Short Comunications

Studies of genetic epidemiology about phenotype with complex determination in the context of a biomedical basic sciences doctoral program

Manuela Herrera-Martínez[1*]

Douglas Fernández-Caraballo[2]

Danay Heredia-Ruiz[2]

Ma Elena de la Torre-Santos[3]

Noel Taboada-Lugo[4]

Lay Salazar-Torres[5]

Lorna González-Herrera[6]

[1]Universidad de Ciencias Médicas de Villa Clara, Laboratorio epidemiología genética.

[2]Universidad de Ciencias Médicas de Villa Clara, Laboratorio Stress Oxidativo, UNIB

[3]Laboratorio citogenética, CPGM VC

[4]Servicio Genética Clínica, CPGM VC

[5]Policlínico Chiqui Gómez Lubian Santa Clara

[6]Dpto. Ecocardiografía prenatal, CPGM VC

[*]Author for correspondence. E-mail: manuelahm@infomed.sld.cu

ABSTRACT

Objective: To identify a core set of research procedures for genetic epidemiology that need to be learnt to obtain a scientific degree in subjects related to basic biomedical sciences.

Methods: We discuss the results of six research projects in the area of genetic epidemiology about complex phenotypes. We show in the studied phenotypes the higher prevalence in relatives compared to the general population, the higher risk among first-degree relatives of affected individuals, and the segregation mode of each phenotype. Meanwhile, to show the influence of environmental factors and various analysis of gene-environment interaction using different models. Finally, we discuss the common research procedures followed during these studies.

Results: research studies in the subject of genetic epidemiology can be divided into three broad categories: (1) those that aim at describing the distribution of a disease or a determinant at the level of a population of interest (familiar hereditary angioedema, reproductive failure); (2) those that attempt to investigate a potential aetiological link between one or more specific determinants and a disease of interest (congenital birth defects, cervix cancer, and longevity) and (3) those aimed at evaluating the effectiveness of an intervention applied to groups of individuals in the population (hypoplasia foetal thymus).

Conclusions: We have identified, a set of theoretical topics as well as study techniques and methodologies for the approach of research projects in genetic epidemiology, could be considered as the general and specific bases for postgraduate teaching in relation to obtaining a doctorate in basic biomedical sciences in our University and could be useful to other Universities of similar contexts.

Keywords: genetic epidemiology, phenotype complex, complex trait genetics, biomedical basic sciences, doctoral program, gene-environment interactions.

INTRODUCTION

Genetic epidemiology was defined by Morton in 1982 as " *a science that deals with the aetiology, distribution and control of disease in groups of relatives; and with inherited causes of disease in populations* ". This is field in which the number of research publications is growing. At the present, a substantial portion of gene expression variations can be explained by both local (cis) and distal (trans) genetic variations. Much progress has been made in uncovering cis-acting expression quantitative trait loci (cis-eQTL), but trans-eQTL have been more difficult to identify and replicate. Nevertheless, there have been considerable advances in the prediction of trans-acting genes and their targets. For instance, in 2019, the Genotype-Tissue Expression Project found evidence of association in 2,356 trans-acting/target gene pairs with high mappability scores. Interestingly, trans-acting genes are more significantly associated with selected complex traits and diseases than target or background genes. This shows how complex is the genetic determination of these diseases and their hereditary nature in a population; and as consequence, the interest of studying such a subject in the doctoral program [1].

METHODS

In this paper, the results of six research projects, which are conducted in the context of the doctoral program of the Medical University *of Villa Clara* (UCMVC), in the area of genetic epidemiology.

Samples

As part of these research projects, we have used so far the following samples:

1. Familiar hereditary angioedema (FHA): 54 cases (for a total of 7 families).
2. Reproductive failures: 364 of different root causes, and 22 new cases (non heredity) of aneuploidies in consecutive reproductive failures.
3. NTD: 36 cases and 72 control individuals.
4. Cervix lesions: 48 cases with LIEBG, 45 cases with LIEAG, 51 cases with epidermoid carcinoma cérvix, and 30 control individuals.
5. A total of 33 cases of longevity (at least nonagenarians), 33 cases of long-lived siblings and 33 intrafamilial controls (spouses or any other person living in the same home, but with no consanguineous link).
6. Fetal thymic hypoplasia: 221 pregnant women with high genetic risk.

To show that the six studied phenotypes have genetic basis, we analyse the existence of: higher prevalence in relatives compared to the general population, higher risk among first-degree relatives of affected individuals, segregation mode of phenotypes, chromosome variants and echography biomarkers. Meanwhile, to show the influence of environmental factors in the pathogenesis of such disorders, we applied interviews with questions about specific risk factors, and conducted analysis of gene-environment interactions with different models to find the type of interaction suggested.

The risk of developing one of the different traits studied as part of our doctoral program are basically linked to genetic influences, environmental conditions or both. In each disorder, the environmental risk differs depending on the nature of the defect. In the case of genetic risks, we always evaluate the positive antecedents in first and second grade relatives (family aggregation) as well as the consanguinity.

The general evaluation of risks due to genetic and environmental interactions was considered in the cases where at least one environmental and one genetic risk factors were present with independence of the risk factor. The risk for specific environmental risks, specific genetic risks and combinations of them is evaluated using the rules shown in the next table.

Table 1. Prevalence of chromosomal aberrations and polymorphisms between reproductive failures

	Cases	CA	PPS	Polimorph	PPS	2 Polimorph	CA+Polimorph
Recurrence Spontaneous miscarriages	170	3	1.8	24	14.1	1	0
Children not live previous with malformation	76	0	0	13	17.1	1	0
Together 1 and 2	37	2	5.4	5	13.5	0	0
Male Infertility	38	5	13.2	5	13.2	2	0
Female Infertility	43	13	30.2	2	4.6	1	1
Total	364	23	6.3	49	13.5	5	1

CA: Chromosal aberration
PPS: Proportion positive of studies

Table 2. Significant risk factors in women with premalignant lesions or cervical cancer with previous HPV infection

Risk Factors	Type of Cervix Lesions					
	Risk Ratio LIEBG vs Control		Risk Ratio LIEAG vs Control		Risk Ratio Epiderm carcinoma vs Control	
Antec CCU +	1.60(1.21-2.12)	*	1.73 (1.31 -2.30)	**	1.66 (1.29 -2.14)	**
Antecedent another gynaecological cancer	1.41 (0.99 -1.98)	ns	1.54 (1.12 -2.13)	*	1.32 (0.92 -1.90)	ns
Smoker	1.70 (1.20- 2.40)	**	1.66 (1.16- 2.36)	**	3.51 (1.91- 6.45)	**
DIU	1.93 (1.12 -3.31)	**	1.88 (1.09 -3.28)	**	2.34 (1.25 -4.40)	**
1ra R sexual<18 ys	2.09 (1.30 -3.35)	**	1.75 (1.15- 2.67)	**	1.75 (1.18 -2.60)	**
More than 3 sexual partners	1.78 (1.23 -2.58)	**	1.62 (1.13- 2.33)	*	1.43 (1.04 -1.97)	*
Unprotected sexual relationship	0.89 (0.62 -1.29)	ns	1.07 (0.74-1.55)	ns	1.63 (1.05- 2.52)	*
Environmental toxicity	0.98 (0.68 -1.42)	ns	1.14 (0.79 -1.64)	ns	1.44 (1.03 -2.05)	*
Genetic and Environmental	1.60 (1.21- 2.12)	*	1.63 (1.20 -2.19)	*	1.61 (1.24 -2.08)	*
Total	48		45		51	

On the other hand, we use the following rules to explore two different models of G-E interaction:

Interaction with additive model: RPge = RPg+RPe

Interaction with multiplicative model: RPge = RPgxRPe

Using the aforementioned rules we obtain the expected risk of interaction that we can then compare to the observed risk of interaction. The statistical analysis for significance of the association were made by means of the chi-squared test, obtaining the Odd ratio, and calculating the risk ratio and its confidence interval at 95 percent, significant differences were considered for 5 and 1%.

RESULTS

From the point of view of the teaching-learning process of research techniques, different aspects that have been taken into account in the training of professionals in the line of research of genetic epidemiology of the doctoral program in basic biomedical sciences from the UCMVC are summarized.

Research studies in the subject of genetic epidemiology can be divided into three broad categories. To put emphasis in those categories, we now present the results of the six research projects split in three parts.

Area 1) Studies that aim at describing the distribution of a disease (phenotype), or a determinant (risk factor) at the level of a population of interest.

In this topic, we strive to correctly establish the prevalence of each disorder. We started from the conceptual basis and practice that to make a study of the prevalence of a genetic disorder, or of a certain phenotype, we must take into account the clinical characteristics of the disease, the diagnostic criteria, whether or not they are certain, if they are of clinical or laboratory basis, their mode of inheritance, age of onset of symptoms, stage of life in which it may manifest itself, and whether or not the base prevalence is known.

In addition, it should be emphasized that prevalence study may not aim at finding the prevalence of the genetic disorder but of its genetic or environmental determinants. This type of study was also explored in the doctoral program projects.

Case of study 1: Familiar hereditary angioedema (FHA)

Situation of study: Establishment of prevalence of FHA in Villa Clara province.

Strategy: Family with Population Base

Design: Prevalence studies in population and families at risk.

Practical training of postgraduate students: The method used in this case was to establish first a clinic, genetic and immune epidemiology register of family at risk. To do so, the criteria had to be established to put the register into operation, since in this case it was about finding the prevalence in the territory for the first time. The registry of the 7 large affected families, with a total of 54

patients, of them 49 alive and 5 deceased (mortality of 9.3%), established a prevalence of 1 every 15 370 individuals of the province.

The immunological studies establish a relative frequency of FHA type I, due to a quantitative deficiency of C1 inhibitor, of 57%; and for type II, due to qualitative deficiency, a relative frequency of 43%.

In affected families there was an average of 8 patients per family, establishing many other data of the characterization of these genealogies, of clinical, genetic and epidemiological interest on this disorder, which allowed the scientific group to make scientific decisions, among them the continuation of the project with a current study to evaluate the possible existence or not of a founding effect.

This allowed the doctoral program to make an incursion in the area of Human Population Genetics, which is one of the edges of research in genetic epidemiology, derived from studies in area 1 of genetic epidemiology.

Case of study 2. Reproductive failure.

Situation of study 1: Environmental Determinant in Reproductive failure

Strategies: Population.

Design: *Case-only design, nested in a longitudinal cohort*

Practical training of postgraduate students: The method used in the first situation from case of study 2, was aimed at finding prevalence of potential environmental determinants of aneuploidy in a study conducted among 22 novo aneuploidies patients in reproductive failure consecutive occurred in 6 years. The most prevalent ones were excessive ingestion of coffee (81.8%), exposure to chemicals (63.6%), use of medications (40.9%), alcohol (22,1%), Cigarette (18.2%) and to a lesser extent the Ionization radiation, Viral Infection, among others.

Situation of study 2: Genetic Determinant in Reproductive failure

Design: Sampling studies design in unrelated individuals.

Strategy: Population.

The second study was also to evaluate the prevalence of determinants, but in this case of genetic type, among the reproductive failures. The results found in summary form are shown in table 1.

Case of Study 3: Congenital birth defects with probable maternal folate deficiency (Shown example NTD)

Strategy: Population

Situation of study 1:

Design: Determination of the adjusted prevalence at birth.

This study focuses on the practical training of postgraduate students on how to obtain the adjusted prevalence at birth. This is an important aspect because the calculation of the prevalence of

congenital defects sometimes requires a specific analysis. This is because, when the disorder has a prenatal diagnosis, the adjusted prevalence criterion must be used for the calculation. Hence, this prevalence must include, in the numerator of the usual quotient, the cases in which it was possible the positive prenatal diagnosis (and therefore are unborn cases since the interruption of pregnancy was performed due to genetic causes). For obvious reasons, live-born should not only be included, since this would imply a falsification of the real prevalence that has been presented at conception in a specific place and time.

Situation of study 2:

Design: Determination of the conglomerates of high prevalence

In addition, the issue of the prevalence of birth defects allowed to perform training for the determination of pure space conglomerates, pure temporal and temporal space, through the Kulldorff method to find high prevalence clusters, technique associated with spatial statistics, very useful in genetic epidemiology.

When applying this method to the data of the province of Villa Clara we observed a temporary conglomerate of DTN in 2013; spatial conglomerates in the municipalities of *Sagua la Grande*, *Ranchuelo* and *Placetas*; and a spatio-temporal conglomerate that year for the municipalities of *Placetas* and Remedios (according to the analysis of historical trend, what was expected for that year was a case and there were seven, for an observed / expected index of 7, RR = 5, 92 (p = 0.048). Figure 1 shows the space-time conglomerate of 2013.

Area 2) Studies to investigate the relationship between one or more potential etiological agents or specific determinants and the disease of interest.

The doctoral program includes projects where the phenotypes of interest are: congenital defects, premalignant lesions / cancer of the cervix, and longevity.

Focusing on these research topics, several working sessions were held in the research group to discuss about: types of research designs of genetic epidemiology of causality, association studies of population and family, univariate and multivariate association studies of both genetic factors such as environmental, gene-gene or gene-environment interaction studies, among others.

The main results are summarized below, sorting according to the case study in question.

Case of study 1: Congenital birth defects with probable maternal deficiency of folate studies. Shown example: NTD (Neural Tube Defects)

Strategy: Population

Design: Studies of association in design longitudinal cohort, 'nested' case-control study

Practical training of postgraduate students: The method used In this doctoral project, which objectives to be consolidated towards study area 2 of genetic epidemiology were aimed at calculating the risk ratio in a study that included all congenital defects that are considered

associated with maternal folate deficiency. In addition, a significant number of genetic and environmental risk factors were evaluated, with sufficient consistency and biological credibility. In this article we are only communicating, for illustrative purposes, the findings of risks that were significantly associated with the appearance of NTD in the pregnant women included.

Among them: Positive Genetics precedent RR = 6.7 (2.14-21.01) p = .001, Mother with DM type 2, RR = 28.448 (1.655 -43.118) p = .010, Smoke RR = 8.846 (2.255-34.7)), p = .001, Hypertermy RR = 23.667 (2.861-195.8), p = .000 and Precedent Genetics and Environmental together RR = 19.783 (4.151-94.28), p = .000, which resulted in a significant GE interaction for the additive model.

The latter served as the basis for testing, during practical training, the different G-E interaction models that have been described. Among all of them, the one that turned out to have a significant G-E interaction, was the one that was tested under the requirements of an additive model.

Case of study 2: Pre-malign lesions (LIEBG, LIEAG) and epidermoid carcinoma of cervix in women with previous infection by human papillomavirus (HPV).

Strategy: Population.

Design: Associations studies in design in unrelated individuals control-case.

Practical training of postgraduate students: In this project, we focus on how to study the possible links between determinants of health and disease. In this case, the peculiarity of the process studied, on which we insisted, is that these determinants are evaluated in different stages of the disease, using cancer as a case study to show the existence of multiple-step diseases. This help us to consolidate the concepts related to genetic epidemiology as both an area of knowledge and a methodology for scientific research. Thus, genetic epidemiology can provide evidence that in many human diseases, as the disease progresses through its clinical stages, we can expect a strongest association between the disease progress and environmental and genetic factors determinants.

Some results that show the aforementioned influence appear, as a summary, in Table 2. Only those determinants with significant associations to all or to some of the stages of the disease are shown.

Control: 30

Case of study 3: Longevity.

Strategies: Family.

Design: studies of associations in design case studies - sick siblings.

Practical training of postgraduate students: This last case of study constitutes the third doctoral project that deals with topics of study area 2 of genetic epidemiology. This is the only research of the doctoral program that is based on a family strategy design. These designs are very important nowadays because they allow to find significant associations between traits and genetic or

environmental factors. This type of study is also interesting because to carry them out you must first find informative families for the phenotype that is going to be studied. Therefore, the project served to strengthen aspects related to research strategies in Genetics aimed at finding families with significant family aggregation.

In this investigation, the case and the control are people who live together, but they are not consanguineous relatives (called cohabitants). Whenever the spouse was alive, this individual was the control employed, otherwise daughters-in-law or sons-in-law were used, which gave homogeneity to the study.

Family design implies that people are studied within the family; that is, with a close relationship of kinship - in this case of first degree. Since we studied brother-brother cases, the basic design was case-sick siblings, with which the possibility of consolidating these designs in the scientific collective was introduced.

Case studies - sick siblings and other combinations of sick relatives, have been appearing in the literature on the subject. In this project, a variation of said designs is carried out. In addition to the common model, we use a non-consanguineal intrafamilial control that we take advantage of to evaluate genetic epidemiology techniques that allow us to assess the specificity of the similarity in terms of environmental influences. This is possible because the control has coexisted at least 5 consecutive years with the case without having genetic endowment of the same origin, which brings interesting challenges to the epidemiological genetic analysis. In summary was a design Case studies - sick siblings - control intrafamily non consanguineous

Another singularity of the project is that it is not a disease but a phenotype, which is also advantageous; therefore, instead of showing the association results as a risk ratio, we show an advantage ratio. In addition, the analysis and design of the contingency tables had to be adapted to these concepts.

Among the findings with significant interest, we discussed the genetic background of longevity in first-degree relatives (Ratio Advantage=7.1,CI=1.07-47.1,p=0.0015); and the joint evaluation of genetic factors and environmental advantages (Ratio Advantage=2.72,CI=1.1-6.7,p=0.006), which was significant when exploring additive models of G-E interaction. The possibility of consolidating practices and knowledge, in the valuation of different theoretical models of environmental genetic interaction, was a result of this project, for the group of young researchers.

Area 3) Studies that formally evaluate the effectiveness of the interventions applied to individuals or groups of individuals in the general population, both already applied and potentially applicable.

Case of study 1: Hypoplasia of the foetal thymus

Strategy: Population

Design: case -only design , 'nested' in a longitudinal cohort.

Practical training of postgraduate students: In this project we analyses of effectiveness of program whit hypothetical application in pregnancy at risk using ultrasound images of the thymus in the view of four vessels in the so-called thymic box, in the superior mediastinum.

This area of study allows for research aimed at studying the effectiveness of prenatal diagnosis programs, neonatal screening, and application of new technologies. With this project in the line of research was achieved by the aspirants to strengthen research methodologies aimed at assessing the effectiveness and effectiveness of the use of technologies and health programs in genetics.

Some results of the project aimed at evaluating the effectiveness of the introduction of sonographic evaluation in the second trimester of gestation of hypoplasia of the foetal thymus, which has been related to other important genetic and obstetric defects, are discussed in the next paragraph.

In order to obtain the proportion of positivity of the study (PPS), we studied the risk factors that were the reason for consultation with greater frequency when thymus-thorax index less than or equal to 0.30 was detected. As a result, we found that these were: suspicious image of heart disease with 5 fetuses with hypoplasia of the gland (33.3% of the total of 16 cases with hypoplasia), followed by maternal hypothyroidism with 4 cases (26.6% of the total number of fetuses with thymic hypoplasia).

When analysing the frequency of thymic hypoplasia according to the risk factor for which the study was performed, the risk factors with higher PPS were: increased nuchal translucency (PPS 33.3%), image of suspected heart malformation in the US of screening (PPS 20.0%), systemic lupus erythematosus (PPS 16.7%) and maternal hypothyroidism (PPE 9.1%). The only risk factor with significantly different PPS was the image with suspected heart malformation.

The predictive capacity for diagnosis of conotroncal heart anomalies was 80%, which had hypoplasia of the gland. The same proportion found in various chromosomopathies (2 trisomy 18, one trisomy 21, and 1 trisomy 13).

The doctoral students could also receive training in other parameters of efficacy, such as sensitivity, specificity, positive and negative predictive value, which we did not break down in this communication, but which were also worked on in collective trainings.

DISCUSSION

In this paper, we analyse aspects concerning the techniques and methodologies of genetic epidemiology, which we believe should be used in research projects in our context, to evaluate in a more adequate and feasible way the behaviour of different complex phenotypes in response to genetic and environmental disturbances. As evidence of the experience gained in the supervision thesis in this line of the doctoral program from the *Universidad de Ciencias Médicas de Villa*

Clara, we show some results achieved with the application of these methodologies in six complex phenotypes, which are studied in the same number of projects.

Area 1) In the projects that pay tribute to Area 1, it is emphasized that the central topic of education is the aspect referred to the determination of the prevalence of genetic disorders or their determinants.

When commenting on the results obtained, and discussing coincidences and singularities; we intend that researchers analyse certainties and validate other alternatives in contrast to the study methodologies used by our research group, which is also part of their learning.

In the case of hereditary angioedema, the frequency of the disease is not established in Villa Clara [2]. Although exact figures on the incidence of HAE are not available, it is estimated that there is a person affected in a range of 10,000 to 50,000 individuals worldwide, although the number of unknown cases is much higher. Our incidence of 1 every 15 370 is high [3]. A determinant of health found was the mortality from the disease that has been described ranges between 15 and 50%. Our mortality of 9.3% is low, although there may be undervaluation.

In the case of reproductive failures, we obtained the frequency of abnormal chromosomal endowments in couples with repeated spontaneous abortions, deceased previous malformed children, male and female infertility, all aspects that are of interest to families and public health. Important complexities should be noted in the prevalence studies in this area, since in fact recognized human fertility is an iceberg, with respect to fertilizations that actually begin. This resulted in the need for the application and teaching of various study techniques and approach to the scientific problem. In this same project, it was possible to obtain the frequency of different polymorphisms in each of the causes of reproductive failures studied, as potential and controversial genetic determinants of the health problem, in relation to possible effects of gene position and differential expression of specific areas of chromatin. On the other hand, we obtained the frequency of potential environmental determinants, considered possible aneugenic agents for aneuploidy again (not inherited), evaluated by previous exposure and during early pregnancy; such as alcohol, chemical substances, viral infections, among others.

In the project referred to congenital defects, where some of their research tasks tax this area of genetic epidemiology, we study aspects related to prevalence; specifically, the search for highly prevalent conglomerates, and temporal-spatial clusters of them were determined. As a way to address prevalence studies when designs, as in this case, are population-based. In these studies, the precise spatial location of the cases is important, as well as the vital statistics, which must be collected according to the spatial unit to be used. The temporal variation of congenital defects may indicate the potential action of environmental agents such as climate, dietary or infectious factors. The spatial variation may indicate risks due to diverse environmental contamination, among others. It should be noted that the municipality of *Sagua la Grande* has important chemical industries in its territory: such as electrochemistry, where chlorine and caustic soda are

obtained, among others. A particular investigation directed towards that spatial conglomerate derives as a contribution of the applied study techniques.

Area 2) In projects related to Area 2 of genetic epidemiology; in which the main topic of research is the search of potential causal or etiological factors in complex diseases or disorders, the line of the doctoral program has worked in training its researchers in the use of different techniques that allow to show, if possible, associations between one or more etiological determinants and the phenotype of interest.

In this area, three disorders or complex characteristics were studied: premalignant lesions and uterine cervical cancer, congenital defects with high association to maternal folate deficiency and human longevity. During these studies, a first and vital step was the creation of family records or the surveys of patients and controls. Regarding this research action, compliance with actions aimed at avoiding information biases, which are a frequent source of errors in these studies and the reporting of false associations, are important. When collecting family pedigrees, it is important that all relatives are recognized both by the mother and father, of the proband or index case, while insisting on establishing the diagnostic criteria to consider, since it is important to take into account the diagnostic certainty, not only of the probands, but of the relatives who will register as sick, and as healthy, as well as the age of onset of the disorder, which avoids taking into account individuals where the disorder has not yet manifested, between other sources of errors.

It has been reported that errors are frequently obtained in the estimation of risk for a person when the family history is taken from the simple counting of the disease among the members of the family. This is because the different components of the variances have to be taken into account in estimating the risk of recurrence, which may be induced by the amount of correlation within the family. Since this correlation can be induced by the accumulated genes that are segregated in it, but also by the accumulated environmental factors, which family members may be sharing. To clarify this aspect, it is also necessary to conduct surveys of environmental factors that thoroughly and exhaustively investigate the environmental exposures to which cases and controls have been submitted. This motivated the insistence on the quality of the elaboration of the elaborated surveys and the methods to be applied in the doctoral study program.

The preparation of suitable family trees guarantees the collection of all the relatives and all the probable patients, and allows a correct estimation of the risk of family recurrence, which is the probability that a person may have an illness, given the report of a family history. This allows adequate calculations of the risk ratio due to positive genetic background.

The application of different methods of family history in the literature shows that the risk of recurrence increases when there is at least one affected relative, as to when there are at least three affected relatives [4]. All these rules were taken into account, in the processing of genealogical

data, and the analysis of the results that in the present, are being carried out in the research topics of the scientific Line.

Considering that the influence of genetic and environmental determinants in complex disorders is generally simultaneous, these considerations are valid to also make a careful management of the environmental exposures to be considered, as well as the interaction between genetic and environmental factors (G-E), which must be evaluated under various hypotheses or models that have been established.

The etiology of many complex diseases involves environmental exposures, inherited genetic predisposition and interactions among them. Gene-environment interaction studies help to explain the interactions, between genetic variation and environmental exposure, that underlie the risk of a disease; and new loci have been identified that suggest evidence of interaction G-E [5]. Common genetic variants have been associated with complex genetic disorders such as cancer, including cervical cancer, congenital defects at birth and longevity. Although they are not reported in this article, they are a main motivation for our research projects in such subjects.

In the same way, these analyses can be extended not only to the association of risks but to the association of advantages, as we have done in the research group with the longevity project.

The heritability of complex disorders, another genetic epidemiology research technique, is often attributed to multiple genetic interactions. In this sense, effects of gene-gene interactions (G-G) have been evaluated, and non-linear, non-additive models such as epistasis are tested. In recent years, several scientific projects have increased the applications of epistasis models to characterize the complex relationships between a large number of genetic variations and exit phenotypes [6].

In addition, case-control association studies can help to understand the pathophysiological mechanisms of complex diseases. It has been considered that studies that include in their analyses the stages of the diseases of interest within the groups of cases obtain greater precision, and that the frequencies of risks generally vary for genetic and environmental variables [7]. As previously shown , these evaluations are useful in the project that studies the progression of clinical stages in cervical cancer, as well as congenital defects.

Area 3) Projects related to Area 3 of genetic epidemiology formally evaluate: (1) the effectiveness of health interventions in genetics, and (2) how they can be applied to individuals or groups of individuals in the general population. In practice, these interventions may already be applied or potentially applicable in the short term. In this last situation, a hypothetical evaluation is made, of its possible effectiveness in case of massive application, or under certain assumptions controlled by the researcher. Several studies with this structure were tested as part of the corresponding learning within the program.

Our research group had previously entered into the evaluation of the effectiveness of five programs of prenatal diagnosis and six of neonatal screening, so that there was a previous experience that was made available to the program.

Pregnant women with suspected heart disease in primary care ultrasound have been internationally recognized as the main criterion for specialized fetal echocardiographic study and therefore frequent finding in the reasons for consulting cases with a lower thoracic thymus index or thoracic thymus index equal to 0.30, which is precisely explained by the relation of congenital heart diseases to thymus disorders, smaller measurements of the thymus should be expected in congenital heart diseases [8].

Although the PPS behaved with high values in pregnant women with increased TN, image of suspicion of heart disease, mothers with SLE and hypothyroidism, the finding that it was only significant in the image of suspected heart disease, shows a high concordance with the literature [9].

Meanwhile, the relationship with chromosomopathies is evident. On the one hand, there are reports that in normal fetuses the thymus uses gene networks in a "*canonical*" way for its functioning. In patients with Down syndrome, however, the existence of "non-canonical" networks has been demonstrated; such networks represent the adaptation of thymus tissues to the genomic dysregulation of trisomy 21 and its functioning under stress conditions. This adaptation is probably driven by epigenetic mechanisms acting at the chromatin level, or by transcriptional miRNA control programs that involve networks of highly hierarchical genes.[10]

To summarize, it has been shown that in the general and specific bases that underlie the teaching-learning process in a doctoral training program in basic biomedical sciences in the line of genetic epidemiology, a combination of theoretical foundations must be strengthened to support the findings obtained in investigative practice.

On the other hand, from the point of view of the required practical skills, the realization of genetic registers through the pedigree of informative families is highly demanding in terms of the rigor of its obtaining and the experience for its interpretation. This forces the carrying out of rigorous genetic, epidemiological and statistical analysis; subjects in which doctoral candidates should be trained in order to obtain reliable prevalence and causality studies with minimal information biases.

The existing designs to carry out studies of association of risk factors or health determinants to complex disorders, can be carried out through population or family strategies, but both by their very nature, show that they are an important topic of study and experimentation that requires consolidation in doctoral training in this area. A topic that requires particular attention in the training processes is that of the methods used to calculate the G-E interactions, calculating the expected risks under the additive or multiplicative model, based on various formulas that have

appeared in the literature and then comparing the observed with the expected, according to these assumptions.

Due to the impact for families, professionals and health systems in Genetics, Training in aspects related to the evaluation of program effectiveness cannot be left behind in the research training of these professionals.

CONCLUSIONS

We have identified in this work, a set of critical theoretical topics as well as study techniques and appropriate methodologies for the approach of research projects in the line of genetic epidemiology. These topics and techniques are relevant in our context, as they have helped in the development of several investigations. Thus, this set of topics could be considered as the general and specific bases for postgraduate teaching in relation to obtaining a doctorate in basic biomedical sciences in our University. They respond to the three main areas of research, which today are addressed internationally by geneticists who investigate in this field of human and medical genetics, and could be useful to other Universities of similar characteristics and contexts.

REFERENCES

1. Wheeler HE, Ploch S, Barbeira AN, Bonazzola R, Andaleon A, Fotuhi Sishpirani A, Saha A, Battle A, Roy S, Im HK. Imputed gene associations identify replicable trans-acting genes enriched in transcription pathways and complex traits. Genet Epidemiol. 2019 Apr 4. doi: 10.1002/gepi.22205)

2. Zanichelli A, Arcoleo F, Barca MP, et al. A nationwide survey of hereditary angioedema due to C1 inhibitor deficiency in Italy.Orphanet J Rare Dis 2015;10:11.

3.Vives Toledo R, SorlíGuerola JV, Sierra Santosa L, García Ribesa M. Angioedema hereditario. REV CLÍN MED FAM 2015;8(1):62-65).

4.Schaid DJ, McDonnell SK, Thibodeau SN. Familial recurrence risk with varying amount of family history.Genet Epidemiol. 2019 Feb 11. doi: 10.1002/gepi.22193. [Epub ahead of print]

5. Kafaie S, Chen Y, Hu T. A network approach to prioritizing susceptibility genes for genome-wide association studies. Genet Epidemiol. 2019 Mar 11. doi: 10.1002/gepi.22198. [Epub ahead of print]

6. Lobach I, Kim I, Alekseyenko A, Lobach S, Zhang L. A simple approximation to bias in the genetic effect estimates when multiple disease states share a clinical diagnosis. Genet Epidemiol. 2019 Mar 19. doi: 10.1002/gepi.22201. [Epub ahead of print])

7.Chen Y, Adrianto I, Ianuzzi MC, Garman L, Montgomery CG, Rybicki BA, Levin AM, Li J. Extended methods for gene-environment-wide interaction scans in studies of admixed individuals with varying degrees of relationships.Genet Epidemiol. 2019 Feb 22. doi: 10.1002/gepi.22196. [Epub ahead of print]

8. Li L, Bahtiyar MO, Buhimschi CS, Zou L, Zhou QC, Copel JA. Assessment of the fetal thymus by two- and three- dimensional ultrasound during normal human gestation and in fetuses with congenital heart defects are useful tools for evaluation of the size and volume of the human fetal thymus through gestation. Ultrasound Obstet Gynecol 2011; 37: 404-409 (wileyonlinelibrary.com). DOI: 10.1002/uog.8853

9. Vigneswaran TV,Homfray T,Allan LD, Simpson JM, Zidere V· Persistently elevated nuchal translucency and the fetal heart J Matern Fetal Neonatal Med. 2018 Sep;31(18):2376-2380. doi: 10.1080/14767058.2017.1342804.).

10. Moreira-Filho CA, Yumi Bando S, Bernardi Bertonha F, Nascimento Silva F, da Fontoura Costa L, Rodrigues Ferreira L Modular transcriptional repertoire and MicroRNA target analyses characterize genomic dysregulation in the thymus of Down syndrome infants Oncotarget,2016 Vol. 7, No. 7, URL:www.impactjournals.com/oncotarget/