bank recruited immigrants in Spain composed of 44 immigrants born in Morocco and 44 donors born in Romania, who were compared with native from Catalonia in Northeast Spain. Paired by age, Moroccan women had higher *waist circumference* than Catalonia women. Moroccan men had by contrast lower blood pressure (diastolic) than Romanian or Catalonians. Higher level for hyperglycaemia was noted in Moroccan and Romanian immigrant women. Significant differences were also found in TG between Romanians and Catalonians. FFQ revealed difference in the use of white bread, potatoes, commercial juices, meat, butter or cream as well as fried food and tea. These data suggested qualitatively that Romanian immigrants living in Catalonia have *different dietary habits* in comparison to their counterparts living in Romania. Therefore, Spanish partner initiated a new study in of at least 52 participants from Romania comparing them with a peer group of Catalans matched by age, and BMI and categorized by *socio-economic level.* The FFQ indicated difference in number of meals, common *added fats* (*olive oil used in Spain compared to sunflower oils*), consumption of vegetables, rice and pasta, fresh juices, fish. More dairy products were consumed in Spain while Romania was characterized by higher consumption of soft drinks and commercial juices, butter and mayonnaise. Increased use of cooking (frying) was also different. 24HDR were already registered by MEDIPAD in these populations and need further analysis.

The French partner was involved in the study of immigrants in France from Algeria and Turkey. France is a traditional land of immigration from *Algeria, Tunisia and Morocco* and studies are complicated by extreme heterogeneity of immigrants and overlapping between several waves of immigration. Information from native countries such as Tunisia and Algeria suggest major difference in diet characterized by a low carbohydrate intake compensated by a high fat and cholesterol intake, data obtained by the Tunisian partner at

IPT. Still unexplained differences were observed in natives from Morocco, where the Moroccan partner studied 528 unrelated subjects recruited from *Medical Biology Center* in IPM. They observed by 24HDR much lower energy intake in MetS patients with significant differences in protein, carbohydrates and lipids as well as total intake in BCAA. Since the number of immigrants in France was so diverse and still insufficient, French partner oriented genetic study on a more in-depth analysis of G x E interaction. In Algerians recruited in France central obesity was found in 73.3%, HBP in 78.5%, hyperglycemia in 86.7%, high TG levels in 28.6% and low LDL in 73.3%. These rates were different from other immigrants. To complete the study, it was estimated at 288 the additional number subjects to perform a (statistical) powered study with less than 30,000 SNPs in the frame of *France Algerian* gate. French and Algerians (natives) were studied by comparatively GWAS targeting insulin resistance. The study indicated positive hits on Chr 2, 8, 9 and 16, among which FTO was a major determinant. However, different SNP were associated either in intron 1 (*French*) or intron 4 and 8 or IRX-3 (*Algerians*). Other associated genes such ALMS1, SLC22A10 and TOX3 were also different between French and Algerians.

Replication in home countries and individual studies

Numerous genes candidate were studied by partners through their leader SNPs (identified in previous GWAS) in several populations in native countries. The experiments were performed in general by the KaspAr method using specific primers. Tunisian partner made a review of these leader SNPs in MetS. Some of these genes were then studied in more detail in from GWAS data considering a much higher number of SNP per gene. There were differences in the involvement of genes by two strategies. Although it's obviously expected that with more dense SNPs to have more chances to find a culprit gene, what was striking that in many cases the leader SNPs were not confirmed in GWAS. In other terms, other SNP in the same gene locus yielded much higher association. There is no a systematic study to explain why in some studies leader SNPs have reached GWAS significance and not in other. However, these studies appear crucial in understanding statistical and ethnic variability between studies. Fusion studies are running in the opposite direction since they ad more and more individuals in GWAS with proportional increase in statistical significance. Equivalent studies from ethnic population are poor and absent from Mediterranean region including North Africa. Therefore, one important issue was to initiate studies in ethnic populations. They were divided by so called "gates" having as goal the recruitment of sufficient number of samples. To be able to compare MetS in immigrants a series of studies were initiated in original native populations, including *North Africa, Balkans, Anatolia and Russia,* although obviously not all recruited new samples were genotyped, having limiting factor the cost of genotyping. Important conclusions were however drawn from individual studies in native populations.

North Africa: Tunisia and Morocco and Algeria. In *Tunisia*, a number of 594 individuals (295 MetS cases and 299 controls) from various regions were studied together with another 528 unrelated subjects recruited from Morocco. Anthropological studies were performed in these populations on mtDNA and well as Chr Y. Results indicated that in Berber and Arab speakers from Morocco (*Essaouira, Benimellal, Tafraoute and Casablanca*) were characterized by *Eurasian, Sub-Saharan and North African* lineages. Therefore, the diversity was 97.9% within the populations whereas relatively little (2.1%) differences there was within the same ethnic group. Moroccan team found strong associations with MetS for APOA5 gene polymorphisms in 175 MetS cases (against 105

controls). Data were replicated in another group of 149 hypertensive subjects, 122 patients with CV diseases (against 135 controls) and revealed the association with risk of arterial hypertension. In *Algeria*, there was also a high diversity on both uniparental and autosomal markers. Thus, in this part of the world, it would be a complex demographic pattern of migration and isolations of populations. Autochthonous component of North Africa would be represented by U6 on mtDNA and E1b1b1b (or M-81) on Chr Y. Moreover, North Africa would be characterized by extensive genetic drift due to *bottlenecks* and *inbreeding*. Since Berber name is in fact a generic name and there are many Berber dialects (*Kabyle, Chaoui, Zenete, Chleuh or Touareg*), studies of immigrants in Europe need much in-depth ethnic and social information of the origin from the native country. Nevertheless, what was observed in a previous epidemiological study was a *bias between sexes*, noted in almost all studies for MetS and obesity. Thus, women in North Africa would be more obese compared to men, reflecting the same difference in prevalence of MetS. Some anthropological studies suggested a sub-Saharan gene flow in females in Algeria while males would be carriers of the ancestral North African component. This pattern was reproduced in our studies in Algeria, although the there may be biases in recruitment of patients in these studies.

Turkey and Central Asia. In *Turkey* and *Central Asia*, the Russian partner observed almost the same diversity and major differences were noted. Turkish partner noted high prevalence of MetS in accordance with previous studies around 28.8%. In the series of patients studied by the Turkish partner on 247 individuals, patients with MetS had high proportion of low HDL and hyperglycaemia (> 82%) while HBP was found at 71%. Data from France on Turkish immigrants are scarce and did not confirm this feature. Although higher energy intake was measured in MetS in Turkish native compared to controls, this difference was not significant. The study from Anatolia was completed by a series of collections from the Russian partner providing clinical and anthropological data from other Turkic-speaking populations such as *Bashkirs* and *Tatars*. Interestingly, in *Bashkirs* although high TG levels were prevalent (87%), all patients had HBP and les prevalent low HDL. Almost the same profile was found in *Tatars*. Therefore Russian partner performed a comprehensive study comparing genetically individuals from Turkey (*Turks*) and Russia (*Bashkirs, Tatars*) based on autosomal and uniparental markers. PCA revealed differences between *Bashkirs* and *Tatars* on one hand, and *Turks* on another hand. In Turks one can see different sub-clades of haplogroup J-M304 while Bashkirs and Tatars lack this haplogroup. Bashkirs and Tatars also show the increased frequency of R1b-P311.

Metabolic syndrome and PCOS – France and Romania. Another bulk of discoveries were coming from the collaboration between Romanian and French partners interested in pathogenesis of insulin resistance and MetS under conditions of hyperandrogeny in women. It was well known that women with PCOS display more severe insulin resistance and prevalence of MetS, although HBP is not prevalent in contrast to dyslipidemia. The enthusiasm was so high that partners from Greece, Italy in Bologna, France and Romania, including other affiliated partners from South Tunisia collected sufficient samples to perform a complete GWAS in PCOS. Results were straightforward and performed separately in European and North African populations on 1474 DNA samples. The GWAS revealed and confirmed the implications of several genes such as DENND1A, by CYP19A1, TMEM55A, although by far the FTO and INSR were involved. In contrast with recent GWAS in Chinese and USA population of European origin, the association with FSHR/LHR receptors was no so strong. More importantly, the French partner tested hypothesis on a collection of patients with more severe insulin resistance and *Acanthosis*

Nigricans (HAIRAN syndrome) and found strong association of key genes in insulin action. These studies are important because revealed another pathway in the pathogenesis of MetS.

Anthropological markers of male baldness and infertile women. French and Romanian partners had the idea to test a new way in looking at how the advent of SNP in the human genome can be used for medical purposes. Thus, the two groups had the idea to test SNP markers for *premature male baldness* (PMB) in women defined as *forensic markers of male baldness* (99 on chromosome Xq12 and 8 on Chr20p11). Interestingly these markers with some variability were associated with PCOS in women thus unraveling one of the major question in endocrinology for several decades. Differences were found in ethnic populations since several PBM markers were associated only in French and some Chr X markers were not associated in Romanians. These associations are interesting defining perhaps common pathways in pathogenesis of MetS because similar markers were found in MetS in Finish populations. These data represent a milestone in MEDIGENE program being able to demonstrate for the first time potential application of *genetic markers in forensic medicine* or *anthropology* to medical genetics.

AVPR1A receptor gene and hypertention. Another novel hypotheses were launched by *Romanian* and *French* groups concerning involvement of *arginine vasopressin* (AVP) particularly the AVPR1A receptor (Chr12). Because Romanian group detected in its national collection high levels of copeptin, we performed a 12q14.2 locus refining at AVPR1A gene in patients from Romania, France and Turkey (222 MetS). While the association with MetS yielded 140 significant SNPs but not sustained by Bonferroni correction, a much stronger signal was detected with arterial hypertension. Thus, one common SNP was associated with HBP, finding reinforced by other multiple SNPs by imputing from 1KGP reference panel. Surprisingly, there was no association with obesity, TG, HDL, cholesterol levels or glycaemia. A fine-scale *haplotype-mapping* entirely confirmed association and explained the interruption of signal downstream of another SNP involved in *autism* and altered behavior in humans. The pathogenic haplotype was also correlated with copeptin levels in Romanian patients. Although the mechanism is not entirely understood, it is of note that the pathogenic SNP was located in a regulatory region on Chr12. At the population level, is tempting to speculate that our findings correlate with recent observations on increased AVP levels in human and alteration in water intake, already demonstrated to be a risk factor of diabetes in other populations.

Role of branched chain amino-acids (BCAA). *French* partner launched another novel hypothesis of potential role of *Branched-chain amino acids* (BCAA) in insulin resistance. Indeed, *leucine* (Leu), *isoleucine* (Ile) and *valine* (Val) are essential amino acids relatively abundant in human diet constituting 15-25 % of the total protein intake. Most of BCAA metabolism occurs in mitochondria and their catabolism produces three ketone bodies, which oxidized would produce in turn three CoA derivatives used in *tricarboxylic acid* (TCA) cycle. Dysregulated BCAA metabolism may have a negative impact on the *anaplerotic* flux into the TCA cycle. Since as shown by Newgard et al. 2009 that it would be a link between BCAA oxidation products including C3 and C5 *acylcanitines*, we consider that it may be a link between byproduct of isoleucine or valine catabolism and insulin resistance. We approached this subject of nutrigenomics by screening firstly all SNPs and their proxies corresponding to 52 genes in BCAA metabolism and then another several thousands SNPs for a dozen most significant genes in the French collection of obesity and MetS. The ACSF3 (*acyl-CoA synthetase family member 3)* was the most significant gene

and associated with HOMA index and degree of obesity. We also studied imputed SNPs in key candidate genes and significant associations were found with MetS. Since these genetic findings suggested potential correlation with clinical parameters including BCAA intake and/or BCAA plasma levels, we further considered nutritional assessments on Albanians immigrants in Greece, Romanians in Spain and native Turkish, whose data were introduced in MEDIPAD. There was a good correlation between HOMA index and BCAA intake, although after correction for energy intake significant differences were found only between extreme quartiles of intake. While the significance of these findings at the population level is not yet known, particularly in the context of high diversity of populations, one of the major outcome of MEDIGENE program is that nutrigenomics field as a new way to understand how researchers can put together knowledge in anthropology, genomics, archeology and medicine.

Data Integration

To integrate tremendous amount of data generated during the program and to understand Gene x Environment (G x E) interaction we have approached *nutrigenomics* field in a holistic manner. The G x E interaction remains underappreciated due to the lack of data in geographically specific populations and still insufficient collaboration between anthropology, ecology as well as sociocultural fields. Evidence has been accumulated for extreme diversity of components of MetS in geographically dispersed populations. Classical examples are for instance variation in the body mass index (BMI) in Asian populations compared to Europeans. Different pathways of MetS exist also in Canadian Amerindians for instance. Therefore, it became more than evident that cut-off values established in Europeans are highly variable to which other environmental factors interact such as diet, body shape, intrauterine milieu, infantile development, all being variable between geographical groups of humans. Moreover, even at the gene level, there are instances in which various genes, for instance in Mexican Americans involving FAT or CD36 and GNAT3 genes are different from Han Chinese where ADIPOQ defect is predominant or Easter European involving HSD1181 genes. To date, among Saudi Arabians, more significant associations were found with eNOS explaining the higher prevalence of MetS. In face of the great diversity of susceptibility genes in geographically different populations and ethnic or ancestral components, we have approach the G x E with the goal to screen the contribution of specific group of genes. We have used databases containing CardioGxE catalog of genes for cardiometabolic traits and selected among 14,957 genes all positive interaction in European populations as function of relation with diet, energy metabolism and Mediterranean diet. We end up to 30 genes in energy consumption, 84 genes for fat intake, 71 genes for *polyunsaturated fatty acids* (PUFA), 18 genes for *mono-unsaturated fatty acids* (MUFA), 25 genes involved in *saturated fatty acids* interaction (SFA), 12 genes in regulation of protein and 25 genes in carbohydrate intake interaction. A number of 91 genes were specific for interaction with Mediterranean diet, including fish oil. Groups of genes were tested in MEDIGENE collections for MetS considering the overlapping with MetS gene in HUGE database (92 genes). Among 116 genes involved in G x E interaction preselected only 16 genes were common with previous results in GWAS of MetS containing 12,427 SNPs of which 3924 were investigated by MEDISCOPE. Strong signals were obtained for TCF7L2 gene for instance confirming previous observation on insulin sensitivity in interaction with dietary fat intake and glucose tolerance. Investigation of the same genes but at a much higher density after imputing at 1KGP, revealed influential role of FTO, IRX3, C2orf43, GCK, LIPC, ADIPOQ, LPL, TCF7L2, APOA1, LEPR on HDL levels. In association, some but not all of these genes were associated with MetS. It was also

interesting to observe that among those genes not included as candidates in previous GWAS for MetS, one of the best associated gene was *Lactase gene* (LCT) by its well known rs4988235 (C/T). Interestingly, we found that LCT gene was strongly associated several components of MetS in nominal association or correlation trend test. Since in human populations the enzymatic digestion of lactose from milk and diary products is expressed in almost all adults of European origin (*Lactose Persistence*) and this genotype is also associated with *obesity* and *fat mass* and MetS, it is of highest importance to understand how North-South gradient in the LCT genes variation may be used in understanding of diversity in Mediterranean populations. The mutation in LCT very likely occurred in Neolithic times in Balkans and Central Europe while in Roman times, it was considered that people from the North are able to better digest unprocessed milk compared to South of Europe.

Along this line, we also tested genes involved in Mediterranean diet known to contribute by a protective effect on obesity and MetS. We have tested in a first step leader SNPs and then 5000 SNPs in in 54 such genes and defined a prioritized list of genes. From all these studies it appears that population diversity more or less mapped on geographical dispersion or movements such as immigration considerably complicates understanding of how genes and environment interact and how we can use this information to ameliorate the medical diagnosis. Partners in Italy (Rome) at IRCCS or Tunisians at IPT are currently investigating more subtle statistical approaches and *meta-analysis* to integrate diverse information from the fields of anthropo-genetics, medical genetics and nutrigenomics. We concluded from MEDIGENE studies that the faith of human populations, including epidemics of chronic diseases and probably life span is pending on complex adaptations to an ecological niche in which population movements, including immigration from one continent to another by social or economic reasons should be carefully investigated at a multidisciplinary level.

Potential impact: socio – economic impact, wider societal implications and main dissemination and exploitation of results

Potential impact. Immigration is a major concern in Europe with medical and societal implications. Immigrants who by themselves designate another virtual continent form more than 3% of the world population. Immigrants in Europe (41 millions) are vulnerable populations raising various healthcare problems. The study of susceptibility for complex disorders such as *diabetes, obesity or metabolic syndrome* is a complicated issue in actual epidemiology since immigrants are carrying genetic susceptibility from their home country interfering with statistical estimation of the overall morbidity and mortality in a geographic territory. Investigators are faced in our days with complex situations to which one should add the potential collision between the genetic predisposition of immigrants and Westernized way of life.

One major impact of the MEDIGENE program was to promote an interdisciplinary research and to *transversally link European laboratories* through an original consortium with complimentary competences in *anthropology, molecular biology, clinical genetics, endocrinology and metabolism*. Very rapidly, from the first year, the annual meetings of MEDIGENE became a forum for multiple discussions integrating historical, anthropological and archaeological data with molecular genetic and endocrine knowledge. The audience was

large and progressively researchers and clinicians developed new concepts offering a novel vision in understanding the pathogenesis of obesity and insulin resistance.

An incredible amount of information is circulating in the scientific milieu in anthropology and despite this, there is very little connection with the clinical practice. Now clinicians in the field of metabolism and endocrinology through MEDIGENE program have another perspective and a complete new vision at the population epidemiological level. Thus, a new *interdisciplinary knowledge* was generated and epidemiologists became aware of the importance of ecological data, climate changes, migrations and admixture of populations in analyzing complex disorders.

The impact was seen in the high quality recruitment of epidemiological data in countries, which was favored by the new database MAGDB. MEDIGENE was able to establish new standards and guidelines for sampling (DNA, blood and tissues) in population to obtain high-quality collections with significance for anthropologists and with complete metabolic and endocrine profiles. Concurrent information was obtained regarding ethnicity, dialect, birthplace, geo-localization (*latitude and longitude*) and parameters of insulin resistance, panel of hormones, 24 h records for diet and considering other epidemiological parameters to which physicians are accustomed. The database is now of extreme value and unique and will represent a continuous source of research in the future.

Along this line, another important impact was created by the development of the software MEDIPAD by a SME in France during the MEDIGENE program. The new software allows imputing of clinical, nutritional and genetic data directly through the web from an iPAD. Face-to-face interview for 24 h dietary records can be also registered through MEDIPAD by the dietician or nutritionist and send data to common server in France from Mediterranean countries. The program was written in different languages (French, English, Romanian and Spanish) and French CIQUAL table of foods items was completed with foods from other countries, including *North Africa, Greece, Turkey, Romania and Spain*, favoring investigation in a unifying way in Mediterranean countries. The impact was felt by opening a dialog between researchers in the field of *ecology, nutrition and genetics*. Accurate measurement of exposure to environmental factors, including nutritional factors will increase the competitiveness of EU research since such type of data are in general missing in genetic studies focused essentially in clinical profiles.

Another challenging aspect, which generated impact in scientific knowledge was the interaction between genes and environment, particularly dietary factors. Although the subject is treated in the literature, MEDIGENE program asked specific questions of how a gene defect leads to pathogenic situations - ineluctably alone - or in interaction with other factors. It is generally accepted that *low physical activity*, *high caloric intake* and *fat diets* parallel the burden of obesity and metabolic syndrome. However, the real question is how these factors interact to the susceptibility genes. Are these interactions constant through time? For instance, ancestral hunters-gatherers in Europe were colonized by gluten-tolerant Neolithic farmers adapted to consume high carbohydrate diets (*cereals*). Lack of this adaptation may lead to diabetes, food allergies, and Celiac or Crohn diseases. Along this line, Indo-Europeans populations colonized aboriginal populations being able to consume milk because of the *lactase gene* (LCT) mutation. These are new aspects that should be integrated in actual epidemiological studies since one can wonder to what extent similar defects would compose susceptibility for insulin resistance in relation to nutritional habits. Therefore, the findings in MEDIGENE will have larger consequences in understanding the

way of life of different people in native countries, to respect their way of life and appropriately to help research in common Public Health concerns.

The Mediterranean regions have been studied for a long time but very often immigration was regarded with a paternalist attitude in European countries – perhaps linked to low-income resources of immigrants or less engagement in preventive *healthcare services*. Examination of several epidemiological paradoxes – such as North Africans in France – strongly suggest that a clear picture of effect of immigration on Health cannot be understood without collaborative research between host and home countries. Collaborative studies between Algeria and France for instance of or between Spain and Romania were very good experiences. The added value was that health problems of immigrants that can not be solved without shared efforts of home and host countries.

Study of Mediterranean countries is an important concern and we considered abnormal that a great bulk of genetic results in these populations is coming from USA. Thus, by common efforts, MEDIGENE program *responded to the EU aim of widening research by integrating partners from the Mediterranean region in the field of diabetes, obesity and cardiovascular complications together with genetic analysis.* Population-based studies in immigrants and comparison with host populations will help to understand how diet, habits or *acculturalisation* contribute to the prevalence of metabolic syndrome and its components. These aspects made the object of disseminated information in various cultural associations and embassies as for instance in Greece and Italy for Albanian immigrants or similar organization in Spain for Romanian immigrants. Impact was also felt in the collaboration with North African countries. Although everybody agrees that nutritional transition is a characteristic of these countries, there is a tendency in Algeria and Tunisia to reproduce findings in European countries. Studies in MEDIGENE impacted in this area and concluded that North African populations are definitively more complex, with a much higher diversity than expected. Examples were multiple as for instance differences between man and women for obesity and metabolic syndrome or regional differences between North and South Tunisia. Definitively, inbreeding in North African population hamper genetic studies of association at a population level. Although these problems were not all solved by MEDIGENE, common efforts were visible from investigators in *Algeria, Tunisia and Morocco*. The specificity of genetic markers in North Africa will continue to impact future studies stimulating research in the field of bioinformatics by creation new reference panels for genetic research. Similar impact was felt in Turkey and Central Asia where the Russian partner initiated comparative studies between Turkic-speaking populations from Caucasus and actual Turkey.

At a more fundamental level, MEDIGENE impacted on the fate of GWAS strategies in complex disorders. For some investigators, GWAS are strategically dead with the belief that information was already obtained in major complex disorders and influential genes are considered as “known genes”. From MEDIGENE results, the conclusion is at the opposite, that GWAS actually started to reveal how various genes became pathogenic in an ethnic population. Major differences found among populations suggest that although genes are known, the pathogenic mechanisms are pending of how these genes defects combine. The study of insulin resistance and the diversity of genes defects in European versus North African populations for instance clearly demonstrated that GWAS enter in a new era. Several examples were offered by MEDIGENE as for instance the variation in the FTO gene between French and Algerian populations or the role of ADRA1 receptor between Turkic-speaking populations, all indicating a high diversity of gene effects in ethnic populations.

Unfortunately, Mediterranean area suffers from lack of standard SNP panels compared to UK or USA databases in the 1kGP or even 10kGP. There is therefore, an urgent need for collaborations and convergent efforts to offer to Mediterranean countries the same level of knowledge and powerful databases. MEDIGENE impacted on this need and engaged investigators in novel strategies in epidemiology (e.g. *genetic binning*) taking the advantage on progresses in molecular biology and bioinformatics.

Main dissemination activities. The project developed a dissemination plan at every stage including partners responsible for specific tasks in their countries. The project dissemination activities included: conferences, teleconferences, meetings, workshops, letters of intent, emails, articles, posters, creation of the MEDIGENE website to reach the largest number of professionals and lay audience, to outline the project aims and to enhance public awareness on the problem of immigration.

Specific activities were performed in the goal to harmonize ethical protocols among countries, anonymization procedures, explanation how studies would benefit for individuals, social implication of potential discoveries and health gain in family members. In working with immigrants investigators should be careful in social and political implications when sampling for instance from minorities and vulnerable populations. The consortium included in *Advisory Board* one expert in ethical aspects –Anne CAMBON-THOMSEN and together with other members several activities were developed with the goal to increase awareness on ethical issues in migrant populations, to detect problems in countries and define high moral attitude in face of the complex problem of immigration. At least three meetings were of particular importance:

Anne Cambon-Thomsen. Ethical aspects of genetic research in human populations and biobanks. Annual MEDIGENE meeting, 27th February-1st March, 2014, Kenzi Tower Hotel, Casablanca, Morocco

Sonia Abdelhac (IPT, Tunisia). Ethical problems solved in Maghrebin countries. Annual MEDIGENE meeting, 19th-20th April, 2013, POIANA BRASOV, Romania

Pr Bouamrane (Algeria). Regard de la religion musulmane: avis du Haut Comité Islamique. Colloque International - Apport des nouvelles technologies dans la compéhension des maladies génétiques. 17-18 May, 2014, Algiers, Algeria

Particular efforts were spent in training activities in the field of Bioinformatics and nutrition. To familiarize scientists with large data sets we organized a training course 20-21st November 2012 in Barcelona (Spain) on "*Data management and analysis for GWAS and locus refining using Next Generation Sequencing for MEDIGENE partners*". The course was a real success, participants being able to communicate through the web interface under condition offered by CIBER. Since MEDIGENE cooperated with three other FP7 SICA projects (RODAM, EPI-MIGRANT and GIFTS), RODAM representatives joined this training course. A meeting in London on 24th May 2013 at Imperial College organized the cooperation with these partners in EU programs.

In the field of nutrition, a workshop was organized in Barcelona, Spain, December 15th 2013, with CIBER, IDIBAPS, UNIBO and INTACTILE as associates. The sessions yielded a nutrition working-group having as referent Elisa Marcato (UNIBO, Italy). The implementation of MEDIPAD software was also the object of several training activities at each Annual Meeting organized by Intactile Design from France.

In addition to these large events, we developed bilateral exchanges with partners to transfer key technologies between participating laboratories. Eight participants from Tunisia (S. Elouej and I. Rejeb), Italy (C. De Lorenzo), Romania (M. Vintila and S. Muraru), Algeria (L. Doubia and M. Aouadi), or Turkey (Y. Tutuncu) were already trained for novel technologies in the French laboratory. Experimented scientists did also some stays in partner's laboratories as S. Litvinov (analysis processes, Montpellier), R. Kefi (ancient DNA, Barcelona and nutritional survey Bologna), F. Grigorescu (genetics, Romania and Croatia) and R. Attaoua (bioinformatics, Alger). Some administratives also took opportunities to come to Montpellier to develop procedures for management of collaborative projects (A. Benyacoub, Montpellier).

On scientific level, MEDIGENE consortium members published and communicate at the most important congresses in the field of Endocrinology, Genetics and Nutrition (see next section). Major impact was felt in *American Diabetes Association* with several published abstracts or EASD meeting (oral and poster presentations).

The coordination team in France, launched the project on 21st January 2012 in Montpellier by the *Kick off meeting* with 59 participants under the auspices of President of UM, event notified in "Lettre d'information de l'Université Montpellier 1, SYNERGIE, Nr 29, 30 January 2012. The project was presented in Brussels in the "*DIABESITY - A World-Wide Challenge: Towards a global initiative on gene-environment interactions in diabetes/obesity in specific populations* " under the auspices of the European Commission, 9-10 February 2012 (http://ec.europa.eu/research/health/events-12_en.html). Romanian Academy organized a meeting "*New genetic approaches in understanding susceptibility for metabolic syndrome in immigrant populations around Mediterranean area*" with the occasion of launching the program in Bucharest, on 3rd February 2012 (www.adsm.ro/anunturi-importante.html). The coordinator (*Florin Grigorescu*) was received as member at Romanian Academy (3 February 2012) and a lay press article notified the VIATA MEDICALA, Nr 3, January 2013. Similar strong relationship was also established between French (FG) and Spanish (RG) and Romanian (MC) partners on the field o nutrition and endocrinology.

Another important communication story of MEDIGENE was the creation through the project of an ancient (a)DNA biobank by ICAC and CIBER. It was relayed by numerous national TV and press in Spain (see next section).

Exploitation of results. MEDIGENE program generated a tremendous amount of data during the 4 years. Besides already published articles, it is expected that major contributions will be published in international high IP journals. Main contributions will include the GWAS of metabolic syndrome, epidemiological studies on immigrants in Europe, locus refining on gene candidates (FTO, INSR, DUSP9, APOA5, AVPR1A), influential genes in MetS in hyperandrogenic women, role of BCAA in diet and gene variability, nutritional studies in immigrant population in Spain, among others. These manuscripts are in preparation and will contain original data of high value. A special work on database protection has been done with *Technology Transfer Officers* of the Montpellier University. This database is the support to customized Affymetrix chip. We also have a special non-disclosure agreement with Affymetrix on the MEDISCOPE to ensure further exploitation of results. We also have contacts with the regional valorisation structure (*Société d'Accélération de Transfert de Technologies* AXLR), to continue to develop the nutrition

software with MEDIPAD and the associated database. For the moment two options are pending, first to present with INTACTILE a FEDER project or the creation of a *spin-off company* with the help of the SATT.

Address of the project website and relevant contact details

PUBLIC WEBSITE : www.medigene-fp7.eu

Institution	Country	Partner	Contact
UM-1 Coordinator Manager	France	Florin GRIGORESCU Christophe NORMAND	florin.grigorescu@inserm.fr christophe.normand@inserm.fr
UCBL	France	Michel PUGEAT	michel.pugeat@chu-lyon.fr
CIBER	Spain	Ramon GOMIS	ramon.gomis@idibaps.org
ICAC	Spain	Josep Maria MACIAS	jmmacias@icac.cat
UMF CDU	Romania	Catalina POIANA	events@bbtr.ro
UNIBO	Italy	Renato PASQUALI	renato.pasquali@unibo.it
UNICZ	Italy	Giorgio SESTI	sesti@unicz.it
IRCCS-CSS	Italy	Vincenzo TRISCHITTA	vincenzo.trischitta@uniroma1.it
TOR VERGATA	Italy	Davide LAURO	d.lauro@med.uniroma2.it
UOI	Greece	Agathocles TSATSOULIS	atsatsou@uoi.gr
IBGU	Russia	Elza KHUSNUTDINOVA	elzakh@mail.ru
IPT	Tunisia	Sonia ABDELHAK	sonia.abdelhak@pasteur.rns.tn
UA	Algeria	Akila ZENATI	akilaze@gmail.com
IPM	Morocco	Abdelhamid BARAKAT	hamid.barakat@pasteur.ma
TU	Albania	Ylli AGRON	ylliagron@yahoo.com
BC PLATFORM	Finland	Timo KANNINEN	timo.kanninen@bcplatforms.com
IU	Turkey	Ilhan SATMAN	satmandiabet@gmail.com
INTACTILE DESIGN SA	France	Yves RINATO	rinato.y@intactile.com
INANTRO	Croatia	Sasa MISSONI	sasa.missoni@inantro.hr