Università degli Studi di Cagliari
Dipartimento di Scienze Biomediche
Sezione di Microbiologia e Virologia

Dottorato di ricerca in
"Sviluppo e Sperimentazione dei Farmaci Antinfettivi"
Ciclo XXVII

Titolo Tesi

Sviluppo e validazione di un Nucleic Acid Test (NAT) multitarget su biochip per lo screening del carcinoma delle cervice uterina

Settore scientifico BIO/19 – Microbiologia generale

Coordinatore Dottorato
Prof.ssa Alessandra Pani

Relatore *Prof.ssa Alessandra Pani*

Candidato *Dott.ssa Antonella De Montis*

2013 – 2014

University of Cagliari
Department of Biomedical Science and Technology
Section of General Microbiology and Virology

Research Doctorate in

Development and Evaluation of Anti-Infective Drugs

Cycle XXVII

Thesis

Development and evaluation of a new biochip Nucleic Acid Test (NAT) multi-target for cervical cancer screening

Scientific Area BIO/19 – General Microbiology

Coordinator of the Doctorate

Prof. Alessandra Pani

Supervisor
Prof. Alessandra Pani

Candidate
Dr. Antonella De Montis

2013-2014

Acknowledgements

This research was made possible thanks to the support of many people. It is really a list that is too long to mention!

Teachers, colleagues, friends and family. People who for various reasons and in various ways that have supported, advised and loved me. People that gave me their time, their knowledge, their dreams, and many moments of their lives. People who have taught me where true happiness is for being human.

Thank You

Abstract

Cervical cancer is the fourth most common cancer in women and the cytological screening represents the most diffuse method of prevention. Human papillomavirus (HPV) is an established essential etiological factor for this cancer. Persistence of HPV infection, particularly by those belonging to the high-risk types (HR-HPV), is associated with an increased risk for cervical cancer development. Low-risk HPV (LR-HPV) types are more often associated with benign warts. Most invasive carcinomas are caused by two HR-HPV types: HPV16 and 18. Recently, HPV tests are used as an adjunct test to decrease the false-negative rate of cytological screening with Papanicolaou test (PAP Test), especially HR-HPV DNA detection tests are useful for primary screening of cervical cancer and for triage of patients with equivocal cytological findings. However, the roles and contributions of other uncommon and rare genotypes remain uncertain, especially in specific geographic areas or populations. Recently, microarray biochip technology has been introduced into the clinical laboratory for HPV detection. One such test is the ProDect® CHIP HPV TYPING KIT (bcs Biotech Srl, Italy), which has the ability to identify 19 HPV types (all HR-HPV and most common LR-HPV types) and detect the generic presence of E6/E7 HR-HPV sequences. The aim of this pilot study was design a new biochip (CHIP PLUS) for the cervical cancer screening able to detect a large number of HPV anogenital types (HR- and LR-HPV) using two regions of the viral genome (L1 and E6/E7 sequences) as targets. This report also presents the results of a preliminary validation study in preparation for an extended clinical validation of the medical diagnostic ProDect® CHIP HPV TYPING PLUS employing the above CHIP PLUS designed both for the simultaneous detection of 31 HPV types (both common and uncommon HPV types) and for the characterization of three E6/E7 consensus sequences belonging to the principal groups of HR-HPV. The preliminary results of the ProDect® CHIP HPV TYPING PLUS KIT validation had higher concordance and/or greater compatibility with those of the reference tests, underlining the importance of searching for uncommon HPV types, enabling good prevention of cervical cancer using HPV DNA as test screening.

Preface

This thesis is subdivided in two chapters on the different research activities I have conducted and about the research programs in which they were entered. All activities described were carried out at the University of Cagliari laboratories and bcs Biotech Srl Research and Development laboratories, a company specialized in the development and production of in Vitro Diagnostic Devices (IVD) and biotech platforms, certified UNI EN ISO 9001: 2008 and UNI CEI EN ISO 13485: 2012, http://www.biocs.it.

The first chapter (Chapter 1), which is also the main topic of discussion for this thesis, with regards to the development and validation of a biochips Nucleic Acid Test (NAT) to be applied to the screening of cervical cancer and to the diagnosis of infection by Human Papillomavirus (HPV). This research commenced using a bcs Biotech biochip kit that can detect 19 HPV genotypes in anogenital tropism, including 14 HPV genotypes considered at high-risk oncogenic and clinically validated as CE IVD. These kits have been used in an International cooperation program funded by the Region of Sardinia (L.R. 1996 ART.19) entitled *Screening of the human papilloma virus (HPV), and prevention of cervical cancer in women of the city of Cotonou and Benin* coordinated by the Cytomorphology Department of the University of Cagliari in which bcs Biotech Srl attended as a partner and the scientific coordinator was Prof. Paola Sirigu. At the end of my study this biochip was implemented with further 12 genotypes, for a total of 31 detectable HPV genotypes, including many medium risk oncogenic genotypes that were not sufficiently investigated by an epidemiological point of view for their transforming power. This new test also allows, on the basis of additional probes, to detect and discriminate the three main groups of Early sequences (E6/E7) HPV, that are related only to these and other common genotypes with high-medium oncogenic risk. In fact, the proteins encoded by these sequences in the HPV genotypes with high oncogenic risk factors would be necessary, although not sufficient, for the neoplastic transformation of the mucosa. The study was suitable for the publication of an article on the first chip F.Piras, M.Piga, A. De Montis, R.F.Zannou Ahissou et.al. *Prevalence of human papillomavirus infection in women in Benin, West Africa*. Virology Journal 2011, 8:514 http://www.ncbi.nlm.nih.gov/pubmed/22074103 and obtained a European Patent, also including the design of the new chip, Perseu S., De Montis A., Lauterio C., Manca I. European Patent EP 1818416B1 http://worldwide.espacenet.com/publicationDetails/originalDocument?FT=D&date=2012122

6&DB=&locale=en_EP&CC=EP&NR=1818416B1&KC=B1&ND=1. The new prototype kit clinical validations on chip are ongoing. In this thesis are only show some preliminary verification data obtained on pre-production batches, before the start of the clinical trials.

The second chapter (Chapter 2) concerns a study in which I have participated at the end of the program POR 3.13 "Creating a center of excellence for bio-informatic technologies applied to personalized medicine" managed by Sardegna Ricerche. In fact, thanks to two different projects funded between 2006 and 2009 under this program entitled respectively "*Development of new bioinformatics methods for the detection of environmental and genetic causes of multifactorial diseases in Ogliastra*" and "*Detection of the genes responsible for autoimmune Hypothyroidism and for Hashimoto Disease in Ogliastra*", at the BCS BIOTECH R & D laboratories was created a bio-bank including DNA from control subjects, verified by anamnestic and instrumental evaluation, as unaffected by the 20 most common multifactorial diseases in Sardinia (including Diabetes I and II, Multiple Sclerosis, Hypothyroidism and Hashimoto's disease, Lupus, Rheumatoid Arthritis etc.) and conversely subjects affected by these diseases. The samples collected in the BCS BIOTECH Bank, include sick individuals as well as healthy ones, carefully selected with a series of fundamental criteria: not related (based on their family tree up to 4th – 5th generation), coming from the same Mendelian Unit Cross (UMI) or geographic area. Starting from this assumption I proceeded to analyze control samples from the bcs Biotech bio-bank selected in the Ogliastra geographical area (central-eastern region of Sardinia) by the Affymetrix Genome-Wide Human SNP Array 6.0 platform localized at the Laboratory of Genomic in the Technology Park of Sardinia (Pula, Cagliari). At the same time, other colleagues from the Department of Experimental Biology, University of Cagliari, coordinated by Prof. Vona, were analyzing DNA of subjects collected in South Sardinia. This type of platform is able to simultaneously determine the genotype of a subject for about 1,000,000 Single Nucletides Polimorphic site (SNPs) and about a million genetic polymorphisms linked to the variation of the number of sequences (Copy Number Variation, CNV) data analysis was performed by using a dedicated software: the Genotyping Console™ software 4.0. The study and the merging of data from the samples collected both in bcs Biotech and from the University of Cagliari has been the subject of the following publication Piras I., De Montis A., Calò CM., Marini M., Atzori M., Corrias L., Sazzini M., Boattini A., Vona G., Contu L. *Genome wide scan with nearly 700.000 SNPs in two sardinian subpopulations suggest some regions as*

candidate targets for positive selection. European Journal of Human Genetics (*EJHG*) 2012, 03:1-7

http://www.ncbi.nlm.nih.gov/pubmed/22535185. The article highlights a sub structure of the Sardinian population, particularly in the area of Ogliastra presenting greater genetic homogeneity compared to a higher genetic variability in the samples from the south of the island. Following this publication I have been involved in an international study regarding the validation of an algorithm structure of geographical population (GPS) that exploits the data using as a reference SNPs 40000-130000, and on this assumption is able to determine the location geographic or genetic origin of any individual. Several research groups around the world have offered their dataset to allow the validation of the new algorithm model. Following the data this new genetic GPS placed 83% of the individuals studied in their country of origin. When applied to the Sardinian samples studied by myself, the GPS placed a quarter of them in their villages, and a majority of others within 50 km of their originating villages. http://www.ncbi.nlm.nih.gov/pubmed/24781250, Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P, Pagani L, Tyler-Smith C, Xue Y, Cucca F, Schurr TG, Gaieski JB, Melendez C, Vilar MG, Owings AC, Gómez R, Fujita R, Santos FR, Comas D, Balanovsky O, Balanovska E, Zalloua P, Soodyall H, Pitchappan R, Ganeshprasad A, Hammer M, Matisoo-Smith L, Wells RS; Genographic Consortium. *Geographic population structure analysis of worldwide human populations infers their biogeographical origins*. Nature Communications. 2014 Apr 29;5:3513.

Table of contents

CHAPTER 1 - DEVELOPMENT AND VALIDATION OF A NUCLEIC ACID TEST (NAT) MULTI-TARGET OF BIOCHIPS TO SCREEN UTERINE CERVIX CANCER 1

1. Introduction 1
   - 1.1 Epidemiology of cervical cancer 3
   - 1.2 HPV infection and the natural history of cervical cancer 4
   - 1.3 Phylogenetic and epidemiologic HPV classification 6
   - 1.4 Virus structure 7
   - 1.5 Virus replication cycle 9
   - 1.6 Molecular diagnosis of HPV infection 14
2. Scope of the research 18
3. Materials and Methods 21
   - 3.1 Biochip preparation 21
   - 3.2 Essay detection biochip 22
   - 3.3 Synthesis of oligonucleotides 23
   - 3.4 Biochip reading system 23
   - 3.5 Molecular diagnosis 24
     - 3.5.1 Samples 24
     - 3.5.2 ProDect® Chip HPV Typing (bcs Biotech S.r.l.) 25
     - 3.5.3 Digene® HC2 HPV DNA Test (Qiagen) 26
   - 3.6 Cell Lines 26
   - 3.7 Cloning DNA HPV types with high and low oncogenic risk 27
4. Results 30
   - 4.1 New HPV CHIP PLUS 30
   - 4.2 CHIP HPV PLUS Sensitivity and Specificity 36
     - 4.2.1 Synthetic DNA 36
     - 4.2.2 DNA extracted from cell lines 37
     - 4.2.3 HPV cloned DNA at high and low oncogenic risk 39
     - 4.2.4 Extracted DNA from cytological samples 47
5. Discussion 66
6. Conclusion 72
7. Bibliography 74
8. Apendix – Raw data of samples tested 87

CHAPTER 2 - THE STUDY OF GENOMES IN SARDINIAN INDIVIDUALS AFFECTED AND NON-AFFECTED FROM MULTIFACTORIAL DISEASES ... 91
1. Background ... 91
1.1. Choice of the geographical areas ... 91
1.2. Selection of the Controls ... 92
1.3. Selection of affected probands ... 92
2. Introduction ... 94
3. Material and methods ... 95
3.1. Samples ... 95
3.2. Genotyping and quality control ... 95
3.3. Genotype calling and dataset filtering ... 96
4. Results ... 97
4.1. Study No. 1, ... 97
4.1.1. Conclusion ... 98
4.1.2. References ... 99
4.2. Study No. 2, ... 102
4.2.1. Conclusion ... 103
4.2.2. References ... 104
5. General conclusions and perspectives ... 105

CHAPTER 1 - DEVELOPMENT AND VALIDATION OF A NUCLEIC ACID TEST (NAT) MULTI-TARGET OF BIOCHIPS TO SCREEN UTERINE CERVIX CANCER

1. Introduction

The discovery of some genotypes of the human papillomavirus (HPV), defined as high-risk oncogenic (HR), at the base of the etiopathogenesis cervical carcinomas[1] laid the foundation for the approach prior to the illness today, to the development of two prophylactic vaccines[2-10]: Gardasil® (Merck & Co.) and Cervarix (GlaxoSmithKline). Randomized controlled trials have demonstrated the efficacy for at least eight years[11] using the first vaccine Gardasil, also known as quadrivalent vaccine, in preventing infection by HPV-16 and HPV-18, together responsible for over 70% of carcinomas, and types HPV-6 and HPV-11, etiologic agents of 90% of anogenital warts, while Cervarix, also called bivalent, was protective about the development of forms precancerous and cancer caused by HPV-16 and HPV-18. Unfortunately, these vaccines do not prove, however, any therapeutic efficacy.

The enhancement of diagnostic techniques, especially on the molecular sector, led till now the discovery of about forty HPV with high tropism towards the anogenital mucosa, including at least a dozen types, with regards to HPV-16 and HPV-18, with an elevated oncogenic propensity.

The importance of these molecular studies in primary screening of cervical cancer is now irreplaceable, often resulting comparable[12-16], when not superior, to the Papanicolaou cytology Test (PAP test), that entrust pathologist to the lesions detection in mucosal according to a scale based on the parameters defined. This test still suffers from the suitability of the sampling and the subjectivity of interpretation, despite having been recently introduced some automation.

In particular, several controlled studies demonstrate the high predictive value of HPV negative DNA, compared to the PAP test, indicating that, if confirmed, would allow the lengthening of the organized screening (3 to 5 years)[17-21] with a relative rationalization of health spending without reduction against the surveillance capabilities of this type of tumor.

The negative predictive value (NPV) of the molecular test takes on great importance because if the cervical cancer recognized as the main cause of persistent infection (ie for several years) by HPV genotypes with high oncogenic power, the probability that a woman developments, relatively quickly (within 5 years), preneoplastic or neoplastic lesions is statistically unlikely if the woman is HPV negative, then is free of any infection.

As an example below there is one of the many screening algorithms proposed[18] which provides as the primary test, the only molecular test, and proposes the PAP test, so the evaluation of cytological abnormalities, only to those women positive for HPV sequences (Fig.1).

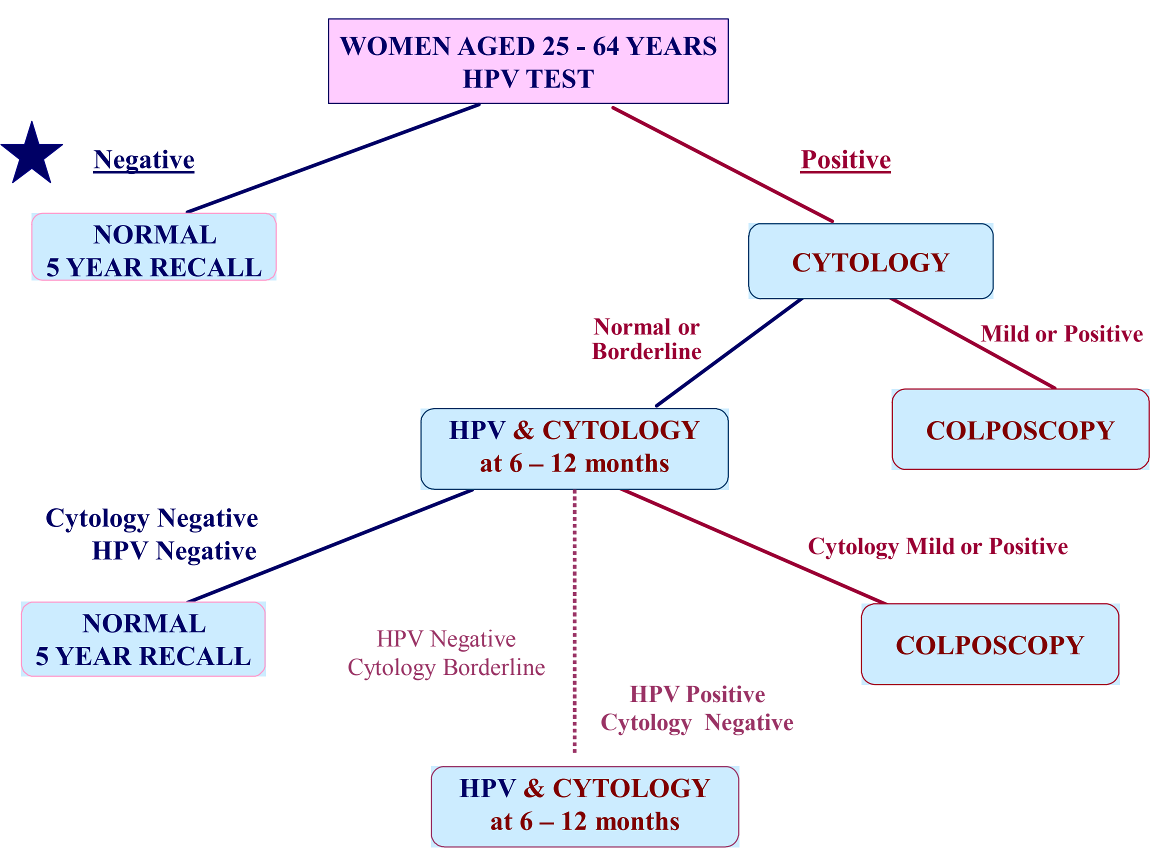


Fig. 1[18] **-** An example of the proposed algorithm to screen for cervical cancer suggests that HPV DNA testing women as the primary test and cytological analysis of women HPV-positive as triage test. (Edited by Cuzick et al, 2008).

However, the high predictive value of a negative HPV DNA test implies that the test used to perform molecular diagnosis can be attributed reliably negative for HPV DNA.

Unfortunately, little is still known about the natural history of HPV in relation to the development of cancer and in many ways very complex, influenced by 1) the ability of the virus to persist in episomal form or integrate into the genome of the host cell, 2) from high number of viral genotypes with variable oncogenic power, 3) from their interaction in the case of co-infection and 4) interaction with various genetic and molecular aspects of the host who would control the course of the disease. This means that many negative molecular diagnosis for HPV infection are not really reliable, (*false negatives*), but affected, as well as unsuitable withdrawals, by low power of the diagnostic tests on the market, especially to those more insidious viral forms, due to integration, co-infection or not widespread, but still equipped with a tumorigenic influences, at the base of recurrence or lesions of uncertain cytology classification (ASCUS), which can already show frank precancerous lesions through focused colposcopic analysis.

The present research project can be included in this context since its general objectives the design, development and validation of a multi-target Nucleic Acid Test (NAT) that complies with CE standards for In Vitro Diagnostics Devices (IVDs), has a high negative predictive value applied to cervical cancer screening with the most modern diagnostic algorithms.

1.1 Epidemiology of cervical cancer

One of the most important achievements in the medical oncology field in the last fifty years, is the huge progress made in reducing mortality and cervical cancer incidents, at least in the countries that have adopted health policies for primary screenings. Slow progressive malignant tumors, specifically invasive that develops in the cervix. This neoplasia until a few decades ago was the most common cancer and the most lethal in the female population and constituted the second tumor, in order of incidents of mortality in general population, second only to lung cancer. The data published by the World Health Organization (WHO)[22] in 2014 showed that it is now the third tumor incidence and the fourth leading cause death from cancer among women.

The basis of this success, on one hand, studies of Professor Georgios Papanicolaou on cervical-vaginal cytological diagnostic during the Second World War that led to the rapid spread from 1960 a budget and poorly invasive screening test, called in his honor PAP test, on

the other hand, the recognition of the fundamental role of the human papillomavirus (HPV) on the neoplastic transformation of cervical mucus cells, hypothesized for the first time by Professor Harald zur Hausen in 1975.

These two factors, together with the progress of the methodologies and the refinement of diagnostic capabilities of uterine cervix pre-neoplastic lesions, as well as the use of large-mass scale screening, have led to a strong reduction in the incidences and mortality of this type of tumor. However, data of the report of the WHO[22] of 2014 referring to 2012, still recorded over half a million new cases, more than 270,000 deaths, with over 85% of cases and deaths in developing countries[22], where it is the leading cause of death in women aged between 15 and 44 years.

1.2 HPV infection and the natural history of cervical cancer

The neoplastic transformation requires a series of complex mechanisms interacting with each other, both endogenous and exogenous types, which act by determining on one hand, a persistent alteration of the viral DNA and cell metabolism and on the other hand, a kind of tolerance by the immune system of the host organism[23-24]. The model currently recognized as the most descriptive HPV infection in course in relation to the development of cervical carcinoma represented in Fig. 2. The woman acquires the infection through sexual contact from an infected partner, with greater prevalence at a younger age due to first sexual encounters or when frequent exposure combined with a lack of prior immunity. The incidence of HPV-infected women is thus particularly high during adolescence and early adulthood and decrease with the progress of age (*incidence*). The contracted infection, however, is confined and of course eliminated (*clearance*) within an average of 2 years. In 90% of cases[25-27] and about 60% of these infections determines a serum conversion, or specific antibody production of the infecting HPV genotype.

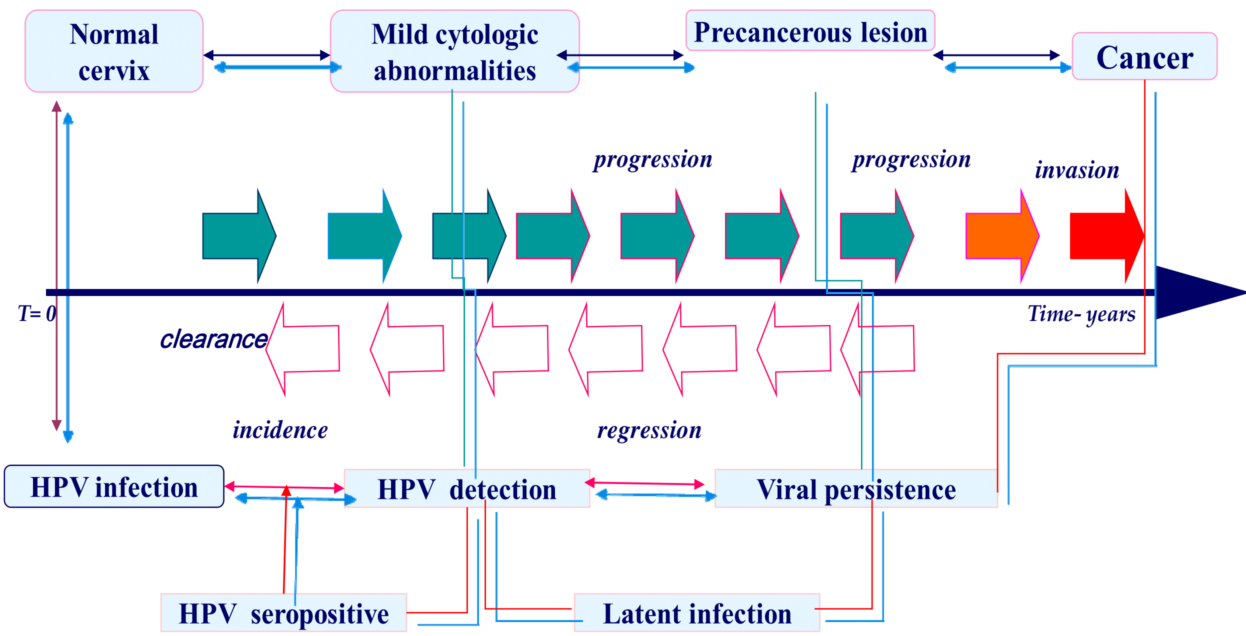


Fig. 2[23] Natural history of cervical carcinogenesis, (Amended by Schiffman M, et al. 2011)

Using a cytology PAP Test (Tab. 1), these women showed cervical dysplasia classified as low or moderate, such as the American classification established in 2001 in Bethesda, *Low-grade Squamous Intraepithelial Lesion* (LSIL) or in the European classification, given by the Norwegian Ralph Richart in 1967, *Cervical Intraepithelial Neoplasm* (CIN I). Only infections that persist have a certain probability to progress to real precancerous lesions, identified as type HSIL or CIN 2-3. Between them only a small proportion, if left untreated, progresses to carcinoma after several years [28]. Among the risk factors that may further predispose more rapidly to cancer (varying between 3 to 20 years is observed) have been highlighted smoking [29-31], prolonged use of oral contraceptives[32] and the number of natural births[33]. It is estimated that approximately 60%, 40% and 33% of CIN1, CIN2 and CIN3 respectively regress spontaneously, a variable amount closed to 20-30% persists without development to more serious lesions, 10-20% progress from CIN1 to CIN3 and from CIN2 to CIN3; and a smaller portion (1% of CIN1, 5% of CIN2, and 12% of CIN3) can, if untreated, evolve to invasive carcinoma[34].

Dysplasia	**Bethesda Classificazion[2001]** **(used in cytology)**	**Cervical intraepithelial neoplasia CIN and CGIN classification**
atypical squamous cells, not classifiable	ASC-US	---
atypical squamous cells, not excluding high-grade lesions	ASC-H	---
low grade, mild	LSIL	CIN1
high grade, moderate to severe	HSIL	CIN2
		CIN3
dysplasia endocervicale with *atypical glandular cells*	AGC differentiated in: *AGC-NOS not otherwise specified* *AGC favor neoplastic*	CGIN (low-grade) (high-grade)
Adenocarcinoma in situ	AIS	CGIN(high-grade)

Tab. 1[34,35] Comparison of the principal terminology used to classify preinvasive intraepithelial lesions

1.3 Phylogenetic and epidemiologic HPV classification

The Papillomavirus derive their name from the Latin "papilla" or pustule and from Greek "oma" which means cancer, in fact discovered only in the twentieth century, the lesions caused by them in humans, as in animals, has been known since ancient times.

Classified in the family of *Papillomaviridae*[36] divided into 16 genera, we find the human papillomavirus, phylogenetically distributed in five genus (Alpha, Beta, Gamma, Mu and Nu), which currently includes around hundred genotypes[37]. Genotypes are distinguished by gene sequence basis of the HPV L1 region, that are dissimilar from each other by more than 10% of the sequence basis, and are identified by a number (HPV16, HPV18, HPV6, etc).

These are viruses that induce hyperproliferative lesions in multilayered epithelia squamous of the skin and mucous membranes resulting in the major cause of elementary

lesions, such as warts and complex pre-cancerous and cancerous lesions[38]. Those with cutaneous tropism are identified as epidermotropic and infect primarily the skin of the hands and feet. The target cells of HPV called mucosals are mainly located at labial mucosa, respiratory tract and anogenital area. The restricted replication to these two types of cells is influenced by the differentiation stage of these cells.

There are about forty HPV genotypes that infect the mucosal and the anogenital area epithelia. Based on the frequency with which they were isolated in anogenital cancerous or precancerous lesions, they are divided into two groups[39-42]:

- The HPV *Low Risk* (LR) with oncogenic low-risk, until now, were only isolated from benign lesions: HPV-6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, CP6108.

- The HPV High Risk (HR) with high-medium oncogenic risk, are frequently associated to the appearance of malignant lesions: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 73, 82. Sequences of HPV16 and HPV18 were found in approximately 50-70% and 10-20% of cervical carcinomas respectively [42].

1.4 Virus structure

Observed under a transmission electron microscope (TEM) the HPV is a small, naked virus, whose capsid is formed by 72 capsomers arranged as a symmetrical icosahedron, encompassing a circular double helix DNA genome, circa 8 kilobases[42] (Fig. 3). Each capsomer is composed of two proteins: 80% from the L1 protein of 57 kD and the remaining 20% from the L2 protein of 43-53 kD. The genome is characterized by an 8-10 coding ORF sequence (Open Reading Frames) and by a non-coding region called LCR (Long Control Region) or URR (Upstream Regulatory Region) regulator of the viral replication and transcription, ranging in size between various genotypes of HPV[44]. The HPV ORF are divided into early regions (E "Early, E1-E8")[45] and late (L "Late" L1-L2)[46] encoding proteins. The L proteins are the viral capsid L1 and L2 proteins, only expressed in infected cells in the final stages of the viral replication cycle in a productive way, while the E proteins are regulators of the duplication process of the viral genome and work directly and indirectly on macromolecular synthesis of the host cell. Among them, the E6 and E7 proteins are critical in inducing the transformation, in the tumor sense of the cell[47-49]. These proteins interact

with several transcription factors, acting cellular proteins; on the one hand the inhibition of mutation repair in the genomic cell DNA reducing the levels of cellular p53 levels[50-51], therefore altering the control and shelter mechanisms that the cell enacts towards any damage to its genetic heritage. On the other hand, inducing cell multiplication through interaction with cellular co-factors (in particular the pRb)[52-54] normally down-regulated if not needed (Fig. 4). The other interactions with various factors for regulation, expression and cell replication always lead to target inactivation/degradation or they involve directly or indirectly stimulation/activation. However, because of viruses are difficult to propagate in vitro, many targets or interactions are still unclear.

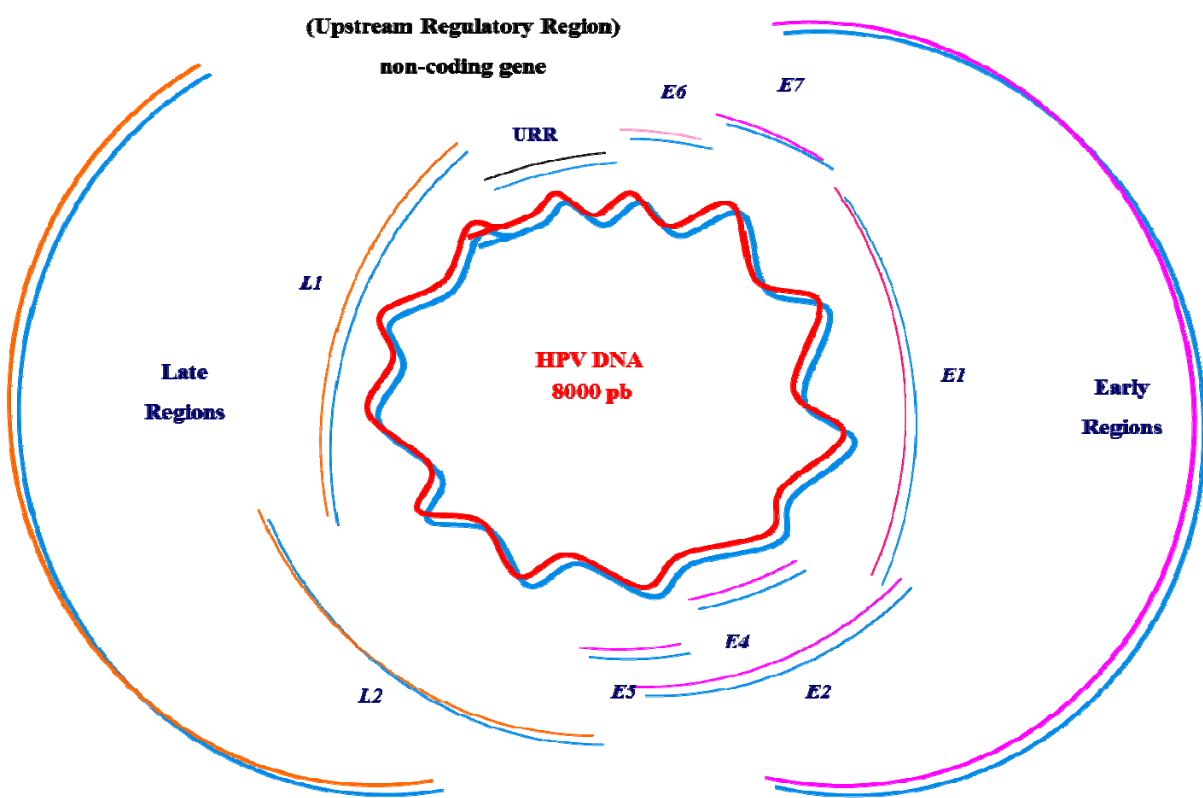


Fig. 3[42] DNA HPV-16 (Modified from Muñoz et.al, 2006)

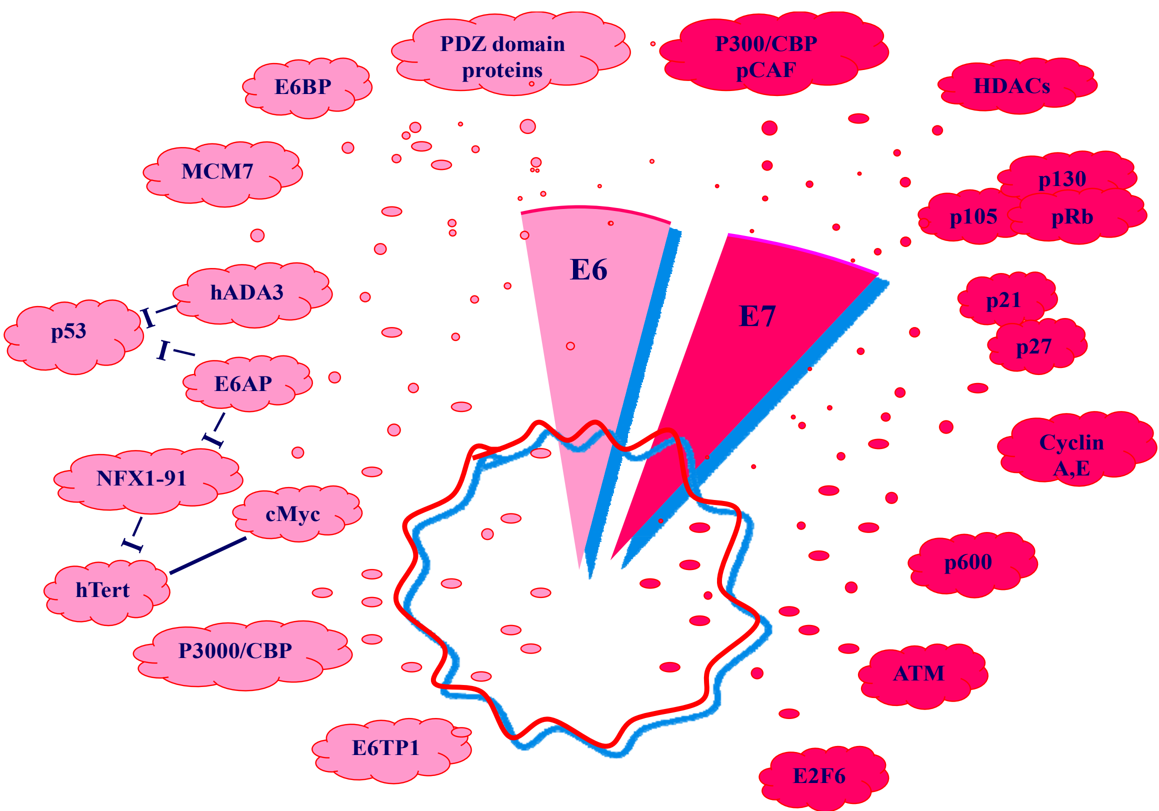


Fig. 4[54] Some target cells for proteins E6 and E7 oncogene (Modified from Bodily et al., 2011)

1.5 Virus replication cycle

From a molecular point of view the replication of HPV arises from sequences placed in the viral genome region Upstream Regulatory Region (URR) (Fig. 5). Two polyadenylation signals define 3 groups of viral sequences, whose expression is regulated by cellular factors during the host cell differentiation specifically: the group of genes E1, E2, E4, E5, and E8 involved in viral DNA replication and in transcriptional control; the group formed by the genes E6 and E7 that maintain and stimulate replication; and lastly the group of L sequences encoding the L1 and L2 capsid structural proteins.

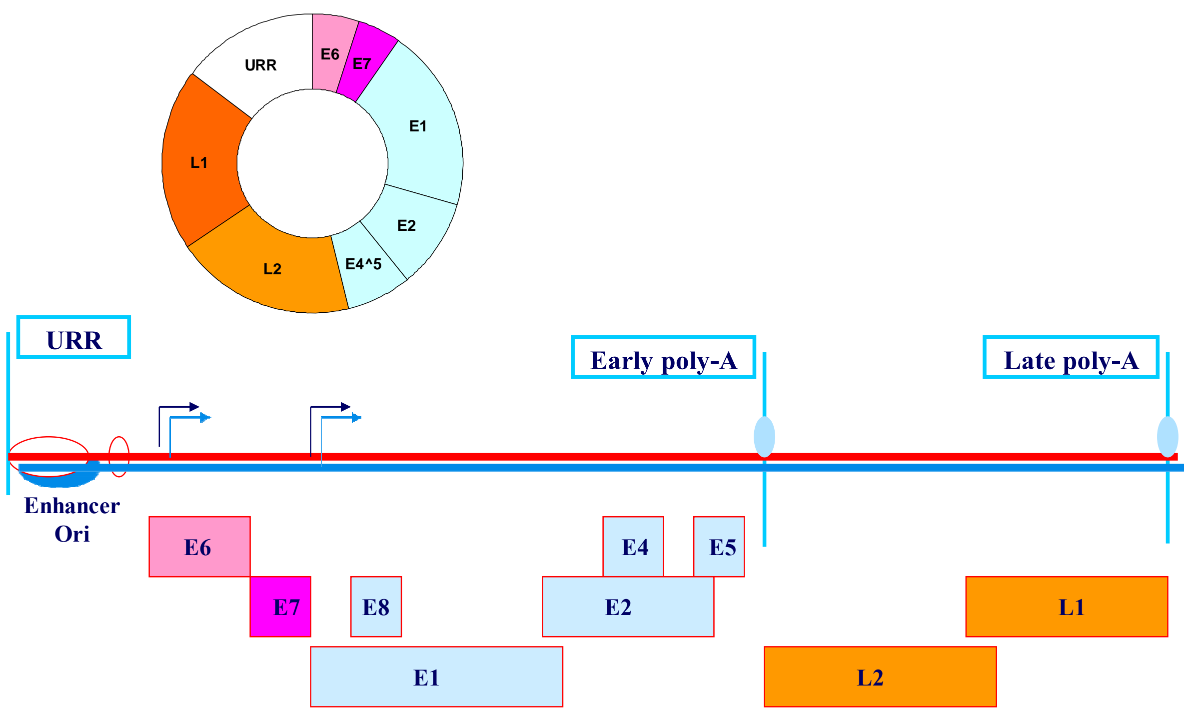


Fig. 5[54] Linearized genome of HPV (Modified from Bodily, 2011)

The replicative cycle of the virus can be subdivided in a *nonproductive stage* called *Early* in which E sequences are transcribed and that lead to the functional synthesis factors, and then there is a *productive stage* called *Late* where the L regions transcription coding for the capsid structural factors.

The non-productive stage is characterized by the stabilization of the viral genome as an episome at the basal layer level of the epithelium in which are localized cells in continuous division. In the cervix most infections are established at the level of the Transformation Zone (TZ) basal layer, the transition zone between the ectocervical stratified epithelium and of the endocervical columnar[54] (Fig. 6). Due to the effect of micro-trauma and to interaction with particular molecules[55], the HR or LR HPV reach the basal layer cells that are essentially of two types[56]: *Transit Amplifying cells* (TA), and proliferating cells capable of differentiating in a vertical direction moving to the mucosa suprabasal layers, and then the *Stem cells* (ST) who have a limited ability to proliferation, with the function of supply type TA cellular pool. Indeed, they represent the epithelium long-term reserve as only one of the two daughter cells derived from their division will become a TA cell while the other will remain ST; the mechanisms underlying this process are still unclear and very controversial. The virus would

infect electively the TA cells[57] replicating at the cell nucleus level its genome up to 50-200 copies depending on the infecting genotype.

The conservation of the episomal forms[58], essential for persistence phenomena establishment, it would be controlled by the sequences E6, E7, E1 and E2[59-62]. In fact, the E1 and E2 proteins cooperate with cellular factors to trigger viral genomes replication and E6 and E7 proteins modulate the same factors so cell remains in active propagation as long as possible[63-64]. At this stage, the viral proteins are expressed at low levels and this contributes to evade the immune response by promoting the persistence. The newly formed viral genomes segregate in an equivalent manner between the two daughter cells at the time of mitotic division as cellular chromosomes[65] under E2 protein control. This protein also coordinates genomes number, which under physiological conditions must not exceed a certain threshold in order not to be lost during mitosis[66]. In this regard, E2 protein has functions of positive and negative control on E6, E7, E1 genes. When TA cell, under the effect of various factors, mainly of hormonal nature, is preparing to leave the basal layer to continue the differentiation, it has as a side effect a raising of transcription of ORFs.

At the spinous layer level[67-69] is observed an impressive increase in E1, E2, E1 ^ E4 and E5 transcripts[70-72] and viral genomic copies from 50-200 copies/cell reach up to about 10^4-10^5 copies per cell[73]. Excess of E1 ^ E4 and E5 protein contributes, however, to the activation of L1 and L2 Late protein synthesis[74-76]. It follows that the capsid proteins, strongly immunogenic, are produced only in the more differentiated epithelial layer[77], that is the one that faces the organ lumen.

The process of virus maturation from the deep layers to the release into the lumen requires long time, in the order of 4-6 months from the moment of infection contraction.

Also the clearance of the virus, which occurs in the majority of infections, it would be a fairly slow process which involves cellular mediated immunity[78] although at present the mechanisms are known only in part[79]. HPV reduces the innate and acquired immune response through different strategies: for example, E6 and E7 cells expressing of HR HPV, repress transcription of many interferons target genes[80-83], while the HPV infected cells produce low levels of cytokines[84] such as IL-1, IL-6, TNF-α and TGF-β and high levels of anti-inflammatory IL-10 reducing the ability of immune cells to infiltrate the infected

tissue[85]. However in a certain proportion of HPV HR infections, viral genome tends to be integrated into the cellular DNA causing a reduced production of viral complete progeny. This integration is at the initialization of the transformation and cellular immortalization mechanisms, usually occurring at the E2 ORF level, resulting in loss of the repressive action on viral E2 factor on the E6 and E7 sequences transcription and consequently the production of relative proteins, fundamental in the mechanisms of carcinogenesis, and therefore considered viral oncoproteins[86-90] (Fig. 7). The interruption of the ORF at E2 and the subsequent integration of the HPV HR genome in the cell, leads to a consequent over-expression of the E1, E6 and E7 proteins and amplification of their effects on the cell; stimulation of proliferation and the inhibition of p53-mediated control/repair mechanisms[89]. Studies conducted on HPV16 have shown that the viral DNA breaking points at the time of integration into the cellular genome can affect most of the virus genome[90]: only the L1-L2 ORF have been with 59 breakpoints characterized and 39 breakpoints in E1-E2-E5 regions. It follows that the smallest region of the viral DNA, which, up to now a loss of sequence has not been observed, but that is fully detectable, both in episomal form and in integrated virus cellular genome, corresponds precisely to the regions URR, E6 and E7.

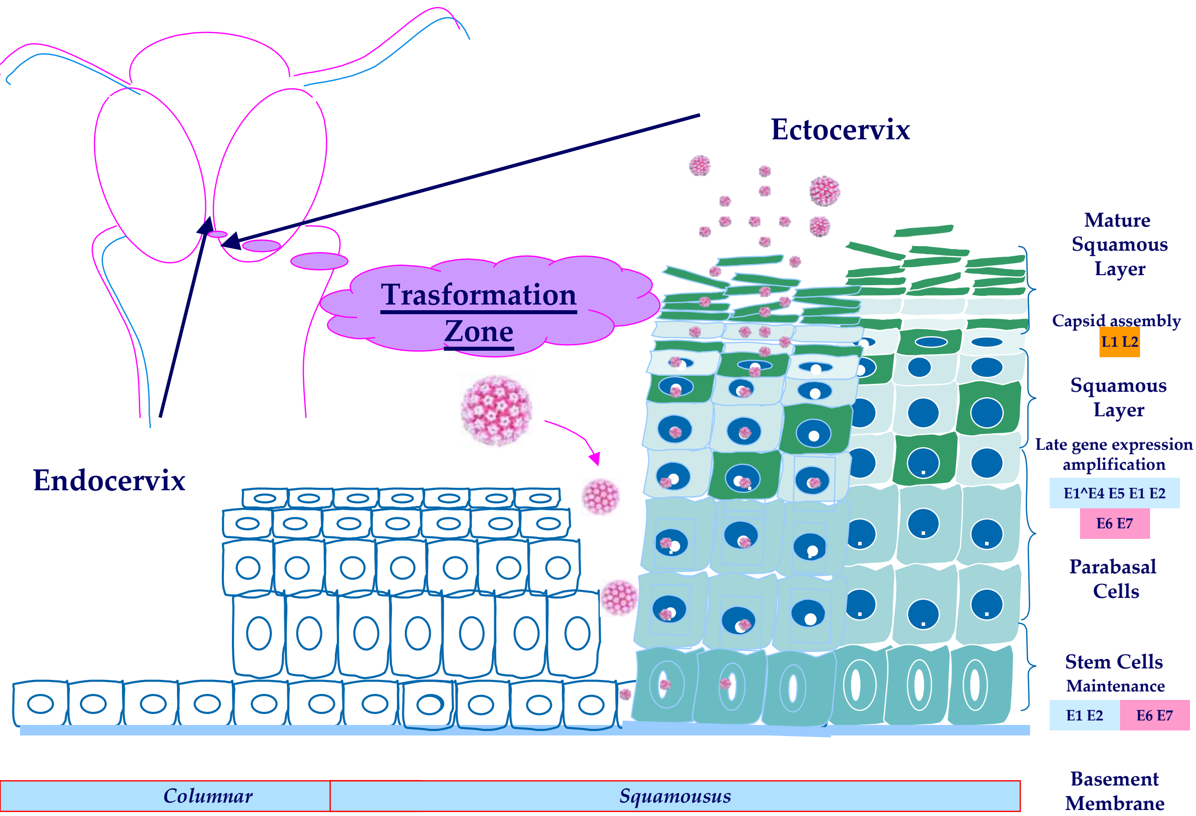


Fig. 6[54] Localization of the infection and different ORFs expressions (Modified from Bodily 2011)

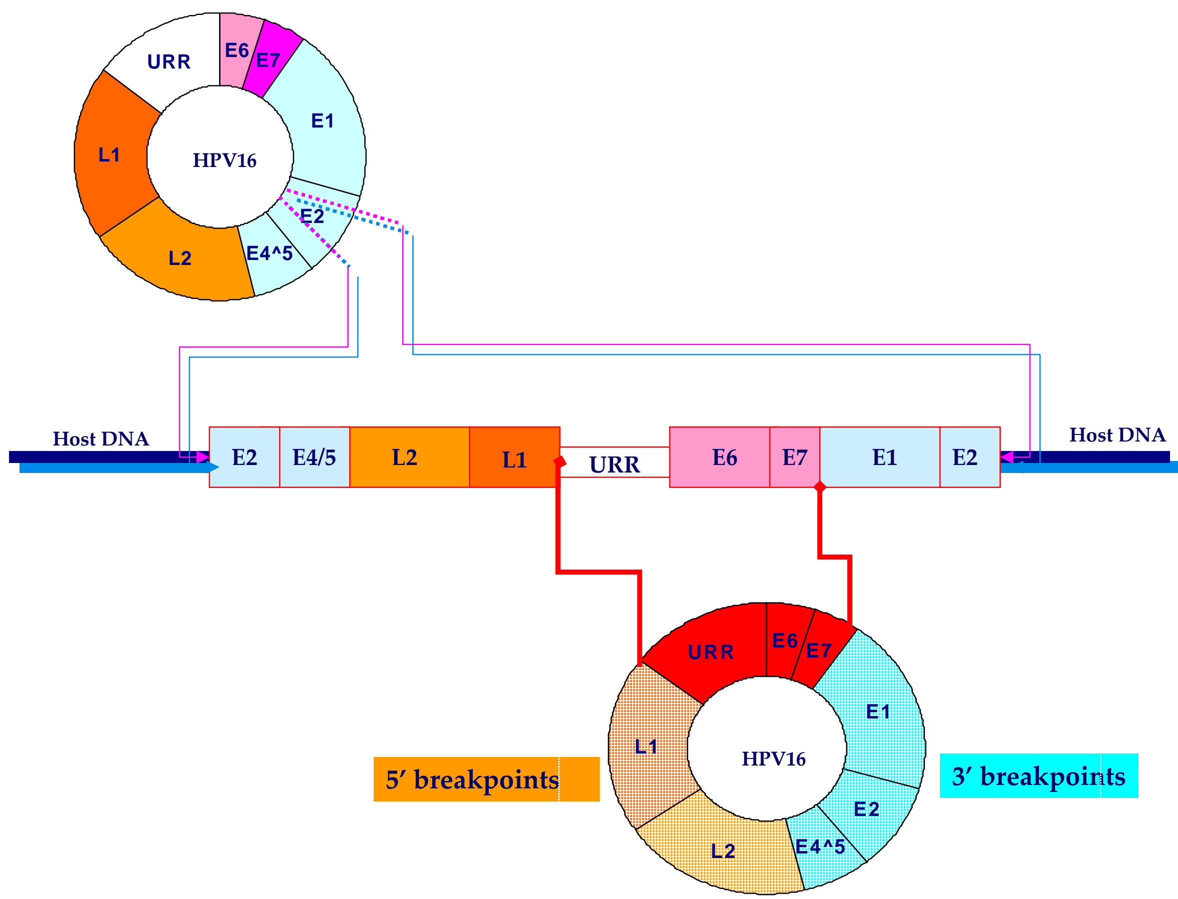


Fig. 7[87,90] Example of the HPV16 integration in the cellular genome with the interruption of the E2 sequence in HPV. Further potential breakpoints areas on the viral genome emphasized in the red region URR-E6 / E7 are apparently not subject to this phenomenon (modified from zur Hausen 2002 and Bo Xu, 2013).

Considering what has been presented, HPV is a complex target from a diagnostic point of view. In fact, the infection may be observed in different stages:

- productive with an elevated number of complete viral genomes
- latent or persistent with controlled production of complete viral genomes
- abortive with integrated and incomplete genomes

Epidemiological data shows that majority of HPV infection heal due to host cell-mediated immune response[91] and only 10% often persist and can potentially cause intraepithelial neoplasia to a various degree, which at a lower percentage may progress further to an invasive carcinoma[92-94]. This progression is due to many factors, among which significantly emerges the state of HPV HR virus integration[95-97] in the cellular genome. However it is not yet clear the influence of this integration in the progression of neoplastic

cell as it would seem a necessary element but not sufficient. In fact, several studies clarify the integrated forms pervasiveness in more advanced lesions[86,98-99], others report the concomitant presence of integrated and episomal forms already in CIN1[100-101], however with an increase of transcripts integrated forms higher in carcinomas than in CIN1[96] and of a certain intensity for high-risk oncogenic genotypes than others (Tab. 2)[97]. Furthermore, it is still controversial with regards to the diagnostic and prognostic role of high level E6/E7 transcripts from integrated and episomal forms, given the increase has also been studied in women with integrated forms of HPV16, HPV18, HPV52 and HPV58, even in the absence of a pathological or normal cervix[102].

	Normal	CIN1	CIN2	CIN3	Carcinoma
HPV16	0	0	6%	19%	55%
HPV18	0	0	0	0	92%
HPV31	0	0	0	10%	14%
HPV33	0	0	0	0	37%
HPV45	0	0	3%	60%	83%

Tab. 2[97] Prevalence transcribed E6 / E7 from HPV integrated forms (modified from Vinokurova 2008)

It is obvious that a diagnostic system in line with the state of the art should be able to detect the virus even when the genomes are relatively few and especially not complete and the poor viral progeny or non-existent as in the case of abortive infections, since this is a condition that can meet these incurred infection with HPV HR *clearance* not naturally resolved by human host .

1.6 Molecular diagnosis of HPV infection

The release into the American/Asian market, rather than European of the *In Vitro Diagnostic Medical Devices* (IVDs) for the molecular diagnosis of HPV infections, is regulated by precise rules established through Control Authorities and sanctioned by a declaration of compliance with these standards, that is present in the documents accompanying the kit/product using two wording: *Food and Drug Administration approval* (***FDA approval***) and *Conformité Européenne* (***CE***).

The IVD applied to diagnostics of cervical cancer can be divided into 4 groups on the basis of the researched molecular marker, infection index, and molecular biology techniques used (Tab. 3)[23]. The first two groups include direct tests to highlight whether the viral nucleic acids, genomic DNA or virus transcripts (mRNA), while the other two groups comprise of indirect tests that detect the presence of HR virus for the effects it has on some metabolic cellular *pathways* related to the progression of CIN to carcinoma.

The GROUP 1 tests have a target specific viral genome sequence that are captured through two general methods: direct hybridization with probes or by target gene amplification (PCR) using appropriate complementary sequences called primers; GROUP 2 tests highlight E6/E7 transcript genes using labeled probes after reverse transcription of messenger RNA (mRNA) into the complementary DNA (cDNA) and their cyclic amplification (TMA, NASBA); GROUP 3 reveal some cellular antigens by immunohistochemistry and GROUP 4 some gene repeated sequences indicated by various studies as predictive markers of neoplastic transformation and progression by *in situ hybridization* (FISH).

Each group of IVDs has their advantages but also the critical issues that limit its use at different diagnostic levels: in cervical cancer screening, in ASCUS triage, in the follow up of patients treated etc.

GROUP 1 tests contemplate direct hybridization in liquid phase between a pool of *full length* complementary RNA probes to viral DNA. They are cheap, easy to use and theoretically detect the virus even in the case of incomplete genomes but actually do not typify the genotype, however they are not able to discriminate between single and multiple infections, nor estimate reliably the viral load that is calculated in a semiquantitative way by a chemiluminescence intensity gradient uncorrelated with the increase in the lesions severity.

The other GROUP 1 tests instead amplify the viral DNA using *primers* more or less *consensus* depending on the PCR protocols, they are generally more expensive than direct hybridization, require dedicated areas for the dangerous environmental contamination and in general are directed on individual regions of the HPV genome, especially on L1 region. This has the advantage of being a highly conserved sequence of the HPV genome, on this base it is possible to assign the genotype. So normally they are test capable of highlighting multiple infections by different types. However, much like the other areas of the HPV genome, the

region L1 may be lost due to integration[90]. It follows, on average, a high specificity test which can detect from 13 HR HPV to 37 HPV (or almost all types in anogenital tropism) in front of a certain proportion of theoretical false negatives due to the integrated genomes potentially undetectable for types HR persistent.

GROUP 2 instead belong to those kits that employ methodologies in Real Time, so in a semiquantitative way, amplify and detect using fluorophores marked probes and post reverse transcription in cDNA, mRNA from the E6/E7 viral oncogenes, whose levels are considered high prognostic index whether arising from integrated forms[97-102]. Although it is an expensive test that has the advantage of being in *total automation* even if in many cases do not define the infecting genotype but only positivity for HR HPV transcripts and are usually used as a second level test, namely in patients with a prior diagnosis and/or in follow-up post treatment.

GROUP 3 includes tests can immunoquantify some cellular antigens on histological samples using specific antibody indirect markers of cellular proliferation. They are the tests that show concomitant presence of $p16^{INK4a}$ pathological cell protein overexpressed under the effect of the E7 HPV protein in samples containing only integrated forms of virus[103-104] and Ki-67 nuclear protein, only observable in cells in active proliferation[105-106].

In fact, under normal conditions there is a concomitant expression of $p16^{INK4a}$ and Ki 67 within a single cell; the contemporary expression of the two markers is indicative of the deregulation of cell cycle control and resulting state of cellular transformation[107].

Lastly GROUP 4 belong to the tests which thanks to the Fluorescence In Situ Hybridization (FISH) technique, are able to identify region amplification mapping the human telomerase gene in locus 26 of the long arm in chromosome 3 (3q26) present only in transformed cells and not in normal tissue[108-109]. The physiological role of the telomerase is to repair chromosome ends that with cellular aging tend to shorten, it follows that a number of copies equal to or greater than 5 (trisomies are para-physiological exhibits and quadrisomies could be due to the cell replication phase) represents a significant marker of CIN progression to carcinoma. Indeed telomerase hyperactivity that causes duplication tends to immortalize the infected cell, with an elevated progression from CIN to cancer and with reduced possibility of lesion spontaneous regression.[110-112]

	Marker	Target	Technology	Test	HPV Typing
GROUP 1	**HPV DNA**	Full genome	Signal amplification Assay	*digene* HC2 HPV DNA Test	NO
				Cervista®HPV HR	NO
		L1	Target amplification assay	AMPLICOR®HPV	NO
				Cobas4800®HPV	partial (HPV16 & HPV18)
				Abbott RealTime High-Risk HPV	partial (HPV16 & HPV18)
				LINEAR ARRAY® HPV	YES
				CLART®HPV2	YES
		E1		Papillocheck®	YES
		L1+E6/E7		ProDect® CHIP HPV TYPING	YES
GROUP 2	**HPV RNA**	E6/E7	Target amplification assay	Aptima ®HPV Assay	NO
				PreTect HPV-Proofer	YES
GROUP 3	**Protein**	p16/Ki-67	Immunostain	CINtec® Plus	-
GROUP 4	**chromosomal aneuploidies**	3q26	(FISH)	oncoFISCH®	-

Tab. 3[23] Some of the IVD for screening cervical cancer and the diagnosis of HPV infection (partly taken from Schiffman 2011)

2. Scope of the research

The overall objective of the research was to investigate the possibility of improving an IVD, the ProDect® Chip HPV Typing[113], once the diagnosis of Human Papillomavirus infection, expanding, the number of genotypes potentially detectable and updating it, comparing them to the latest knowledge about the molecular biology and persistence mechanisms due to some high-risk oncogenic genotypes. Both aspects are in fact substantial for a kit that will present itself as a screening test for cervical cancer in the near future with regards to the prophylactic strategies towards some genotypes carried out worldwide through systemic vaccination programs, and the need to verify the real oncological risk of some HPV types, so far, defined as medium or low risk, but within particular geographical, social or health conditions may express a higher oncogenicity.

The molecular platform in question involves three phases: 1) total nucleic acid (DNA) extraction from the cervical-vaginal sample, 2) gene amplification by Multiplex polymerase chain reaction (PCR) with three pairs of biotinylated primers sequences directed on human β-globin gene, L1 region and the HPV E6/E7 region, respectively [114], 3) detection of amplicons obtained, labeled with biotin, through reverse hybridization with specific probes. In fact in the appropriate conditions the amplifications are denaturing and hybridized with complementary probes adhering to a plastic support said *biochip* or simply *chip*; follows a colorimetric reaction, streptavidin-mediated, which leads to the formation of a brown precipitate at the reactive probe. Each positivity creates different designs on the biochip (*patterns*) that are captured by a reader, the *ProDect® BCS Biochip Reader*, analyzed and interpreted by a dedicated software that explicit the diagnostic implications related to the different results obtained, in a report [115-117].

Assuming that the bcs Biotech Srl (www.biocs.it) platform IVD ProDect® CHIP HPV TYPING[113] employs a biochip type 5x5, or with 25 positions available for the same number of probes, the aim of the research was to design, develop and validate a biochip type 6x6, with 36 positions available for additional gene targets, hereinafter referred to as HPV CHIP PLUS, all details will be described in the results section.

For this project it was necessary to deepen the basic knowledge of HPV molecular diagnostics and implement a series of technological innovations and procedures for the new biochip

preparation. An increased number of genotypes detected, genomic regions characterizable with equal genotypes at high oncogenic risk, and the cross reactions between similar gene target have been disposed of, minimizing the nonspecific, without reducing the new platform sensitivity. Sensitivity and specificity are, in fact, two reverse concepts and in a multi target system must strike a point of compromise, especially if the platform must have a high negative predictive value (NPV), which must be attributed, in a sample, with reasonable certainty negativity for HPV infection.

The research was conducted in accordance to the: Directive 98/79/CE that indicates the criteria of the *Conformité Européenne (CE Mark)* for in Vitro Diagnostic Medical Devices (IVD)[118], UNI EN ISO 9001:2008 relating to quality management systems, UNI CEI EN ISO 13485:2012 relating to research and development of medical devices and UNI EN ISO 14971: 2012, concerning the application of management risk to medical devices.

The research validation step was designed, even though with reduced numbers, inspired by the European Commission Decision of 27 November 2009 amending Decision 2002/364/CE on Common Technical Specifications (CTS) for in vitro diagnostic medical devices in which establishes the need to test, in parallel with the test already CE Mark, any new kit contemplating a number of negative samples, positive for the sought target and/or potential interfering with it, such as the commensal flora or other pathogens due to infection.

The study was divided into three main phases in order to achieve the specific objectives and results reported below more extensively (Tab. 4).

Phase	Specific Objective	Activity	Result
I	Ideation 6x6 chip with 36 positions available for the deposition of as many informative probes complementary to HPV sequences and human	Design/verify HPV sequence System calibration for the spotting chip and acquisition patterns Probe calibration [ ] / Positive synthetic controls (PC) Verify and test chip 6x6 batches Check complete platform of cell lines	Biochip 6x6 prototype, 36 dots Define standard spotting conditions / reading software for 6X6 Chip performance with PC Pre-tested batches to run search Performance platform complete with cell lines
II	Cloning HPV genomic sequences from clinical isolates (L1 - E6 / E7), acquisition WHO Standard	Selection clinical samples HPV+ Cloning target sequences Check inserts Performance verification testing assays of cloned products from scratch and WHO standard	18 cloned HPV: 13 cloned HPV L1 with E6 / E7 5 cloned HPV L1 Performance platform complete in detecting multiple sequences of cloned HPV and WHO standard
III	Preliminary validation of the diagnostic capabilities Test	171 samples from Benin (Africa)	Preliminary data regarding the Diagnostic Performance

Tab. 4 - Articulation schematic research stages, objectives, activities and results

3. Materials and Methods

3.1 Biochip preparation

On a white polystyrene support, appropriately chemically modified (biochip), width the dimensions of a 96 well microplate, filed with aqueous oligonucleotide sequence solutions between 30-40 nucleotides (*probes*) in length, according to a precise *pattern*. Microdrops of these solutions, with sized and calibrated volumes, were deposited by means of a spotting machine (SCIENION AG Volmerstraße 7b D-12489 Berlin, Germany) equipped with a piezoelectric system[119], or an upright micro needle immersed in an electromagnetic field which reduces the phenomena of surface tension limiting the diffusion of each drop outside of the squares contours, constituting an ideal grid deposition, bounded in red (Fig. 10A-B), in which the black dots represent download areas, 36 in the present project.

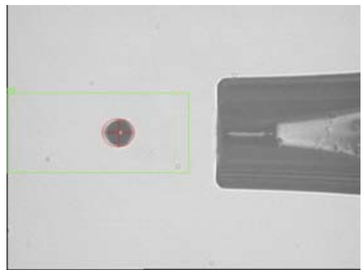


Fig. 8A Droplet from the needle

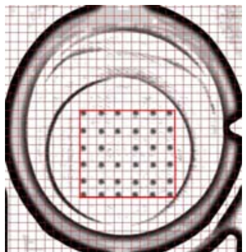


Fig. 8B Grid positions on the chip probes

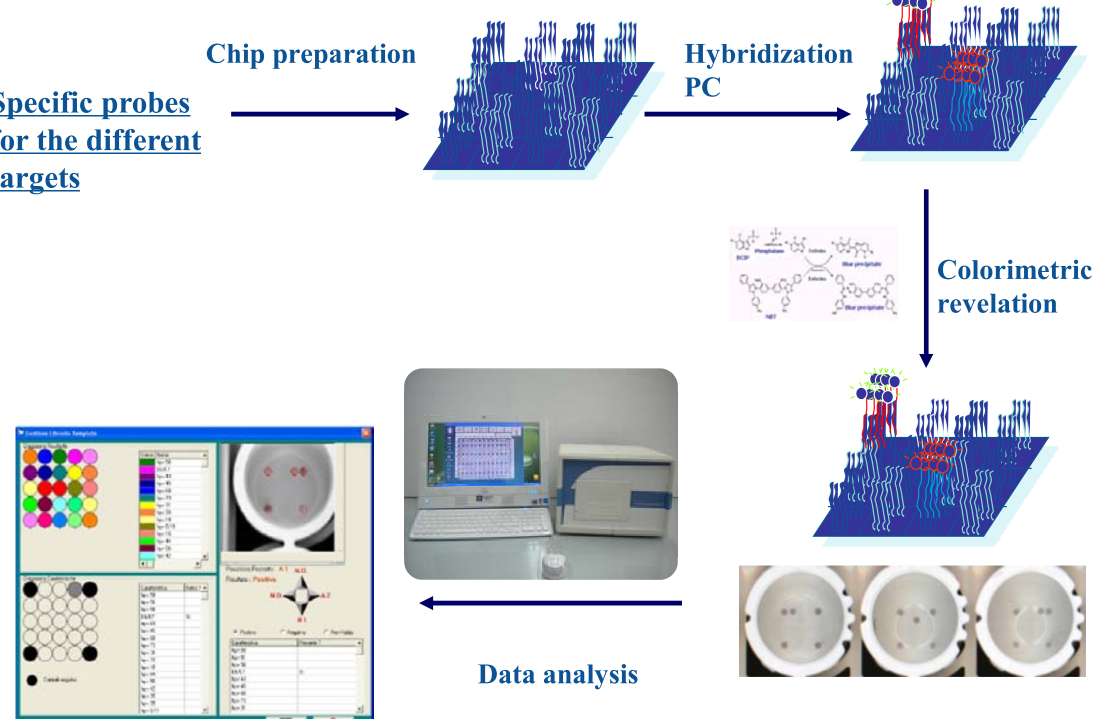
All HPV probes, covalently bound to the well, are designed to be complementary to specific *L1 region* sequences of the most prevalent high, medium and low HPV oncogenic risk for a total of 31 genotypes characterizable, and *E6/E7 region* sequences of the 3 major high and medium HPV oncogenic risk groups. On the two opposite grid corners, a probe placed in duplicate, reveals an internal sequence of the *human β-globin gene*, which acts as a process control region for the total DNA extraction of the original sample represented by the cervical-vaginal spatula. At the other two opposite corners, another probe called *Hybridization Control*, was designed to have no homology with human origin or viral infection sequences, and serves as a control for the physical-chemical conditions of the chip colorimetric assay detection.

3.2 Essay detection biochip

The optimal concentration [ ] for each probe has been defined on the basis of the specific signal obtained by hybridizing, under appropriate stringency conditions, scalar concentrations of complementary nucleotide sequences to the various probes having a biotin molecule in the 5th position, referred to as the Positive Controls (PC), so as to detect the hybrid *Probe: PC* using traditional colorimetric methods, as shown schematically in Fig. 9.

Briefly, 20 µl of a Tris EDTA solution at pH8 (TE pH8) of PC and 100µl of hybridization buffer are incubated for 1 hour at controlled temperature and shaking on the chip. After washing, followed by the addition of a Streptavidin Alkaline Phosphatase conjugated (AP) solution (30 'incubation at room temperature), which, after further washing, facilitation of the colorimetric reaction required for detection due to the addition of 5-bromo, 4-chloro, 3-indolylphosphate (BCIP)/Nitro-Blue Tetrazolium (NBT) enzyme substrate for 15 minutes at room temperature and obscured from light. The complementary sequence presence to the probe (PC) or the target in the original sample prior target amplification by Multiplex PCR is shown by the formation of a brown color precipitate (dot) corresponding to the relative specific probe.

Fig. 9. Colormetric revelation, pattern analysis via ProDect® BCS Biochip Reader



3.3 Synthesis of oligonucleotides

All oligonucleotide sequences used were designed based on the sequences deposited at the National Institute of Health (NIH Nucleotide GeneBank: http://www.ncbi.nlm.nih.gov/entrez/). Sequences with a > 90% homology degree, compared to all sequences reported for that region and in particularly for that HPV genotype in the database, have been synthesized by AB Applied Biosystems 3400 DNA Synthesizer multichannel system on columns to 1µM, desalted overnight in a 30% ammonium solution, and then purified by HPLC (Gilson 360) obtaining solutions with purity degrees of OD 260:280 = 1.8-2.0 and an early one [ ] of 100 ng/µl. For the PC sequences were employed similar synthesis procedures, taking care to place them in 5th position, as a modified base, a biotin molecule. Starting from the stock to 100 ng/µl for Positive Controls were prepared scalar solutions of each one (2.5 ng/µl, 0.5 ng/µl, 0.125 ng/µl, 0.05 ng/µl, 0.0115 ng/µl, 0.00125 ng/µl, 0.000125 ng/µl) which then were evaluated individually on all probes adhering to the biochip in order to verify the specificity and sensitivity of each probe and the set of probes deposited on the chip. After colorimetric detection on each chip was evident the specific signal expected and have been shown any signs of cross-hybridization with additional probes on the chip. For each probe was defined both the PC minimum detectable concentration and the PC maximum concentration which does not determine the occurrence of non-specific signals.

3.4 Biochip reading system

For the acquisition and analysis of the results or patterns that are formed on the chip as a result of the precipitation of the dye in correspondence to the reactive probes was an integrated platform ProDect® BCS Biochip Reader (bcs Biotech Srl). It is formed by a sliding carriage and two optical modules that moves on Cartesian axes XY, formed by reflecting lens and focusing, matched to the same number of the linear Charge Couple Device (CCD), that are independent and symmetrical. Through a process of controlled scanning by a Central Processing Unit (CPU), acquires the images in parallel and high resolution[116]. These are transmitted from an interface such as a USB port to a computer, where a software will reassemble them into a single image which is processed and analyzed with respect to predefined patterns of interpretation.

3.5 Molecular diagnosis

3.5.1 Samples

Amplified DNA extracted from cervical-vaginal spatulas were tested on the prototype biochip developed. All samples were exposed to previous molecular screenings for the detection of HPV with various CE Marked tests and/or sequencing for confirmation of the genotype, as well as cytological screening by PAP TEST. The degree of dysplasia was attributed according to the Bethesda classification 2001[120], namely: "*atypical squamous cells*" (ASC) subdivided in 1) ASC of "*undetermined significance*" (ASC-US), e 2) "*cannot exclude high-grade squamous intraepithelial lesion*" (ASC-H); "*low-grade squamous intraepithelial lesion*" (LSIL); "*high-grade squamous intraepithelial lesion*" (HSIL); "*squamous cell carcinoma*" (SCC); *adenocarcinoma in situ* (AIS); "*atypical glandular cells*" (AGC), the latter further classified into atypical "*not otherwise specified*" (*AGC-NOS* AGC-NOS source endocervical, endometrial, or glandular) or probably neoplastic (*AGC favor neoplastic* origin endocervical or glandular). The amplifications were acquired by PCR multiplex with the ProDect® Chip HPV Typing kit (bcs Biotech Srl). In short, each reaction was performed in duplicate on a volume of 50 µl: 5 µl of the DNA extracted with the ProDect® HPV Extraction kit and 45 µl mixture containing KCl, Tris-HCl, $MgCl_2$, Taq DNA polymerase, dNTPs, and three pairs of biotinylated primers. GP5+/GP6+ pair of primers amplified fragments of the 143 bp from the L1 region of HPV, the 1M/2R pairs of the E5/E7 region from the HPV HR (whose target depending on the genotype vary from 233-268 bp) and the third pair of primers amplifying a sequence of 178 bp of the gene of the human β-globin. PCR applied protocol, reagents and amplified storage conditions follow kit recommendations. Only one aliquot, after PCR, was used for the diagnosis on the 5x5 chip contained in the ProDect® Chip HPV Typing kit, the other one was stored at a -20°C controlled temperature to reduce degradation/fragmentation risk and it was used subsequently for the preliminary validation of 6x6 biochip in development, hereinafter referred to CHIP PLUS. This new platform used to reanalyzed 171 amplified obtained from African origin samples, already collected and studied in an International cooperation project funded by the Autonomous Region of Sardinia. The project was entitled *Screening of human papilloma virus (HPV), and prevention of cervical cancer in women from city of Cotonou and Benin* coordinated by the Department of Cytomorphology at the University of Cagliari in which bcs Biotech Srl participated as a partner [113]. All histological, cytological and molecular preliminary samples data, used in this research were received and processed by the

Cytomorphology Department of the University of Cagliari. The study protocol was approved by the local ethics committee of Benin and informed consent has been given to all parties in accordance with the World Medical Association, Declaration of Helsinki.

In order to confirm, when possible, the results obtained with the ProDect® Chip HPV Typing [113], amplification products were also sequenced using the Big Dye Terminator 3.1 kit (Applied Biosystems). The sequences were read on the ABI PRISM 3100-Avant sequencer (Applied Biosystems).

3.5.2 ProDect® Chip HPV Typing (bcs Biotech S.r.l.)

The data obtained by the new biochip were compared with those obtained using the ProDect® Chip HPV Typing Kit[113], IVD analogous method, which detects, besides the human β-globin gene as a control, 19 genotypes on the basis of the sequence of HPV L1 (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, 6/11, 42, 43, 44) and some HPV at high-medium oncogenic risk[23] on the E6/E7 region, through a probes pool laid in a single dot. The different targets are amplified by Multiplex PCR. 5µl of total DNA extracted from biological samples are added to 45 µl of ready to use mix (1X PCR buffer set, 0.2 mM dNTP's, 2 mM $MgCl_2$, and 2.5 U of Taq polymerase) also containing three pairs of primers: one complementary to the L1 region, one to E6/E7 region of high-medium oncogenic risk types and the last one to a human β-globin gene. Each amplification was also assessed by electrophoresis (80V - 48mA) on 2% agarose gel (Agarose, MP Biomedicals, Inc.) obtaining for each PCR reaction three different amplicons: 143 bp for the L1 region, 178 bp for the β-globin and between 233-268 bp for the high-medium oncogenic risk types of E6/E7 region. At the end of the PCR, 20 µl of the amplified denatured are then dispensed in each well containing 100 µl of a Hybridization Buffer (80 µl) and Hybridization Control (20 µl) solution. Followed by 1 hour at 45°C in a shaking controlled incubation, then 3 biochip washes with 100 µl of Wash Buffer, at the end of which a solution is dispensed consisting of 100 µl of Blocking Buffer 1X and 0.25 ml of streptavidin conjugated with an alkaline phosphatase (Strep-AP) for 30 minutes at room temperature. After another 3 washes proceeded with the addition of 100 µl of Revelation Buffer comprising chromogenic substrate (BCIP/NBT). Once an incubation for 15 minute at room temperature and protected from light, and a final wash, the test result is displayed on the well bottom which constitutes the biochip.

Indeed, the presence of any target sequence of the three primers pairs in the starting sample, is shown by the formation of a brown color dot in relation with the probe which has hybridized the complementary sequence to one or more amplicons between the three, which possibly formed during Multiplex PCR from each DNA.

3.5.3 Digene® HC2 HPV DNA Test (Qiagen)

The samples included in this study were also evaluated using Digene® HC2 HPV Test (Qiagen)[121-124] as the reference test, considered the gold standard for HPV infection primary screening. It is a signal amplification method in a liquid phase which starts to release total DNA contained in the biological sample by alkaline lysis and makes it suitable for hybridization with RNA complementary pool probes of 13 HR HPV types [pool HR: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68] and 5 LR HPV types (LR pool: 6, 11, 42, 43 and 44). RNA/DNA hybrid resulting are captured on the wells surface of a microtiter plate coated with specific hybrids antibodies. Follows disclosure with the same antibodies conjugated with alkaline phosphatase. Each antibody binds various molecules of alkaline phosphatase and more conjugated antibodies bind to each captured hybrid. The enzyme substrate addition mediate a chemiluminescent reaction; radiation is quantified in Relative Light Units (RLU). The emitted radiation intensity indicates DNA target presence or absence in the sample compared to a threshold value (cut off) related to suitable calibrators supplied with the test: to cut off ≤ 1 sample is negative, to cut off >1 is positive for HPV DNA sequences. The test, however, lacks a compliance audit of sampling thus only owns calibrators feature test based to HPV positive controls with respect to which is assigned positivity/negativity signal, while no calibrator investigates the proper specimen collection in terms of quality and quantity. It is an analysis with high reproducibility even if the signal intensity that is obtained cannot be correlated with the viral load or the cytological lesion severity. The test is also not distinguishing between the different HPV genotypes, except between groups of oncogenic risk, cannot establish any co-infections between different types within the same group.

3.6 Cell Lines

Three cell lines[125-129] were purchased from ATCC/LGC catalog to verify the diagnostic capabilities of the complete platform, rather from the extraction phase: C-33 A (ATCC® CRM-HTB-31™), HeLa (ATCC® CRM-CCL-2™), Ca Ski (ATCC® CRL-1550™). Each line was propagated following the indication in the accompanying form provided. Scalar dilutions

with a 1:10 ratio of the various lines were prepared in a cytological transport medium (as ThinPrep®) in order to obtain, in principle, 10^6, 10^5, 10^4, 10^3, 10^2 and 10^{-1} cells/ml. From each dilution, and for each line, were extracted total DNAs, according to the ProDect® HPV Extraction kit (bcs Biotech Srl) protocol that exploits the osmotic cells lysis, protein digestion with Proteinase K and total DNA alcoholic precipitation, or according to the purification and elution column Qiamp DNA Mini Kit (QIAGEN) protocol. The extracted DNA was submitted to the diagnostic procedure listed in the ProDect® CHIP HPV TYPING (bcs Biotech Srl) kit protocol. All these lines were derived from uterine cervix carcinomas, however C33A cells are a line obtained and stabilized from a uterine cervix carcinoma negative for HPV DNA/RNA, Ca Ski contains about 600 copies of HPV-16 genomes per cell and Hela about 50 copies of HPV-18 genomes per cell. Using these numbers, conventionally, 50 ng of DNA (about 1.510^4 cell) extracted from Ca Ski and Hela cells contain circa 9 x 10^6 copies of HPV16 and 7.5 x 10^5 copies of HPV18, respectively. It shows that the dilutions tested theoretically contains from 610^8 to 600 total copies of HPV16 and from 7.510^5 to 50 total copies of HPV18, respectively.

3.7 Cloning DNA HPV types with high and low oncogenic risk

All clones, nominated BCS clones, were prepared in the bcs Biotech Srl confined area, authorized by the Ministry of Health for the production of Genetically Modified Microrganism (MOGM). Particularly in the context of this research, plasmid vectors were inserted into bacterial cultures of E.coli, containing a viral DNA amplified insert[130]. 5 µl of extracted DNA from positive samples for individual infections by ProDect® CHIP HPV TYPING kit were amplified with primers of the specific infected genotype. Primer pairs were designed with reference to the genotype sequences deposited at the National Institute of Health (NIH Nucleotide GeneBank: http://www.ncbi.nlm.nih.gov/entrez/), in order to obtain amplification products, and then inserts, where possible, > 3.0 KB for high-medium risk types which, clockwise, included from ORF L to ORF E6/E7 or for some low-risk types a > 300 bp region internal to the L1 sequence. Each amplification reaction contained: 0.5 µM of each primer, 1X PCR buffer set, 0.2 mM of dNTPs, 2 mM of MgCl2, and 2.5 U of Taq polymerase. Each amplified was assessed by electrophoresis (80 V-48 mA) on 2% agarose gel (Agarose, MP Biomedicals, Inc.). 1 µl of the amplification product was then cloned into a pCR2.1 plasmid vector of *E.coli* by TA Cloning® kit (Invitrogen Life Technologies). The Invitrogen technology relies on the ability of T4 DNA Ligase to mediate ligation between an

amplification product with cohesive ends (blunt end) and a pCR2.1 vector, with complementary and cohesive sites creating a larger insertable plasmid in competent *E.coli* cells. Following seeding in agar plates containing a selective differential medium, then bacterial colonies containing the plasmid with the desired insert, may be isolated and appear in translucent white compared to those containing only the vector without the amplified insert (Fig. 10). This screening method is called white/blue screening. The color difference is due to the gene LacZα in the recombinant colonies, present in the pCR®2.1 vector, the interruption due to the amplified sequence insertion are no longer able to metabolize the galactose analog of X-Gal.

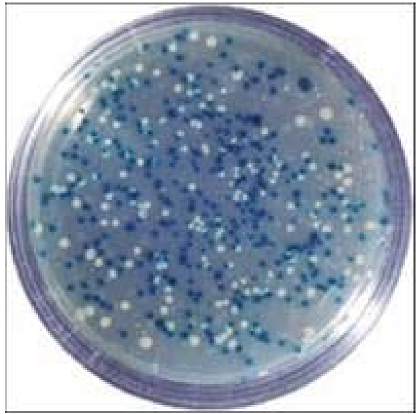


Fig. 10. Colonies including or excluding the plasmid with the insert according to the white / blue screening method

For the plasmid DNA purification was used the GenElute™ HP Plasmid Miniprep Kit (Sigma-Aldrich). Each clone was quantified by spectrophotometer (Thermo Electron Corporation - Biomate3) after dilution 1: 400 (2.5 µl extract plasmid + 997.5 µl of TE). No. of copies/ µl of plasmid preparations has been calculated as follows:

$$\frac{copies}{\mu l} = \frac{OD}{(PM_i + PM_p)} x\, 181 x\, 10^{14}$$

where PM_i = is the number of bp in the insert

PM_p = is the number of plasmide bp (pCR®2.1 è di 3912 bp)

This formula was derived by mathematical treatment by the following formula:

$$\frac{\text{copies}}{\mu l} = \frac{\frac{g}{\mu l}}{PM} x \text{ NA}$$

where:

g/µl of extracted DNA = OD x f.d. x 50 x 10^{-9}

PM = (PM_i + PM_p)x2x330*

NA° = 6 x 10^{23}

*weight takes into account that DNA is double-stranded and assumes the weight of a nucleotide equal to 330 daltons

° Avogadro's number = number of molecules contained in a mole

The equivalent copies are calculable by multiplying the number of copies/µl *ml (1000) but in the case the clones, with the exception of the WHO clones, it was preferred to express the concentration [ ] in copies/µl. The isolated clones were evaluated by Restriction Endonuclease EcoRI and PstI (Roche Diagnostics) on 2% agarose gels and gene sequencing. For each clone a restriction map were designed using the reference sequences and Webcutter 2.0 software. Once the expected sequence accuracy confirmed, a maxiprep or bacterial culture expansion and plasmid DNA extraction and purification containing at least 10^{10} copies/µl of target sequence for each clone was prepared. Besides the prepared clone by bcs Biotech, the HPV CHIP PLUS developed, was also evaluated using the Standard cloned HPV16 and HPV18 of the World Health Organization (WHO) prepared according to the guidelines given in the WHO Human Papillomavirus laboratory manual ed. 2009, arranged and distributed by the National Institute for Biological Standards and Control (NCBI) http://www.nibsc.org/. These standards, called henceforth WHO16 and WHO18 were diluted and manipulated as described in the accompanying form[131-132]. The ProDect® CHIP HPV TYPING amplified, under the External Quality Audits (EQA) of WHO HPV LabNet Global Reference Laboratory in 2011 distributed by Equalis AB[133] program were evaluated on CHIP HPV PLUS[134].

4. Results

Over three years, this research led to the development of a diagnostic system for innovative research for broad spectrum and molecular characterization of the main Human Papillomavirus infecting the anogenital area, potential etiologic agents alone, or in co-infection among them, of preneoplastic and neoplastic lesions. In summary, the path in several stages brought:

- design of the new chip, hereinafter referred to HPV CHIP PLUS, to be used in combination with the amplification mix of the ProDect® CHIP HPV TYPING[113], kit already CE marked, and promoted by bcs Biotech Srl for cervical cancer screening. This system currently types only 19 HPV genotypes and generally reveals the presence of E6/E7 sequences related to the virus at high-medium oncogenic risk (HR) represented by a single dot on the chip for this region.

- the creation of a pre-series batches HPV CHIP PLUS prototype

- verification of diagnostic capabilities using synthetic DNA targets (PC), cloned DNA, DNA extracted from immortalized cell lines and DNA amplification by cytology samples from women with different *grading* of cervical dysplasia, infected or not by HPV.

Below description of the individual results and the evidence to support them.

4.1 New HPV CHIP PLUS

Specific genotype probes with HPV high anogenital tropism and for the major subtypes have been designed in order to realize a new prototype chip capable to typify on the basis of the L1 region sequence no.21 in high-medium oncogenic risk genotypes and no.10 low-risk types, alone or in co-infection between them, (HPV 6, HPV 11, HPV 16, HPV 18, HPV 26, HPV 30/40, HPV 31, HPV 33, HPV 35, HPV 39, HPV 42, HPV 43, HPV 44, HPV 45, HPV 51, HPV 52, HPV 53, HPV 56, HPV 58, HPV 59, HPV 66, HPV 67, HPV 68, HPV 70/72, HPV 73, HPV 81, HPV 82, HPV61/CP6108) and at the same time to monitor the presence of sequences in the HPV E6/E7 region (always preserved even in the case of viral integration) of high-medium oncogenic risk types being able to discriminate them in three main groups:

E6/E7-HPV18/45, E6/E7- HPV33/52/58/56/66/67/82/59, E6/E7-HPV16/35/31. Three dots on the chip, in fact, represents a consensus of probes with different mixtures and/or specifications capable of reacting with the amplified E6/E7 region of the main HPV high-medium oncogenic risk types involved in cervical carcinoma. The presence of a positivity in the chip for one or more of these *dots*, in the absence of a specific typing with direct probes on the L1 virus region highlight, anyway, the positivity of the sample for high-medium risk HPV and allows increased levels of attention on cytologic evaluation or patient follow-up.

The complete prototype test, named ProDect® CHIP HPV TYPING PLUS, includes the necessary reagents to amplify and reveal via reverse hybridization for each sample:

1. L1 region of 31 DNA HPV (typing region),

2. E6/E7 region main high-medium oncogenic risk types (which highlights the presence even in the absence of a L1 region signal, guiding the user to one or more different risk groups)

3. Human β-globin, essential in order to have complete control of the entire process (monitoring of any extractions performed in a non-suitable or substances inhibiting PCR). It is calibrated to obtain evident signals only in the case of clinical samples (levy) whit sufficient DNA for the virus detection (productive, latent or abortive infection). The amplification threshold/detection of human DNA corresponding to at least 10^5-10^6 GenEqu/ml.

The interpretation of the chip can be performed by naked eye (the dot are colored and their position is clear) or with by *ProDect®BCS Biochip Reader* whose software has been suitably adapted in the context of the present research to read a 6x6 pattern. The system is designed to establish, automatically and objectively, if the result that appears on the chip is *valid* or *invalid* based on the probes or pool positions as they were placed in the ideal 6x6 grid (Fig. 11).

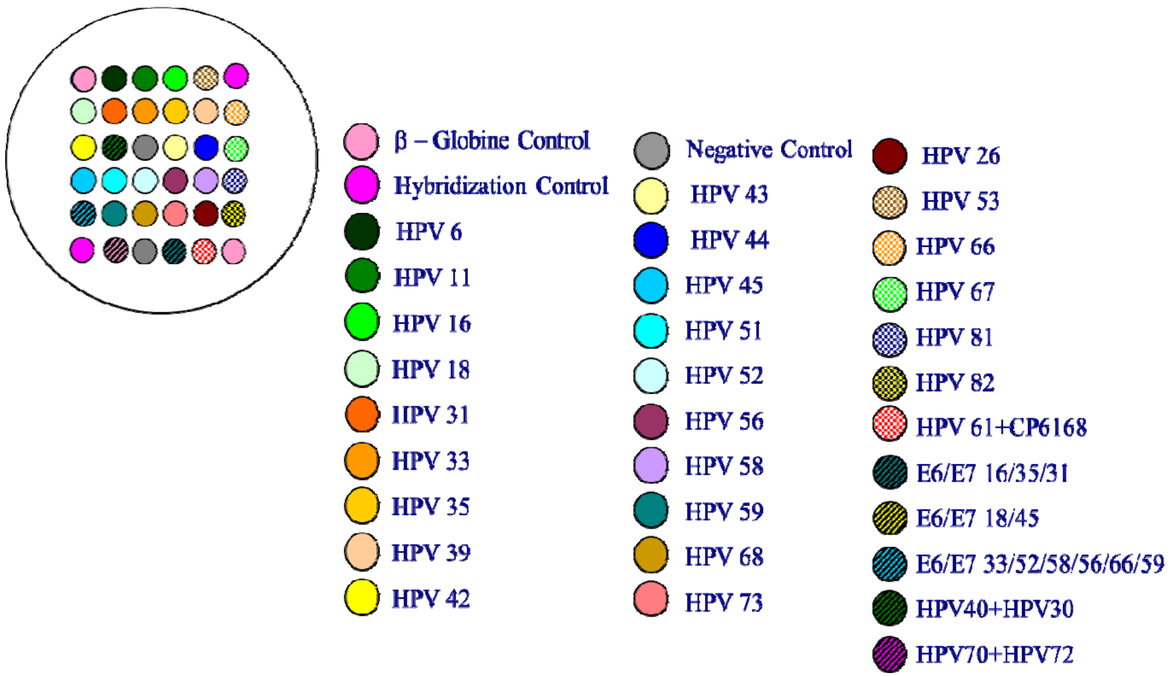


Fig. 11 – The new CHIP HPV PLUS scheme: Probe positions with respects to their genetic targets

A “VALID RESULT " can be due to different situations:

1. Negative sample in the presence of a withdrawal/DNA conforms to the analysis, the chip has only 4 signals (*dot*) at the four corners: N° 2 dot for β-globin gene, which acts as a control sequence for the monitoring of extraction/amplification and N° 2 dot relating to the hybridization control for the detection step monitoring (Fig. 12A).

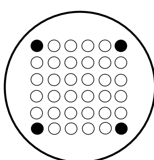


Fig. 12A – Negative sample

2. Positive Sample when, in the presence of a levy/compliant DNA, on the chip another *dot* appears, over 4 concerning levy/DNA compliance, which indicate the presence of HPV DNA in the starting sample, be characterized on the basis of the L1 and/or E6/E7 sequences. These dot can be single or multiple depending on the type of infection and from the different viral genotypes which support it (Fig. 12B).

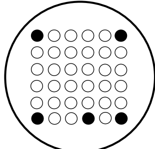


Fig. 12B – Positive sample

3. Positive Sample when, in the presence of a not conform specimen/DNA, chip presents 2 hybridization control *dot*, one or more signals relating to sequences/HPV types, but not any dot related to the β-globin (Fig. 12C). In this case the sample is positive even if the sample is incongruous (<10^5 GenEqu/ml), having evidently very high viral load.

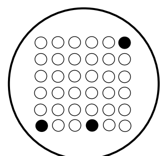


Fig. 12C - Positive sample

In conditions of a valid result and sample positivity for HPV are expected similar pattern to those reported below, which schematize possible single or multiple infections, Fig. 13.

Note that the developed prototype chip, due to the probes mixture designed and placed in 3 chip different positions having as target the HR HPV E6/E7 region, is able to detect these genotypes presence in different conditions, or when:

1) the ratio between HR HPV L1: E6/E7 copies number is much greater than 1, only L1 region, with its virus typing, can be amplified and detected.

2) is close to 1, it is possible to amplify and detect of both L1 region and E6/E7 region of the same virus or of different genotypes.

3) the ratio between HR HPV L1: E6/E7 copies number is much less than 1, it has amplification and detection of the single E6/E7 region with the corresponding classification of risk group (Group HPV18/45, Group 16/35/31, Group HPV33/52/58/56/66/67/81/59) on which basis three E6/E7 probe groups react,

4) in the case of co-infection between LR HPV, which generally show productive benign infection always typable on the L1 region as HPV6 or HPV11, and HR HPV persistent or in already integrated form viruses (see examples in Fig. 13).

The "INVALID result" can occur in different situations listed below.

Lack of one or more controls. If the software is not able to detect the *dots* of the four corner related to the controls (2 Hybridization Control + 2 for β-globin) it considers the outcome and the diagnostic test as "invalid". Positivity for only one point of each control affects the compliance of the test. The only exception might be made if the test presents 2 hybridization control signals, but not the β-globin ones, and one or more signals relating to sequences/HPV types. In this case the sample is certainly positive even if it is incongruous (<10^5GenEqu/ml) as the viral load is obviously high elevated (see example Fig. 12C).

Negative Control Signal. A precipitate at the relative negative control *dot* shows an incorrect hybridization. It follows that a signal in this position must not be present in any type of result. If operator or reader displays its presence, result will be "invalid". The test is considered INVALID even in the case of positivity for HPV sequences (Fig. 12D).

Signal attributable to identical HPV in all analyzed chips. The test is to be considered INVALID even in the case of the same attributable signal as the HPV probe is obvious in all samples, processed in parallel, in the same run. In this case the *carry-over* possibility exists or the dragging of the amplified HPV from one chip to another by the operator during the test execution. In this case the process needs to be restarted from the extract DNAs and repeat the session.

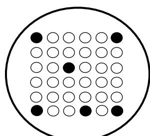


Fig. 12D – Example of an invalid sample

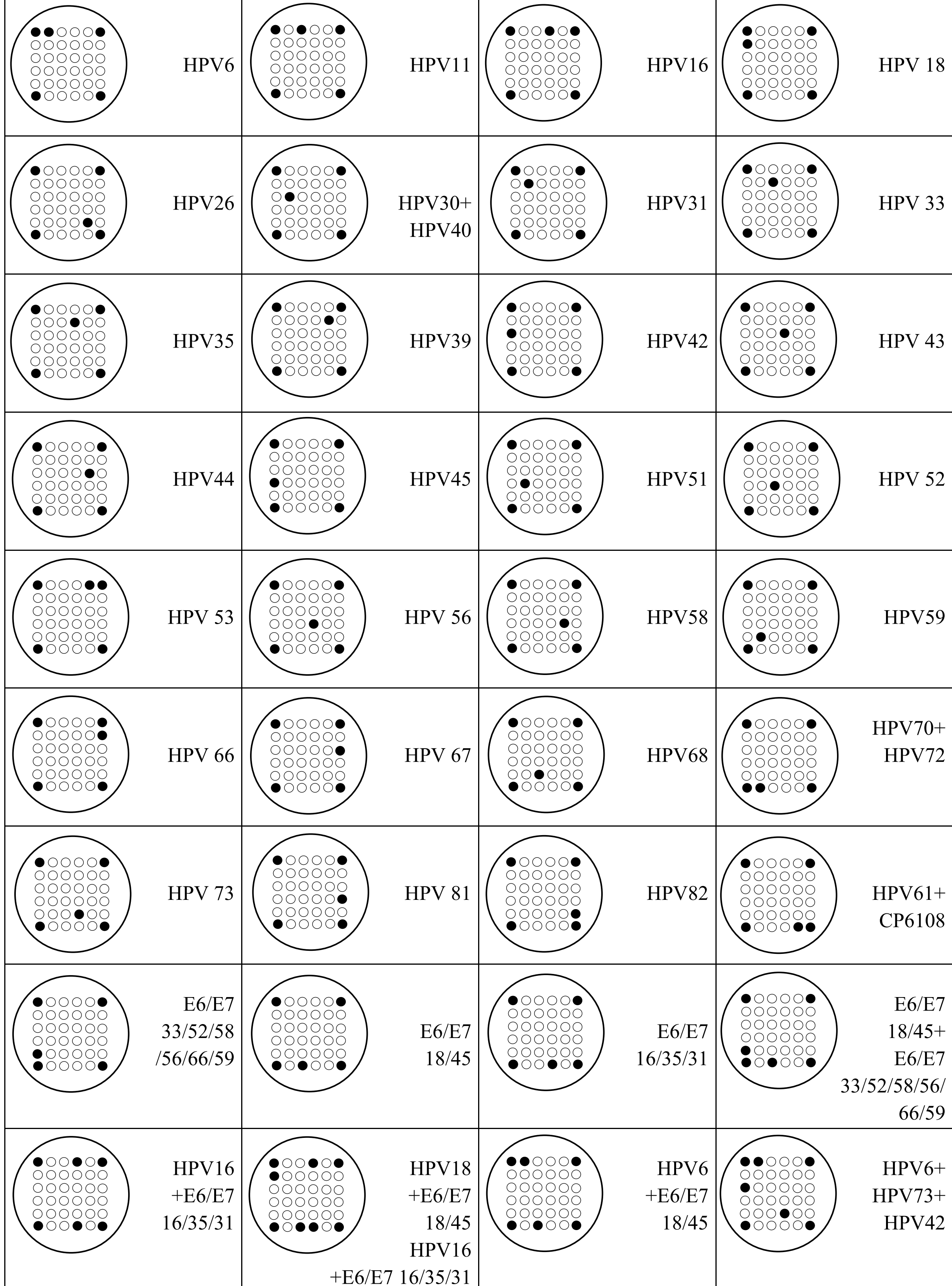


Fig. 13 - Examples of positive patterns on the CHIP HPV PLUS

4.2 CHIP HPV PLUS Sensitivity and Specificity

Eight preproduction batches of the new prototype CHIP HPV PLUS were prepared. Each one has been identified by a code and it has been evaluated separately from the other to verify the new chip performance and the variability between manufactured products at different times and with different probe batches in order to establish the method robustness also in relation to reliability requirements of the individual components, as required by the standard 98/79/CE for IVD.

4.2.1 Synthetic DNA

Evaluating chip sensitivity and specificity in detecting scalar concentrations of the Positive Controls (PC), synthetic complementary DNA to the probes immobilized on it, were highlighted by the following characteristics (Tab. 5):

- an analytical sensitivity equal to 0.0115 ng/µl for each probe, meaning the minimum detectable amount of each PC for all deposited probes is <11.5 pg/µl, this is due to some probes being able to detect up to 0.00115 ng/µl (~1 pg).

- an analytical specificity equal to 97.53%, with a standard deviation equal to ± 0.37 and 0.38% Variation Coefficient

This specificity arises from a medium of all specific information obtained by the informative probes in the panel, testing PCs at an elevated concentration, equal to 2.5 ng/µl. Note that in the detection assay a further 20 µl of PC was added to the concentration of 2.5 ng/µl, corresponding to ~50 ng total synthetic DNA single strand place to hybridize on each chip. This concentration was chosen in order to imitate an amount of HPV DNA target potentially high, comparable to that defined in the literature, to be contained in 50 ng of total DNA extracted from 1.510^4 Ca Ski and HeLa cells, or 9 x 10^6 HPV 16 copies and 7.5 10^5 HPV 18 copies, respectively. This assay condition was developed and put in to place to detect possible cross-reaction between probes and not complementary genotypic sequences. Indeed, in the case of high viral loads, an example referring to a single HPV type, any cross-reaction between probes would lead to an incorrect diagnosis of the sample, attributed to the presence of multiple signals, in multiple infections to a woman with an individual infection. Specificity

increases, 99.65%, for PC concentrations equal to 0.5 ng/µl, theoretically corresponding to containing ~10 ng of total DNA, if derived from Ca Ski ~1.8×10^6 copies of HPV 16 and 1.5×10^5 copies of HPV 18 if derived from HeLa cells, respectively.

Batch CHIP HPV PLUS	Analytical Sensibility	Analytical Specificity PC **2.5 ng/ µl**	Analytical Specificity PC **0.5 ng/ µl**
12132	0.0115 ng/ µl	97,44	99,8
11234	0.0115 ng/ µl	98,32	100
12136	0.0115 ng/ µl	97,22	99,2
12137	0.0115 ng/ µl	97,22	99,2
12138	0.0115 ng/ µl	97,29	99,5
12144	0.0115 ng/ µl	97,36	99,5
12145	0.0115 ng/ µl	97,73	100
12146	0.0115 ng/ µl	97,66	100
Media	**0.0115 ng/ µl**	**97.53**	**99.65**
DS	**-**	**0.37**	**0.32**
C.V.	**-**	**0.38%**	**0.32%**

Tab. 5 Analytical performance from different CHIP HPV PLUS lots using PC

4.2.2 DNA extracted from cell lines

The complete diagnostic platform, denominated ProDect® CHIP HPV TYPING PLUS, has been verified on cell lines using 8 pre-series batches of CHIP HPV PLUS product in combination with as many number of primer mix as batches. As shown in Tab. 6, C33A cells, HPV-uninfected, for dilutions with low cellularity $<10^4$ cells/ml (10^3-10^2-10 cells/ml) result as "invalid" samples, i.e. for all 8 batch chips, the DNA extracted, amplified and detected on the C33A chip, did not show a corresponding *dot* to the human β-globin. This conditional test is designed and implemented to alert the user when the test is not running a significant and sufficient amount of total DNA. This limit of the minimum necessary cells to assign a levy compliance takes into account different forms of HPV infection (productive, latent, abortive). Indeed samples with a fair amount of cells, i.e. ~10^5-10^6 cells/ml, are more representative of the cytological situation but especially allow greater statistical probability of virus detection in its more subtle forms (latent and integrated). When the same dilutions of C33A cells (10^3-

10^2-10 cells/ml) were evaluated on Hybrid Capture® 2 (HC2) HPV Test platform they were defined as negative. Note, however, that in this type of test there is no levy compliance check, or minimum start sample cellularity; the negativity diagnosis are attributable to the HPV sequences absence and not to the fair amount of collected and subjected to lysis material. ProDect® CHIP HPV TYPING PLUS platform, when tested using the HPV infected cell lines, detected viral sequences with the same diagnostic sensitivity and specificity with regards to the Hybrid Capture® 2 (HC2) HPV Test. No cross reaction (non-specific) occurrence was also detected with the extract containing 10^6 cells/ml both in the Ca Ski cells than on HeLa, while HPV sequences (both the L1 region of HPV-16, as well as the E6/E7region) were detectable in extracted DNA, theoretically from a dilution with a single Ca Ski cell with 10-100 HeLa cells (assumed that one Ca Ski cell contains 600 genomes of HPV16 and 1 HeLa cell 50 genomes of HPV18). When cells same dilutions were evaluated on Hybrid Capture® 2 (HC2) HPV test platform, reproducible positive signals were obtained with DNA extracted from at least 10 Ca Ski cells or 10 HeLa cells. The diagnostic sensitivity of this platform was then 6000 and 500 supposed genomes of HPV-16 and HPV-18, respectively. Some examples of biochips detection of DNA extracted from these cell lines (Fig. 14).

ProDect®CHIP HPV TYPING PLUS			
Line	**Sensitivity cells/ml**	**Result**	**Copies/ µl**
C33A	$<10^4$	Not Valid	0
C33A	$\geq 10^4$	Negative	0
Ca Ski	1	Positive HPV16	600
HeLa	10	Positive HPV18	500
Hybrid Capture ® 2 (HC2) HPV Test			
Line	**Sensitivity cells/ml**	**Result**	**Copies/ µl**
C33A	$<10^4$	Negative	0
C33A	$\geq 10^4$	Negative	0
Ca Ski	10	Positive HPV HR	6000
HeLa	10	Positive HPV HR	500

Tab. 6 The analytical performance of the different CHIP PLUS HPV lots using lines C33A, CaSki, Hela

Fig. 14. In the biochip C33A cells superior (10^5/ml) HPV result negative, note the only points of conformity of the samples and the test.
In the biochip Ca Ski cells inferior (10^5 cells/ml) HPV16 result positive, note the positivity for the region corresponding to the type HPV16 L1 and for the region E6/E7 corresponding to the high-risk group E6/E7-HPV16/35/31

4.2.3 HPV cloned DNA at high and low oncogenic risk

In this research context 18 plasmids containing viral DNA sequences of variable size were prepared, formed by as many clinical specimens HPV types (Tab. 7). 13 of 18 genotypes, including 11 HR-HPV, inserts above 3Kb comprising regions L1, URR, E6 and E7 were obtained. At the basis of this activity is the need to have available unlimited quantities of double-stranded DNA of high molecular weight similar to viral genomes found in real samples, these are to be used as additional controls for calibration of the amplification conditions and detection of multiple different gene targets test: L1 and E6/E7 regions of HPV and human β-globin. After checking the accuracy of each insert sequence, DNA clone dilutions in TE were tested undiluted or with the addition of ThinPrep®Cytolyt Solution (Hologic) preservatives solutions containing a [ ] of C33A cells equal to 10^5-10^6 cells/ml, in order to check sensitivity and specificity of similar calibrators of the diagnostic system for the quantity and quality of the biological samples. The assays were performed in parallel using the HPV-16 and HPV-18 WHO standards, whereby [ ] is expressed in International Units (IU) or Equivalents Genomes (GenEqu). As analytical sensitivity was shown detection limit of each control corresponding to its minimum concentration on the specific probe (Tab. 8). In many cases, the detection threshold was expressed as interval and not as absolute number considered that cloned title, or No. copies/μl was measured by spectrophotometric reading (which is then considered margin of error). Reported sensitivity refer to both L1 and E6/E7

genetic targets and represent the minimum amount of copies/µl or IU/ML of cloned detected in 100% of the 24 tests carried out for each scale (Fig. 15). Some cloned scales were also tested on the Hybrid Capture® 2 (HC2) HPV reference method, which does not employ gene target amplification (Tab. 9). As expected, a greater amount of HPV DNA target is needed for a positive signal by Hybrid Capture® 2 (HC2) HPV Test compared to the ProDect® CHIP HPV TYPING PLUS. Note that ProDect® CHIP HPV TYPING PLUS, reveals two gene sequences of HPV genome for all HR clones tested, confirming both HPV genotype and the belonging group with respect to the E6/E7sequence, while Hybrid Capture® 2 (HC2) HPV Test defines only positivity for HPV HR cutoff > 1.

The analytical specificity was calculated using samples obtained by mixing high titer clones with one or more low titer clones with 10^6 GenEqu/mL of C33A cells, referring to the threshold analytical sensitivity of each genotype, in order to evaluate at the gene amplification level and detection on the chip, inhibitions or higher affinity of the platform for some genotypes than others in the case of co-infection. In fact, a Multiplex PCR can arise interference phenomena between one or more genotypes[132]. As demonstrated by some examples below (Fig. 16) all HPV expected sequences, regardless of the clones mixed between them, if they were above the test detection limit, they were found on the new chip; different genotypes were correctly classified on the basis of L1 region direct probes. No cross-reaction was highlighted with genome sequences belonging to viruses, bacteria, fungi or protozoa often present in the female genital tract. (Tab. 10).

HPV type	ORFs	Genebank Accession Number	PCR product
6	L1	AF092932.1	426 bp
11	L1	M14119.1	448 bp
16	L1- E6/E7	K02718.1	3532 bp
18	L1- E6/E7	X05015.1	3647 bp
30	L1- E6/E7	X74474.1	3641 bp
31	L1- E6/E7	J04353.1	3523 bp
35	L1- E6/E7	M74117.1	3427 bp
42	L1- E6/E7	M73236.1	3394 bp
45	L1- E6/E7	X74479.1	3500 bp
51	L1- E6/E7	M62877.1	3535 bp
52	L1- E6/E7	X74481.1	3064 bp
56	L1- E6/E7	X74483.1	3554 bp
58	L1- E6/E7	D90400.1	3585 bp
59	L1- E6/E7	X77858.1	3544 bp
66	L1	U31794.1	448 bp
70	L1	U21941.1	454 bp
73	L1- E6/E7	X94165.1	3474 bp
81	L1	AJ620209.1	450 bp

Tab. 7 HPV sequences cloned list and insert size achievable with specific HPV type primers

ProDect® CHIP HPV TYPING PLUS			
Clone#lot	**Stock copies/µl**	**Sensitivity copies/µl**	
		reg.L1	**reg.E6/E7**
HPV16#01A	6.910^{10}	69-690	69-690
HPV18#03A	2.810^{10}	28-280	2800
HPV45#05	$2.0\ 10^{10}$	2000	200000
HPV6#02	$1.0\ 10^{11}$	100-1000	ND
HPV11#01	$1.3\ 10^{11}$	130-1300	ND
HPV35#19A	$4.7\ 10^{10}$	5-10	500-1000
HPV31#09A	$2.4\ 10^{10}$	$1\text{-}510^{4}$	240
HPV73#02A	$3.3\ 10^{10}$	33-330	ND
HPV51#08	$4.6\ 10^{10}$	4800	10000
HPV56#01A	$5.0\ 10^{10}$	5-50	50000
HPV59#11Al	$1.0\ 10^{10}$	10000	100000
HPV52#09	$6.9\ 10^{10}$	6900- 69000	6.9-69
HPV58#11	9.610^{10}	9600- 96000	9.6- 960
HPV66#02	$1.3\ 10^{11}$	10000	ND
HPV81#03	$7.7\ 10^{10}$	100-770	ND
HPV70#03	$1.0\ 10^{11}$	100-1000	ND
Clone#lot	**Stock UI/ML**	**Sensitivity UI/ML**	
WHO HPV16	10^{7}	10^{2}	510^{3}
WHO HPV18	10^{7}	10^{3}	10^{4}

Tab. 8 Test sensitivity of new kit with the bcs Biotech S.r.l cloned and WHO standards

Hybrid Capture ® 2 (HC2) HPV		
Clone#lot	**Stock copie/µl**	**Sensitivity copies/µl**
HPV16#01A	6.910^{10}	$1\text{-}3.5\ 10^{4}$
HPV18#03A	2.810^{10}	$2.8\text{-}7\ 10^{4}$
HPV35#19A	$4.7\ 10^{10}$	$4.7\ 10^{4}\text{-}10^{5}$
Clone#lot	**Stock UI/ML**	**Sensitivity UI/ML**
WHO HPV16	10^{7}	1.510^{4}
WHO HPV18	10^{7}	310^{4}

Tab. 9 Sensitivity of the reference test with some cloned bcs Biotech S.r.l and WHO standards

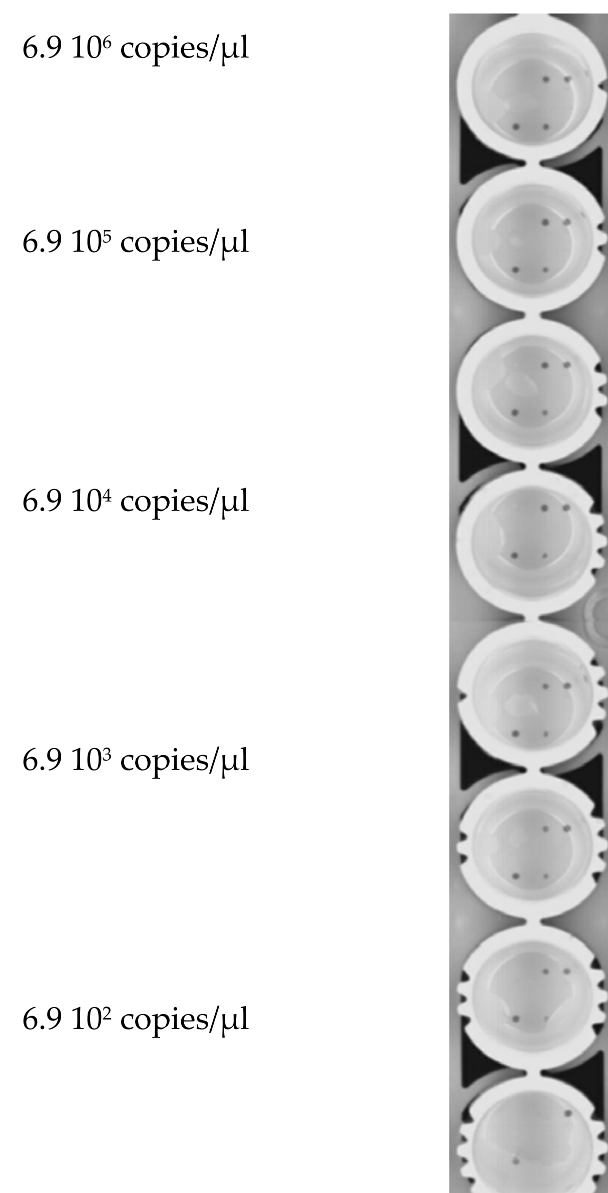


Fig. 15 Example of analytical sensitivity batch n° HPV CHIP PLUS 12144 assessed scale cloned HPV16 BCS

Organism		
Acinetobacter	Neisseria gonorrhoeae	Chlamydia tracomatis
Klebsiella pneumoniae	Gardnella vaginalis	Candida albicans
Bacteroides fragilis	Mobilincus curtisii	Tricomonas vaginalis
Bacteroides ureolyticus	Enterococcus faecalis	Herpes simplex virus 1
Bifidobacterium spp	Escherichia coli	Herpes simplex virus 2
Lactobacillus spp	Staphylococcus aureus	Cytomegalovirus
Mycoplasma hominis	Streptococcus agalactiae	Adenovirus
Mycoplasma genitalium	Corynebacterium spp	Staphilococcus epidermidis

Tab. 10 List of microorganisms that potentially could interfere

Clones: copies/µl	Clones: copies/µl	Observation
HPV59: 10^8 +	**HPV16**: 10^3	
HPV59: 10^8 +	**HPV18**: 10^4	
HPV59: 10^8 +	**HPV35**: 10^4 , **HPV45**: 10^4, **HPV56**: 10^4	
HPV18: 10^8 +	**HPV11**: 1.3 10^8 1.3 10^6 1.3 10^4 1.3 10^2	
HPV16: 10^8 +	**HPV59**: 10^8 10^6 10^4 10^2	

Fig. 16. Examples of cloned mixtures to mimic co-infections between the various HPV types with a high number of copies close to the detection threshold of the test

4.2.4 Extracted DNA from cytological samples

A preliminary assessment of the platform diagnostic performance was made on a total number of 171 amplified cytology specimens (Appendix - List 1) collected in Benin, Africa between 2009-2012 under the cooperation International project funded by the Autonomous Region of Sardinia, directed by Prof. Paola Sirigu, with the cooperation of bcs Biotech as a partner. Analyzed samples of PAP TESTs undertaken resulted with a normal cytological status of 68.4% (117), while 31.5% (54) had morphological alterations with variable gradation (Tab. 10), with a prevalence of high grade dysplasia among women aged between 30 and 50 years (Fig. 17).

Cytology	Age range					Total
	15-25	25-35	35-45	45-55	55-65	
Normal	14	29	33	28	13	117
LSIL	1	1	0	0	0	2
ASCUS	5	14	9	4	1	33
ASC-H	0	3	1	2	0	6
HSIL	0	1	4	1	0	6
AGC	0	0	4	2	0	6
SSC	0	1	0	0	0	1
Total	20	49	51	37	14	171

Tab. 11 Distribution of samples with respect to age and cytology

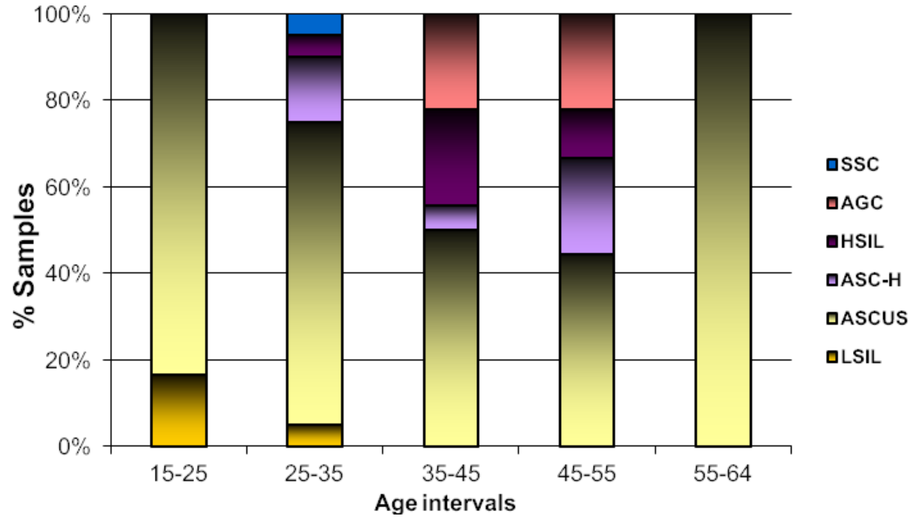


Fig. 17 Prevalence of dysplasia by age in the population sample analyzed

All samples were screened for HPV sequences using IVD ProDect® CHIP HPV TYPING kit, Hybrid Capture® 2 (HC2) HPV Test as a reference method and the ProDect® CHIP HPV TYPING PLUS platform under development.

The following data refer to the infection detection ability by the three kits with two different approaches:

- limit HPV types common to all three tests (Tab. 12A, 13A, 14A)
- with reference to the ability to attribute sample negativity/positivity for HPV infection, compared to HPV circulating globally in the analyzed population (Tab. 12B, 13B, 14B).

Limiting performance evaluation test only to common HPV types (18 genotypes detected by the probes pool of Hybrid Capture® 2 (HC2) HPV Test, and only to the positively infected samples by HPV HR: 16, 18, 31 , 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68, and LR HPV 6, 11, 42, 43 and 44) test on chip showed high indices of correlation with the reference test, Hybrid Capture® 2 (HC2) HPV Test, demonstrating comparable performance.

ProDect® CHIP HPV TYPING presents a Concordance Index (CI) by the reference standard method for sharing HPV types of 97.6% (167/171), a diagnostic sensitivity of 100% (64/64 agree) and a specificity of 96.2% (Tab. 12A).

ProDect® CHIP HPV TYPING PLUS prototype showed a Concordance Index (CI) by the reference standard method for sharing types of 98.8% (169/171), a diagnostic sensitivity of 100% (64/64 agree) and a specificity of 98.1% (Tab. 13A).

Obviously the correlation between the tests on CHIP, ProDect® CHIP HPV TYPING and ProDect® CHIP HPV TYPING PLUS is even higher, with reference to the common HPV sequences (19 genotypes detectable HPV HR: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 - LR HPV: 6, 11, 42, 43 and 44, and HPV E6/E7) equal to CI = 73 + 97/171x100 = 99.4%, (Tab. 14A).

	Hybrid Capture® 2 (HC2) HPV Test		Total
ProDect® CHIP HPV TYPING	+	-	
+	**64**	**4**	**68**
-	**0**	**103**	**103**
Total	**64**	**107**	**171**

Tab. 12A. Screening samples for 18 common types of HPV, ProDect®CHIP HPV TYPING vs Hybrid Capture ® 2 (HC2) HPV Test

	Hybrid Capture® 2 (HC2) HPV Test		Total
ProDect® CHIP HPV TYPING PLUS	+	-	
+	**64**	**2**	**66**
-	**0**	**105**	**105**
Total	**64**	**107**	**171**

Tab. 13A Screening samples for 18 common types of HPV, HPV ProDect® CHIP TYPING PLUS vs Hybrid Capture 2 (HC2) HPV Test

	ProDect® CHIP HPV TYPING PLUS		Total
ProDect® CHIP HPV TYPING	+	-	
+	**73**	**1**	**74**
-	**0**	**97**	**97**
Total	**73**	**98**	**171**

Tab. 14A Screening samples for 19 common types of HPV and region E6 / E7, ProDect® CHIP HPV TYPING PLUS vs ProDect® CHIP HPV TYPING

For all three tests parting from data contingency table, the Cohen's Kappa was calculated, using as a reference the interpretation guidelines that define each K value a different concordance grading between tests (K <0.2 = poor concordance; 0.21<K<0.40 = modest concordance, 0.41<K<0.60 = moderate; 0.61<K<0.80 = good; K> 0.81 = excellent).

All comparison between tests showed an excellent correlation (0.95 < Cohen's Kappa < 0.97). However, analyzing the sample on the basis of all positivity detected, and then expands the number of genotypes to be considered as actually circulating in the population, a number

of negative samples highlighted with the reference methods, are, instead, positive for HPV infection with the new prototype ProDect® CHIP HPV TYPING PLUS, even if the correlation between the various tests is still good or excellent (0.79 < Cohen's Kappa <0.89).

In light of such evidence, re-evaluating the ability of the different tests to diagnose HPV infection in the analyzed population, independently of its genotype, shows that:

- ProDect® CHIP HPV TYPING presents a concordance index (CI) compared to the standard reference molecular method Hybrid Capture® 2 (HC2) HPV Test of 94%, CI (161/171), a diagnostic specificity of 90.6% (97/107 coincident negative) and a diagnostic sensitivity of 100% (64/64 coincident positive). In particular, the IVD chip, employing the pattern 5x5, revealed 10 further positive samples classified as negative by Hybrid Capture® 2 (HC2) HPV Test, including 3 with positive PAP TEST (3 ASCUS) and 7 with apparently normal cytology , (Tab. 12B). Cohen's Kappa is equal to 0.875.

	Hybrid Capture® 2 (HC2) HPV Test		Total
ProDect® CHIP HPV TYPING	+	-	
+	**64**	**10**	**74**
-	**0**	**97**	**97**
Total	**64**	**107**	**171**

Tab. 12B. Screening infection: ProDect®CHIP HPV TYPING vs Hybrid Capture ® 2 (HC2) HPV Test

- Concordance Index between ProDect® CHIP HPV TYPING PLUS and the reference standard method Hybrid Capture® 2 (HC2) HPV Test was equal to 90% (154/171 concordant). In fact, 17 samples were classified as positive by the prototype; of these, 4/17 had a positive PAP TEST (1 HSIL-carcinoma and 3 ASCUS) and 8/17 showed positivity related to HPV genotypes cannot be classified by the Hybrid Capture® 2 (HC2) HPV Test without the specific probes, (Tab. 13B). Cohen's Kappa is equal to 0.79.

	Hybrid Capture ® 2 (HC2) HPV Test		Total
ProDect®CHIP HPV TYPING PLUS	+	-	
+	**64**	**17**	**81**
-	**0**	**90**	**90**
Total	**64**	**107**	**171**

Tab. 13B Screening infection: ProDect® CHIP HPV TYPING PLUS vs Hybrid Capture ® 2 (HC2) HPV Test

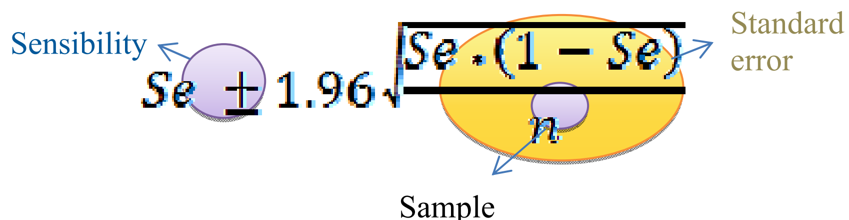
- Concordance Index between ProDect® CHIP HPV TYPING and ProDect®CHIP HPV TYPING PLUS in development was 95% (162/171 concordant). Also in this case 8/8 samples assigned as positive from the prototype showed HPV types undetectable by the IVD test without the specific probes (Tab. 14B). Cohen's Kappa is equal to 0.89.

	ProDect®CHIP HPV TYPING PLUS		Total
ProDect®CHIP HPV TYPING	+	-	
+	**73**	**1**	**74**
-	**8**	**89**	**97**
Total	**81**	**90**	**171**

Tab. 14B Screening infection: ProDect® CHIP HPV TYPING PLUS vs ProDect®CHIP HPV TYPING

Given the reduced sample size of the population, and being impossible to attribute sample positivity with the certainty on the basis of the full-blown disease, from the moment that an infection can occur, even in the presence of a normal cytologic condition, and could not rule out other HPVs circulating in the population that were not included among the 31 types detected by the prototype, individual platform performances, comparisons between tests were calculated at a confidence interval of 95%, according to the below exemplified formula and considering healthy negative and some infected samples resulting with at least two of the three methodologies used. We proceeded to consider a total of 89 samples as negative, including 76 normal, 13 seemingly pathological (1 HSIL, 1AGC, 2 ASCH, 9 ASCUS) by PAP TEST and 82 samples as positive of which 41 apparently with a normal PAP TEST. All positives were confirmed by gene primers specific amplification followed by reverse

hybridization with specific probes and/or sequencing. For each kit were calculated the diagnostic performance (Tab. 15,16,17).



$$Se \pm 1.96\sqrt{\frac{Se \cdot (1 - Se)}{n}}$$

Formula for calculating Confidence Interval of 95%

	Infected		Healthy	
Positive	**a:**	**64**	**b:**	**0**
Negative	**c:**	**18**	**d:**	**89**
True Positive	64	a		
False Positive	0	b		
True Negative	89	d		
False Negative	18	c		
Total Samples	171	a+b+c+d		
Sensitivity (SE) ± 0.064 (IC95)	**0,7805**	a/(a+c)		
Specificity (SP)	**1,0000**	d/(b+d)		
Prev. POS. Values (VPP)	**1,0000**	a/(a+b)		
Prev. NEG. Values (VPN)	**0,8318**	d/(c+d)		

Tab. 15 Hybrid Capture ® 2 (HC2) HPV Test performance

	Infected			Healthy
Positive	**a:**	**74**		**b:** **0**
Negative	**c:**	**8**		**d:** **89**
True Positive		74	a	
False Positive		0	b	
True Negative		89	d	
False Negative		8	c	
Total Samples		171	a+b+c+d	
Sensitivity (SE) ± 0.064 (IC95)		**0,9024**	a/(a+c)	
Specificity (SP)		**1,0000**	d/(b+d)	
Prev. POS. Values (VPP)		**1,0000**	a/(a+b)	
Prev. NEG. Values (VPN)		**0,9175**	d/(c+d)	

Tab. 16 ProDect®CHIP HPV TYPING performance

	Infected			Healthy
Positive	**a:**	**81**		**b:** **0**
Negative	**c:**	**1**		**d:** **89**
True Positive		81	a	
False Positive		0	b	
True Negative		89	d	
False Negative		1	c	
Total Samples		171	a+b+c+d	
Sensitivity (SE) ± 0.033 (IC95)		**0,9878**	a/(a+c)	
Specificity (SP)		**1,0000**	d/(b+d)	
Prev. POS. Values (VPP)		**1,0000**	a/(a+b)	
Prev. NEG. Values (VPN)		**0,9889**	d/(c+d)	

Tab. 17 ProDect® CHIP HPV TYPING PLUS performance

Performance calculated in this way means, therefore, the capacity of the individual tests to evaluate the actual biological sample negativity for HPV sequences (Tab. 18) refer to 31 genotypes classified by the prototype HPV ProDect® CHIP HPV TYPING PLUS, able to detect a further 14 and 13 genotypes respectively, with regards to those identified by the Hybrid Capture® 2 (HC2) HPV Test and with ProDect® CHIP HPV TYPING.

Note that in the analyzed population, there is a strong prevalence of HPV that fluctuates, depending on the test used for screening, between 25% for the Hybrid Capture® 2 (HC2) HPV Test and 47% for the ProDect® CHIP HPV TYPING PLUS in development (Fig. 18A, B, C). In a sample like this, given the high probability of finding HPV sequences, high prevalence of infection enhance the positive predictive value of all tests.

Performance of 31 HPV types	Diagnostic Sensitivity IC95	Diagnostic Specificity IC95	K	PPV	NPV
Hybrid Capture ® 2 (HC2) HPV Test	78.0%	100%	0.78	100%	83.0%
ProDect® CHIP HPV TYPING	90.2%	100%	0.90	100%	91.7%
ProDect® CHIP HPV TYPING PLUS	98.7%	100%	0.98	100%	98.8%

Tab. 18 Summary of the performance tests used for screening the same population

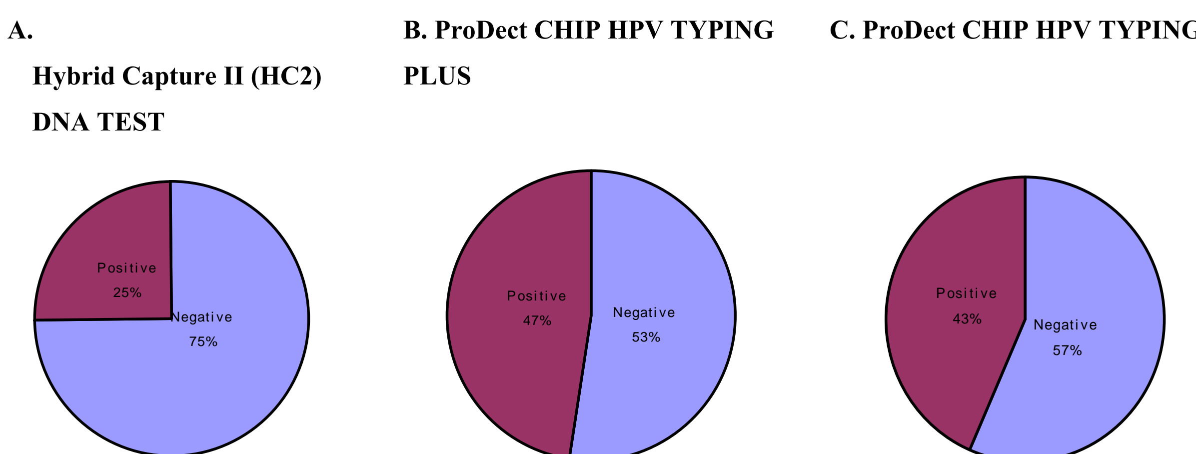


Fig. 18A,B,C Prevalence of apparent HPV infection in the population as a part of the screening test

Sample analysis by screening Hybrid Capture® 2 (HC2) HPV Test , widely used in all world for its practicality, in this study divides the population of HPV-infected samples into three subgroups: those infected with HR HPV, those with LR HPV, and those with an apparently sustained HPV infection at high and low risk, HR + LR (Fig. 19).

It should be noted that this test does not perform the individual genotypes characterization, but attributes sample positivity compared to the two major risk groups: a HR positivity indicates sample positivity for one or more high-risk HPV oncogenic types between 13 HR whose probes are contained in the test (HPV16, 18, 31, 33, 35 ,39 ,45, 51, 52, 56, 58, 59, 68), while LR positive is an infection index for one or more low risk types that can be detect by the test (HPV6, 11, 42, 43, 44).

It follows that the only detectable multiple infections are potentially the result of co-infections between HR HPV and LR HPV. For infections caused by HR or LR cannot be established without further testing, whether it is a single or multiple infections.

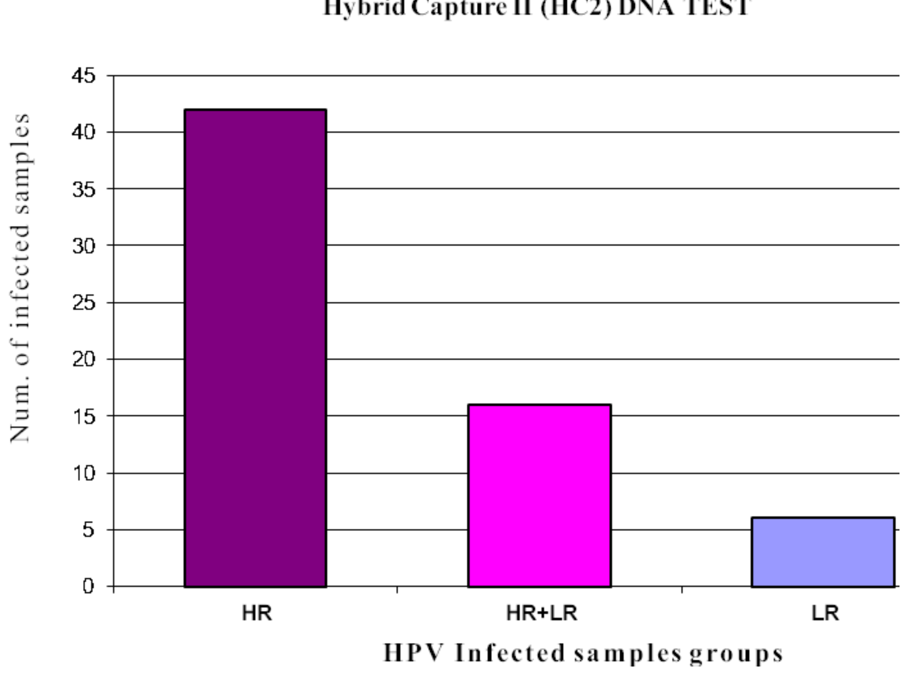


Fig. 19 Analysis of positive samples by type of infection with Hybrid Capture ® 2 (HC2) HPV Test

The population samples analyzed utilizes the chip, and therefore typifies contextually the genotypes in different ways, these may be divided into two subgroups (Fig. 20), i.e. samples with single infections and those with multiple infections, involved in most cases 2 or 3 viral types but, in some cases, also 5 different types (Fig. 21A & 21B).

The amount of co-infections detected increases if the analysis is performed by ProDect® CHIP HPV TYPING PLUS kit in development due to the greater number of genotypes

characterized by the test and not due to cross-reaction occurrences between probes, excluded in the earlier stages of calibration test .

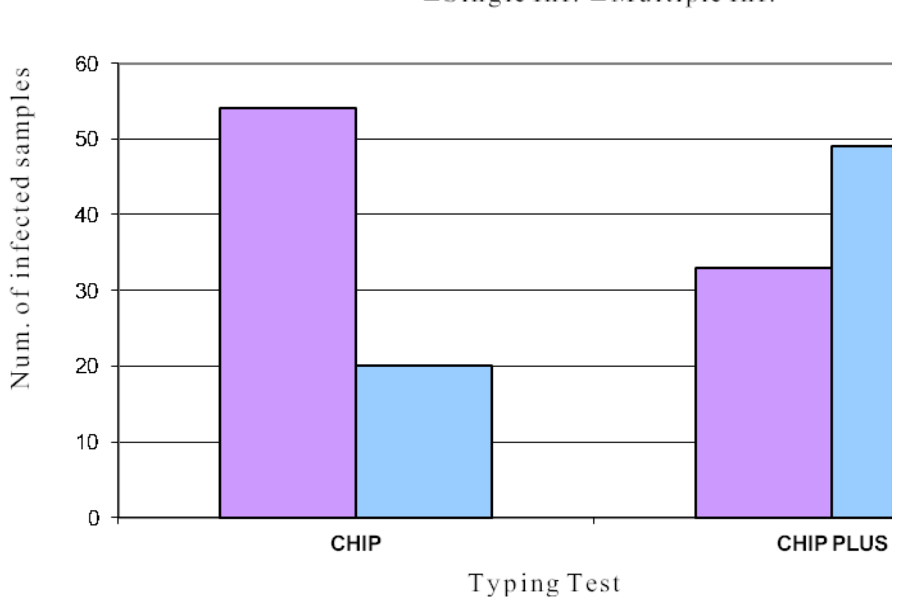


Fig. 20 Analysis of positive samples by infection type with HPV ProDect® CHIP TYPING (CHIP) and the ProDect® CHIP HPV TYPING PLUS (CHIP PLUS)

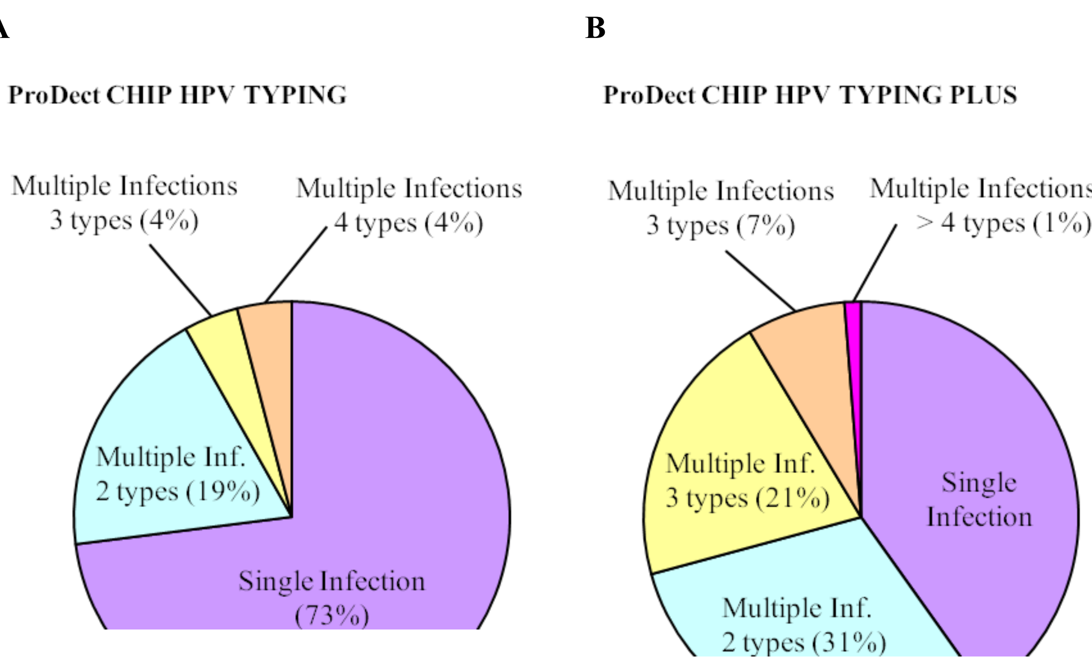


Fig. 21 Single and multiple infections detected with ProDect® CHIP HPV TYPING (A) and with ProDect® CHIP HPV TYPING PLUS (B)

Employing the ProDect® CHIP HPV TYPING the predominance of high-risk types compared to low risk was 70% when calculated on observed positivity (Figure 22); the most frequent in absolute value (type/total positivity) and in decreasing order are types: HPV-16 (15/103), HPV59 (14/103), HPV18 (10/103), HPV35 (8/103) and HPV58 (8/103). Some viral types, such as HPV-16, support in equal measure both single and multiple infections, however other genotypes are found almost exclusively in single infections, as HPV56 and HPV31, or vice versa prevail only in co-infection with others, such as HPV33 and HPV45.

In 19% of infections it was observed that individual E6/E7 sequences attributable to non-typable virus and that by sequencing and the new HPV CHIP PLUS developed were connected to the genotypes HPV16 (5 of 11 cases), HPV52 (3/11) and HPV35 (3/11). The positivity for the E6/E7 region in absolute values for the HR HPV constitutes 10% (11/103) of the positive signals and refer to sequences of high HPV oncogenic risk evidently not typable easily by the different tests. It is worth noting that, Hybrid Capture® 2 (HC2) HPV Test did not detect 6 of the 11 HPV positivities of these samples and 5 of the 11 cases were in agreement to the presence of HR HPV.

In multiple infections, instead, prevalent genotype is HPV35 followed by HPV16, HPV59 and HPV45 (Fig. 23).

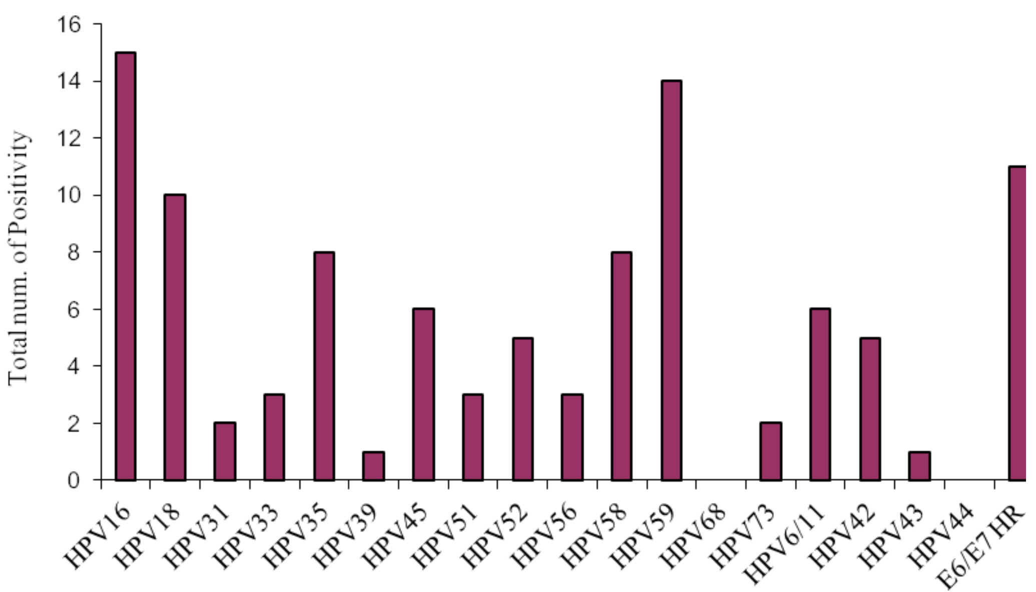
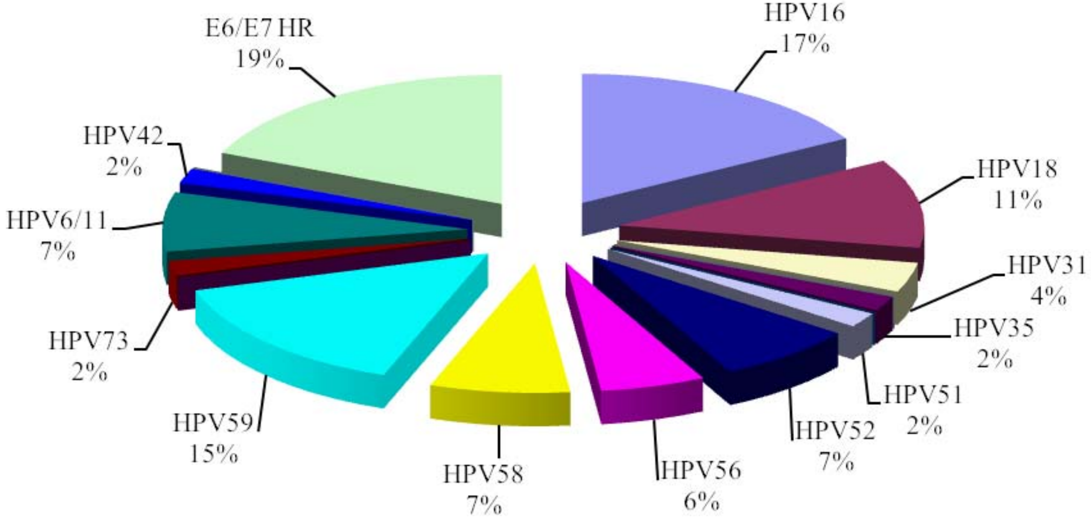


Fig. 22 HPV genotypes detected with ProDect®CHIP HPV TYPING



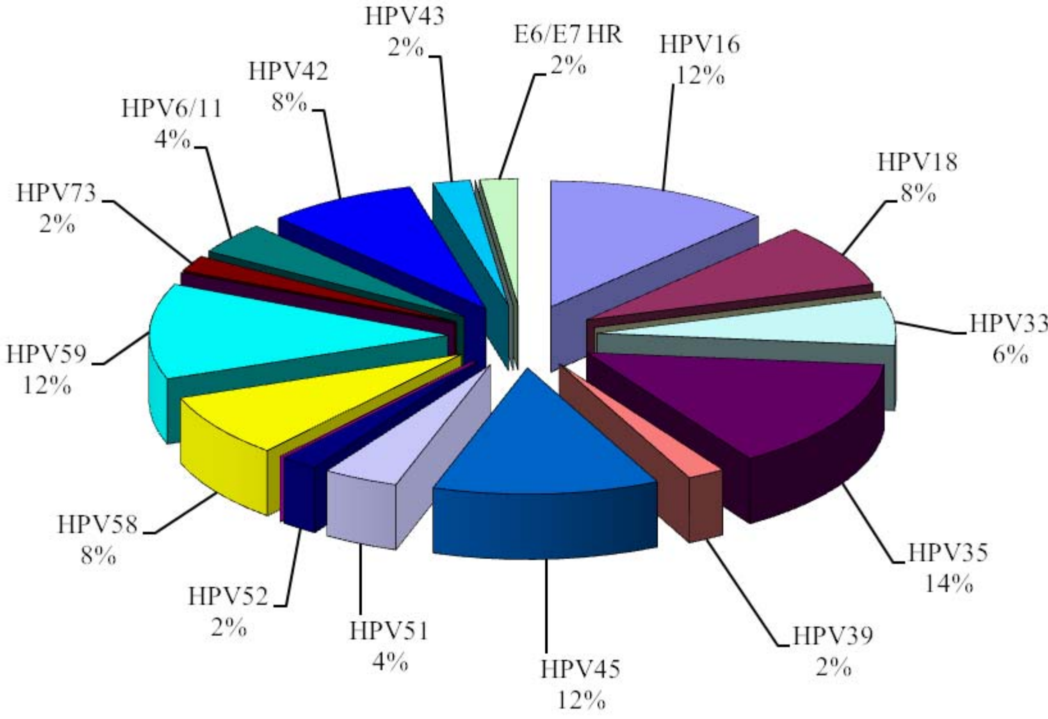


Fig. 23 Prevalence of HPV types detected with ProDect® CHIP HPV TYPING - single and multiple infections

The same population analyzed with ProDect®CHIP HPV TYPING PLUS substantially confirms the positivity observed even with the ProDect® CHIP HPV TYPING, but having a greater number of probes, features a number of types and viral sequences not discriminable with the other test , shown as a row columns in Fig. 24A-B.

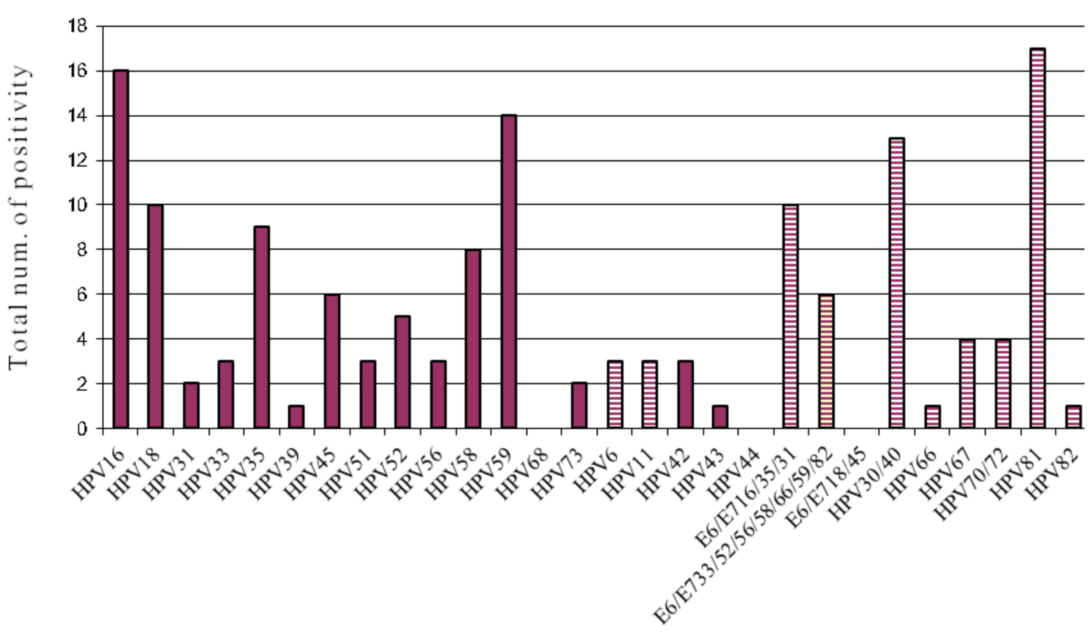


Fig. 24A All HPV genotypes detected with ProDect® CHIP HPV TYPING PLUS in the population

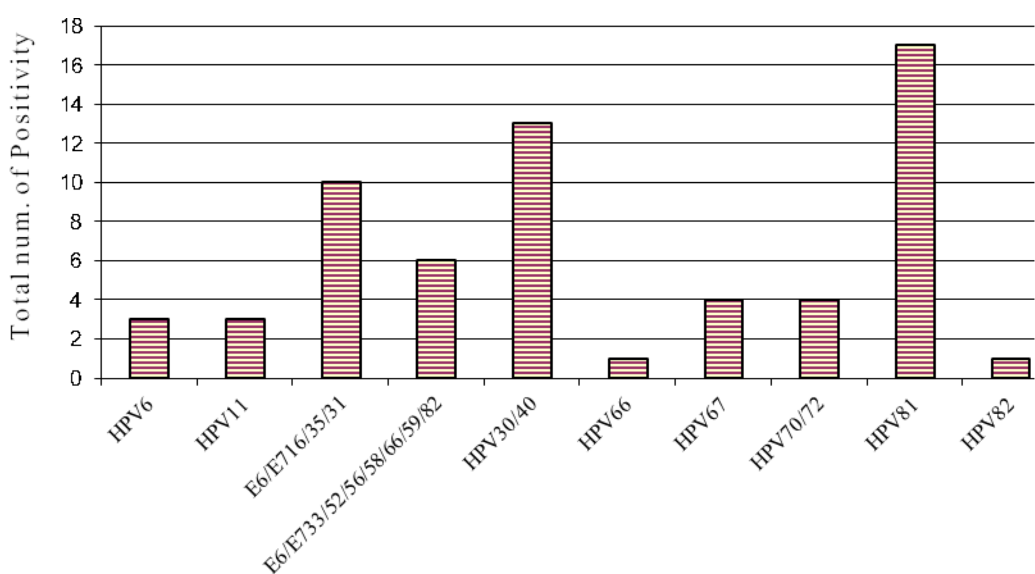


Fig. 24B Particular genotypes / HPV sequences not characterized by ProDect® CHIP HPV TYPING

Among these, the numerically relevant are positive for the HPV81 type and the pool of probes that detect HPV30 and HPV40 were however, not able to distinguish them. At the moment it is considered low-risk oncogenic genotypes.

Significantly, from the point of view of a diagnostic and prognostic value of test, even the data obtained from a single DNA sample analyzed resulted as a carcinoma (SCC-HSIL)

by PAP Test, belonging to a woman of 27 years (ID004TG), negative with others tests employed, result, instead, positive for HPV82 and HPV66 by ProDect® CHIP HPV TYPING PLUS analysis, where, above all the type 82 is now considered a high oncogenic risk while 66 types as medium risk.

It should be highlighted that the probes in the ProDect® CHIP HPV TYPING PLUS for HPV6 and HPV11 occupy distinct positions on the chip and has allowed the positivity allocation and more specifically for the two genotypes.

In particular, the four patients presenting prior positivity to infection with HPV6/11 with ProDect® CHIP HPV TYPING, and treated by warts cauterization, to 2012 follow-up showed altered cytologic state (Tab. 19) and still positive for HPV6 (3/4) and HPV11 (1/4).

Sample ID	**Age**	**PAP Test**	**HPV**
48 HO	24	LSIL	HPV6
26 PN	37	ASCUS	HPV6
2012.001LK	42	HSIL	HPV6
2012.155SL	26	ASCUS	HPV11

Tab. 19 Follow up of some cases treated for condyloma

From the above explanation it is evident that the HPV epidemiology is strongly conditioned by the screening test used. In fact, with the ProDect® CHIP HPV TYPING PLUS the number of infections caused by a single genotype is drastically reduced compared to ProDect® CHIP HPV TYPING (31 versus 54). In this type of infection, the prevalent genotype is HPV52, followed by HPV6, while 19% of positivity is attributable to individual E6/E7 sequences belonging to the probes pool for HPV16/35/31 types (Fig. 25).

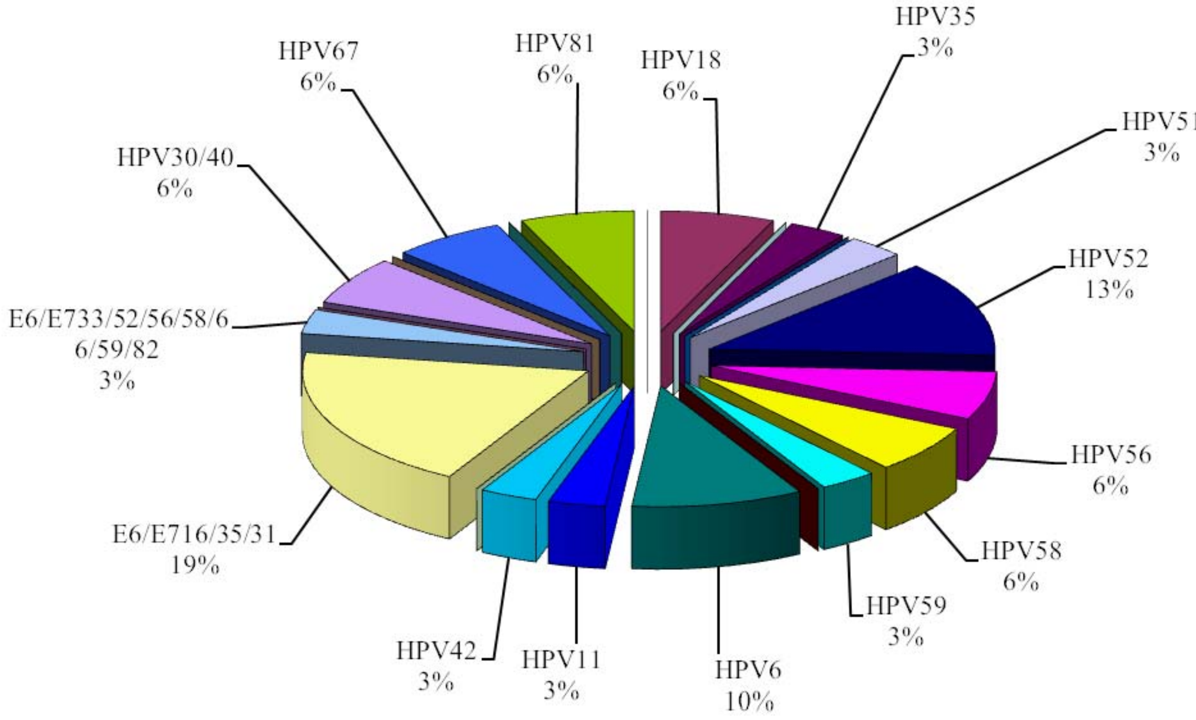


Fig. 25 Prevalence of HPV types detected with ProDect® CHIP HPV TYPING PLUS - single infection

As already reported multiple infections are more frequent if you run screening by ProDect® CHIP HPV TYPING PLUS (Figure 23B) and a certain amount of samples has actually often 3, 4 and more genotypes involved in the infection.

Example from this point of view, the subpopulation of women who are HIV positive (11/171) that analyzed by ProDect® CHIP HPV TYPING were 3 negative and 8 were positive for HPV, and between the 8 positive, 5 were infected by a single HPV type. Analyzed by ProDect® CHIP HPV TYPING PLUS in development only 1/8 showed a single infection (Figure 26 A,B) supported by a HPV35 while all the others showed multiple infections where in 4 out of 8 cases was involved HPV16 genotype.

A **ProDect CHIP HPV TYPING**

B **ProDect CHIP HPV TYPING PLUS**

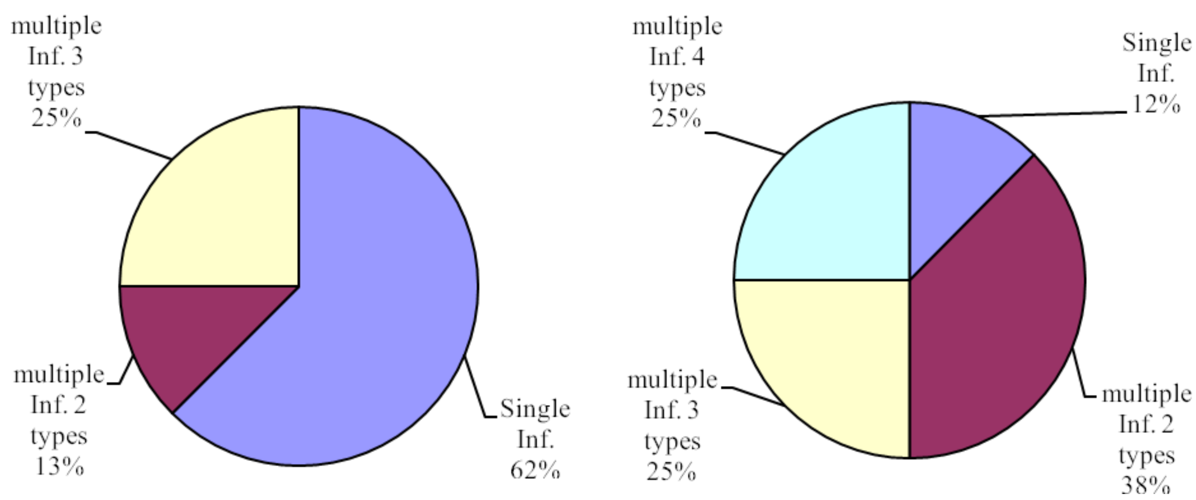


Fig. 26A-B Single and multiple infections detected with ProDect® CHIP HPV TYPING (A) and with ProDect® CHIP HPV TYPING PLUS (B) in the sub-sample of HIV-infected women

Stratifying sampled HPV positive population with respect to the PAP test, it is observed a classic pyramid distribution (Fig. 27), with a very small proportion of high-grade lesions compared with a higher incidence of infection. It reveals a wide variety of genotypes in women who have undergone a normal PAP TEST (Fig. 28), which, however, is substantially reduced with the progression of the cytologic severity as expected on the basis of epidemiological data reported in the literature. However, at least in this population, when searched, are highlighted in the ASC-H, HSIL, AGC and SSC condition, uncommon genotypes or considered low oncogenic risk (HPV 52,59,81,82,66,67,6,30/40), in co-infection with another one or with the most frequent HR HPV types (HPV16, 18, 45), Fig. 29.

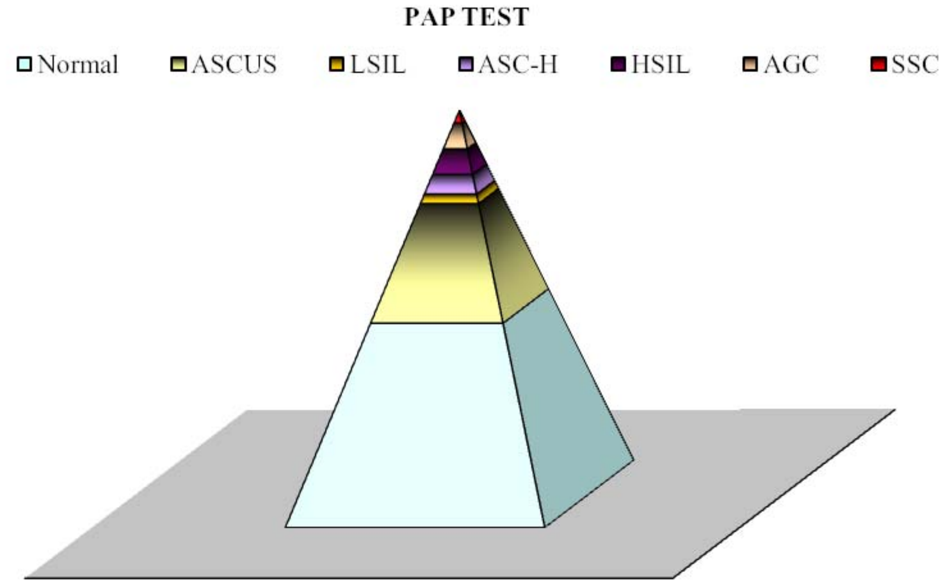


Fig. 27 HPV infected samples with respects to cytological PAP Test

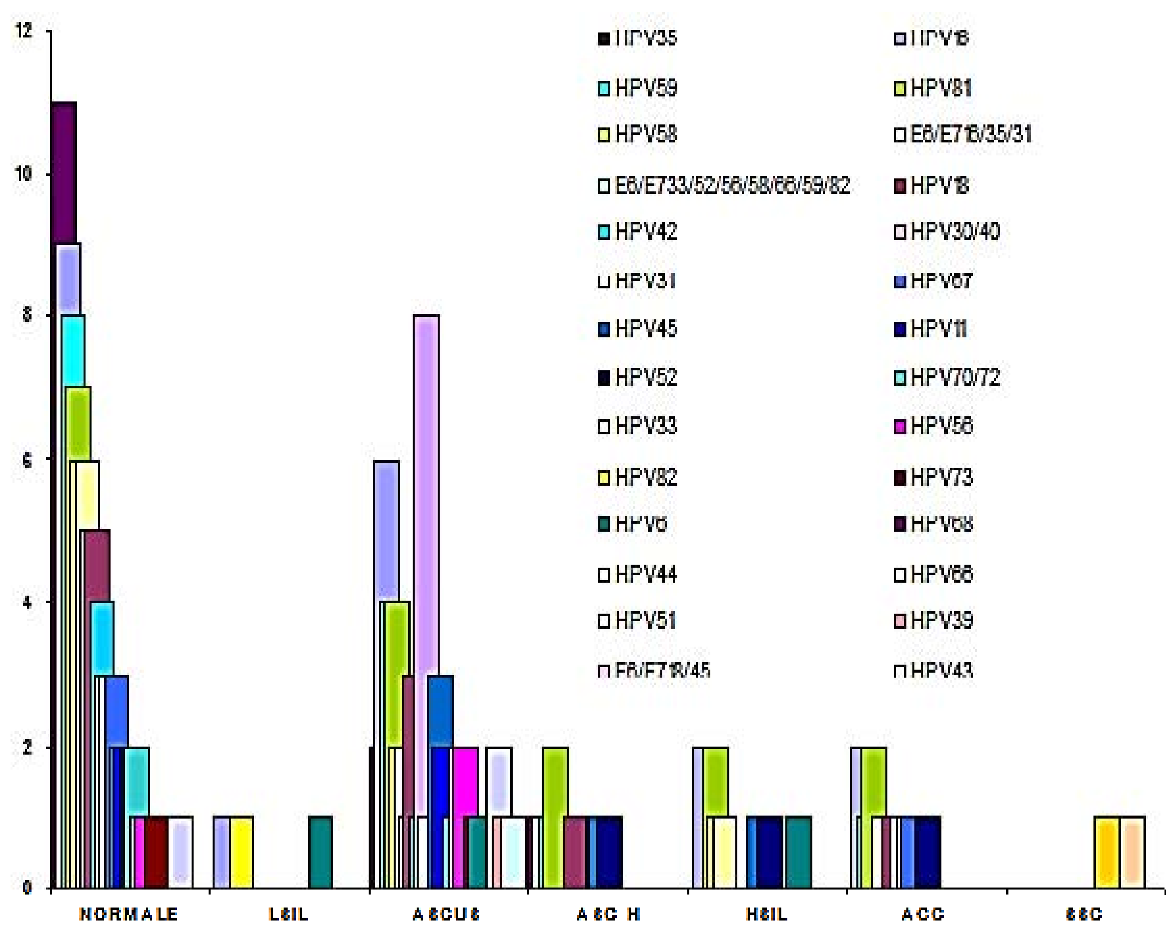


Fig. 28 HPV types present in the infected samples with respects to cytological PAP Tests

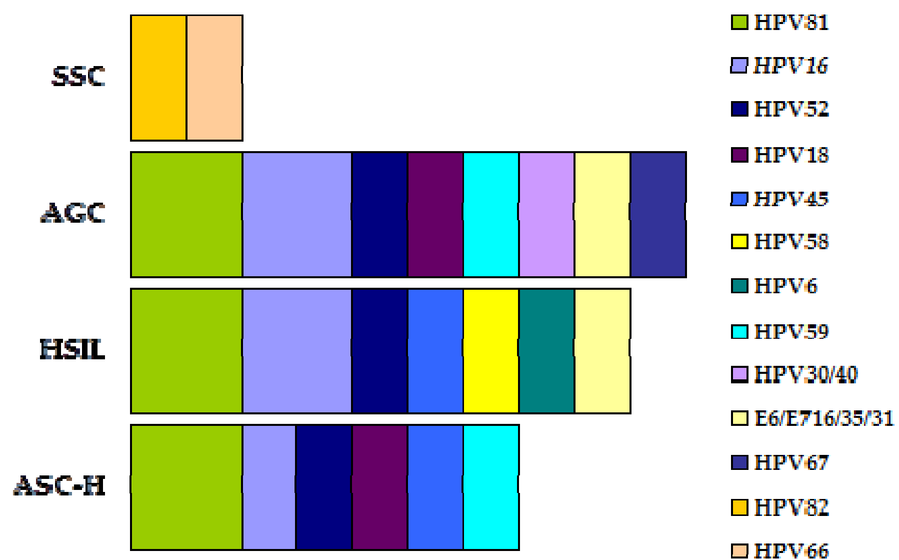


Fig. 29 Prevalence of HPV types in cytological high-grade dysplasia

5. Discussion

Up to the last decade, cervical cancer had been second most common cancer in women for incidence and mortality after breast cancer. At the base of the neoplastic process the establishment of a persistent infection by certain genotypes of the human Papillomavirus (HPV). Organized screening programs for cytological and molecular testing, at least in most developed countries, have, however, contributed significantly to the reduction of the cases in terms of both incidence and mortality. Currently it is the 4th leading cause of death from cancer among women worldwide.

To complicate the tasks of screening and prevention, there are some significant issues:

From a cytological point of view, the presence of so-called cadres "ASC-US" addressed subsequently in colposcopies are often already CIN2 or more severe. In Italy the National Observatory Screening records last report in 2014 that, based on the data reported by the regional centers for screening, the national average, with probability of having a CIN2 histological lesion or more severe among women who have undergone a ASCUS cytology is 15.8%, the national average, with an appreciable variability as a result of geographic areas (16.8% in the North, 21.2% in the Centre and 10% in the South and the Islands)[135]

From a molecular point of view, there is an incomplete knowledge of the HPV-host mechanisms interaction at the base of the infection, its course and its possible persistence. In fact there is great heterogeneity of circulating HPV genotypes, of which more than thirty are high tropism anogenital, classified on the epidemiological database for high-medium or low oncogenic risk types, capable of multiplying only in multi-layered epithelia (therefore, difficult to grow in vitro) and where establish infections characterized by productive and latent stages, and where the genomes can be in the episomal and/or integrated form.

The development and spread of prophylactic vaccines against HPV16, HPV18, recognized at the moment as etiological agents, for about 70% of cervical cancers, hopefully in the near future could lead to a further substantial reduction of CIN2 and CIN3 cases due to these genotypes, but proposes a substantial problem today, to make a careful assessment of the virological condition of the woman at the time of vaccination, and to start, in the most general sense, to closely monitor circulating types. In fact, with the vaccination expansion

other genotypes at high oncogenic risk, which current vaccines have little or no effect, might tend to spread. To confirm what has been said, surveillance needs to be increasingly precise and especially the improved of techniques in molecular biology are leading to a continuous re-classification of the medium and low risk types to higher levels of oncogenic risk, especially if detected in particular population categories. In fact, the persistent infection of some HPV is necessary, but clearly not sufficient for tumor development, and other genetic factors and host immune systems, as well as environmental, contribute to the disease onset and progression, or vice versa, contribute to its spontaneous resolution both in terms of infection and precancerous lesion.

Important changes are therefore in progress in the cervical cancer screening approach, and many other HPV related cancers, so that different nations and also Italian regions[136] have already defined the HPV test as the primary test and the PAP test as a triage, to be applied only to HPV positive women, in order to distinguish those which have obvious cytologic atypia and thus a greater risk of disease progression, from those that can, with time, resolve the infection spontaneously in the absence of atypia or moderate dysplasia[137]. This non secondary aspect will allow a greater period of time for monitoring woman, much more extended in the presence of a negative HPV test. In this regard WHO[138] defining new criteria for organized screening that:

• the interval between controls (frequency of screening) should not be less than five years, when using HPV testing as a primary test,

• women less than 30 years old are not screened if not HIV infected or from geographic areas with a high HIV prevalence

• at national level programs, as a minimum, screening should involve women aged between 30-49 years

An in vitro diagnostic device that is proposed as a molecular test to be used in the cervical cancer screening, so for next 15-20 years, must obviously taking into account all these aspects. It must support clinical, and health organization upstream in the development of algorithms that address the specific problems of various subpopulations in a risk women

sample, or potentially all women aged >15 years (about 3.45 billion of women in the world) that actually include various subpopulations. In light of the foregoing, the HPV test should ideally be able to:

1. diagnose the state or at least the infection of the woman or even if the virus was integrated or at least in a latent form,

2. discriminate individual from multiple infections

3. characterize largest number of genotypes with anogenital tropism in order to:
    a. diagnose infections with other HR HPV not covered by the vaccine in women vaccinated a while before,

    b. monitor course of women already infected with HPV to which current vaccines are not therapeutic,

    c. assesses the real absence of HPV sequences in women undergoing on treatment. In fact, it is still significant relapse rate[139-143]: around 20% at 5 years in the absence of lymph node localization to the first treatment, and up to 70% within 10 years in case of lymph node metastases,

    d. also characterize HPV rare or defined as low-risk types that in certain population such as immunosuppressed women for several factors (HIV+, therapies, genetic conditions etc. etc.) could alone or in co-infection with other types and/or pathogens enhance speeding and severity of neoplastic disease progression.

In this context the realization of this research project was born and founded, aimed to design a diagnostic kit that would by the first to meet the substantial prerequisites for a molecular test to be used in primary screening programs with frequency no less than five years, or to discriminate, with reasonable certainty, women non HPV infected from those infected, being able to recall not infected after five years after a subsequent control, and to start with the infected by cytological triage to characterize and/or monitor any lesions.

The developed prototype, ProDect® CHIP HPV TYPING PLUS provides amplification starting from an extract nucleic acid of three genetic substantial targets for the proper

management of a woman included in a screening organized program: the human β-globin gene, as a marker of the DNA appropriate amount and then of the sample collected cellularity, the L1 region for the typing of more than 30 HPV types with high anogenital tropism and the E6/E7 region for targeted monitoring and further genotypes with high and medium oncogenic risk that could escape to a possible detection on the L1 region for low viral load, reduced number of copies or loss of viral genomic sequences as a result of virus integration into the cellular genome.

A preliminary assessment of the diagnostic test performance, once individual components concentrations were calibrated and 8 pre series batches prepared, the assessment was performed on the clones, cell lines and clinical samples. The tests were conducted in parallel using two kits already on the market, ProDect® CHIP HPV TYPING and Hybrid Capture® 2 (HC2) HPV Test, in order to evaluate the actual contribution of this new test to a more accurate HPV infection diagnosis. For the prototype preliminary validation a female population at particular risk of disease from a geographic area with high prevalence of HPV but also HIV was chosen, Benin, with reduced economic resources devoted to the screening programs organization and therefore in need of accurate and decisive test with high negative predictive values. In fact, in these socio-health, given the reduced instrumental and logistics means, it is of utmost importance to identify with reasonable certainty HPV negative women, compared to those HPV positive, in order to rationalize spending in full health protection for women.

This study found that in terms of sensitivity for HPV common types, the ProDect® CHIP HPV TYPING PLUS shows comparable performance to the reference tests used: ProDect®CHIP HPV TYPING and Hybrid Capture® 2 (HC2) HPV Test . Compared to the Hybrid Capture® 2 (HC2) HPV Test results to be 10-100 times more sensible in detecting bcs Biotech cloned containing L1-E6/E7 regions in the same insert. There is no evidence of nonspecific signs to the highest tested cloned concentrations (10^6-10^7 copies/μl) in any of the three tests. However, calculating the diagnostic capabilities compared to the virus presence in the population can be seen, as a certain sample proportions, the infection has been underestimated. In fact, subject to the reference tests used from 10% to 20% of the samples resulting as negative (false negatives) and many infections are incurred by individual genotypes, indeed several are co-infections genotypes. So, even though all are excellent screening tests (in practice are considered good screening test those who have >80%

sensitivity and specificity) the new prototype in this preliminary study demonstrates a higher VPN (ProDect® CHIP HPV TYPING PLUS VPN = 98.8%). It was exhibited a very high VPP (100%), with a strong HPV prevalence the analyzed population, supported in part by genotypes undetectable with other tests utilized(which in fact are therefore less sensitive than expected, not covering these viral types). This aspect is relevant from an epidemiological point of view: in fact in some contexts such as African or in particular portions of the population at high risk/prevalence of infection, using tests that do not cover a large types, underestimated numbers in the total infections and the role of some oncogenic genotypes that in specific socio-environmental-health have low oncogenic potential, but in different contexts, and in co-infections with other pathogens or specific HPV types, may be contrarily highly oncogenic. The data, although preliminary, opens up further general discussions of the desirability in adopting as the primary cervical cancer screening, being the only test revealing, often without genotyping, high oncogenic risk HPV. It appears intuitive that even if such an approach involves a certain reduction in costs related to screening, moreover to prove, a test that also genotype all anogenital genotypes and maybe not in manual procedure may also be provided at lower cost, with time and natural epidemiological changes to all infections caused by viral genotypes undergo, it may not be sufficient, with savings in the short term but an increase in health care costs in the medium to long term. In the near future prophylaxis, in industrialized countries, and with well-organized health organizations, will further reduce the disease incidence and the prevalence of the major HR HPV, in primis HPV16 and HPV18, while the use of only HR screening tests in populations or segments of the population with high prevalence of HPV infection in the presence of concomitant infections (e.g. HIV, HCV, Chlamydia etc.), immune depression or other pathologies can lead to a significant underestimation of types circulating and their oncogenic potential, given that many HPV medium risk types or types being not well classified, as little or no research in a certain contexts may have greater significance instead. If what emerged, was shown that the reduced sample size reflected that of wider studies, it would be advised not to limit the possibility of molecular testing as an organized screening program for HR genotypes only, but vice versa, trying to cover the prevalence of the infection, types actually circulating in the population sample, in order to establish effective vaccination plans especially in developing or developed countries, but not yet well organized for capillary primary screening, as some Eastern European State members, where unfortunately incidence and mortality due to cervical cancer data are still very high, with a high infection and genotypes prevalence of no clear classification. The study also highlights how sample conformity is a concept for a substantial

molecular screening test aimed at cervical cancer diagnosis. In fact, from the anatomo-pathological point of view characteristics have been established to define a suitably PAP test in terms of representativeness of the different layers of the endocervical mucosa and epithelial transition, it should not be sufficient for a molecular test to establish the conformity based on the amplificability of a DNA control, both viral or human sequence. It would substantial to determine and identify a cellularity cutoff and consequently nucleic acid obtained from the biological sample taken with reasonable certainty, and with any molecular method used, and would guarantee the clinician a probable negativity of the sample for HPV sequences. Indeed, the probability of highlighting the virus in not-replicative phase or integrated and/or with incomplete genomes has as a prerequisite that a reasonable amount of cells ($> 10^5$) extracted with a fair and qualitatively good amount of total nucleic acid, in order to reveal even a few viral copies (10-100 copies), without ever dropping below sensitivity threshold in the method used (theoretical 10 copies for PCR), in order to avoid an incorrect negativity sample diagnosis for HPV sequences, obviously less represented with respects to the human gene or internal control considered by most diagnostic tests on the market. Even this is also not a negligible aspect if you look at the cases of women who have a negative molecular test result (not on the basis of cellularity but often by the absence of HPV sequences) and a positive PAP test, show obvious lesions by colposcopy or following up after a first molecular negative test have recurrent disease often incurred by the same HR genotypes, or from different types usually underestimated in primary screening considered at low-medium risk and are often in co-infection with HR types. A molecular test characterized by an accurate monitoring of levy adequacy would probably be able to further reduce the number of false negatives due to withdrawals of inadequate diagnostic sampling and the type of virus being searched for.

6. Conclusion

The multi-target approach is the new frontier of molecular diagnostics since the acquisition of more and more in-depth knowledge about the etiology of cervical cancer compels manufacturers of medical devices to focus on the prevention, diagnosis and treatment of this disease and keep up to date in state of the art and continuously innovate their platforms.

In this context, the applied research conducted in these three years has led to the design, development and production of a new test prototype multi-target, able to diagnose HPV infection in cervical-vaginal spatula samples having two genomic target sequences of the virus: the L1 region essential for the genotyping of the virus and the E6 / E7 region to detect the presence of higher risk oncogenic HPV, even in cases with the loss of more or less extensive portions of the genome as in the case of the integrated viral forms.

The prototype attributes the negativity or positivity of each sample based on a quantity/quality of DNA extracted from cells harvested at the time of sample collection (at least 10^5 cells/ml) so that we can define compliance for sampling and guarantees a cell representation that will allow us statistically, using gene amplification techniques, amplify of a few copies of the HPV target sequences, if the infection is latent or abortive, to whether or not the virus has integrated into the cellular genome.

The first tests on the complete diagnostic platform (extraction, amplification and detection) show that the system has a specificity and diagnostic sensitivity similar to those of the reference test, the Hybrid Capture® 2 (HC2) HPV Test, currently the most utilized in screening programs worldwide, presents, in addition, the advantage of a procurement compliance control as a function of cellular sampling and not the sole function of testing for the presence/absence of the virus. It also allows with respect to the reference test an immediate classification of the viral genotype over two sequences of genotype in question, at least for the HPV-HR (reg.L1 and reg.E6/E7), moreover being able to characterize up to 31 HPV types, against 13 HPV-HR of the reference test in a single session per sample.

The ProDect® CHIP HPV TYPING PLUS is a natural and necessary evolution of ProDect® CHIP HPV TYPING, diagnostic kit already CE marked, additionally presents, the ability to characterize on a chip a higher density of probes, more anogenital genotypes over 19 already

typable with the kits currently on the market, and to characterize the region of E6/E7 of HPV-HR defining the homology with three pool of probes which react to the same number of groups of high-risk HPV (group 1: E6 / E7HPV16,31,35 , 2nd group E6 / E7HPV18,45, 3rd group E6 / E7HPV33 / 52/58/56/66/67/82/59).

Preliminary data show that the inclusion of new probes on the chip did not have an increase in cross reaction nor the difficulties of acquisition and analysis by the reading system chip, whose software was simply implemented with the new pattern of interpretation results.

In conclusion, the main objective has been reached, that is, to produce a test with high specificity (comparable to that of the two reference tests) but at the same time with a higher Negative Predictive Value. The validation of 171 cervical spatulas samples made it possible to recover a significant amount of negative data samples with regards to the reference tests employed (22% with respects to Hybrid Capture®2 (HC2) HPV Test and 10% compared to ProDect® CHIP TYPING HPV), which conversely analyzed with the new prototype resulted in positive for the sequences for medium and low risk types of HPV.

These encouraging results, as well as the grant for the patent EP1818416 B1 "Methods and Kits for the detection of HPV" dated 26.12.2012 (EPO Bulletin 2012/52) and its nationalization of the same in different countries as of 2013, laid the foundations for the opening of independent clinical trials and multi-centers, as required by rule concerning the risk assessment and management for in Vitro Diagnostic Devices, before the CE markings on the kit and entering in to the commercial market, which are the subject of a proposal from the bcs Biotech Srl currently under evaluation by the EU Commission, as part of the invitation HORIZON2020, SMEInst-2014/15: Clinical research for the validation of biomarkers and / or diagnostic medical devices.

7. Bibliography

[1] Walboomers JM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999;189(1):12–19.

[2] Lu B, et al. Human papillomavirus (HPV) 6, 11,16, and 18 seroprevalence is associated with sexual practice and age: results from the multinational HPV Infection in Men Study (HIM Study). Cancer Epidemiol Biomarkers Prev. 2011;20(5):990–1002.

[3] Munoz N, et al. Impact of human papillomavirus (HPV)-6/11/16/18 vaccine on all HPV-associated genital diseases in young women. J Natl Cancer Inst.2010;102(5):325–339.

[4] Garland SM, Smith JS. Human papillomavirus vaccines: current status and future prospects. Drugs.2010;70(9):1079–1098.

[5] Schwarz T. Clinical update of the AS04-Adjuvanted human Papillomavirus-16/18 cervical cancer vaccine, Cervarix. Adv Ther. 2009;26(11):983–998.

[6] Dunne EF, Datta SD, E Markowitz L. A review of prophylactic human papillomavirus vaccines: recommendations and monitoring in the US. Cancer.2008;113(10 suppl):2995–3003.

[7] Einstein M, et al. Comparison of the immunogenicity and safety of Cervarix and Gardasil human papillomavirus (HPV) cervical cancer vaccines in healthy women aged 18–45 years. Hum Vaccin. 2009;5(10):705–719.

[8] Muñoz N, et al. Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine in women aged 24-45 years: a randomised, double-blind trial. Lancet. 2009;373(9679):1949–1957.

[9] Paavonen J, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a doubleblind, randomised study in young women. Lancet. 2009;374(9686):301–314.

[10] Kjaer SK, et al. A pooled analysis of continued prophylactic efficacy of quadrivalent human papillomavirus (types 6/11/16/18) vaccine against highgrade cervical and external genital lesions. Cancer Prev Res (Phila). 2009;2(10):868–878.

[11] Romanowski B. Long term protection against cervical infection with the human papillomavirus: review of currently available vaccines. Hum Vaccin. 2011;7(2):161–169.

[12] Ronco G, et al. Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. Lancet Oncol. 2010;11(3):249–257.

[13] Leinonen M, et al. Age-specific evaluation of primary human papillomavirus screening vs conventional cytology in a randomized setting. J Natl Cancer Inst. 2009;101(23):1612–1623.

[14] Mayrand MH, et al. Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. N Engl J Med. 2007;357(16):1579–1588.

[15] Naucler P, et al. Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening. J Natl Cancer Inst. 2009;101(2):88–99.

[16] Sankaranarayanan R, et al. HPV screening for cervical cancer in rural India. N Engl J Med. 2009; 360(14):1385–1394.

[17] Khan MJ, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. J Natl Cancer Inst. 2005;97(14):1072–1079.

[18] Cuzick J, Szarewski A., Mesher D., Cadman L. et al. Long term follow-up of cervical abnormalities among women screened by HPV testing and cytology. Results from the Hammersmith study. International Journal of Cancer , Vol.122, No.10 (May) p.2294-300.

[19] Kjaer SK, Frederiksen K, Munk C, Iftner T. Long term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus infection: role of persistence. J Natl Cancer Inst. 2010;102(19):1478–1488.

[20] Dillner J, et al. Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: joint European cohort study. BMJ. 2008;337:a1754.

[21] Grant LA, Dunne EF, Chesson H, Markowitz LE. Considerations for human papillomavirus (HPV) vaccination of mid-adult women in the United States. Vaccine. 2011;29(13):2365–2370.

[22] Bruni L, Barrionuevo-Rosas L, Serrano B, Brotons M, Albero G, Cosano R, Muñoz J, Bosch FX, de Sanjosé S, Castellsagué X. ICO Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in the World. Summary Report 2014-08-22. [Data Accessed]

[23] Schiffman M, Wentzensen N, Wacholder S, Kinney W, Gage JC, Castle PE. Human papillomavirus testing in the prevention of cervical cancer. J Natl Cancer Inst. 2011;103(5):368–383.

[24] Bosch FX, et al. Epidemiology and natural history of human papillomavirus infections and type-specific implications in cervical neoplasia. Vaccine. 2008;26(suppl 10):K1–K16.

[25] Franco EL, et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. J Infect Dis. 1999;180(5):1415–1423.

[26] Woodman CB, et al. Natural history of cervical human papillomavirus infection in young women: a longitudinal cohort study. Lancet. 2001; 357(9271):1831–1836.

[27] Munoz N, et al. Incidence, duration, and determinants of cervical human papillomavirus infection in a cohort of Colombian women with normal cytological results. J Infect Dis. 2004;190(12):2077–2087.

[28] McCredie MRE, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. Lancet Oncol. 2008;9(5):425–434.

[29] Plummer M, et al. Smoking and cervical cancer: pooled analysis of the IARC multi-centric case-control study. Cancer Causes Control. 2003;14 (9):805–814.

[30] Collins S, Rollason TP, Young LS, Woodman CBJ. Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: a longitudinal study. Eur J Cancer. 2010;46(2):405–411.

[31] McIntyre-Seltman K, Castle PE, Guido R, Schiffman M, Wheeler CM. Smoking is a risk factor for cervical intraepithelial neoplasia grade 3 among oncogenic human

papillomavirus DNA-positive women with equivocal or mildly abnormal cytology. Cancer Epidemiol Biomarkers Prev. 2005;14(5):1165–1170.

[32] Appleby P, et al. Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. Lancet. 2007;370(9599):1609–1621.

[33] Munoz N, et al. Role of parity and human papillomavirus in cervical cancer: the IARC multicentric casecontrol study. Lancet. 2002;359(9312):1093–1101.

[34] European Guidelines for quality assurance in cervical cancer screening - Second Edition 2008

[35] Kavita et al. Accuracy of the Papanicolau test in screening for and follow-up of cervical cytologic abnormalities: a systematic review " 2002

[36] De Villiers E-M, Fauquet C, Broker TR, et al. “Classification of Papillomaviruses”. Virology. 2004; 324:17-27

[37] Peter M, Howley B. Warts, Cancer And Ubiquitylation:Lessons From The Papillomaviruses. Transactions of The American Clinical and Climatological Association, Massachusetts 2006; Vol. 117.pp.35-57.

[38] Scheurer ME, Tortolero-Luna G, Adler-Storthz K. “Human papillomavirus infection: biology, epidemiology and prevention”. International Journal of Gynecological Cancer 2005.

[39] Schiffman M, Castle PE. “Human papillomavirus: epidemiology and public health.” Arch Pathol Lab Med. 2003;127(8):930-4.

[40] Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders P.J. , Meijer CJ “Epidemiological classification of human papillomavirus types associated with cervical cancer” N. Engl. J. Med. 2003; 348 :518-527.

[41] Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S.Human papillomavirus and cervical cancer. Lancet. 2007; 370 (9590): 890-907. Review.

[42] Muñoz N, Castellsagué X, de González AB, Gissmann L. HPV in the etiology of human cancer. Vaccine. 2006; 24 Suppl 3: S3/1-10. Review.

[43] Bonnez W, Reichman R. Papillomaviruses. “Principles and Practice of Infectious Diseases” 6th Mandell, Bennett, Dolin (eds)2005; 1841-1856.

[44] Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA, Howley M. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. J Virol 1987; 61:962-67.

[45] Sanclemente G., Gill D.K. 2002 “Human papillomavirus molecular biology and pathogenesis”. JEADV 16: 231-240

[46] Fehrmann F., Laimins L.A., “Human papillomavirus: targeting differentiating epithelial cells for malignant transformation”. Oncogene 2003; 22: 5201-5207

[47] Scheurer ME, Tortolero-Luna G, Adler-Storthz K. “Human papillomavirus infection: biology, epidemiology and prevention”. International Journal of Gynecological Cancer 2005

[48] Bravo IG, Alonso A. Phylogeny and evolution of papillomaviruses based on the E1 and E2 proteins. *Virus Genes*. 2007; 34: 249-262.

[49] Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 an E7 proteins. *Cancer Sci*. 2007; 98:1505-1511.

[50] zur Hausen H. “Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis”.J Natl Cancer Inst. 2000 May 3;92(9):690-8. Review

[51] Tindle RW. “Immune Evasion in Human Papillomavirus-Associated Cervical Cancer”. Nature Reviews Cancer. Jan 2002; 2(1):59-64.

[52] Ciccolini, F., G. Di Pasquale, F. Carlotti, L. Crawford, and M. Tommasino. “Functional studies of E7 proteins from different HPV types”. Oncogene 1994; 9:2633–2638

[53] Fehrmann F., Laimins L.A., “Human papillomavirus: targeting differentiating epithelial cells for malignant transformation”. Oncogene 2003; 22: 5201-5207

[54] Jason Bodily and Laimonis A. Laimins. Persistence of human papillomavirus infections: keys to progression. Trends Microbiol. 2011 January ; 19(1): 33–39

[55] Bossis I., Roden R. B., Gambhira R. et al. Interaction of tSNARE syntaxin 18 with the papillomavirus minor capsid protein mediates infection. *J. Virol.* 2005; 79, 6723–6731.

[56] Jones PH, et al. Sic transit gloria: farewell to the epidermal transit amplifying cell? Cell Stem Cell.2007; 1:371–381. [PubMed: 18371376]

[57] Kines RC, et al. The initial steps leading to papillomavirus infection occur on the basement membrane prior to cell surface binding. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:20458–20463.

[58] Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nature Reviews Cancer. 2010; 10:550–560.

[59] Stubenrauch F, et al. Differential requirements for conserved E2 binding sites in the life cycle of oncogenic human papillomavirus type 31. Journal of Virology. 1998; 72:1071–1077.

[60] Thomas JT, et al. Human papillomavirus type 31 oncoproteins E6 and E7 are required for the maintenance of episomes during the viral life cycle in normal human keratinocytes. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96:8449–8454.

[61] Park RB, Androphy EJ. Genetic analysis of high-risk E6 in episomal maintenance of human papillomavirus genomes in primary human keratinocytes. Journal of Virology. 2002; 76:11359–11364.

[62] Frattini MG, et al. In vitro synthesis of oncogenic human papillomaviruses requires episomal genomes for differentiation-dependent late expression. Proceedings of the National Academy of Sciences of the United States of America. 1996; 93:3062–3067.

[63] Frattini MG, Laimins LA. Binding of the human papillomavirus E1 origin-recognition protein is regulated through complex formation with the E2 enhancer-binding protein. Proceedings of the National Academy of Sciences of the United States of America. 1994; 91:12398–12402.

[64] Sedman J, Stenlund A. Co-operative interaction between the initiator E1 and the transcriptional activator E2 is required for replicator specific DNA replication of bovine papillomavirus in vivo and in vitro. The EMBO Journal. 1995; 14:6218–6228.

[65] McBride AA, et al. Partitioning viral genomes in mitosis: same idea, different targets. Cell Cycle (Georgetown), Tex. 2006; 5:1499–1502.

[66] Steger G, Corbach S. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. Journal of Virology. 1997; 71:50–58.

[67] Peh WL, et al. Life cycle heterogeneity in animal models of human papillomavirus-associated disease. Journal of Virology. 2002; 76:10401–10416.

[68] Middleton K, et al. Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. Journal of Virology. 2003; 77:10186–10201.

[69] Ruesch MN, Laimins LA. Human papillomavirus oncoproteins alter differentiation-dependent cell cycle exit on suspension in semisolid medium. Virology. 1998; 250:19–29.

[70] Hummel M, et al. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. Journal of Virology. 1992; 66:6070–6080.

[71] Ozbun MA, Meyers C. Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b. Journal of Virology. 1997; 71:5161–5172.

[72] Ozbun MA, Meyers C. Temporal usage of multiple promoters during the life cycle of human papillomavirus type 31b. Journal of Virology. 1998; 72:2715–2722.

[73] Bedell MA, et al. Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation. Journal of Virology. 1991; 65:2254–2260.

[74] Wilson R, et al. Role of the E1^E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. Journal of Virology. 2005; 79:6732–6740.

[75] Peh WL, et al. The viral E4 protein is required for the completion of the cottontail rabbit papillomavirus productive cycle in vivo. Journal of Virology. 2004; 78:2142–2151

[76] Fehrmann F, et al. Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. Journal of Virology. 2003; 77:2819–2831.

[77] Schwartz S. HPV-16 RNA processing. Front Biosci. 2008; 13:5880–5891.

[78] Fichorova RN, Anderson DJ. Differential expression of immunobiological mediators by immortalized human cervical and vaginal epithelial cells. Biol Reprod. 1999; 60:508–514.

[79] Stanley M. Immunobiology of HPV and HPV vaccines. Gynecologic Oncology. 2008; 109:S15–21.

[80] Samuel CE. Antiviral actions of interferons. Clin Microbiol Rev. 2001; 14:778–809.

[81] Nees M, et al. Papillomavirus type 16 oncogenes downregulate expression of interferon-responsive genes and upregulate proliferation-associated and NF-kappaB-responsive genes in cervical keratinocytes. Journal of Virology. 2001; 75:4283–4296.

[82] Chang YE, Laimins LA. Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. Journal of Virology. 2000; 74:4174–4182.

[83] Stanley MA. Immune responses to human papilloma viruses. The Indian Journal of Medical Research. 2009; 130:266–276.

[84] Arany I, Tyring SK. Status of local cellular immunity in interferon-responsive and –nonresponsive human papillomavirus-associated lesions. Sex Transm Dis. 1996; 23:475–480.

[85] Alcocer-Gonzalez JM, et al. In vivo expression of immunosuppressive cytokines in human papillomavirus-transformed cervical cancer cells. Viral Immunol. 2006; 19:481–491.

[86] Peitsaro P., Johansson B., Syrjanen S. Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique. *J Clin Microbiol 2002,* 40(3), pp. 886-891

[87] zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer. 2002; 2 (5): 342-50. Review.

[88] Münger K, Baldwin A, Edwards KM, et al. "Mechanisms of Human Papillomavirus-Induced Oncogenesis". J Virol. Nov 2004; 78:11451-11460

[89] Doorbar J (2006) Molecular biology of human papillomavirus infection and cervical cancer. Clin Sci (Lond) 110: 525–541.

[90] Bo Xu, S. Chotewutmontri, S.Wolf, U.Klos, M.Schmitz, M.Durst, E.Schwarz. Multiplex identification of Human Papillomavirus 16 DNA integration in cervical carcinomas. PlosOne 2013, Vol 8, Issue 6,e66693

[91] Moscicki AB, Schiffman M, Kjaer S, Villa LL. Chapter 5: updating the natural history of HPV and anogenital cancer. Vaccine 2006;24 Suppl 3:S42–51.

[92] Schiffman M, Herrero R, Desalle R, et al. The carcinogenicity of human papillomavirus types reflects viral evolution. Virology 2005;337:76–84.

[93] Snijders PJ, Steenbergen RD, Heideman DA, Meijer CJ. HPV-mediated cervical carcinogenesis: concepts and clinical implications. J. Pathol 2006;208(2):152–164.

[94] Ostör AG. Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Pathol.* 1993;12:186-92.

[95] Cullen AP, Reid R, Campion M, Lorincz AT. Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. J. Virol 1991;65(2):606–612.

[96] Klaes, R. et al. 1999. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. *Cancer Res* 59(24), pp. 6132-6136.

[97] Vinokurova, S. et al. 2008. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Res* 68(1), pp. 307-313.

[98] Kulmala, S. M. A. et al. 2006. Early integration of high copy HPV16 detectable in women with normal and low grade cervical cytology and histology. *Journal of Clinical Pathology* 59(5), pp. 513-517.

[99] Li, W. et al. 2008. The physical state of HPV16 infection and its clinical significance in cancer precursor lesion and cervical carcinoma. *Journal of Cancer Research and Clinical Oncology* 134(12), pp. 1355-1361.

[100] Cheung, J. L. et al. 2006. Viral load, E2 gene disruption status, and lineage of human papillomavirus type 16 infection in cervical neoplasia. *The Journal of infectious diseases* 194(12), pp. 1706-1712.

[101] Huang, L. W. et al. 2008. Integration of human papillomavirus type-16 and type-18 is a very early event in cervical carcinogenesis. *J Clin Pathol* 61(5), pp. 627-631.

[102] Ho, C. M. et al. 2011. Integration of human papillomavirus correlates with high levels of viral oncogene transcripts in cervical carcinogenesis. *Virus Res* 161(2), pp. 124-130.

[103] Klaes R., Friedrich T., Spitkovsky D. et al. Over expression of p16INK4A as a specific marker for dysplastic and neoplastic epithelial cells of cervix uteri. *Int. J. Cancer:* 92**,** 276–284 (2001)

[104] Missaoui N, Hmissa S, Frappart L, et al. p16INK4A overexpression and HPV infection in uterine cervix adenocarcinoma. *Virchows Arch.* 2006;448:597-603.

[105] Petry KU, Schmidt D, Scherbring S, et al. Triaging Pap cytology negative, HPV positive cervical cancer screening results with p16/Ki-67 Dual-stained cytology. *Gynecol Oncol.*2011;121:505-509.

[106] Schmidt D, Bergeron C, Denton KJ, et al. European CINtec Cytology Study Group. p16/Ki-67 Dual-Stain cytology in the triage of ASCUS and LSIL Papanicolaou cytology: results from the European equivocal or mildly abnormal Papanicolaou cytology study. *Cancer Cytopathol.*2011;119:158-166.

[107] Singh M, Mockler D, Akalin A, et al. Immunocytochemical colocalization of p16(INK4a) and Ki-67 predicts CIN2/3 and AIS/adenocarcinoma: pilot studies. *Cancer Cytopathol.*2011;120:26-34.

[108] Ferber, M. J. et al. 2003a. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. *Oncogene* 22(24), pp. 3813-3820.

[109] Nancy PC, Khanna A, Marilyn D, Guo M, Guo N, Lin E, Katz RL: Gain of the 3q26 region in cervicovaginal liquid-based pap preparations is associated with squamous intraepithelial lesions and squamous cell carcinoma. *Gynecol Oncol* 2008, 110**:**37-42

[110] Alameda F, Espinet B, Corzo C, Muñoz R, Bellosillo B, Lloveras B, Pijuan L, Gimeno J, Salido M, Solè F, Carreras R, Serrano S. 3q26 (hTERC) gain studied by fluorescent in situ hybridization as a persistence-progression indicator in low-grade squamous intraepithelial lesion cases. *Hum Pathol.* 2009; 40:1474-8.

[111] Reza Jalali G. Amplification of the hromosome 3q26 region shows high negative predictive value for nomalignant transformation of LSIL cytologic finding. *Am.J.Obstet.Gynecol.* 2010; 202.

[112] Qisang G, Long S, Youji F: Cervical cancer screening: hTERC gene amplification detection by FISH in comparison with conventional methods. *Open J Obstet Gynecol* 2012, 2**:**11-17

[113] F.Piras, M.Piga, A. De Montis, R.F.Zannou Ahissou et.al. Prevalence of human papillomavirus infection in women in Benin, West Africa. Virology Journal 2011, 8:514

[114] Perseu S., De Montis A., Floris M.. International Patent N° PCT/IB2001/000771 Methods and means for identifying HPV VIRUS. N° PCT/WO/01/85994

[115] Perseu S., De Montis A., Lauterio C., Manca I European Patent N° EP1609874 B1 Chip System for research and detection of pathogenic organisms

[116] De Montis A., Perseu S., Lauterio C., Deiana L., Lostia S., Musiu S., Desogus M., Perseu F. WO2010081536(A1) - A biochip reader for qualitative and quantitative analysis of images, in particular for the analysis of single or multiple biochips.

[117] Perseu S., De Montis A., Lauterio C., Manca I. European Patent EP 1818416 A3 Method and kits for detection of HPV.

[118] Decreto lgs. 8 Settembre 2000, n. 332, emendato col D. lgs. 25.01.2010,n.37 - Recepimento Direttiva 2007/47/CE

[119] Sumerel J, Lewis J, Doraiswamy A, Deravi LF, Sewell SL, Gerdon AE, Wright DW, Narayan RJ. Piezoelectric ink jet processing of materials for medical and biological applications. Biotechnol J. 2006 Sep;1(9):976-87.

[120] Solomon D, Davey D, Kurman R et al. The 2001 Bethesda System: Terminology for Reporting Results of Cervical Cytology. *JAMA*. 2002;287(16):2114-2119

[121] De Cremoux P, Coste J, Sastre-Garau X, et al. Efficiency of the Hybrid Capture 2 HPV DNA test in cervical cancer screening.Am J Clin Pathol. 2003;120:492-499.

[122] Castle PE, Wheeler CM, Solomon D, et al. Interlaboratory reliability of Hybrid Capture 2. Am J Clin Pathol. 2004;122:238-245.

[123] WHO - Immunization, vaccines and biologicals. Human papillomavirus laboratory manual.1st edition. 2009.

[124] Meijer CJ, Berkhof J, Castle PE, Hesselink AT, Franco EL, Ronco G, Arbyn M, Bosch FX, Cuzick J, Dillner J, Heideman DA, Snijders PJ (2009). Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. Int J Cancer 124(3): 516–520.

[125] Yee C, et al. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. Am. J. Pathol. 119: 361-366, 1985.

[126] Schwarz E, et al. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature 314: 111-114, 1985.

[127] Pater MM, Pater A. Human papillomavirus types 16 and 18 sequences in carcinoma cell lines of the cervix. Virology 145: 313-318, 1985. PubMed: 2992153

[128] Baker CC, et al. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. J. Virol. 61: 962-971, 1987

[129] Cubie HA, Seagar AL, McGoogan E, et al. Rapid real time PCR to distinguish between high risk human papillomavirus types 16 and 18. *Mol Pathol 2001*;*54*:*24*–9.

[130] Mead, D. A., Pey, N. K., Herrnstadt, C., Marcil, R. A., and Smith, L. M. (1991) A Universal Method for the Direct Cloning of PCR Amplified Nucleic Acid. Bio/Technology 9: 657-663.

[131] WHO International Standard 1st WHO International Standard for Human Papillomavirus (HPV) Type 16 DNA NIBSC code: 06/202 Instructions for use (Version 4.0, Dated 26/04/2013)

[132] WHO International Standard 1st WHO International Standard for Human Papillomavirus (HPV) Type 18 DNA NIBSC code: 06/206 Instructions for use (Version 3.0, Dated 11/04/2013)

[133] http://www.equalis.se/en/start.aspx

[134] Eklund C.et al, (2014) Global Improvement in Genotyping of Human Papillomavirus DNA: the 2011 HPV LabNet International Proficiency Study. Journal of Clinical Microbiology. February 2014 52:2 449-459

[135] Rapporto breve 2013. http://www.osservatorionazionalescreening.it/content/i-numeri-degli-screening.

[136] Deliberazione della Giunta Regionale 23 aprile 2013, n. 21-5705-Regione Piemonte 16/05/2013.Approvazione della modifica del programma regionale di screening oncologico, "Prevenzione Serena" per i tumori della cervice uterina. Aggiornamento dell'allegato A della D.G.R. n. 111-3632 del 02.08.2006.

[137] Ronco G, Biggeri A, Confortini M, et al. Health Technology Assessment-Ricerca del DNA di papilloma virus umano (HPV) come test primario per lo screening dei precursori del cancro del collo uterino. Epidemiol Prev 2012; 36 (3-4) suppl1.

[138] http://www.who.int/reproductivehealth/publications/cancers/screening_and_treatment_of_precancerous_lesions/en/index.html WHO guidelines for screening and treatment of precancerous lesions for cervical cancer prevention, 2013

[139] Delgado G, Bundy B et al. Prospective surgical-pathological study of disease-free interval in patients with stage IB squamous cell carcinoma of the cervix: a Gynecologic Oncology Group study. Gynecol Oncol 1990; 38:352-7.

[140] Stehman FB, Bundy BN et al. Carcinoma of the cervix treated with irradiation therapy. I. A multi-variate analysis of prognostic variables in the Gynecologic Oncology Group. Cancer 1991; 67:2776-85

[141] Perez CA, Grigsby PW et al. Tumor size, irradiation dose, and long-term outcome of carcinoma of the uterine cervix. Int J Radiat Oncol Biol Phys 1998; 41:307-17

[142] Quinn MA, Benedet JL et al. Carcinoma of the cervix uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. Int J Gynaecol Obstet 2006 Nov; 95 Suppl 1:S43-103.

[143] Gospodarowicz MK, O'Sullivan B et al. Prognostic Factors in Cancer. 3rd edition. WILEY-LISS, 2006

8. Apendix – Raw data of samples tested

SUBJECT	AGE	PAP TEST	ProDect®CHIP HPV TYPING	ProDect®CHIP HPV TYPING PLUS	HC2	SEQ	FINAL DIAGNOSIS
015 BIS	21	ASCUS	HPV 18, E6/E7	HPV18,81	HR	HPV18,81	HPV18,81
019 BIS	25	Negative	HPV 16	HPV81, 11, E6/E716/35/31	HR+LR	HPV16	HPV16,81,11
001SL	38	Negative	HPV31 E6/E7	HPVE6/E716/35/31	HR+LR	HPV31	HPV31
004SL	18	Negative	Negative	Negative	Negative	neg	Negative
008SL	44	Negative	Negative	Negative	Negative	neg	Negative
017SL	28	ASCUS	HPV16, 45	HPV 31,16	HR+LR	HPV16	HPV16,45,31
042SL	27	Negative	Negative	Negative	Negative	neg	Negative
048SL	29	H-SIL	HPVE6/E7	HPVE6/E716/35/31	HR	HPV16	HPVE6/E716/35/31
049SL	44	Negative	Negative	Negative	Negative	neg	Negative
065SL	50	ASC-H	HPV18,45,E6/E7	HPV18,40	HR+LR	HPV18	HPV18,45,40
067SL	36	ASCUS	HPV18,35	HPV40,E6/E7	HR	HPV18,35	HPV18,35,40
113SL	40	ASCUS	HPV73	HPV81,70	HR+LR	HPV73	HPV 73, 81,70
008 PK	42	Negative	HPV59	HPV81,35,E6/E7	HR	HPV59,35	HPV59,35,81
019PK	29	Negative	HPV35,E6/E7	HPV35,40,E6/E716/35/31	HR	HPV35	HPV35,40
021PK	35	ASCUS	HPV59	HPV16,40,E6/E7	HR	HPV16,59	HPV59,40,16
030PK	28	Negative	HPV59	HPV81,16, E6/E7	HR	HPV16,59	HPV81,16,59
033PK	38	AGC	HPV59	HPV16,40,67,70,81,E6/E7	HR+LR	HPV16,59,67	HPV59,16,40,67,70,81
041PK	50	Negative	HPV 6/11,E6/E7	HPV 35, 58, E6/E7	HR+LR	HPV35	HPV35, 58, 6/11
14HO	27	Negative	HPV35,51,58,59,E6/E7	HPV35, E6/E716/31/35	HR+LR	HPV35	HPV35,51,58,59
15HO	46	Negative	HPV16,35,59,45,E6/E7	HPV35,16,59,E6/E7	HR	HPV16,35	HPV16,35,59,45
18HO	34	Negative	HPV35,59,18	HPVE6/E7	Negative	HPV35	HPV35,59,18
19HO	29	Negative	HPV35,58,42,59	HPVE6/E716/31/35	HR	HPV35	HPV35,58,42,59
20HO	25	Negative	HPV58	HPVE6/E733/58	HR	HPV58	HPV58
21HO	58	Negative	HPV73,59	HPV 40	HR+LR	HPV73	HPV73,59,40
23HO	35	ASCUS	HPVE6/E7	HPVE6/E716/35/31	Negative	HPV31	HPVE6/E716/35/31
24HO	52	Negative	Negative	HPV81, E6/E733/58	Negative	HPV81	HPV81,E6/E733/58
25HO	49	Negative	HPVE6/E7	HPVE6/E716/35/31, 33/58/52	Negative	HPV16	6/E716/35/31,E6/E733/58
1LK	26	ASCUS	HPV16 E6/E7	HPV16, 33, 40, E6/E7	HR	HPV16,33	HPV16, 33, 40
2LK	38	Negative	Negative	Negative	Negative	neg	Negative
5LK	28	ASC-H	Negative	Negative	Negative	neg	Negative
7LK	57	Negative	HPV42	HPV42	HR+LR	HPV42	HPV42
11LK	31	Negative	HPV42, 52	HPV42,52,70	LR	HPV42,52	HPV42,52,70
27LK	20	ASCUS	HPV58	HPV58,81,E6/E7 33/58/52	HR	HPV58	HPV58,81
39LK	40	Negative	HPV33,16,E6/E7	HPV16,33,67,81,E6/E7	HR	HPV16,33	HPV16,33,67,81
40LK	33	ASCUS	HPV59	HPV 59,16,E6/E7	HR	HPV59,16	HPV 59,16
41LK	44	Negative	HPVE6/E7	HPVE6/E716/35/31	Negative	HPV35	HPVE6/E716/35/31
42LK	32	AGC	HPV18,E6/E7	HPV18,31, E6/E7	HR+LR	HPV18	HPV18,31
43LK	54	Negative	HPV58,59,18	HPVE6/E733/58	HR	HPV58	HPV58,59,18
44LK	33	AGC	HPV16,E6/E7	HPV16,81,E6/E716/35/31	HR	HPV16	HPV16,81
45LK	33	ASC-H	Negative	Negative	Negative	neg	Negative
47LK	37	H-SIL	HPV52, E6/E7	HPV40,81,16,E6/E7	HR	HPV52,16	HPV52,16,40,81
49LK	35	Negative	Negative	Negative	Negative	neg	Negative
50LK	30	ASCUS	HPV59	HPV59,35,40,E6/E7	HR	HPV59	HPV59,35,40

51LK	47	ASC-H	HPV59	HPV81,E6/E7 33/58/52	HR	HPV59	HPV59,81
52LK	43	ASCUS	Negative	Negative	Negative	neg	Negative
001TG	47	ASCUS	Negative	Negative	Negative	neg	Negative
002TG	50	Negative	HPV16,E6/E7	HPVE6/E716/35/31	HR	HPV16	HPV16
003TG	26	Negative	HPV35, 42	HPV35, 42	HR+LR	HPV35,42	HPV35, 42
004TG	27	H-SIL SCC	Negative	HPV66,82,E6/E7	Negative	HPV66,82	HPV66,82
005TG	24	Negative	HPVE6/E7	HPV31, E6/E733/58	Negative	HPV31	HPV31, E6/E733/58
016TG	38	ASCUS	Negative	Negative	Negative	neg	Negative
019TG	33	ASCUS	Negative	Negative	Negative	neg	Negative
020TG	35	ASCUS	HPV16, 6/11	HPV11,81,40,16,E6/E7	LR	HPV16,11	HPV16,11,81,40
031TG	39	Negative	HPV 16,35,E6/E7	HPV16,35,E6/E7	HR	HPV16,35	HPV16,35
034TG	31	Negative	Negative	Negative	Negative	neg	Negative
036TG	36	Negative	Negative	Negative	Negative	neg	Negative
041TG	39	Negative	Negative	HPV40	Negative	HPV40	HPV40
012AB	50	H-SIL	HPV16,45, E6/E7	HPV16, 45,E6/E716/35/31	HR	HPV16,45	HPV16,45
049AB	44	H-SIL	HPV58	HPV81,E6/E7 33/58	HR	HPV58,81	HPV58,81
03PN	55	Negative	Negative	Negative	Negative	neg	Negative
05PN	63	Negative	Negative	Negative	Negative	neg	Negative
06PN	59	Negative	Negative	HPV81	Negative	HPV81	HPV81
07PN	28	Negative	HPV 16	HPVE6/E716/35/31	Negative	HPV16	HPV16
08PN	45	Negative	Negative	Negative	Negative	neg	Negative
10PN	23	Negative	HPVE6/E7	HPVE6/E716/35/31, E6/E733/58	Negative	HPV16	6/E716/35/31,E6/E733/58
11PN	54	Negative	Negative	Negative	Negative	neg	Negative
12PN	39	Negative	HPV 16	HPVE6/E716/35/31	HR	HPV16	HPV16
13PN	15	ASCUS	HPV 56	Negative	Negative	HPV56	HPV56
14PN	31	ASCUS	HPV 16, E6/E7	HPVE6/E716/35/31	HR	HPV16	HPV16
15PN	44	Negative	Negative	Negative	Negative	neg	Negative
17PN	35	Negative	Negative	Negative	Negative	neg	Negative
18PN	39	ASC-H	HPV 52	HPV35,E6/E716/35/31, 33/58/52	HR+LR	HPV52,35	HPV52,35
20PN	58	Negative	Negative	Negative	Negative	neg	Negative
21PN	20	Negative	Negative	Negative	Negative	neg	Negative
22PN	40	Negative	Negative	Negative	Negative	neg	Negative
24PN	25	Negative	Negative	Negative	Negative	neg	Negative
26PN	37	ASCUS	HPV6/11	HPV6	LR	HPV6	HPV6
27PN	49	Negative	HPVE6/E7	HPVE6/E733/58	Negative	HPV52	HPVE6/E733/58
28PN	55	Negative	Negative	Negative	Negative	neg	Negative
29PN	39	Negative	Negative	HPV67	Negative	HPV67	HPV67
31PN	46	Negative	Negative	Negative	Negative	neg	Negative
33PN	22	ASCUS	HPV56	HPV56,40, E6/E7	Negative	HPV56	HPV56,40
37PN	54	Negative	Negative	Negative	Negative	neg	Negative
39PN	34	Negative	Negative	Negative	Negative	neg	Negative
40PN	30	ASCUS	Negative	Negative	Negative	neg	Negative
43PN	49	Negative	Negative	Negative	Negative	neg	Negative
44PN	50	AGC	HPVE6/E7	HPVE6/E716/35/31	HR	HPV35	HPVE6/E7 16/35/31
50PN	26	Negative	Negative	Negative	Negative	neg	Negative
01 DJ	39	Negative	Negative	Negative	Negative	neg	Negative

02 DJ	20	Negative	Negative	Negative	Negative	neg	Negative
03 DJ	22	Negative	Negative	Negative	Negative	neg	Negative
04 DJ	24	Negative	Negative	Negative	Negative	neg	Negative
05 DJ	21	Negative	Negative	Negative	Negative	neg	Negative
06 DJ	27	Negative	Negative	Negative	Negative	neg	Negative
07 DJ	20	Negative	Negative	HPV67	Negative	HPV67	HPV67
08 DJ	42	Negative	Negative	Negative	Negative	neg	Negative
09 DJ	28	Negative	Negative	Negative	Negative	neg	Negative
10DJ	35	Negative	Negative	Negative	Negative	neg	Negative
13 DJ	27	Negative	HPV56	HPVE6/E733/56/58	HR	HPV56	HPV56
15 DJ	25	ASCUS	Negative	HPV40	Negative	HPV40	HPV40
4 AT	28	LSIL	HPV16,E6//E7	HPV16,58,E6/E716/35/31,33/58	HR	HPV16	HPV16,58
7 AT	21	Negative	HPVE6/E7	HPVE6/E716/35/31	HR	HPV35	HPVE6/E716/35/31
33 AT	41	ASCUS	HPV 42, 33	HPVE6/E733/58	HR+LR	HPV33,42	HPV42,E6/E733/58
35 AT	31	ASCUS	HPV 58, 45	HPVE6/E733/58	HR	HPV58	HPV58,45
27 HO	36	Negative	Negative	Negative	Negative	neg	Negative
29 HO	47	Negative	Negative	Negative	Negative	neg	Negative
30 HO	43	Negative	Negative	Negative	Negative	neg	Negative
31 HO	30	Negative	Negative	Negative	Negative	neg	Negative
32 HO	47	AGC	Negative	Negative	Negative	neg	Negative
33 HO	45	Negative	Negative	Negative	Negative	neg	Negative
34 HO	34	ASCUS	Negative	Negative	Negative	neg	Negative
38 HO	62	Negative	Negative	Negative	Negative	neg	Negative
39 HO	33	Negative	Negative	Negative	Negative	neg	Negative
40 HO	54	Negative	Negative	Negative	Negative	neg	Negative
41 HO	59	Negative	Negative	Negative	Negative	neg	Negative
46 HO	51	Negative	Negative	Negative	Negative	neg	Negative
48 HO	24	LSIL	HPV 6/11	HPV6	LR	HPV6	HPV6/11
50 HO	33	Negative	Negative	Negative	Negative	neg	Negative
53 PN	58	Negative	Negative	Negative	Negative	neg	Negative
56 PN	62	Negative	Negative	Negative	Negative	neg	Negative
57 PN	57	Negative	Negative	Negative	Negative	neg	Negative
58 PN	55	Negative	Negative	Negative	Negative	neg	Negative
62 PN	54	Negative	Negative	Negative	Negative	neg	Negative
63 PN	55	Negative	Negative	Negative	Negative	neg	Negative
64 PN	53	Negative	Negative	Negative	Negative	neg	Negative
69 PN	35	Negative	Negative	Negative	Negative	neg	Negative
70 PN	53	Negative	Negative	Negative	Negative	neg	Negative
71 PN	52	Negative	Negative	Negative	Negative	neg	Negative
72 PN	39	Negative	Negative	Negative	Negative	neg	Negative
75 PN	54	Negative	Negative	Negative	Negative	neg	Negative
76 PN	38	Negative	Negative	Negative	Negative	neg	Negative
78 PN	31	ASCUS	Negative	Negative	Negative	neg	Negative
80 PN	36	Negative	Negative	Negative	Negative	neg	Negative
81 PN	44	Negative	Negative	Negative	Negative	neg	Negative
84 PN	31	Negative	Negative	Negative	Negative	neg	Negative

85 PN	57	Negative	Negative	Negative	Negative	neg	Negative
86 PN	42	Negative	Negative	Negative	Negative	neg	Negative
87 PN	36	ASCUS	Negative	Negative	Negative	neg	Negative
88 PN	38	Negative	Negative	Negative	Negative	neg	Negative
89 PN	55	Negative	Negative	HPV81	Negative	HPV81	HPV81
90 PN	59	ASCUS	Negative	Negative	Negative	neg	Negative
93 PN	47	ASCUS	HPV 33,45,43	HPVE6/E733/58	HR	HPV45,33	HPV 33,45,43
96 PN	43	Negative	Negative	Negative	Negative	neg	Negative
97 PN	40	Negative	Negative	Negative	Negative	neg	Negative
2012.137SL	52	ASCUS	HPV18,E6/E7	HPV18,30/40,E6/E718/45, 16/35	HR	HPV18	HPV18, 30/40
2012.139SL	50	Negative	Negative	Negative	Negative	neg	Negative
2012.140SL	46	Negative	HPV18,E6/E7	HPV18,E6/E718/45	HR	HPV18	HPV18
2012.146SL	34	Negative	Negative	Negative	Negative	neg	Negative
2012.151SL	26	Negative	Negative	Negative	Negative	neg	Negative
2012.152SL	25	Negative	Negative	Negative	Negative	neg	Negative
2012.155SL	26	ASCUS	HPV6/11	HPV11	LR	HPV11	HPV11
2012.156SL	28	Negative	Negative	Negative	Negative	neg	Negative
2012.158SL	40	ASCUS	HPV59, E6/E7	HPV59	HR	HPV59	HPV59
2012.161SL	31	Negative	HPV58, E6/E7	HPV58, E6/E733/52/58/56/66/59	HR	HPV58	HPV58
2012.162SL	41	AGC	HPV52	HPVE6/E733/52/58/56/66/59	HR	HPV52	HPV52
2012.163SL	43	ASCUS	HPV51	HPV51	HR+LR	HPV51	HPV51
2012.164SL	46	ASCUS	HPV39, HPV51, E6/E7	HPVE6/E716/35/31	HR	HPV39,51	HPV39, HPV51
2012.165SL	38	Negative	Negative	Negative	Negative	neg	Negative
2012.166SL	57	Negative	HPV52	HPVE6/E733/52/58/56/66/59	HR	HPV52	HPV52
2012.026BIS	28	ASC-H	HPV16,E6/E7	HPV16,81,E6/E716/35/31	HR	HPV16,81	HPV16,81
2012.032BIS	32	Negative	Negative	Negative	Negative	neg	Negative
2012.071BIS	62	Negative	Negative	Negative	Negative	neg	Negative
2012.001LK	42	H-SIL	HPV6/11	HPV6	LR	HPV6	HPV6
2012.002LK	40	Negative	Negative	Negative	Negative	neg	Negative
2012.003LK	40	H-SIL	Negative	Negative	Negative	neg	Negative
2012.004LK	43	Negative	Negative	Negative	Negative	neg	Negative
2012.005LK	47	Negative	Negative	Negative	Negative	neg	Negative
2012.006LK	36	Negative	HPV31, E6/E7	HPV31,45,70/72, E6/E716/35/31	HR	HPV31,45	HPV31,HPV45,HPV70/72
2012.008AB	30	Negative	HPV18,E6/E7	HPV18, E6/E718/45	HR	HPV18	HPV18
2012.009AB	48	Negative	HPVE6/E7	HPVE6/E716/35/31	HR	HPV16	HPVE6/E716/35/31
2012.010AB	45	Negative	HPV18,E6/E7	HPV18, E6/E718/45, 16/35/31	HR+LR	HPV18	HPV18, E6/E716/35/31

CHAPTER 2 - THE STUDY OF GENOMES IN SARDINIAN INDIVIDUALS AFFECTED AND NON-AFFECTED FROM MULTIFACTORIAL DISEASES

1. Background

This work was aimed at searching for genetic factors responsible for common multifactorial diseases. The strategy used was an indirect association using a case-control method. An essential requisite for this type of study is that the controls and patients pertain to the same population. Moreover, this population must be homogeneous and without ethnic stratifications to ensure that associations among causal disease variants and the genetic markers in the vicinity are due to linkage disequilibrium and not to other factors.

The Sardinian population is one of the most homogeneous populations in Europe with a millenary history of scarce and lowly significant external gene influx. The island distribution of the population has long been influenced by historical (invasions and wars) and climatic-environmental factors (malaria) but above all by the particular orographical structure of the territory. This gave rise to numerous and more or less vast and populated genetic isolates that were necessarily faced with phenomena of casual genetic drift and selection (ß°39 thalassemia, G-6-PD deficiency), all of which led to a certain degree of heterogeneity of the population in different geographical areas.

Based on studies of numerous genetic polymorphisms, including microsatellites and the HLA loci, the genetist[1-8] divided the Island of Sardinia into 32 geographical areas characterized by high internal genetic homogeneity and microheterogeneity among areas (FST= 0,0068 ± 0,0015). Only for the mutations ßth and G6PD¯(adaptive), heterogeneity (FST= 0,0430) clearly existed between the populations of the highlands (not exposed to malaria) and those of the lowlands (exposed to malaria).

1.1. Choice of the geographical areas

In order to reduce to a minimum the possibility of bias due to dyshomogeneity or stratification within the population groups to be selected for the study of association, we made a major effort to establish if and which populations residing in the previously individuated 32 geographical areas were substantially mendelian breeding units (MBU). In each geographical area we chose 2-4 villages representative of the geographical population and number of

inhabitants, for a total of 94 villages. In each village, we analyzed all the matrimonies celebrated over the past 150 years, distinguishing between the place of origin of the married couples (from the same village, other villages of the same area, villages of other areas). We then evaluated the mean endogamy index for each village and each area. We also evaluated the relative migration rates.

For the study, we selected villages or geographical areas with a mean endogamy index of ≥0,50 and a mean immigration rate of ≤0,085, which we considered as MBU. The number of inhabitants of the MBUs selected for the study ranged from ≥15.000 to ≤ 65.000.

1.2. Selection of the Controls

A preliminary choice of the controls in each MBU was made through the local community anagraphical registrars with a primary selection of residents aged ≥ 35 year pertaining to different family nuclei and born to parents of the same MBU. After obtaining informed consent for participation to the study, we constructed family trees up to the 4° – 5° generation, including, for each participant, name and surname, date and place of birth, eventual cause of death and the presence or absence of common multifactorial diseases. All of the family trees of a given MBU lacking the presence of multifactorial diseases underwent analyses with a special program specifically designed to identify and exclude: 1) trees with ancestors originating from different MBUs and 2) trees with one or more members related to members of other trees. In this way, the individuals selected as controls in each MBU were: 1) descendents of ancestors all originating from the same MBU up to the 4° – 5° generation, 2) not related among each other, and 3) presumably not affected by common multifactorial diseases.

1.3. Selection of affected probands

A preliminary choice of probands affected by common multifactorial diseases in the selected MBUs was made with the assistance of general practitioners and specialist services such as the Antidiabetes Centers and Services of Rheumatology, Dermatology, Neurology and Endocrinology etc. After obtaining informed consent for participation to the study, we completed the family trees for each proband up to the 4°-5° generation and then proceeded with the final selection of the affected probands in each MBU taking care, as for the controls,

that the selected probands were not related with other probands and that their ancestors all originated from the same MBU up to the 4°-5° generation.
The number of controls and affected subjects is given in Tab. 1 below.

Biobank DNAs were collected as part of various projects, of which the last two funded by Sardegna Ricerche via BCS BIOTECH between 2006 and 2009 years entitled respectively, "*Development of new bioinformatics methods for the detection of environmental and genetic causes of multifactorial diseases in Ogliastra*" and "*Detection of the genes responsible for autoimmune Hypothyroidism and for Hashimoto Disease in Ogliastra*", Prof.L.Contu was the Scientific Head for both projects.

	Samples**	
*DNA from →	**528**	Controls
	201	Type 1 Diabetes
	93	Hashimoto Disease
	96	Multinodular Goiter
	24	Schizophrenia
	143	Longevity
	81	Myopia
	573	Other Diseases

Tab. 1. Samples of controls and probands affected by multifactorial diseases that are currently stored in BCS BIOTECH Biobank. *Many patients have 2 or much more diseases.
** The samples were collected in different areas of the south, center and north of Sardinia, as well as in Ogliastra.

2. Introduction

The first project goal, still in progress, is to create a dataset of genetic data related to molecular characterization of the whole genome in a population formed by a sufficiently large enough sample of healthy subjects, or rather have not been affected by common and multifactorial pathologies, and are of a known Sardinian origin to be used in association studies. In fact, large number of studies have demonstrated the monophyletic genetic origin and Sardinian peculiarities[1-9] as well as their substantial genetic homogeneity, despite microheterogeneity between different regions[10-11a] due to the combined effects of random genetic drift and migration between villages. The population is large enough and there have been 32 homogeneous sub regions identified with a total population of about a million inhabitants. The analysis of the genetic variation of the Sardinian population, considered a founder population, is particularly useful for association studies on complex diseases, especially autoimmune diseases, such as Type I diabetes and multiple sclerosis, which are highly represented in the island[12-14]. Therefore, having at the disposal a dataset of genetic healthy controls of a Sardinian population represents a fundamental tool for the study of the most common diseases, not only in Sardinia but in the world, and offers the possibility of an estimate of what could be the differences created by geographical location of subjects/patients groups being analyzed.

3. Material and methods

3.1. Samples

Biobank kept at the R&D laboratories of bcs Biotech, from which my research started, containing 528 healthy controls and 900 patient DNAs with one or more complex diseases, all of certain Sardinian origin, selected in different geographical Sardinia subareas and in particular 6 Ogliastra municipalities. For each subject the reconstruction of the genealogical trees up to 4°-5° generation was performed, and steps taken to exclude all the other probands relatives (sick and controls). All samples were managed in an anonymous way, they were genotyped for loci: HLA A-B-C-DQB1-DRB1, SNP-137; SNP-397; SNP-351; Haplotype -397/-351; NALP1; HLA G; FTO (data not shown) and 234 controls were analyzed by Affymetrix platform. Currently being undertaken is the genotyping of patients distributed by disease, Tab. 1.

3.2. Genotyping and quality control

The DNA of 234 control subjects exclusively originating from Ogliastra was genotyped by Affymetrix Genome-Wide Human SNP 6.0 Array (Affymetrix, Inc., Santa Clara, CA, USA), which includes more than 906,000 SNPs and more than 946,000 probes for the *Copy Number Variation* determination[15].

The experiments were performed in the genotyping laboratory within the “Cluster” project, organized by Sardegna Ricerche within the scientific and technological park in Pula addressed at public and private research laboratories participants at the "*Creating a centre of excellence in bio-informatic technologies applied to personalised medicine technologies*" program funded by the Region of Sardinia following the recommended protocol as described in the Affymetrix manual. Briefly, total genomic DNA (500 ng) was digested with restriction enzymes Nsp I and Sty I, linked to the adaptors and amplified using a primer sequences that recognizes adaptors. Amplified DNA was fragmented, labeled and hybridized to oligonucleotide probes bound to the surface of the array in the GeneChip Hybridization Oven 640 (Affymetrix, Inc., Santa Clara, CA, USA), followed by washing and staining procedures performed in the GeneChip Fluidics Station 450 (Affymetrix, Inc., Santa Clara, CA, USA). Arrays were then scanned by GeneChip Scanner 3000 7G (Affymetrix, Inc., Santa Clara, CA, USA). Samples intensity files were analyzed with programs included in Affymetrix Power

Tools (APT version 1.12.0). Quality control was performed by Constrast QC (CQC) algorithm, which quantifies the separation of the intensity signals for each allele in three different clusters in a contrast space, defined as the two-dimensional space projection of the allele intensity in a single informative dimension. Analyses were performed using APT-GENO-QC executable, set with default parameters.

3.3. Genotype calling and dataset filtering

The genotypes used for the following analysis were determined by the Birdseed v2 algorithm, implemented in the APT-PROBESET-GENOTYPE program, using a *score confidence threshold* equal to 0.1. Analysis by Birdseed algorithm was conducted with all samples that had passed the quality control, as Affymetrix recommended the execution of the algorithm with at least 44 samples to get a more accurate genotypes determination.

4. Results

The data derived from the analysis of 234 control subjects was provided for two studies in collaboration.

4.1. Study No. 1,

In collaboration with researchers from the following public and private Italian institutions: *Shardna Life Sciences, University of Cagliari – Department of Experimental Biology, University of Bologna – Laboratory of Molecular Anthropology, Department of Experimental Evolutionary Biology* and that ended in to the following publication Piras I., De Montis A., Calò CM., Marini M., Atzori M., Corrias L., Sazzini M., Boattini A., Vona G., Contu L. *Genome wide scan with nearly 700.000 SNPs in two sardinian subpopulations suggest some regions as candidate targets for positive selection*. European Journal of Human Genetics (*EJHG*) 2012, 03:1-7 http://www.ncbi.nlm.nih.gov/pubmed/22535185

In short, to increase genomic information on the Sardinian population, the analysis of both the genetic structure at the sub-regional level and the consequences of natural selection in two distinct Sardinia subpopulations. We analyzed 321 healthy individuals from two different Sardinian regions. In particular, sample datasets from Ogliastra preserved in BCS BIOTECH biobank have been compared to 125 Southern Sardinian sample datasets, conserved in the Department of Experimental Biology biobank, University of Cagliari, these samples were from 20 different villages distributed in three historic-neighboring geographical regions in central and southern Sardinia (Sulcis, Campidano and Trexenta). The information from nearly 700,000 informative autosomal SNPs genotyped by Affymetrix platform 6.0, were analyzed with several statistical approaches to determine the genomic differences at individual SNPs. The more differentiated regions were further analyzed using a test based on *Extending Haplotype Homozygosity (EHH)16*, which is able to detect the occurrence of potential "selective sweeps" in the populations genomes examined.

The principal component analysis (PCA) and *admixture* analysis suggested a differentiation between the two subpopulations, as confirmed by AMOVA (Fst = 0.011; P = 0.001). In addition, we have identified 40 genomic regions with significant differences, particularly in the chromosomes 1, 9, 12 and 13. These 40 regions were further analyzed with

the *Long Range Haplotype* test, which showed statistically significant values of REHH for rs11070188 and rs11070192 SNPs in the Ogliastra population. These markers are located on chromosome 15 (15q15), located less than +5 kb from the rs7181250 SNP *peak*, and show higher values of Fst. In light of SNPs analyzed, characteristics and correlation between the different methods used, the region characterized by these SNPs can be considered an important candidate that have been subjected to a population specific selective event. This region is defined by C15orf54 (-81 kb) and THBS1 (+244 kb) genes. In particular, the gene coding for the THBS1 Thrombospondin I, is a multimodular secreted protein that is linked to the extracellular matrix and exerts a variety of biological functions, such as platelet aggregation, angiogenesis, and *tumorigenesis*. One of these functions is also represented by the capacity to act as a receptor for red blood cells infected by *Plasmodium falciparum*, adhesion of which to the capillaries increases the parasite virulence. Moreover, the region characterized by these genes has been previously identified as a selective pressure target in two distinct papers[17-18]. Because of the SNPs location, in particular for the THBS1 gene presence, it is possible to speculate on a potential selective action by malarial infection, which has characterized Sardinia up to its complete eradication, which was held in the mid of last century[19]. Referring to the data on malaria morbidity for each village[20-21], morbidity average, calculated with a weighted average based on the number of samples from each village, would be 57.95% for Ogliastra compared to 28.56% of South Sardinia. These data suggest a significant presence of malaria in Ogliastra, supporting, as a working hypothesis, a selective event associated to malarial infection, which still needs further verification.

4.1.1. Conclusion

The published results emphasize a presence of internal differentiation within the Sardinian population, and support the identification of a *selective sweep* at a micro-geographic level that pertain to the Ogliastra population, such as discovery of other highly differentiated regions in addition to be potential candidates for selective pressure event, may represent substantial functional differences. Though many other studies have been performed such as genome-wide for the detection of regions subject to natural selection, the population utilized have always shown high levels of divergence, while at a micro-geographic level analysis, are quite rare. In this paper we report a rather plausible evidence of selective pressure right at micro-geographic level. Finally, it is possible to speculate that these high differences in genome specific regions between the two samples tested could play a role in susceptibility to complex disease[22].

These results could have a predictive role for future association studies of complex diseases in the same subpopulations.

4.1.2. References

[1] Contu, L., Deschamps, I., Lestradet, H., Hors, J., Schmid, M., Busson, M. *et al.* (1982) HLA haplotype study of 53 juvenile insulin-dependent diabetic (I.D.D.) families. *Tissue Antigens*, 20, 123.

[2] Contu, L., Carcassi, C. & Trucco, M. (1991) Diabetes susceptibility in Sardinia. *Lancet*, 6, 338, 65.

[3] Piazza, A., Mayr, W.R., Contu, L., Amoroso, A., Borelli, I., Curtoni, E.S., Marcello, C., Moroni, A., Olivetti, E., Richiardi, P., et al. (1985) Genetic and population structure of four Sardinian villages. *Ann Hum Genet.* 49, 47.

[4] Di Rienzo, A., Wilson, A.C. (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA *Proc Natl Acad Sci U S A*. 88 (5),1597.

[5] Contu, L., Arras, M., Carcassi, C., La Nasa, G., Mulargia, M. (1992) HLA structure of the Sardinian population: a haplotype study of 551 families. *Tissue Antigens,* 40 (4),165.

[6] Cavalli Sforza, L.L., Menozzi, P., Piazza, A. (1994) The history and geography of human genes. Princeton, Princeton University Press.

[7] Crouau-Roy, B., Bouzekri, N., Carcassi, C., Clayton, J., Contu, L., Cambon-Thomsen, A. (1996) Strong association between microsatellites and an HLA-B, DR haplotype (B18-DR3): implication for microsatellite evolution. *Immunogenetics*, 43 (5), 255.

[8] Piazza, A., Cappello, N., Olivetti, E., Rendine, S. (1988) A genetic history of Italy. *Ann Hum Genet.* 52, 203.

[9] Cappello, N., Rendine, S., Griffo, R., Mameli, G.E., Succa, V., Vona, G., Piazza, A. (1996) Genetic analysis of Sardinia: I. data on 12 polymorphisms in 21 linguistic domains. *Ann Hum Genet*. 60(2),125.

[10] Contu, L., Carcassi, C., Orrù, S., Mulargia, M., Arras, M., Boero, R., Gessa, S., Loizedda, A.L., Lai, S., Floris, L. (1998) HLA-B35 frequency variations correlate with malaria infection in Sardinia. *Tissue Antigens,* 52 (5), 452.

[11] Lampis, R., Morelli, L., Congia, M., Macis, M.D., Mulargia, A., Loddo, M., De Virgiliis, S., Marrosu, M.G., Todd, J.A., Cucca, F. (2000) The inter-regional distribution of HLA class II haplotypes indicates the suitability of the Sardinian population for case-control association studies in complex diseases. *Hum Mol Genet*., 9 (20), 2959.

[12] Songini, M., Casu, A. (2005) Epidemiology of childhood diabetes. *Acta Biomed.,*76, 19.

[13] Karvonen, M., Tuomilehto, J., Libman, I., La Porte, R. (1993) A review of the recent epidemiological data on the worldwide incidence of type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* 36, 883–892.

[14] Pugliatti M, Rosati G, Carton H et al: The epidemiology of multiple sclerosis in Europe.Eur J Neurol 2006; 13: 700–722.

[15] McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., et al. (2008) Integrated detection and population genetic analysis of SNPs and copy number variation. *Nat. Genet.,* 40, 1166-1174.

[16] Sabeti PC, Varilly P, Fry B et al: Genome-wide detection and characterization of positive selection in human populations. Nature 2007; 449: 913–918.

[17] Simonson TS, Yang Y, Huff CD et al: Genetic evidence for high-altitude adaptation in Tibet. Science 2010; 329: 72–75.

[18] Williamson SH, Hubisz MJ, Clark AG et al: Localizing recent adaptive evolution in the human genome. Plos Genet 2007; 3: e90.

[19] Tognotti E: La Malaria in Sardegna – Per Una Storia Del Paludismo Nel Mezzogiorno (1880–1950), In Franco Angeli (ed). Milano, 1996.

[20] Fermi C: Le Regioni Malariche. Decadenza e Risanamento E Spesa ‘Sardegna’,Vol I: Roma: Tipografia editrice di Roma S.A, 1934.

[21] Fermi C: Provincia Di Nuoro. Malaria, Danni Economici, Risanamento E Proposte Per Il Suo Risorgimento, Vol II:Sassari: Stamperia della libreria italiana e straniera, 1938.

[22] Myles S, Davison D, Barrett J et al: Worldwide population differentiation at disease-associated SNPs. BMC Med Genomics 2008; 1: 22.

4.2. Study No. 2,

In collaboration with researchers from the following international institutions: Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK - Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA - Department of Pediatrics, Keck School of Medicine and Children's Hospital Los Angeles, University of Southern California, Los Angeles, California 90027, USA - T.T. Chang Genetic Resources Center, International Rice Research Institute, Los Ban˜os, Laguna , Philippines. - Department of Sciences of Life and Environment, University of Cagliari Italy, - Department of Biology, University of Pisa, Italy, - Department of Science of Nature and Territory, University of Sassari, Italy, - The Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK, - Department of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA, - Departamento de Toxicologı ´a, Cinvestav, San Pedro Zacatenco, CP 07360, Mexico. 12 Instituto de Gene´tica y Biologı ´a Molecular, University of San Martin de Porres, Lima, Peru, - Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, - Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciences de la Salut i de la Vida, Universitat Pompeu Fabra, 08003 Barcelona, Spain, - Vavilov Institute for General Genetics: Moscow, Russia, - Research Centre for Medical Genetics, Moscow, Russia, - The Lebanese American University, Chouran, Beirut, Lebanon, National Health Laboratory Service, Johannesburg, South Africa, - The Genographic Laboratory, School of Biological Sciences, Madurai Kamaraj University, Madurai, India, Department of ecology and evolutionary biology, University of Arizona, Tucson, Arizona 85721, USA, - Department of Anatomy, University of Otago, Dunedin 9054, New Zealand, - National Geographic Society,Washington, District of Columbia 20036, USA, that ended with the article http://www.ncbi.nlm.nih.gov/pubmed/24781250, Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P, Pagani L, Tyler-Smith C, Xue Y, Cucca F, Schurr TG, Gaieski JB, Melendez C, Vilar MG, Owings AC, Gómez R, Fujita R, Santos FR, Comas D, Balanovsky O, Balanovska E, Zalloua P, Soodyall H, Pitchappan R, Ganeshprasad A, Hammer M, Matisoo-Smith L, Wells RS; Genographic Consortium. *Geographic population structure analysis of worldwide human populations infers their biogeographical origins*. Nature Communications. 2014 Apr 29;5:3513.

The paper proposes an algorithm to identify the geographical origin of the individuals using their biological data, and thus trying to overcome one of the most complex issues of population genetics and anthropology. In fact the nature of biological diversity, makes it hard to tell where one stops and the other population starts watching the simple spatial distribution of a feature fenotipica[1-4] but at the same time important studies[5-6] have established a strong relationship between the genetic characteristics of different populations and geographical distances separating them. These observations, has stimulated the development of biogeographic methods. The different biogeographical applications currently in place are not very precise and therefore the percentage of individuals correctly assigned to their country of origin, is relatively low for Europeans and very low for non Europeans[7]. The work was then hypothesized and a new analysis model that is based on the fundamental approach of mixing and genetic analysis including 100,000 ancestral informative markers (AIMs), reported on platforms GenChip.

This analysis tool identifies the smallest number of markers sufficient to adequately differentiate two genetically distinct populations. It was applied to the set of public and private data relating to the sampled populations in various parts of the world, in order to validate the potential application. About 600 individuals collected in Genographic Project and circa 1000 DNA collected at different Genomes Project in several geographical areas around the world, including very heterogeneous populations such as rural populations of Kuwait[8], Puerto Rico and Bermuda[9], as well as the community of the same country and therefore theoretically closer: such as Peruvians from Lima and highland indigenous Peruvians. When applied to 200 Sardinians[10], the GPS Placed a quarter of them in their villages and most of the rest within 50km of their villages.

4.2.1. Conclusion

The work for which we have made available a part of the data set of genetic data concerning the Ogliastra population samples analyzed by us, describes a method of geographical population structure (GPS) based on mixing which aims to predict the biogeographical origin of individuals worldwide from local residents. The validation performed on our sample and those of many other international groups valid the new algorithm Geographic Population Structure (GPS) by placing approximately 83% of worldwide individuals in their country of origin. The analysis led to the formation of clusters

for those subjects from the same geographic area and allowed the distinction of populations from different geographical areas. For each individual was possible to assign with relative accuracy the geographical coordinates and the distance of the supposed place of origin.

4.2.2. References

[1] Harcourt, A. H. Human Biogeography (University of California Press, 2012)

[2] Darwin, C. The Descent of Man and Selection in Relation to Sex (John Murray London, 1871).

[3] Rowe, J. H. The Renaissance Foundations of Anthropology. American Anthropologist 67, 1–20 (1965).

[4] Cavalli-Sforza, L. L. L., Menozzi, P. & Piazza, A. The History and Geography of Human Genes (Princeton university press (1994).

[5] Eller, E. Population substructure and isolation by distance in three continental regions. Am. J. Phys. Anthropol. 108, 147–159 (1999).

[6] Relethford, J. H. Global analysis of regional differences in craniometric diversity and population substructure. Hum. Biol. 73, 629–636 (2001).

[7] Ramachandran, S. et al. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl Acad. Sci. USA 102, 15942–15947 (2005).

[8] Alsmadi, O. et al. Genetic substructure of Kuwaiti population reveals migration history. PLoS One 8, e74913 (2013).

[9] Price, A. L. et al. A genomewide admixture map for Latino populations. Am. J.Hum. Genet. 80, 1024–1036 (2007).

[10] Piras, I. S. et al. Genome-wide scan with nearly 700,000 SNPs in two Sardinian sub-populations suggests some regions as candidate targets for positive selection. Eur. J. Hum. Genet. 20, 1155–1161 (2012).

5. General conclusions and perspectives

The data obtained in the second phase of my research analyzing the Biobank created in Ogliastra have highlighted already important characteristics of this population, useful for anthropological studies and validation of new tools / algorithms to be applied to the study of population genetics. However, the goal remains, start a genotyping population control, healthy, for the study of the genes responsible for many common multifactorial diseases. The homogeneity of the Ogliastra population, all selected subjects (healthy controls and patients) strictly belonging to a single population without ethnic stratification, with the absence of consanguinity between until 4th or 5th generation, the absence among the controls, of family members with common complex diseases, and the minimum risk that some subject included among the controls have a latent predisposition to one or more complex diseases, along with the number of subjects that make sampling of controls and that among some of the most important complex diseases, and the complete anonymity of the subjects included in the Biobank, suggest that this is an ideal sample for the research of genes and gene combinations involved in complex diseases, through case-control association studies.

The dataset of DNA samples of all feature controls will be the reference for the study of the most common multifactorial diseases in Ogliastra and the world. In fact they are stored in the same way the DNA samples of patients with multifactorial diseases, which is currently in progress along the whole genome genotyping and analysis of related datasets.

Using combinations of genotypic susceptibility and/or protection it emerged that, you can identify allelic combinations or haplotype high predictive values for the detection of both individuals at risk or protected of disease. For example it is well known that children of parents with diabetes have a very high risk of developing the disease during the first decades of life. It would be very important to recognize early in families at risk as children have genetic conditions of high risk T1D, and those who have a genetic conditions of non-risk or protection. The prospects for effective preventive interventions in individuals at risk for T1D become thanks to this research increasingly concrete.