UNIVERSITÀ DEGLI STUDI DI PADOVA

Department of Pharmaceutical and Pharmacological Sciences

Ph.D. COURSE IN PHARMACOLOGICAL SCIENCES
CURRICULUM: PHARMACOLOGY, TOXICOLOGY AND THERAPEUTICS
SERIES: XXXI

Development of a NGS workflow for diagnostic applications in oncology

Thesis written with the financial contribution of DOTT. DINO PALADIN Company

Coordinator: Prof. Piero Maestrelli
Supervisor: Prof. Sara De Martin
Co-Supervisor: Dr. Dino Paladin

Ph.D. student: Massimiliano Gianni Cattelan

"Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza"

Dante Alighieri, Divina Commedia,
Inferno canto XXVI, vv 118-120

TABLE OF CONTENTS

1 ABSTRACT 1

2 INTRODUCTION 3

2.1 CANCER OVERVIEW 3

2.1.1 EPIDEMIOLOGY 3

2.1.2 CANCER CHARACTERISTICS 4

2.2 OVARIAN CANCER 6

2.2.1 EPIDEMIOLOGY 6

2.2.2 CLASSIFICATIONS 6

2.3 HIGH-GRADE SEROUS OVARIAN CANCER 9

2.3.1 BIOLOGICAL CHARACTERISTICS 9

2.3.1.1 MACROSCOPICAL FEATURES 9

2.3.1.2 HISTOPATHOLOGICAL FEATURES 9

2.3.1.3 GENETIC PROFILE 10

2.3.1.4 GENETIC SUSCEPTIBILITY 11

2.3.1.5 PROGNOSTIC AND PREDICTIVE FACTORS 11

2.3.2 DIAGNOSIS 12

2.3.3 TREATMENTS AND FOLLOW UP 12

2.3.3.1 SURGERY 12

2.3.3.2 PHARMACOLOGICAL THERAPY 13

2.3.3.3 FOLLOW UP 14

2.4 HERCULES PROJECT 15

2.5 NEXT GENERATION SEQUENCING IN CLINICAL ONCOLOGY 18

2.5.1 NGS OVERVIEW 18

2.5.2 ILLUMINA TECHNOLOGY 20

2.5.3 NGS ONCOLOGICAL APPLICATION 23

2.5.3.1 TARGETED SEQUENCING 24

2.5.3.2 TARGETED PANEL DEFINITION 24

2.5.3.3 TARGETED SEQUENCING WORKFLOW 24

2.5.3.3.1 SAMPLE PREPARATION 25

2.5.3.3.1.1 TISSUE CONSIDERATIONS 25

2.5.3.3.1.2 DNA EXTRACTION CONSIDERATIONS 25

2.5.3.3.2 TARGETED LIBRARY PREPARATION 26

2.5.3.3.2.1 HYBRID CAPTURE TECHNOLOGY 26

2.5.3.3.2.2 AMPLICON TECHNOLOGY 27
2.5.3.3.3 SEQUENCING 28
2.5.3.3.4 DATA ANALISIS 28
2.5.3.3.4.1 SEQUENCE GENERATION 28
2.5.3.3.4.2 SEQUENCE ALIGNMENT 29
2.5.3.3.4.3 VARIANT CALLING 29
2.5.3.3.4.4 VARIANT FILTERING 29
2.5.3.3.4.5 VARIANT ANNOTATION 29
3 AIM OF THE WORK 32
4 MATERIALS AND METHODS 34
4.1 PATENT SEARCH ANALISYS 34
4.2 BIOLOGICAL SAMPLES COLLECTION 35
4.2.1 BIOLOGICAL SAMPLES 35
4.2.2 FFPE SAMPLES PREPARATION 35
4.3 SAMPLES NOMENCLATURE 36
4.4 NUCLEIC ACID EXTRACTION KITS 37
4.4.1 COMMERCIAL NUCLEIC ACID EXTRACTION KIT TECHNOLOGY 37
4.4.1.1 SPIN COLUMN-BASED NUCLEIC ACID PURIFICATION TECHNOLOGY 37
4.4.1.2 MAGNETIC BEAD BASED NUCLEIC ACID PURIFICATION TECHNOLOGY 38
4.4.2 NUCLEIC ACID KITS LIST 39
4.5 NUCLEIC ACID EXTRACTION KITS SCREENING CRITERIA 41
4.5.1 NUCLEIC ACID EXTRACTION PROTOCOLS 42
4.6 NUCLEIC ACID QUANTIFICATION 42
4.6.1 ABSORBANCE: THERMOFISHER SCIENTIFIC NANODROP 2000 42
4.6.2 FLUORESCENCE: THERMOFISHER SCIENTIFIC QUBIT 3 FLUOROMETER 44
4.7 FRAGMENT SIZE DISTRIBUTION: AGILENT 2200 TAPESTATION SYSTEM 44
4.8 SAMPLE QUALITY CONTROLS FOR NGS LIBRARY PREPARATION: QIASEQ DNA QUANTIMIZE SYSTEM 45
4.9 NGS LIBRARY GENES PANELS 47
4.9.1 ILLUMINA TRUSEQ CUSTOM AMPLICON LOW INPUT 47
4.9.2 QIASEQ TARGETED DNA CUSTOM PANEL 48

4.10 LIBRARIES QUANTIFICATIONS FOR NGS ANALYSIS: KAPA LIBRARY QUANTIFICATION KIT – ILLUMINA PLATFORMS 51
4.11 NGS ANALYSIS: ILLUMINA MISEQ 52
4.11.1 LIBRARIES POOL LOADING SOLUTION PREPARATION 53
4.11.2 QIASEQ A READ 1 PRIMER I PREPARATION 53
5 RESULTS AND DISCUSSION 54
5.1 PATENT SEARCH ANALISYS 56
5.1.1 DISCUSSION 56
5.2 NGS LIBRARY PREPARATION KIT FOR HGS-OvCa 58
5.2.1 GENE PANEL DESIGN 58
5.2.2 DISCUSSION 60
5.3 NUCLEIC ACID EXTRACTION KITS SCREENING 63
5.4 NUCLEIC ACID EXTRACTION KITS PERFORMANCES ANALYSIS 68
5.4.1 DNA QUANTIFICATION TESTS 68
5.4.1.1 ANALYTIK JENA - BLACKPREP FFPE DNA KIT 68
5.4.1.2 BIOCHAIN - FFPE TISSUE DNA EXTRACTION KIT - COLUMN 69
5.4.1.3 MACHEREY-NAGEL - NUCLEOSPIN DNA FFPE XS 71
5.4.1.4 MAGBIO GENOMICS - HIGHPREP FFPE TISSUE DNA KIT 73
5.4.1.5 OMEGA BIO-TEK - E.Z.N.A. FFPE DNA KIT 75
5.4.1.6 OMEGA BIO-TEK - MAG-BIND FFPE DNA KIT 77
5.4.1.7 PROMEGA - MAXWELL RSC DNA FFPE KIT 79
5.4.1.8 PROMEGA - RELIAPREP FFPE GDNA MINIPREP SYSTEM 79
5.4.1.9 QIAGEN - ALLPREP DNA/RNA FFPE KIT 82
5.4.1.10 QIAGEN - GENEREAD DNA FFPE KIT 83
5.4.1.11 PERFORMANCES OVERVIEW 87
5.4.1.12 DISCUSSION 87
5.4.2 RNA QUANTIFICATION TESTS 92
5.4.2.1 MACHEREY-NAGEL - NUCLEOSPIN TOTALRNA FFPE XS 92
5.4.2.2 OMEGA BIO-TEK - E.Z.N.A. FFPE RNA KIT 94
5.4.2.3 OMEGA BIO-TEK - MAG-BIND FFPE RNA KIT 95
5.4.2.4 QIAGEN - RNEASY FFPE KIT 96
5.4.2.5 PERFORMANCES OVERVIEW 98
5.4.2.6 DISCUSSION 98
5.5 DNA EXTRACTION FROM FFPE SAMPLES 100

5.5.1 BIOCHAIN - FFPE TISSUE DNA EXTRACTION KIT - COLUMN 100
5.5.2 PROMEGA - RELIAPREP FFPE GDNA MINIPREP SYSTEM 101
5.5.3 QIAGEN - GENEREAD DNA FFPE KIT 101
5.5.4 EXTRACTIONS OVERVIEW 102
5.5.5 DISCUSSION 102
5.6 QUALITY CONTROLS FOR NGS LIBRARIES PREPARATION 104
5.6.1 REAL TIME PCR QUANTIFICATION ANALYSIS 104
5.6.1.1 DISCUSSION 105
5.7 NGS LIBRARIES QUANTIFICATION ANALYSIS 107
5.7.1 FLUORESCENT-BASED NGS LIBRERIES QUANTIFICATION AND AMPLICON DISTRIBUTION ANALYSIS 107
5.7.1.1 DISCUSSION 112
5.7.2 REAL TIME PCR QUANTIFICATION ANALYSIS 113
5.7.2.1 LIBRARIES QUANTIFICATION 113
5.7.2.1.1 DISCUSSION 116
5.7.2.2 POOL QUANTIFICATION 117
5.7.2.2.1 DISCUSSION 118
5.8 NGS ANALYSIS 119
5.8.1 DISCUSSION 119
6 CONCLUSIONS 121
7 REFERENCES 125
8 ACKNOWLEDGEMENTS 131
9 SUPPLEMENTARY MATERIALS 132
9.1 APPENDIX A 132

1 ABSTRACT

The Ph.D. research work is an integral part of a Horizon 2020 project, called **HERCULES** project (Compre**HE**nsive cha**R**acterisation and effe**C**tive combinatorial targeting of high-grade sero**U**s ovarian cancer via sing**LE**-cell analysi**S**).

The topics of the Horizon 2020 project are to comprehensively characterize high grade serous ovarian cancer (HGS-OvCa) by integrating and modeling clinical and biological data (e.g., genetics, transcriptomics, protein binding, drug screens) from primary, metastatic and relapsed tumors from various anatomical sites of HGS-OvCa patients, and establish combinatorial treatment modalities that effectively kill HGS-OvCa tumor cell subpopulations.

The role of the Company in HERCULES project is to develop and validate a marketable prototype biomarker kit for predicting HGS-OvCa patients response to combinatorial therapeutic modalities.

In detail, the kit hypothesized for this application is a NGS gene panel, based on an Illumina platform and technology, capable to predict the outcome of a pharmacological therapy using high grade serous ovarian cancer subpopulation genetic biomarkers.

A further aim of the Ph.D. project is more generally to study the processes required for developing an *in vitro* diagnostic (IVD) workflow finalized to a next generation sequencing analysis for oncology applications, not only focused on high grade serous ovarian cancer, having diagnostic, prognostic and predictive purposes.

In order to achieve this goal, the following activities were carried out:

1. Evaluation of the "state of art" regarding the presence of patents relevant for the HERCULES project.
2. Definition of a gene panel design based on several genes involved in HGS-OvCa.
3. Identification and selection of FFPE DNA and RNA extraction kits having features, in terms of nucleic acid yields, quality and purity, that are compatible with the NGS downstream analysis and an IVD workflow.
4. Study of the processes necessary to sequence 12 human FFPE samples, having both KRAS wild type and mutated, on the Illumina platform using amplicon technology.

As a result of this work three commercial column-based DNA extraction kits have been identified as the most effective when included in a NGS workflow based on Illumina technology.

The study and the development of the NGS workflow were done employing available clinical FFPE sections of human colorectal cancer with already determined KRAS mutational status using established IVD assays.

The results obtained with the NGS analysis have also demonstrated that the gene panel design for the library preparation kit, which uses both amplicon and UMI (unique molecular identifier)

technologies, provides high UMI incorporation, high coverage depth of the regions of interest. This are fundamental aspects required for the identification of false-positive and mutations expressed at very low levels.

These results will be used for developing clinical custom genes panel, intended to be used by pathologists and oncologists, capable to provide:

- Prediction cancer onset.
- Characterization of different tumor types.
- Precise information about the exact therapeutic treatment for different cancer types (HGS-OvCa included).
- Prognosis.

2 INTRODUCTION

2.1 CANCER OVERVIEW

2.1.1 EPIDEMIOLOGY

In 2012, the cancer was one of the main causes of morbidity and mortality worldwide; in fact, there were approximately 14.068 million new cancer cases recorded, 8.202 million cancer deaths and 32.455 million people living with cancer (within 5 years of diagnosis) worldwide (considering both sex) [1] (**Figure 1**).

Estimated numbers (thousands)	Men			Women			Both sexes		
	Cases	Deaths	5-year prev.	Cases	Deaths	5-year prev.	Cases	Deaths	5-year prev.
World	7410	4653	15296	6658	3548	17159	14068	8202	32455
More developed regions	3227	1592	8550	2827	1287	8274	6054	2878	16823
Less developed regions	4184	3062	6747	3831	2261	8885	8014	5323	15632
WHO Africa region (AFRO)	265	205	468	381	250	895	645	456	1363
WHO Americas region (PAHO)	1454	677	3843	1429	618	4115	2882	1295	7958
WHO East Mediterranean region (EMRO)	263	191	461	293	176	733	555	367	1194
WHO Europe region (EURO)	1970	1081	4791	1744	852	4910	3715	1933	9701
WHO South-East Asia region (SEARO)	816	616	1237	908	555	2041	1724	1171	3278
WHO Western Pacific region (WPRO)	2642	1882	4493	1902	1096	4464	4543	2978	8956
IARC membership (24 countries)	3689	1900	9193	3349	1570	9402	7038	3470	18595
United States of America	825	324	2402	779	293	2373	1604	617	4775
China	1823	1429	2496	1243	776	2549	3065	2206	5045
India	477	357	665	537	326	1126	1015	683	1790
European Union (EU-28)	1430	716	3693	1206	561	3464	2635	1276	7157

International Agency for Research on Cancer
World Health Organization

Figure 1: Cancer estimated Incidence, Mortality and Prevalence Worldwide in 2012.
http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx

Focusing on United States of America (USA) and European Union (EU) countries, two of the more developed regions, it is possible to see that:

- The number of new cancer cases is about 47% higher in EU than USA (2.635 million vs 1.604 million);
- The number of deaths is more than double in EU respect to USA (1.276 million vs 0.617 million).

Statistical projections estimate that the number of new cancer cases in USA and EU will increase by respectively about 48% and 30% over the 2035 [2].

More recent analysis underlines that cancer is the second leading cause of death globally, and was responsible for 8.8 million deaths in 2015. Globally, nearly 1 in 6 deaths is due to cancer [3].

All these data suggest that the economic impact of cancer worldwide is significant and is still increasing, in fact, considering only the United States, the national expenditures for cancer care was approximately 125 billion dollars in 2010 and can reach potentially $156 billion in 2020 [4].

Considering all these social and economic implications, researchers, clinicians and in particular diagnostic/pharmaceutical companies are investing many resources in terms of personnel, time and

money for reaching scientific knowledge and instruments capable to cope with and, possibly, eradicate cancer as an invalidating and lethal disease.

2.1.2 CANCER CHARACTERISTICS

Cancer is a disease that can affect various organs in which normal cells lose the capability to regulate their growth, leading to an abnormal proliferation. Tumor cells acquire also the ability to migrate to other organs in the body where they can occupy new tissues and proliferate uncontrollably.

These malignant cells have various effects on normal tissue; typically, they are able to modify the homeostatic system and organ functions, potentially until the complete loss of tissue function and leading it to degradation or necrosis [5].

Carcinogenesis is a multistep process in which normal cells collect genomic alterations (for example in nucleotide sequences coding for essential proteins involved in normal cell growth, regulation and homeostasis) that gradually drive the progressive transformation into highly malignant derivatives. Tumor cells are genetically, phenotypically and morphologically distinguishable from normal tissue [6].

In the last decades, scientists have found that cancer cells need normal cells (including immune system cells) and different types of molecules/hormones for growing and acquiring malignant characteristics.

Indeed, tumors are very complex tissues composed of cancer cells and multiple distinct homeostatic cell types (named tumor-associated stroma) that actively take a part in the tumorigenesis phenomenon [7].

In order to provide a logical framework to rationalize the neoplastic disease complexity, scientists have established specific biological mechanisms (and respective molecular pathways) that allow cancer cells to survive, proliferate, and disseminate.

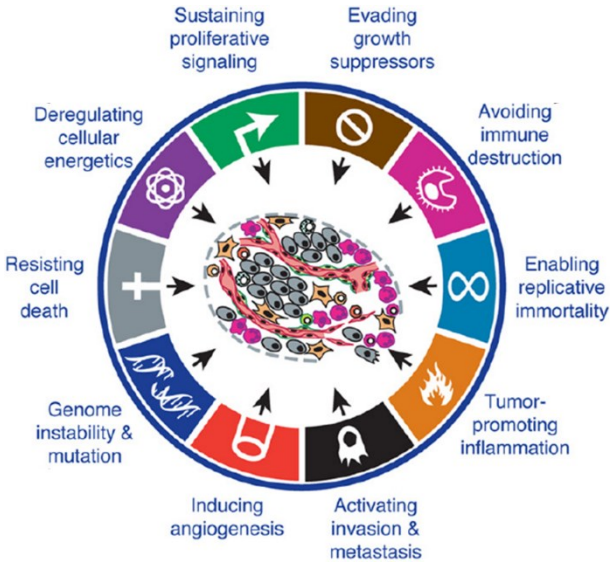


Figure 2: The picture shows the tumor characteristics summarized from Hannan an Weinber in 2000 (and confirmed also in 2011). Hanahan D., Weinberg R. A. "Hallmarks of cancer: the next generation." Cell. 2011.

Only with the identification and the understanding of genes and proteins (in particular their mutations/alterations) that play a key role into each molecular pathway and mechanism responsible for cancer promotion and generation, it will be possible to detect, treat and prevent the tumor onset correctly.

Among all tumors affecting humans, ovarian cancer is a severe and deadly gynecological tumor; many scientists are focusing their efforts to understand the biological and pathological features of this very heterogeneous disease in order to find the best diagnostic biomarkers and therapeutic treatments to reduce mortality as much as possible.

2.2 OVARIAN CANCER

2.2.1 EPIDEMIOLOGY

Ovarian cancer is one of the most common cancer diagnosis among women worldwide.

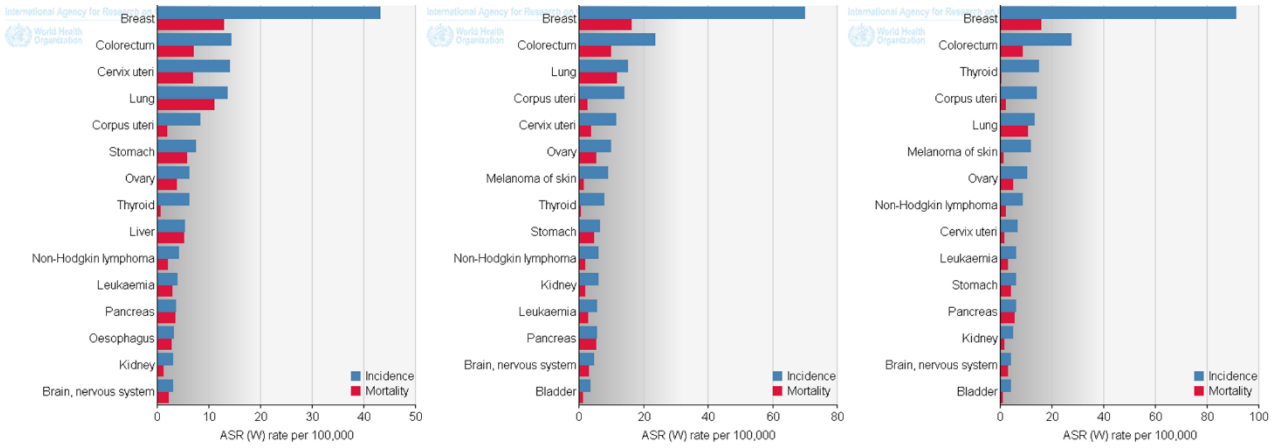


Figure 3: Incidence and mortality of most common women cancer. The three bar charts shows the 15 tumors affecting women organized on incidence rate expressed as age-standardized rate. Data represent world (left), European Union (centre) and Italy (right) overviews in 2012. http://globocan.iarc.fr/Pages/fact_sheets_population.aspx.

In detail, World Health Organization (WHO) statistics show that in 2012 ovarian cancer is (**Figure 3**):

- The seventh cancer type with higher incidence in the world (considering only women population, **Figure 3, left**). The incidence world rate is 6.1 per 100,000 women (238,719 affected and 151,917 died, with a 5-year prevalence of 586,624 women, equal to 22.6 per 100,000 women) [1].
- The sixth cancer type affecting women with higher incidence and, at the same time, the third most frequent gynecological cancer in Europe (**Figure 3, centre)**. The European incidence rate is 9.4 per 100,000 women (70,320 women affected and 45,945 died, with a 5-year prevalence of 168,878 women, equal to 43.5 per 100,000 female) [1].

Focusing on Italy (**Figure 3, right)**, ovarian cancer has an incidence approximately of 3.6% of all tumors affecting Italian women (238,719 on 6,657,518 total cases), with a mortality close to 4.0 per 100,000 women [1].

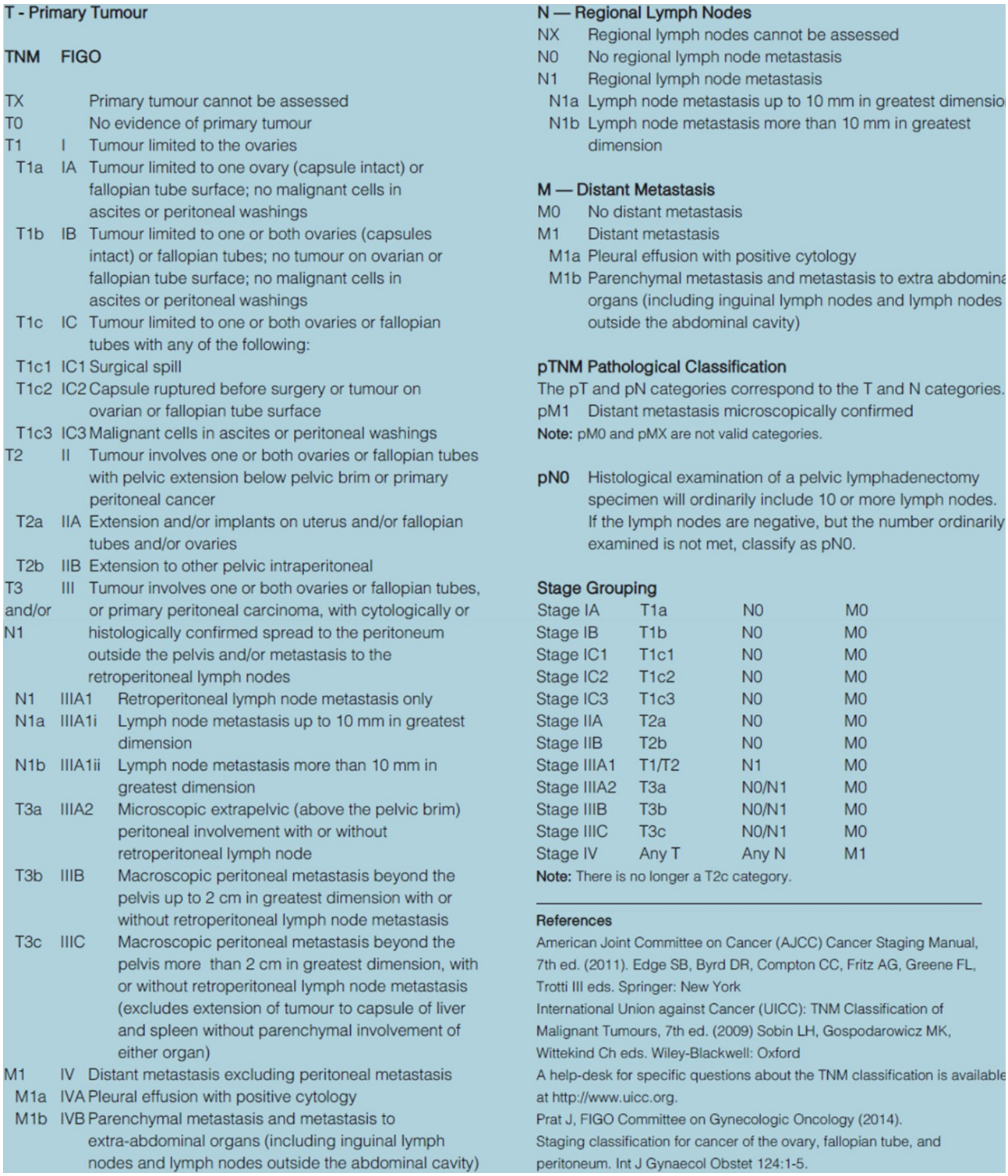
2.2.2 CLASSIFICATIONS

Ovarian cancer is a disease consisting of a heterogeneous group of malignant tumors with different features in terms of etiology, molecular biology, pathogenesis, prognosis and pathology [8].

The stage and grade characterizations of ovarian cancer play a fundamental role in the diagnosis, prognosis and the tumor treatment.

The term “grade” refers to cell morphology within the tumor and the term “stage” describes the onset and distance between primary and secondary tumor site, derived from metastasis colonization. Stage involves the primary tumor location, the tumor size, lymph node involvement and metastasis at diagnosis. Ovarian cancer staging normally occurs through surgery and pathological analysis of tissue cancer fragments [9].

Clinicians and scientists use mainly the following classification systems for analyzing ovarian cancer at diagnosis: the WHO Classification and the Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) system, based on the Tumor Node Metastasis (TNM) system (**Figure 4**).



T - Primary Tumour

TNM	FIGO	
TX		Primary tumour cannot be assessed
T0		No evidence of primary tumour
T1	I	Tumour limited to the ovaries
T1a	IA	Tumour limited to one ovary (capsule intact) or fallopian tube surface; no malignant cells in ascites or peritoneal washings
T1b	IB	Tumour limited to one or both ovaries (capsules intact) or fallopian tubes; no tumour on ovarian or fallopian tube surface; no malignant cells in ascites or peritoneal washings
T1c	IC	Tumour limited to one or both ovaries or fallopian tubes with any of the following:
T1c1	IC1	Surgical spill
T1c2	IC2	Capsule ruptured before surgery or tumour on ovarian or fallopian tube surface
T1c3	IC3	Malignant cells in ascites or peritoneal washings
T2	II	Tumour involves one or both ovaries or fallopian tubes with pelvic extension below pelvic brim or primary peritoneal cancer
T2a	IIA	Extension and/or implants on uterus and/or fallopian tubes and/or ovaries
T2b	IIB	Extension to other pelvic intraperitoneal
T3 and/or N1	III	Tumour involves one or both ovaries or fallopian tubes, or primary peritoneal carcinoma, with cytologically or histologically confirmed spread to the peritoneum outside the pelvis and/or metastasis to the retroperitoneal lymph nodes
N1	IIIA1	Retroperitoneal lymph node metastasis only
N1a	IIIA1i	Lymph node metastasis up to 10 mm in greatest dimension
N1b	IIIA1ii	Lymph node metastasis more than 10 mm in greatest dimension
T3a	IIIA2	Microscopic extrapelvic (above the pelvic brim) peritoneal involvement with or without retroperitoneal lymph node
T3b	IIIB	Macroscopic peritoneal metastasis beyond the pelvis up to 2 cm in greatest dimension with or without retroperitoneal lymph node metastasis
T3c	IIIC	Macroscopic peritoneal metastasis beyond the pelvis more than 2 cm in greatest dimension, with or without retroperitoneal lymph node metastasis (excludes extension of tumour to capsule of liver and spleen without parenchymal involvement of either organ)
M1	IV	Distant metastasis excluding peritoneal metastasis
M1a	IVA	Pleural effusion with positive cytology
M1b	IVB	Parenchymal metastasis and metastasis to extra-abdominal organs (including inguinal lymph nodes and lymph nodes outside the abdominal cavity)

N — Regional Lymph Nodes

NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis
N1a	Lymph node metastasis up to 10 mm in greatest dimensio
N1b	Lymph node metastasis more than 10 mm in greatest dimension

M — Distant Metastasis

M0	No distant metastasis
M1	Distant metastasis
M1a	Pleural effusion with positive cytology
M1b	Parenchymal metastasis and metastasis to extra abdomina organs (including inguinal lymph nodes and lymph nodes outside the abdominal cavity)

pTNM Pathological Classification

The pT and pN categories correspond to the T and N categories.

pM1 Distant metastasis microscopically confirmed

Note: pM0 and pMX are not valid categories.

pN0 Histological examination of a pelvic lymphadenectomy specimen will ordinarily include 10 or more lymph nodes. If the lymph nodes are negative, but the number ordinarily examined is not met, classify as pN0.

Stage Grouping

Stage IA	T1a	N0	M0
Stage IB	T1b	N0	M0
Stage IC1	T1c1	N0	M0
Stage IC2	T1c2	N0	M0
Stage IC3	T1c3	N0	M0
Stage IIA	T2a	N0	M0
Stage IIB	T2b	N0	M0
Stage IIIA1	T1/T2	N1	M0
Stage IIIA2	T3a	N0/N1	M0
Stage IIIB	T3b	N0/N1	M0
Stage IIIC	T3c	N0/N1	M0
Stage IV	Any T	Any N	M1

Note: There is no longer a T2c category.

References

American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 7th ed. (2011). Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti III eds. Springer: New York

International Union against Cancer (UICC): TNM Classification of Malignant Tumours, 7th ed. (2009) Sobin LH, Gospodarowicz MK, Wittekind Ch eds. Wiley-Blackwell: Oxford

A help-desk for specific questions about the TNM classification is available at http://www.uicc.org.

Prat J, FIGO Committee on Gynecologic Oncology (2014). Staging classification for cancer of the ovary, fallopian tube, and peritoneum. Int J Gynaecol Obstet 124:1-5.

Figure 4: TNM and FIGO classification of tumors of the ovary, fallopian tube and primary peritoneal carcinoma. FIGO staging system uses information obtained after surgery. The alphanumeric TNM system describes the cancer stage, which derives from a solid tumor. In detail, the system analyzes: 1) primary tumor size (T), including contiguous tissues invasion; 2) regional lymph nodes involvement (N) 3) Metastasis formation (M). Each single TNM value is combined in order to generate the diseases stage, which are compatible with FIGO stage classification [11]. Kurman R.J., Carcangiu M.L., Herrington C.S. et al "WHO Classification of Tumours of Female Reproductive Organs". WHO Classification of Tumours. 4th Edition. 2014.

The WHO classifies the ovarian cancer based on the histological diversity of tumor samples. In detail, the ovarian cancer is classified in 14 macro-histological categories [10].

Among the different ovarian cancer histological groups, over 90% of malignant ovarian tumors are epithelial cancers [10] and most deaths (approximately 70%) are patients presenting high-grade serous ovarian cancer [12], for this reason the Ph.D. thesis focuses on this particular epithelial ovarian cancer subtype.

2.3 HIGH-GRADE SEROUS OVARIAN CANCER

High-grade serous ovarian cancer (HGS-OvCa) represents the most common and lethal histotype of epithelial ovarian cancer. This high mortality rate is due in large part to poor early detection methods leading to late diagnosis after the disease has metastasized [13] [14].

2.3.1 BIOLOGICAL CHARACTERISTICS

2.3.1.1 MACROSCOPICAL FEATURES

HGS-OvCa shows variable size and the ovaries are typically either normal or display surface nodules of tumor smaller than 1 cm. HGS-OvCa is frequently bilateral, exophytic and shows solid and papillary growth and fluid-filled cysts. The solid regions are tan-white and typically contain extensive necrosis and hemorrhage. The fallopian tube may be involved. A small, firm, polypoid lesion could be present at its fimbriated end. The fimbriae may be splayed over the tumor surface [10].

2.3.1.2 HISTOPATHOLOGICAL FEATURES

HGS-OvCa is generally constituted by solid masses of cells with slit-like spaces, papillary, glandular and cribriform areas, where necrosis is typically present.

The tumor nuclei are large, prominent, hyperchromatic and pleomorphic, frequently with multinucleated or atypical shapes. Furthermore, copious and anomalous mitoses structures are present. Epithelial cells urothelial carcinoma-like that form a papillary pattern or one of thick undulating bands may be present in the tumor [10].

Immunohistochemistry studies show that the WT1 nuclear expression can be considered a useful marker for HGS-OvCa, because it is not present in most endometrial serous carcinomas.

The expression of p53 protein is fundamental to demonstrate two different patterns of HGS-OvCa, which are related to TP53 mutations.

The typical and the most common HGS-OvCa pattern shows a TP53 strong diffuse nuclear staining in approximately 60% of cells or greater and it correlates with a missense mutation.

The other pattern shows complete absence of TP53 protein staining, which correlates with a nonsense mutation resulting in a truncated protein that is not detected by the p53 antibody.

Both patterns correlate with a TP53 mutation and since mutated TP53 occurs approximately in all HGS-OvCa, this stain is considered helpful in confirming the diagnosis of HGS-OvCa [10] [12].

2.3.1.3 GENETIC PROFILE

HGS-OvCa is characterized by high genetic instability and high frequency of DNA copy number gain or losses.

The most prominent genetic mutations include TP53 alterations in almost all tumors (96%), and inactivation (through germline or somatic mutation or promoter methylation) of BRCA1 (about 12.5%, 9% germline mutation and 3.5% somatic mutations) and BRCA2 (about 11.5%, 8% germline mutations and 3.3% somatic mutations) [10] [15].

Genes that are recurrently mutated in HGS-OvCa are CSMD3 (6%), NF1 (4%), CDK12 (3%), GABRA6 (2%) and RB1 (2%) [12] [15].

The focal copy number aberrations are much more frequent compared to the gene mutations (113 copy number alterations were identified). The main focal amplifications are CCNE1 (Cyclin E1), MYC and MECOM (each is amplified in more than 20% of cases) [15]. Furthermore, about 20 genes are potential therapeutic targets, including MECOM, MAPK1, CCNE1 and KRAS, which are amplified in at least 10% of the cases [12] [15].

Copy number changes and/or gene expression alterations are present in important pathways of HGS-OvCa: RB1 and PI3K/RAS pathways are downregulated in 67% and 45% of cases respectively; and NOTCH signaling pathway is altered in 22% of cases (**Figure 5**) [15].

HGS-OvCa is also characterized by the alteration of homologous recombination pathway in 51% of cases. In detail:

- 20% of cases shows germline or somatic mutations in BRCA1 and BRCA2.
- 11% of cases shows the loss of BRCA1 expression through DNA hypermethylation (this methylation abnormality is mutually exclusive of BRCA1 mutations).
- 8% of cases shows amplification of EMSY, focal deletion or mutation of PTEN.
- 3% of cases shows hypermethylation of RAD51C.
- 2% of cases shows mutation of ATM or ATR.
- 5% of cases shows mutation of Fanconi anemia genes (**Figure 5**).

This observation provides a rationale for clinical trials based on the use of PARP inhibitors targeting HSG-OvCa homologous-recombination-related aberrations [10] [12] [15].

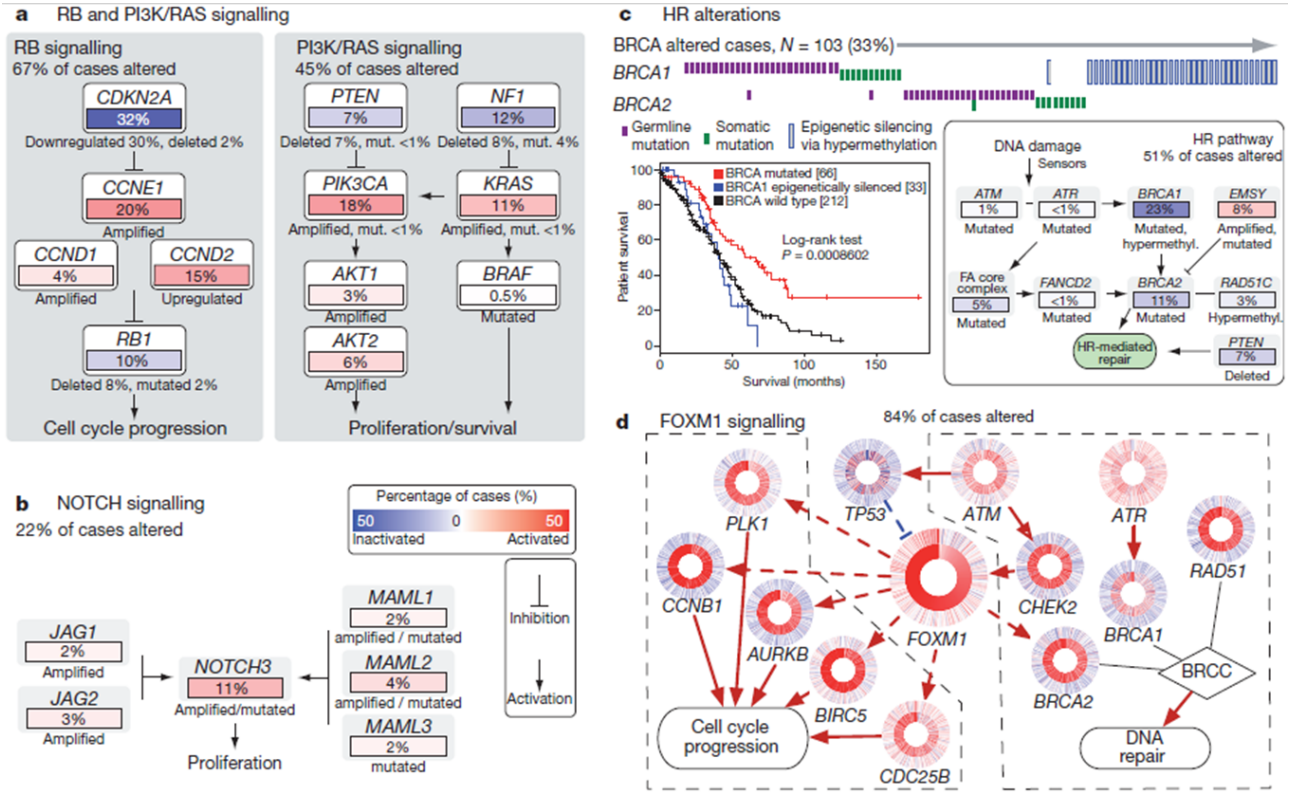


Figure 5: The picture shows the altered pathways in HGS-OvCa: a) RB and PI3K/RAS pathways, b) NOTCH pathway, c) HR gene alterations an d) FOXM1 signaling alterations| "The Cancer Genome Atlas Research Network. Integrated genomic analysis of ovarian carcinoma". Nature. 2011.

These and other findings suggest a model in which HGS-OvCa preferentially evolves as a consequence of initial disruption of DNA repair mechanism, followed by chromosomal instability, copy number change, and segregation into molecular subtypes [16].

2.3.1.4 GENETIC SUSCEPTIBILITY

HGS-OvCa present a hereditary predisposition in up to 15-20% of new cases due to germline mutations in BRCA1 (17q21.31) and BRCA2 (13q13.1) genes. Mutations correlated to BRCA1 gene provoke a 50% lifetime risk of ovarian cancer (20%–65%) in women with average age at diagnosis of 49–53 years. In contrast, the risk associated to BRCA2 alterations is lower than BRCA1 (11–37%, average age at diagnosis of 49–53 years) [17].

2.3.1.5 PROGNOSTIC AND PREDICTIVE FACTORS

The main HGS-OvCa prognostic factor is FIGO stage. Since the patient becomes symptomatic, advanced stage disease occurs in 75–80% of cases, and less than 25% of patients with stage III/IV will respond to the current therapies [18]. In advanced stage patients, the amount of residual tumor

after staging and debulking is the most important prognostic factor. In fact, patients where the disease is completely resected, have a better prognosis [19].

Nowadays, HGS-OvCa prognostic molecular markers are not present in routine clinical practice. The presence of tumor infiltrating lymphocytes, specifically CD3+ or CD8+ T-cells, are a marker of favorable prognosis in patients with HGS-OvCa [20].

Since 2016, the only molecular biomarker approved for pharmacological therapy response is the BRCA gene. In fact, patients with recurrent high-grade serous ovarian cancer and a germline or tumor BRCA mutation can be treated with olaparib (PARP inhibitor) after a response to platinum-based chemotherapy [21].

2.3.2 DIAGNOSIS

The HGS-OvCa diagnosis, which occurs normally when the patients have a mean age equal to 63 years, starts with the identification of abnormal mass in the gynecological examination. Distinctive symptoms are not present, they are typically correlated to gastrointestinal disease, including nausea, anorexia, early satiety, abdominal distension, bloating, pain, tenesmus and constipation as well as back pain and urinary frequency [22]. After, the clinician does pelvic and abdominal examinations, transvaginal ultrasound and blood analysis for CA-125, HE-4 associated to CEA and CA 19.9 markers, in order to exclude gastroenteric diseases. Sometimes a rectovaginal exam is used to help plan a surgery. The diagnosis must be confirmed with surgery to inspect the abdominal cavity, take biopsies for histological, cytological and molecular analysis, and look for cancer cells in the abdominal fluid. This helps to determine if an ovarian mass is benign or malignant [23].

The HGS-OvCa is typically identified late due to a lack of symptoms in initial disease phases, it has a poor-prognosis (5-year survival rate 40%) and it is classified FIGO III/IV (**Figure 4**) 75% of cases [9] [23].

2.3.3 TREATMENTS AND FOLLOW UP

The management of HGS-OvCa patients consists in 2 phases.

2.3.3.1 SURGERY

Surgery by laparoscopy technique is the first treatment choice because it is also part of the diagnostic procedure (biopsy).

In case of malignant tumor classified as FIGO I/II, the entity of the second surgery depends mainly on woman age and tumor size, in fact clinicians may decide to remove partially or fully the reproductive apparatus.

In presence of tumor classified as FIGO III/IV (over 90% of HGS-OvCa cases), the surgery intervention is focused on total removal of reproductive apparatus, including lymph nodes. The procedure includes also the surgery intervention after pharmacological therapy [23] [24].

2.3.3.2 PHARMACOLOGICAL THERAPY

As described before, the main early ovarian cancer treatment, in particular HGS-OvCa, is surgery debulking but the risk of recurrence is statistically 25-30%. For this reason, after an accurate diagnosis and FIGO classification, oncologists starts with a pharmacological therapy (**Figure 6**).

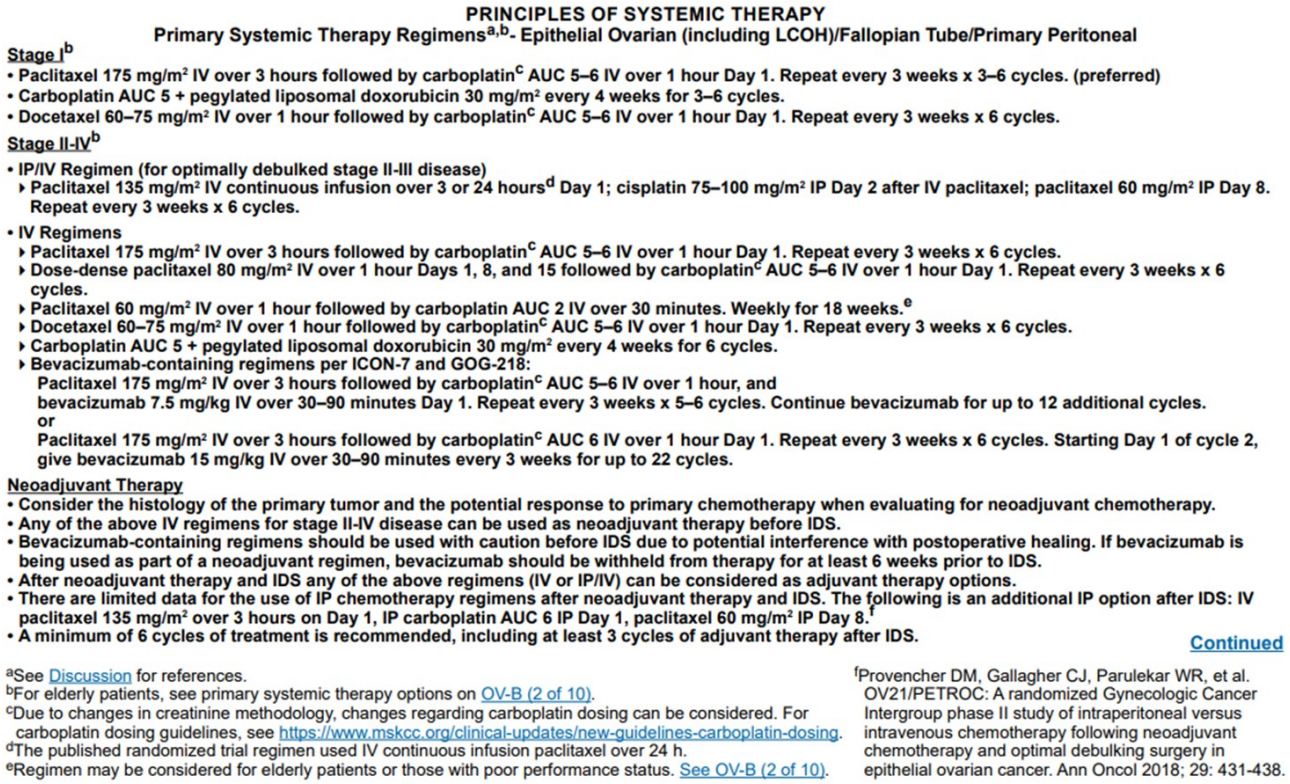
The first line FIGO I/II pharmacological therapy is constituted by carboplatin AUC 6 single agent for 6 cycles or by the carboplatin AUC 5/taxol 175 mg/mq combination for 6 cycles [23] [25].

In presence of tumor classified as FIGO IIIa (HGS-OvCa in almost all cases), the treatment is the combination of carboplatin AUC 5/taxol 175 mg/mq e.v. (3 hours of infusion every 21 days). In patients that are not subject to taxol therapy, taxol is changed with carboplatin/ liposomal doxorubicin [26].

Only in case of HGS-OvCa FIGO IIIb-c/IV, the first choice therapy is the combination of carboplatin AUC 5/taxol 175 mg/mq e.v. and bevacizumab (VEGF-A antibody) for 6 cycles, and a maintaining therapy with only bevacizumab for 15 months [27] [28].

Unfortunately, although HGS-OvCa is sensitive to platinum based drugs, it shows a strong drug resistance (even for platinum derivates) during the time because HGS-OvCa presents high levels of genomic instability, due to the mutations of genes included in DNA damage and homologous recombination pathways [29].

The current therapy for platinum resistant HGS-OvCa is based on doxorubicin, taxol and topotecan in association to bevacizumab, but the outcome is poor [23] [29].



PRINCIPLES OF SYSTEMIC THERAPY

Primary Systemic Therapy Regimens[a,b]- Epithelial Ovarian (including LCOH)/Fallopian Tube/Primary Peritoneal

Stage I[b]

- Paclitaxel 175 mg/m² IV over 3 hours followed by carboplatin[c] AUC 5–6 IV over 1 hour Day 1. Repeat every 3 weeks x 3–6 cycles. (preferred)
- Carboplatin AUC 5 + pegylated liposomal doxorubicin 30 mg/m² every 4 weeks for 3–6 cycles.
- Docetaxel 60–75 mg/m² IV over 1 hour followed by carboplatin[c] AUC 5–6 IV over 1 hour Day 1. Repeat every 3 weeks x 6 cycles.

Stage II-IV[b]

- IP/IV Regimen (for optimally debulked stage II-III disease)
  - Paclitaxel 135 mg/m² IV continuous infusion over 3 or 24 hours[d] Day 1; cisplatin 75–100 mg/m² IP Day 2 after IV paclitaxel; paclitaxel 60 mg/m² IP Day 8. Repeat every 3 weeks x 6 cycles.
- IV Regimens
  - Paclitaxel 175 mg/m² IV over 3 hours followed by carboplatin[c] AUC 5–6 IV over 1 hour Day 1. Repeat every 3 weeks x 6 cycles.
  - Dose-dense paclitaxel 80 mg/m² IV over 1 hour Days 1, 8, and 15 followed by carboplatin[c] AUC 5–6 IV over 1 hour Day 1. Repeat every 3 weeks x 6 cycles.
  - Paclitaxel 60 mg/m² IV over 1 hour followed by carboplatin AUC 2 IV over 30 minutes. Weekly for 18 weeks.[e]
  - Docetaxel 60–75 mg/m² IV over 1 hour followed by carboplatin[c] AUC 5–6 IV over 1 hour Day 1. Repeat every 3 weeks x 6 cycles.
  - Carboplatin AUC 5 + pegylated liposomal doxorubicin 30 mg/m² every 4 weeks for 6 cycles.
  - Bevacizumab-containing regimens per ICON-7 and GOG-218:
    Paclitaxel 175 mg/m² IV over 3 hours followed by carboplatin[c] AUC 5–6 IV over 1 hour, and
    bevacizumab 7.5 mg/kg IV over 30–90 minutes Day 1. Repeat every 3 weeks x 5–6 cycles. Continue bevacizumab for up to 12 additional cycles.
    or
    Paclitaxel 175 mg/m² IV over 3 hours followed by carboplatin[c] AUC 6 IV over 1 hour Day 1. Repeat every 3 weeks x 6 cycles. Starting Day 1 of cycle 2, give bevacizumab 15 mg/kg IV over 30–90 minutes every 3 weeks for up to 22 cycles.

Neoadjuvant Therapy

- Consider the histology of the primary tumor and the potential response to primary chemotherapy when evaluating for neoadjuvant chemotherapy.
- Any of the above IV regimens for stage II-IV disease can be used as neoadjuvant therapy before IDS.
- Bevacizumab-containing regimens should be used with caution before IDS due to potential interference with postoperative healing. If bevacizumab is being used as part of a neoadjuvant regimen, bevacizumab should be withheld from therapy for at least 6 weeks prior to IDS.
- After neoadjuvant therapy and IDS any of the above regimens (IV or IP/IV) can be considered as adjuvant therapy options.
- There are limited data for the use of IP chemotherapy regimens after neoadjuvant therapy and IDS. The following is an additional IP option after IDS: IV paclitaxel 135 mg/m² over 3 hours on Day 1, IP carboplatin AUC 6 IP Day 1, paclitaxel 60 mg/m² IP Day 8.[f]
- A minimum of 6 cycles of treatment is recommended, including at least 3 cycles of adjuvant therapy after IDS.

Continued

[a]See Discussion for references.
[b]For elderly patients, see primary systemic therapy options on OV-B (2 of 10).
[c]Due to changes in creatinine methodology, changes regarding carboplatin dosing can be considered. For carboplatin dosing guidelines, see https://www.mskcc.org/clinical-updates/new-guidelines-carboplatin-dosing.
[d]The published randomized trial regimen used IV continuous infusion paclitaxel over 24 h.
[e]Regimen may be considered for elderly patients or those with poor performance status. See OV-B (2 of 10).
[f]Provencher DM, Gallagher CJ, Parulekar WR, et al. OV21/PETROC: A randomized Gynecologic Cancer Intergroup phase II study of intraperitoneal versus intravenous chemotherapy following neoadjuvant chemotherapy and optimal debulking surgery in epithelial ovarian cancer. Ann Oncol 2018; 29: 431-438.

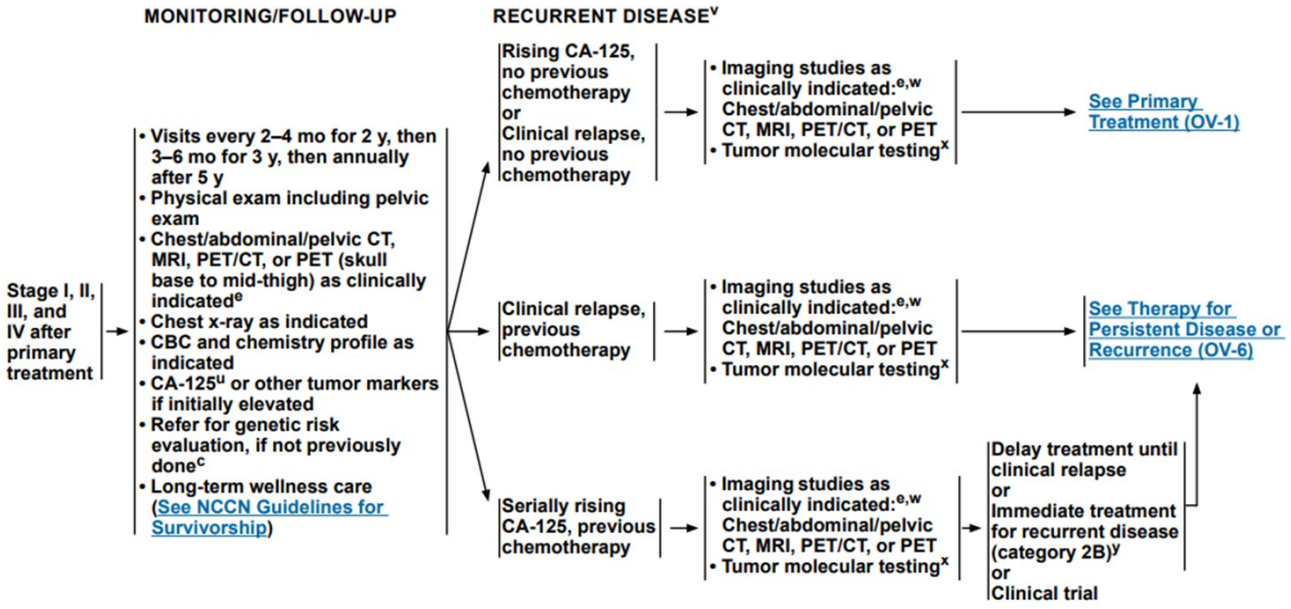
Figure 6: The picture shows the pharmacological treatments for HGS-OvCa. National Comprehensive Cancer Network (NCCN). Guidelines Epithelial Ovarian Cancer (including Fallopian Tube Cancer and Primary Peritoneal Cancer). Version 2.2018. https://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf.

Recently, both FDA and EMA have approved olaparib as monotherapy, at a recommended dose of 400 mg taken twice per day [30], but they differ in the application.

FDA approves the olaparib use in germline BRCA mutated advanced ovarian cancer that has received three or more prior lines of chemotherapy [30] [31]. Instead, EMA approved the use of olaparib "not later than 8 weeks after a course of platinum-based medicines, when the tumor was diminishing in size or had completely disappeared" [32] as maintenance therapy in platinum-sensitive relapsed HGS-OvCa [21][30].

2.3.3.3 FOLLOW UP

The monitoring and follow-up after surgery and pharmacological treatments (**Figure 7**) is done clinically (for example pelvic and abdominal examinations), biochemically (measurement of blood CA-125 levels, which have to be low) and with imaging (for example transvaginal ultrasound, X-ray, PET) [33].



[c]See NCCN Guidelines for Genetic/Familial High-Risk Assessment: Breast and Ovarian and NCCN Guidelines for Genetic/Familial High-Risk Assessment: Colorectal.
[e]Imaging performed with contrast unless contraindicated.
[u]There are data regarding the utility of CA-125 for monitoring of ovarian cancer after completion of primary therapy. See The Society of Gynecologic Oncology (SGO) position statement and Discussion.
[v]Consider symptom management and best supportive care. See NCCN Guidelines for Palliative Care. Refer for palliative care assessement, if appropriate.
[w]Surveillance imaging may be indicated when tumor markers are considered unreliable, the physical exam is unreliable, and/or there is a high risk of recurrence.
[x]Validated molecular testing should be performed in a CLIA-approved facility using the most recent available tumor tissue. Testing should include at least: *BRCA1/2*, homologous recombination pathway genes, and microsatellite instability or DNA mismatch repair.
[y]See Acceptable Recurrence Therapies (OV-B, 5 of 10).

Figure 7: The picture shows the guidelines for monitoring/follow-up of HGS-OvCa. National Comprehensive Cancer Network (NCCN). Guidelines Epithelial Ovarian Cancer (including Fallopian Tube Cancer and Primary Peritoneal Cancer). Version 2.2018. https://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf.

2.4 HERCULES PROJECT

Being HGS-OvCa constituted by genetically distinct cell subpopulations, some of them have features that render them resistant to the current treatments, such as platinum. Thus, after the treatments, resistant subpopulations are enriched making the relapsed tumor chemorefractory, which typically leads to death of the patient within a few months [25] [34].

Improved understanding of how HGS-OvCa cell subpopulations evolve through the disease course is an area of unmet clinical and scientific need and a prerequisite for successful personalized targeted therapies.

HERCULES Project (compre**HE**nsive cha**R**acterisation and effe**C**tive combinatorial targeting of high-grade sero**U**s ovarian cancer via sing**LE**-cell analysi**S**) is a European research project (HORIZON 2020) that integrates cutting-edge expertise on measurement technology, systems medicine modeling, and translational and clinical ovarian cancer expertise in Europe.

Different scientific centers are involved in the HERCULES project:

- University of Helsinki, Research Programs Unit (Coordinator) and Institute for Molecular Medicine Finland, FIMM (Finland);
- Turku University Hospital, Department of Obstetrics and Gynecology (Finland);
- University of Turku, Department of Pathology (Finland);

- University of Trieste, Department of Medical Sciences (Italy);
- Istituto Superiore di Sanità, Department of Haematology (Italy);
- Karolinska Institute, Department of Biosciences and Nutrition (Sweden);
- Institut Pasteur, Systems Biology Lab (France);
- AB ANALITICA (Italy).

The main objectives of this proposal are to comprehensively characterize HGS-OvCa by integrating and modeling clinical and biological data (e.g., genetics, transcriptomics, protein binding, drug screens) from primary, metastatic and relapsed tumors from various anatomical sites of HGS-OvCa patients, and establish combinatorial treatment modalities that effectively kill HGS-OvCa tumor cell subpopulations.

These will be achieved through a combination of the prospective and longitudinal sample collection from HGS -OvCa patients with disseminated disease, and the use of state-of-the-art fluorescence and mass cytometry, single-cell deep sequencing, novel computational and mathematical models, and a cutting-edge ex vivo drug screening approach.

By the end of the project, the novel predictive biomarkers and combinatorial drug combinations will be validated with additional HGS-OvCa sample cohorts, patient-derived xenograft models and proof-of-principle individualized trial with consented chemorefractory HGS-OvCa patients lacking treatment options and establish a CE-IVD marked predictive biomarker kit able to select the best drug combination for individual patients (personalized medicine approach).

The project is built on four closely connected subprojects that will be successfully completed by close collaboration between the participating groups (**Figure 8**):

1. Comprehensive characterization and modeling of the spatial and temporal landscape of tumor cell subpopulations during disease progression and therapy using prospective collection of multiple subsections of freshly acquired tumors.
2. Identification of genetic and transcriptomic biomarkers for tumor cell subpopulations, in particular drug resistant cell subpopulations, using mass cytometry and single-cell deep sequencing.
3. Development and application of integrative computational tools that will predict efficient combination therapy modalities among molecularly-targeted anticancer compounds that are able to kill the identified subpopulations ex vivo and in vivo.
4. Development and validation of a marketable prototype biomarker kit for predicting HGS-OvCa patients response to combinatorial therapeutic modalities based on formalin-fixed, paraffin-embedded (FFPE) sample from a tumor.

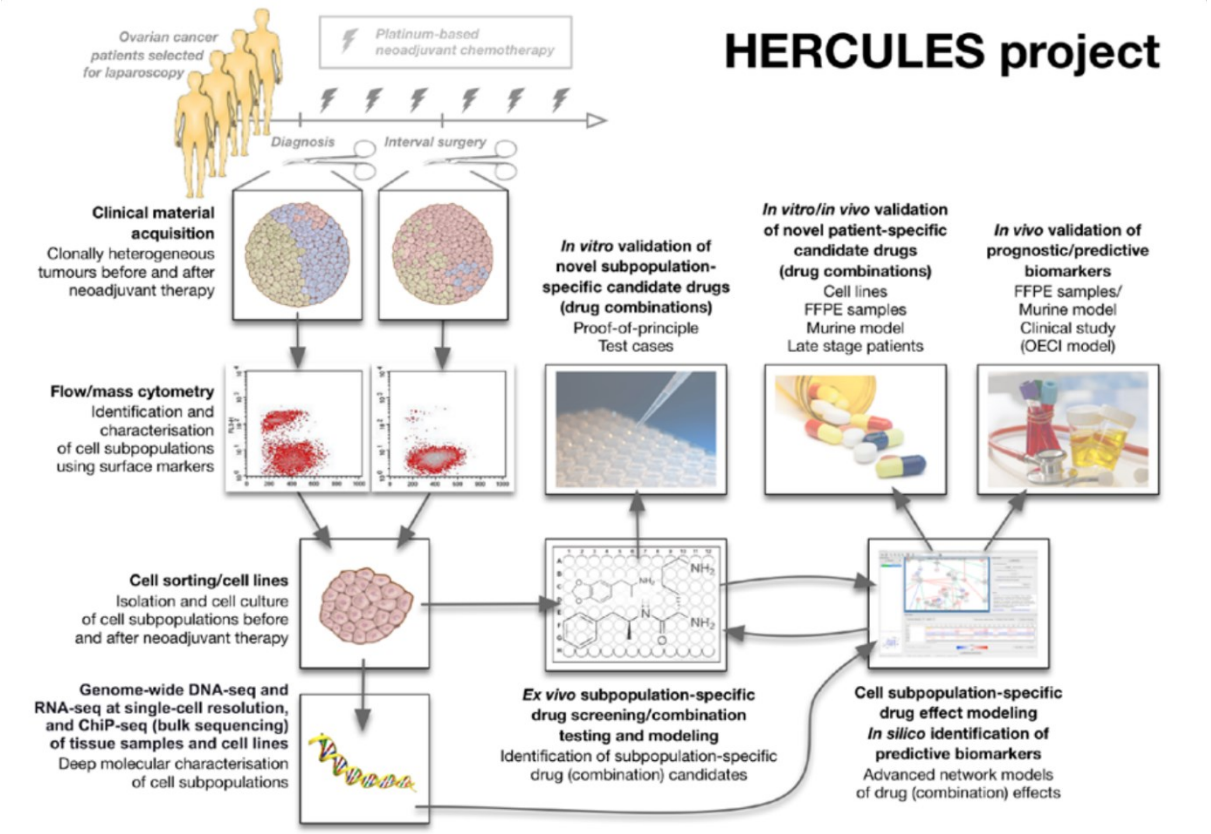


Figure 8: Overall concept of the HERCULES project. Prospective ovarian cancer patient samples before and after chemotherapy will be used to unravel cell subpopulations that drive tumor progression and drug resistance with mass cytometry. Single-cell sequencing will be used to characterize genomic and transcriptomics landscape of the subpopulations. An ex vivo drug screening approach combined with mathematical network-based modelling will discover the most efficient single and combination treatments. The most promising therapeutic regimens will be tested in preclinical models and in the clinic, and the models are iteratively improved based on the validation experiments. Network analysis and machine learning-based prediction algorithms will be used to identify a predictive biomarker panel, which will be validated using a large number of retrospective ovarian tumor samples. The predictive biomarker panel will be developed into a commercially available predictive test kit.

These subprojects will provide the most detailed understanding of HGS-OvCa tumor composition and characterization of drug resistant cell subpopulations to date, as well as computational tools to translate the unique single-cell deep sequencing data into biological knowledge and medical benefits. The clinically most actionable treatment suggestions from our modeling and computational analysis efforts will be translated to HGS-OvCa patient care. A major aim of the HERCULES project is to translate the herein produced scientific results into commercially available predictive test kit that facilitates choosing effective treatment regimens for ovarian cancer patients. Thus, HERCULES will strengthen the position of EU in systems medicine research and business by being a prime example of an EU-based consortium that both produces science at the highest level and exploits the results commercially.

Nowadays, targeted next generation sequencing (NGS) provides to scientists and clinicians an instrument to implement the testing of multiple genetic alterations and aberrations in diagnostic

pathology practice, which is necessary for personalized cancer treatment [35]. For this reason, AB ANALITICA, in collaboration with the DOTT.DINO PALADIN Company as a provider of contract research services, which funded this Ph.D. project, has chosen the NGS approach for studying the processes required for developing an in vitro diagnostic (IVD) workflow finalized to a next generation sequencing analysis with clinical application.

2.5 NEXT GENERATION SEQUENCING IN CLINICAL ONCOLOGY

2.5.1 NGS OVERVIEW

High-throughput sequencing technologies are improving quickly in recent years, moving from first generation Sanger sequencing to new high performance sequencing instruments.

Starting from the discovery of the biochemical structure of DNA, scientists have developed increasingly sophisticated techniques able to analyze the nucleic acid sequences of DNA molecules. In particular, the continuous improvement of the polymerase chain reaction, the availability of high-quality nucleic acid–modifying enzymes, and the development of fluorescent automated DNA sequencing enabled the Human Genome Project to provide the first complete human genome sequence draft [36].

Next generation sequencing, also known as massively parallel sequencing, is a method able to do simultaneously sequencing of millions DNA fragments or complementary DNA and to analyze several genes, gene regions or genomes with a single test compared to the traditional first generation method (Sanger DNA sequencing) [37]. DNA is the fundamental molecule used for NGS sequencing, RNA cannot be used for direct sequencing.

Thanks to the technological advances in NGS, significant improvements in the "omic" field (genomics, transcriptomics, epigenomics, and metagenomics) have been reached within different international projects having both research and clinical purposes [38].

The more recent NGS technologies are quite different from the Sanger method for the following reasons (**Figure 9**):

- NGS platforms sequence from millions to billions of nucleotides.
- The NGS sequencers are able to do massively parallel DNA sequence analysis.
- NGS systems have extremely high-throughput from multiple samples at the same time.
- Reduced costs for each test
- NGS instruments provide data that require complex bioinformatics tools for the elaboration and the interpretation [39].

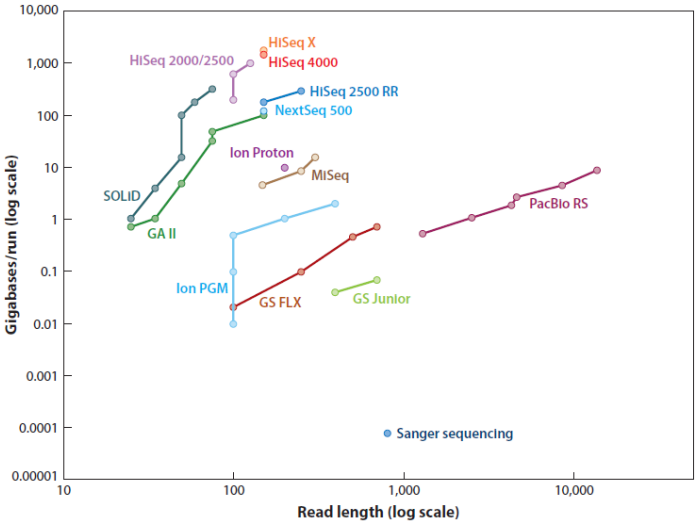


Figure 9: The graph shows the sequencing performances in terms of read length and depth among Sanger method and the main NGS platforms. Levy S.E., Myers R.M. "Advancements in Next-Generation Sequencing". Annu. Rev. Genom. Hum. Genet. 2016.

The NGS technologies employ different sequencing biochemistries and methods for the amplification of the input DNA molecules, which leads to different advantages and disadvantages in terms of read length, quality and throughput.

The fundamental workflow shared by all high-throughput sequencing platforms is [40]:

- Library preparation:
  - high molecular weight DNA molecules are mechanically or enzymatically fragmented into appropriate platform-specific size range.
  - End polishing step to generate blunt ended DNA fragments.
  - Platform specific adapters are ligated to the 3′ and 5′ fragment ends by either A/T overhang or direct blunt ligation.
- Library amplification: each library molecule is amplified in a way that the original molecules and all their copies stay clustered in the same position.
- Sequencing process: sequencing is then performed by alternating cycles of nucleotides addition and detection.

All high-throughput sequencing instruments require that the DNA molecules are processed into a library suitable for sequencing.

The NGS platforms can be classified using the following criteria:

- **Number of detectable molecules**: Pacific Biosciences and Oxford Nanopore platforms are able to detect just one DNA molecule per reaction well or sensor. Illumina, Ion Torrent, and Roche 454 and other platforms require clonally amplified DNA for the analysis.

- **Detection technology**: the NGS instruments can use optical or non-optical detection technologies to make sequencing base calls.

  Both Illumina and Pacific Biosciences platforms employ optical detection to measure the fluorescently modified nucleotides, instead Roche 454 instrument detects the light via pyrosequencing.

  IonTorrent and Oxford Nanopore platforms measure the release of H+ during a polymerization reaction via a solid-state sensor or the translocation of DNA through a nanopore sensor respectively (both non-optical detection strategies).
- **Sequencing chemistry**: the NGS platforms can use the polymerase or the ligation process to produce sequencing data.

  Illumina, Ion Torrent, Pacific Biosciences, and Roche 454 platforms for example employ sequencing-by-synthesis reactions in which the products of the reactions are measured to obtain the massively sequence information whereas the former Applied Biosystems SOLiD and the Polonator platforms use a ligation-mediated synthesis.

  Direct measurement of DNA molecules is performed by the Oxford Nanopore platform.

Each commercially available platform has similarities and differences relative to the others depending on the chemistries and detection methods used [36].

Today, Illumina technology is one of the most successful sequencing system available, with a claimed >70% dominance of the market, particularly with the HiSeq and MiSeq platforms [39]. The Company provides seven sequencing machines with performance ranges between 1.2Gb-120Gb for benchtop instruments to 120Gb-6,000 Gb for production-scale sequencers, the highest yield of error-free reads, and average quality score value (Q-score) higher than 30 [41]. The Q-score is a prediction of the probability of an error in base calling, the higher the value of Q, the lower the probability that the identified base is incorrect.

Among the Illumina instruments portfolio, the MiSeq sequencer is also available in the CE-IVD marking version. So the MiSeq coupled to kits having diagnostic-compatible features can be used in the clinical field.

Considering the reasons described above and because DOTT.DINO PALADIN Company has the MiSeq instrument in its laboratory, the experiments and the studies carried out in this thesis are based on the Illumina technology.

2.5.2 ILLUMINA TECHNOLOGY

The Illumina MiSeq instruments use the sequencing by synthesis (SBS) technology to identify the nucleotide sequence of DNA molecules.

The Illumina NGS workflow is based on four steps [41] [42]:

1. **Library Preparation**: The DNA library is prepared by random fragmentation of the DNA or cDNA sample, followed by 5′and 3′ adapter ligation. The adapters are complementary to the DNA sequences tied on flowcell internal surface. The library double strands DNA molecules are then PCR amplified, gel purified and denatured for obtaining single strands DNA molecules (**Figure 10**).

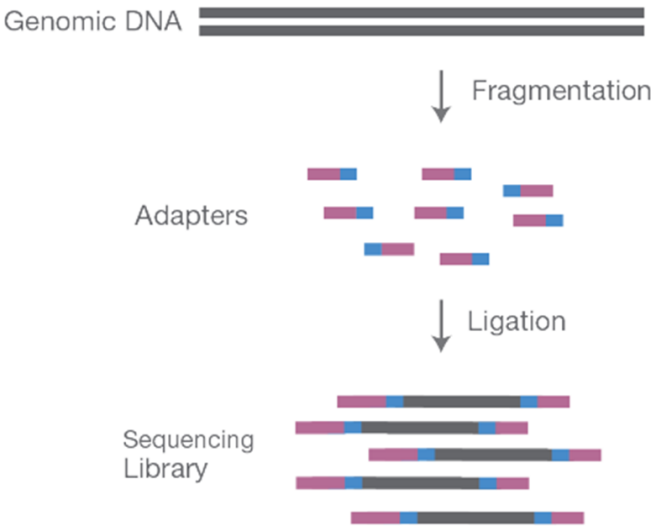


Figure 10: The library is obtained through the fragmentation of genomic DNA molecules and subsequent ligation with dedicated adapters to both fragment ends. "An Introduction to Next-Generation Sequencing Technology" 2017. https://www.illumina.com/Documents/products/Illumina_Sequencing_Introduction.pdf pp: 1-16.

2. **Cluster Generation**: The denatured library is loaded into a flowcell (it is a slide that contain 8 channels, the so-called lanes, in which the sequencing takes place) where the single strand DNA sequences hybridize with surface-bound oligos complementary to the library adapters (**Figure 11, 1**). Here the flowcell sequences are used as primer to create a complimentary copy of the initial template molecule (**Figure 11, 2**). Each fragment is then amplified into distinct, clonal clusters through bridge amplification (**Figure 11, 3-5**). The cluster generation ends when one strand of the dsDNA fragments is removed using the cleavable site in the surface oligo and the 3′ ends are bound with ddNTP in order to prevent the otherwise open 3′ ends to act as sequencing primer sites on neighboring library sequences (**Figure 11, 6**).

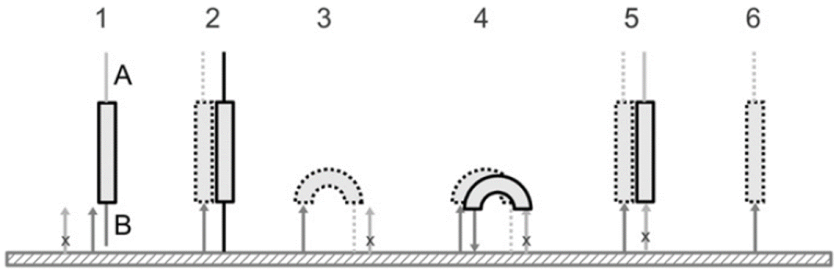


Figure 11: When the library is loaded into a flow cell, the DNA sequences hybridize to the flow cell surface oligonucleotides. Each bound fragment is amplified into a clonal cluster through bridge amplification. Buermans H.P.J., den Dunnen J.T. "Next generation sequencing technology: Advances and applications". Biochimica et Biophysica Acta. 2014. https://www.ncbi.nlm.nih.gov/pubmed/24995601 , pp: 1932–1941.

3. **Sequencing**: Illumina SBS technology employs a proprietary fluorescent reversible terminator–based method able to detect single bases when they are included into DNA template strands.

   During sequencing phase, each cluster is read one nucleotide at a time in repetitive cycles. In detail, during each sequencing cycle, all four reversible terminator–bound dNTPs species (A, C, T, G) are present, but only one is incorporated into the growing DNA fragment. The four dNTP contain fluorescent labels that allows the nucleotide identification and act as a reversible terminator to prevent multiple extension events. After each imaging analysis, the fluorescent group is removed, the reversible terminator is de-activated and the template sequences are ready for the next incorporation cycle. The sequence is read by following the fluorescent signal per extension step for each cluster (**Figure 12**).

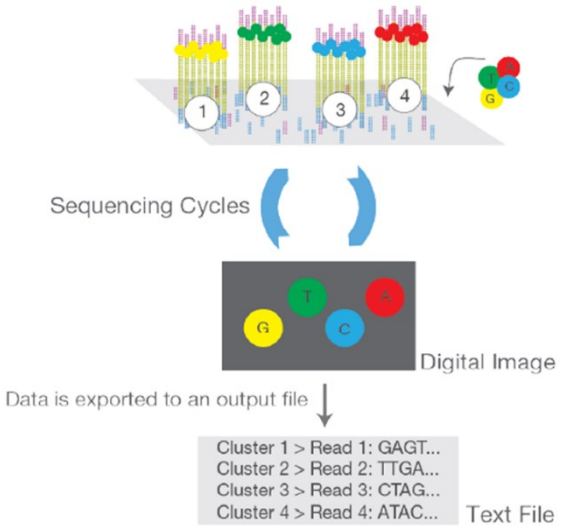


Figure 12: All sequencing reagents, including fluorescently labeled nucleotides, are added and removed at each elongation step. When a nucleotide is incorporated, the flowcell is imaged and the emission from each cluster is recoded. The emission wavelength and intensity are used to identify the included base. Each elongation step is repeated "n" times to create a read length of "n" bases. "An Introduction to Next-Generation Sequencing Technology" 2017. https://www.illumina.com/Documents/products/Illumina_Sequencing_Introduction.pdf pp: 1-16.

4. **Data Analysis**: During data analysis and alignment, the newly identified sequence reads are aligned to a reference genome. Following alignment, many variations of analysis are possible, such as single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, read counting for RNA methods, phylogenetic or metagenomic analysis, and more (**Figure 13**).

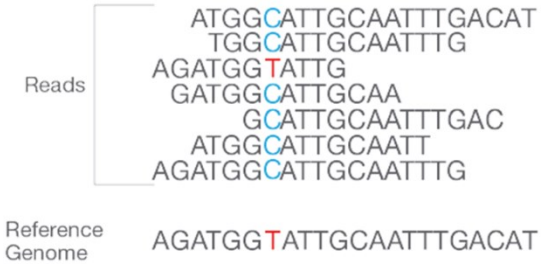


Figure 13: The reads obtained from sequencing are aligned to reference sequence using the bioinformatics software. After the aligment, it is possible to identify the differences between the reference genome and the each read sequenced. "An Introduction to Next-Generation Sequencing Technology" 2017. https://www.illumina.com/Documents/products/Illumina_Sequencing_Introduction.pdf pp: 1-16.

2.5.3 NGS ONCOLOGICAL APPLICATION

The diagnostic labs are employing standard molecular biology technologies (Real Time PCR, melt curve analysis, Sanger and pyrosequencing) to detect mutations, alterations and aberrations of distinct genes involved in different solid tumor samples [43].

Although these techniques are well known and reliable, in general, they need relevant cancer tissue DNA quantities for the analysis of each gene and this amount is in many cases limited due to the restricted availability of good quality tumor tissue, in particular from FFPE samples [44]. Furthermore, the turn-around time (TAT) from tumor resection to results is estimated at 1 up to 4 weeks, but can be prolonged due to subsequent analysis done using different hotspots for confirmation [45].

In recent years, the multiplex parallel analysis performed by next generation sequencing are flanking and, in some cases, replacing these standard assays [43].

Targeted NGS assays seem to be the best choice for diagnostic laboratories because they permit to clinicians the simultaneously detection of numerous specific genes alterations (from hotspot mutations to copy number variations, deletions, insertions) employing low amounts of DNA with high depth of coverages. Therefore, targeted NGS panel, in respect to whole genome or exome sequencing assays, allow to obtain run, analysis and interpretation much faster, easier and cheaper without compromising on its associated clinical value [43] [46].

2.5.3.1 TARGETED SEQUENCING

In cancer tissue (both fresh and FFPE samples) as well as in body fluids (in particular blood), malignant mutations prognostic and predictive of therapy indications may be present in a very low fraction of DNA molecules due to contamination or mutations collected in a small subset of tumor cells. The detection of these very low allele fraction mutations requires high deep sequencing coverage since mutations need to be detected from sufficient number of reads to pass a fixed variant calling threshold [47] [48].

Targeted sequencing allows reaching high coverage of the region of interest (ROI), particularly for cancer alterations expressed at low levels, while keeping the cost of sequencing and the complexity of data interpretation manageable.

2.5.3.2 TARGETED PANEL DEFINITION

For designing the NGS panel for diagnostic applications, it is fundamental to know the intended use, namely for diagnosis, prognosis or pharmacological treatment (precision medicine), and the genes involved in the cancer type desired.

The targeted NGS panels can be designed to detect single nucleotide variants (SNVs), small insertions and deletions (indels), copy number alterations (CNAs), and structural variants (SVs), or gene fusions [47].

Basing on the intended use, the panel is designed to cover hotspot regions of a single gene or to cover the entirety of the coding and noncoding sequences relevant to a given gene or SV. So, the targeted NGS panels can range from hotspot panels focused on individual codons to more comprehensive panels that include the coding regions of hundreds of genes [47].

Furthermore, the panel size may influence the sequencing depth, the sequencing reagent cost, the laboratory productivity and complexity of analytical and clinical interpretation [47] [49].

Guidelines recommend employing only genes having sufficient scientific evidence for the disease diagnosis, prognostication, or treatment [47] [50].

According to the previous observation, part of the Ph.D. activities involved the development of HGS-OvCa and colorectal cancer targeted panels.

2.5.3.3 TARGETED SEQUENCING WORKFLOW

The targeted NGS workflow includes four phases: sample preparation, library preparation, sequencing, and data analysis [50].

Each phase has to be analyzed carefully for developing a CE-IVD NGS workflow for oncology application.

2.5.3.3.1 SAMPLE PREPARATION

2.5.3.3.1.1 TISSUE CONSIDERATIONS

In a clinical diagnostics laboratory, fresh tissue is the preferred specimen for most molecular biology tests since the processes and procedures that may alter the sample DNA integrity are limited.

Nevertheless, the fresh or frozen-frozen tissues are rare due to the logistical complexities of collecting and storing samples. Consequently, almost all the molecular pathology laboratories use formalin-fixed, paraffin embedded (FFPE) tissue blocks, derived from the operating theater [51].

The FFPE specimens have numerous advantages, such as easier storage, but it is well known that formalin fixation results in DNA damage. In fact, formaldehyde reacts with DNA and proteins to form labile hydroxymethyl intermediates, DNA-DNA, DNA-RNA, and DNA-protein covalent bonds, oxidation and deamination reactions and formation of cyclic base derivatives [52] [53].

It is fundamental to standardize the FFPE sample procedures in order to obtain tissues with the same features (in particular regarding the potential DNA alterations, where the purpose is to reduce the nucleic acid damage) for the following DNA extraction phase and diagnostic analysis.

The FFPE tumor firstly has to be analyzed microscopically by an appropriately trained and certified pathologist before NGS testing. The microscopic review ensures [47]:

- Tumor cells fraction enrichment, which allows the NGS analysis and increase the sensitivity for gene alterations.
- Absence of necrotic tumor areas that can affect the analysis.
- Precise estimation of the tumor cell fraction, which is critical information when interpreting mutant allele frequencies and CNAs.

This process guarantees that FFPE slices used for DNA extraction have standard characteristics and that the DNA is derived mainly from tumor cells.

2.5.3.3.1.2 DNA EXTRACTION CONSIDERATIONS

The nucleic acid yields obtained from small samples, particularly FFPE samples, are a critical aspect for the NGS analysis due to nucleotides chemical modifications, which can significantly increase background noise in the final NGS reads [54].

For these reasons, the CE-IVD NGS workflow requires an in-depth study of the nucleic acid extraction process in order to standardize the conditions and minimize transfers and loss of material through multiple steps [47] [54].

To comply with the conditions described above, the experiments reported in this thesis have been carried out with FFPE samples provided by clinical partners or collaborators.

2.5.3.3.2 TARGETED LIBRARY PREPARATION

There are mainly two technologies for targeted NGS analysis of oncology specimens: hybrid capture-based and amplicon-based approaches (**Figure 14**).

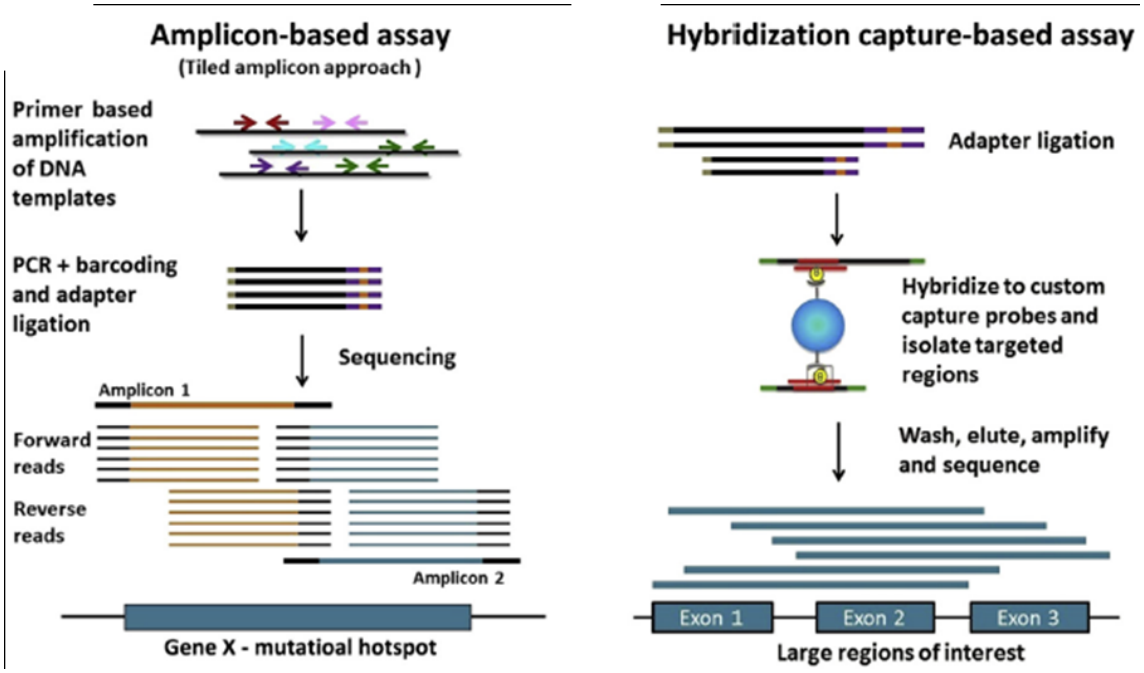


Figure 14: Target enrichment workflow for amplicon and capture hybridization NGS assays. Jennings L.J. et al. "Guidelines for Validation of Next-Generation SequencingeBased Oncology Panels". The Journal of Molecular Diagnostics, 2017. pp: 341-365.

2.5.3.3.2.1 HYBRID CAPTURE TECHNOLOGY

The method uses oligonucleotide probes, which are complementary to specific gene sequences, for isolating target DNA molecules. The probes are biotinylated oligonucleotide sequences that are designed to hybridize and capture the regions intended in the design.

The advantages of this library preparation method are:

- The probes are able to hybridize with target sequences although mismatches in the probe-binding site are present thanks to their long nucleotide sequences.
- The probes allow the isolation of target regions and their neighboring regions because the target sequences are included between the flanking sequences.
- The probes capture short fragments with high specificity.
- The longer fragments captured would be expected to map to the reference sequence with less ambiguity than shorter reads.

The disadvantages are:

- The overall coverage in the regions of interest are reduced due to the neighboring regions (in some cases, they can originate from far genomic areas with respect to the target sequence due to rearrangement events) isolation that occur with the target sequences capture.
- The shearing and/or other fragmentation methods affect significantly the library fragment sizes, and this will have a large influence on the outcome of the assays.
- The probes capture long fragments with lower specificity than short fragment [47].

2.5.3.3.2.2 AMPLICON TECHNOLOGY

The amplicon or amplification-based library preparation technique employs the multiplex PCR amplification reactions for enriching the target sequences.

The sequences of interest are tagged with sample-specific indexes (they identify the biological sample of origin) and sequencing adaptors used to hybridize the amplicons to complimentary oligonucleotides embedded in the platform's sequencer (in this thesis Illumina flow cell internal surfaces) before the sequencing process.

The advantages of this library preparation method are [47] [50]:

- This method allows only the amplification of the sequences of interest, without including off target regions.
- This technique is versatile and scalable.
- Low DNA input requirement.
- The protocol is simple and is typically shorter than for hybridization capture methods.

The disadvantages are:

- The amplicon technologies are affected from the typical PCR chemistry issues, mainly the errors accumulated during PCR amplification due to DNA polymerases error rate of 10^{-6} per base [55].
- The amplification-based library preparation is less likely to work effectively for genes with high guanine-cytosine content or regions with highly repetitive sequences [56].

With the purpose of reducing the PCR duplication and biased amplification issues in NGS analysis, researchers introduced molecular barcoding technology (random nucleotides sequences) for improving the detection accuracy of NGS quantification [57]. Here, each original DNA sequence, isolated during extraction phase, is tagged with a unique molecular index (UMI) and amplified with PCR reaction. After the sequencing run, the NGS data analysis software is able to identify the molecular barcodes contained in each read, allowing each read to be traced back to its original DNA sequence.

This is not to be confused with sample barcodes commonly used in current NGS workflows for identifying the samples in a libraries pool [47].

By measuring the number of UMIs instead of the number of total reads obtained, it is possible to quantify the template molecules, which reduces PCR resampling bias and improves quantification accuracy. Furthermore, UMIs allow the identification of sequencing errors by comparing across the reads containing the same barcode [59].

In accordance with the above, in the Ph.D. experiments the sample libraries were prepared using the amplicon and UMI technologies included in the Qiagen QIASeq kit.

2.5.3.3.3 SEQUENCING

In general, when deciding on a clinical sequencing platform, several aspects have to be considered. The most important are the expected testing volume; required test turnaround time; availability of bioinformatics support; provider's degree of technological innovation, platform flexibility, and scalability; and laboratory resources, technical expertise, and manufacturer's level of technical support. Choice of sequencing method will also depend highly on the number of genes required for the panel and the specific needs for the regions of interest [49].

Nowadays, the most commonly used NGS platforms in a clinical laboratory are the Illumina series and ThermoFisher's Ion Torrent series.

For the reasons described in the "NGS OVERVIEV" section, Illumina MySeq is used in this Ph.D. work.

2.5.3.3.4 DATA ANALISIS

The NGS instruments produce a large amounts of raw data that require multiple computationally bioinformatics algorithms for appropriate analysis and interpretation [60] (**Figure 15**).

The NGS bioinformatics pipeline processes extensive sequence data, obtained from the NGS run, and their metadata using various software components, databases, and operation environments (hardware and operating system) [61].

The operations executed by NGS bioinformatics software (usually NGS platform based) are shown in the following sections [61].

2.5.3.3.4.1 SEQUENCE GENERATION

This process allows the conversion of the signals recorded by the instrument (optical and non-optical) into the respective nucleotides that form the sample sequences. Each nucleotide is associated to a quality score, specific for each NGS platform. The nucleotide sequences analyzed and the

corresponding quality scores are noted into the FASTQ file, which is a de facto standard for representing biological sequence information [61] [62].

2.5.3.3.4.2 SEQUENCE ALIGNMENT

This process allows to identify the position of the analyzed sequences (each generally <250 bp) through alignment to the reference genome. The algorithms provide the mapping quality score and the genomic context (location in the reference genome) of each sequence reads. These info can be used to calculate the proportion of mapped reads and depth (coverage) of sequencing for the region of interest.

The data obtained during this process are noted a de facto standard binary alignment map (BAM) file format, which is a binary version of the sequence alignment/map format [61].

2.5.3.3.4.3 VARIANT CALLING

This process allows the accurate identification of variations that occur between the sample and the reference genome sequences, which are aligned in the BAM format file. The variant calling analysis employs various computational algorithms for identifying the sequence variants, such as single nucleotide variants (SNVs), small insertions and deletions (indels), copy number alterations, and large structural alterations (insertions, inversions, and translocations).

Being the variant calling accuracy strictly correlated to the quality of called bases and aligned reads, local realignment around the expected indels and base quality score recalibration or other more specific operations are employed for guaranteeing an efficient process.

The variant call format (VCF) is the de facto standard file used for storing the info obtained in this process [61].

2.5.3.3.4.4 VARIANT FILTERING

This process, which normally occurs after the variant calling phase - but in some bioinformatics tools it can be incorporated in this step -, has the purpose of filtering the original VCF file using different sequence alignment and variant calling associated metadata (eg, mapping quality, base-calling quality, strand bias, and others) for detecting the sequencing artifacts [61].

2.5.3.3.4.5 VARIANT ANNOTATION

This process employs computational algorithms to query different variant and sequence databases to describe the called variants obtained during the NGS analysis with a set of metadata such as variant

location, predicted cDNA and amino acid sequence change (HGVS nomenclature), minor allele frequencies in human populations, and prevalence in different variant databases.

The results collected with this operation are used to prioritize or filter variants for classification and interpretation [61].

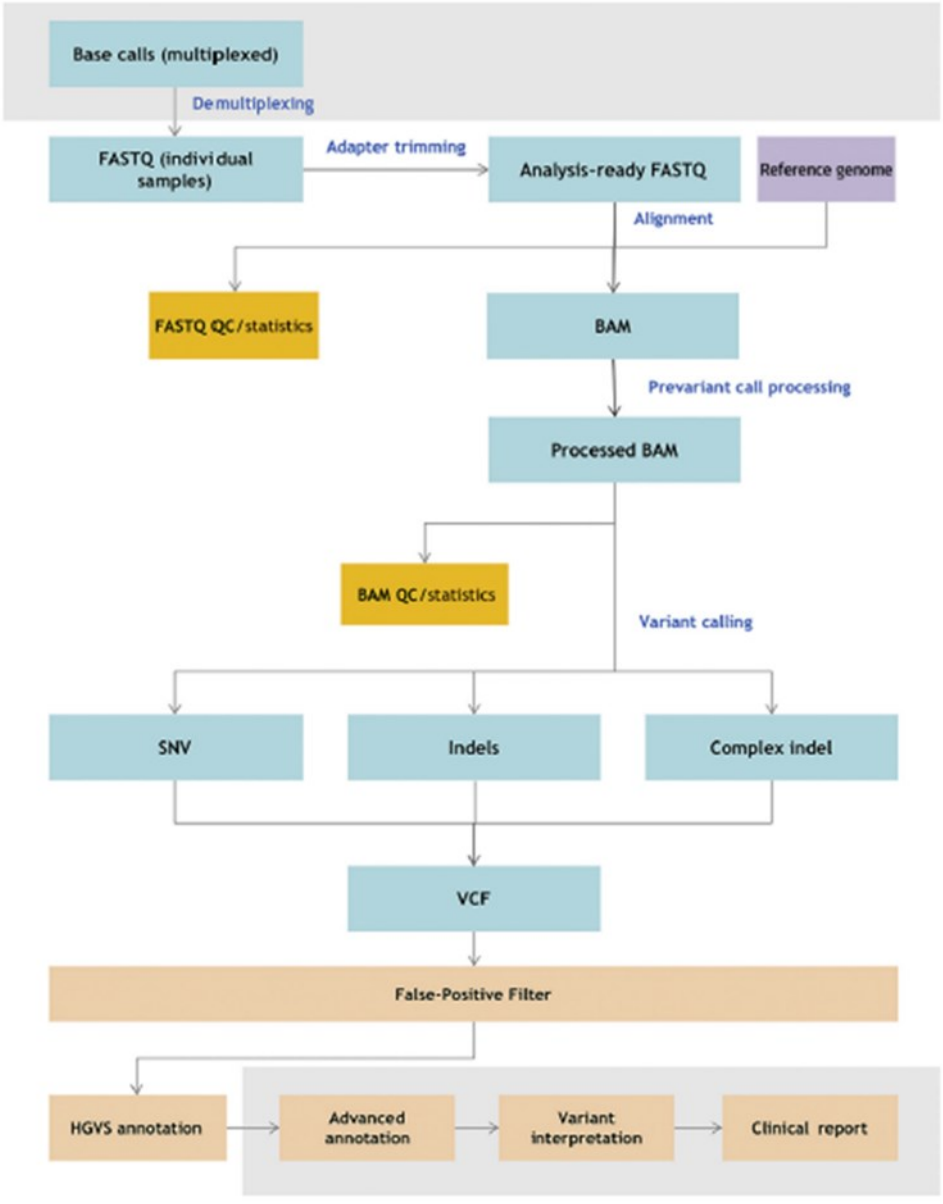


Figure 15: The processes involved in the NGS bioinformatics pipeline. BAM, binary alignment map; HGVS, Human Genome Variation Society; indel, insertion/deletion; QC, quality control; SNV, single nucleotide variant; VCF, variant call format. Roy S. et al. "Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines". J Mol Diagn. 2018. pp: 4-27.

The bioinformatics tools are an indispensable part of the NGS analysis; in fact the detection of genomic alterations has significant impact on disease management and patient care.

The members of the global molecular genetics and pathology community are involved in establishing guidelines for the validation of bioinformatics pipelines.

This work requires extreme care because if the bioinformatics tools for clinical use are improperly developed, validated, and/or monitored, they may generate inaccurate results that may have negative consequences for patient care.

The next generation sequencing technology is evolving not only in research but also in different clinical fields.
In oncology, the targeted NGS approach provides an enormous advantage for the simultaneous detection of multiple alterations using a single platform and analysis. In fact, this technology allows obtaining information useful for prediction of response to targeted therapies, disease diagnosis, and patient prognostication.
Nowadays, clinical laboratories are choosing the targeted NGS technology as test for the detection of these genetic alterations, but they require a validated and certified workflow to ensure the high quality of sequencing results for the subsequent clinical application.

The thesis reports the studies conducted for planning and developing an IVD NGS workflow with oncological application.
The studies involved the patent research analysis and the design of a gene panel for HGS-OvCa.
The subsequent studies and experiments are involved in the NGS workflow steps ranging from the selection of FFPE nucleic acid extraction kits to the NGS run, including the quality control tests necessary to understand if the DNA sample characteristics are compatible with the sequencing analysis.

3 AIM OF THE WORK

This Ph.D. research work is an integral part of a Horizon 2020 project, called HERCULES project (Compre**HE**nsive cha**R**acterisation and effe**C**tive combinatorial targeting of high-grade sero**U**s ovarian cancer via sing**LE**-cell analysi**S**).

The objectives of the Horizon 2020 project are to comprehensively characterize high grade serous ovarian cancer (HGS-OvCa) by integrating and modeling clinical and biological data (e.g., genetics, transcriptomics, protein binding, drug screens) from primary, metastatic and relapsed tumors from various anatomical sites of HGS-OvCa patients, and establish combinatorial treatment modalities that effectively kill HGS-OvCa tumor cell subpopulations. This will be achieved through a combination of the prospective and longitudinal sample collection from HGS -OvCa patients with disseminated disease, and the use of state-of-the-art fluorescence and mass cytometry, single-cell deep sequencing, novel computational and mathematical models, and a cutting-edge ex vivo drug screening approach. By the end of the project, the novel predictive biomarkers and combinatorial drug combinations will be validated with additional HGS-OvCa sample cohorts, patient-derived xenograft models and proof-of-principle individualized trial with consented chemorefractory HGS- OvCa patients lacking treatment options.

The company role in HERCULES project is to develop and validate a marketable prototype biomarker kit for predicting HGS-OvCa patients response to combinatorial therapeutic modalities.

In particular, the kit hypothesized for this application is a NGS genes panel (Illumina platform and technology based) able to predict the outcome of a pharmacological therapy using high grade serous ovarian cancer subpopulation genetic biomarkers.

AB ANALITICA, has chosen to develop a kit based on NGS technology for the following reasons:

- HGS-OvCa is such a heterogeneous tumor from genetic and genomic points of view that it requires a technique capable of analyzing a wide genomic range in a single analytical session.
- NGS is a technology that is entering and acquiring increasing importance in the molecular diagnostic field.
- NGS is a technology that AB ANALITICA is interested in developing and mastering, alongside those currently in use (End Point PCR, Reverse Line Blot and Real Time PCR). This allows it to propose a new generation of technologically advanced products to clinical diagnostics.

Taking the above into consideration , the aim of the Ph.D. research project is to study the processes required for developing an *in vitro* diagnostic (IVD) workflow finalized to a next generation

sequencing analysis for clinical application (not necessarily focused on high grade serous ovarian cancer) having diagnostic, prognostic and predictive purpose.

The research activities are performed DOTT.DINO PALADIN Company, which funded this Ph.D. work.

In order to achieve this goal, the following activities were carried out:

1. Evaluation of the "state of art" regarding the presence of patents relevant for the HERCULES project.
2. Definition of a gene panel design based on several genes involved in HGS-OvCa.
3. Identification and selection of FFPE DNA and RNA extraction kits having features, in terms of nucleic acid yields, quality and purity, that are compatible with NGS application and IVD workflow.
4. Studying the processes necessary to sequence human FFPE samples, with KRAS wild type and with specific mutation, on the Illumina platform using amplicon technology on well-defined target genes panel. As the study of NGS workflow the object of this research activity, which can also be used for applications other than HGS-OvCa, samples with mutated KRAS were used because they were samples available to achieve this purpose. Once we have mastered the NGS method, we can use it to HGS-OvCa.

The results obtained with this research work will be used by the company R&D division for developing a clinical custom genes panel, intended to be used by pathologists and oncologists, able to provide:

- Prediction cancer onset.
- Characterization of different tumor types.
- Precise information about the exact therapeutic treatment for different cancer types (HGS-OvCa included).
- Outcome prognosis.

4 MATERIALS AND METHODS

4.1 PATENT SEARCH ANALISYS

The purpose of this search activity was to take stock of the situation about the "state of art" regarding the presence of patents or patent applications relevant for the commercial exploitation of the results expected from the HERCULES project.

A patent or patent application is considered relevant if it deals with or includes

- Biomarkers related to pharmacological treatments of high-grade serous ovarian cancer, and more generally ovarian cancer or ubiquitous in cancer.
- Methods and protocols for the detection of biomarkers as defined above.
- Reagent kit formulations used in the detection of biomarkers as defined above.

The patent and patent application search was made by using on-line open tools:

- EPO (European Patent Office) ESPACENET PATENT SEARCH in Smart Search mode on worldwide domain.
- GOOGLE PATENTS.

The searches were performed using keywords and respective online search tools:

- ovarian cancer therapy (GOOGLE PATENTS).
- cancer combinatorial therapy (GOOGLE PATENTS).
- ovarian cancer combination (ESPACENET, GOOGLE PATENTS).
- cancer pharmacogenomic kit (ESPACENET, GOOGLE PATENTS).
- ovarian population assay (GOOGLE PATENTS).
- PCR cancer therapy (ESPACENET, GOOGLE PATENTS).
- Real Time PCR cancer treatment (GOOGLE PATENTS, ESPACENET).
- Real Time PCR ovar* pharmac*(GOOGLE PATENTS, ESPACENET).
- cancer assay diagn* therap* (GOOGLE PATENTS, ESPACENET).
- ovar* cancer assay diagn* therap*(GOOGLE PATENTS, ESPACENET).
- ovarian population assay trea*(GOOGLE PATENTS).
- ovarian biomarker* pcr popul* (ESPACENET,GOOGLE PATENTS).
- ovarian biomarker* pcr subpopul*(ESPACENET, GOOGLE PATENS).
- predictive biomarker* ovar*(GOOGLE PATENTS).
- ovari* popul* marke* kit (ESPACENET, GOOGLE PATENS).
- ovari* marke* drug (GOOGLE PATENTS).
- ovari* popul* biomar*(GOOGLE PATENTS).

- "high-grade serous ovarian cancer" drug screening (GOOGLE PATENTS).

Where the * symbol (wildcard) stands for any combination of characters following the characters preceding the symbol, according to the syntax accepted by the search engine.

Patent application status was checked using the INPADOC legal status, EP register and Global register services available from the ESPACENET search tool.

4.2 BIOLOGICAL SAMPLES COLLECTION

The Ph.D. research experiments employed two distinct types of biological samples for the following distinct purposes:

- Identification of the nucleic acid extraction kits having features compatible with a NGS-based diagnostic workflow.
- Identification and characterization of tumor mutations using Illumina NGS technology.

4.2.1 BIOLOGICAL SAMPLES

The biological samples were all formalin-fixed paraffin-embedded (FFPE) tissues. The biological samples used for the Ph.D. experiments are:

- Tissue obtained from the non-tumoral human material of a colon resection surgery provided by a partner of the HERCULES project. The research work has required ninety FFPE sections derived from a FFPE samples block.
- Series of FFPE samples obtained by residual samples from a previous company development project on KRAS mutation detection by real time PCR. Each tube contains five 10 μm thick sections of that were processed together.

4.2.2 FFPE SAMPLES PREPARATION

The FFPE samples, derived from non-tumoral tissue, were used in order to study and compare the results obtained by different extraction kits in standard conditions, employing biological samples with the same characteristics regarding tissue matrix, nucleic acid quantity and FFPE preparation protocol. The protocol adopted for FFPE samples preparation is described below:

- Non-tumoral tissue was isolated from a colon colorectal cancer resection.
- Human tissue fixation within 30 minutes from resection in following conditions: formaldehyde 10%, pH 7.0, 18h, 1:20 sample to formalin volume ratio.
- Standard dehydration cycles, infiltration and embedding, using Tissue Processor Diapath Donatello 52PRO01.

- The FFPE block was stored at room temperature.
- The spatial dimensions of each FFPE slice: 5 μm (depth), approximately 2 cm (height) and 1.5 cm (width).
- The samples were cut two and twelve weeks (FR1 series and FR2 series, respectively) after block preparation.
- The total ninety FFPE sections were stored in refrigerator.

On the other hand, the FFPE samples derived from colorectal tumor tissue were used in order to test the extraction protocols with a number of sections and tissue amount which was estimated to be markedly higher than in the first series. The aims are:

- Simulating clinical conditions in the workflow of an anatomo-pathological laboratory on actual tumor samples.
- Verifying the limits of selected commercial extraction kits in presence of high paraffin and tissue quantities.
- Comparing the results with analytical techniques for variant detection which are already established in clinical routine

They are samples of clinical origin prepared according to the procedures in use in the clinical laboratory from which the samples come.

4.3 SAMPLES NOMENCLATURE

With the purpose of identifying the samples during each experiment session, a coding convention has been adopted, where the sample series, sample ID in the series, type of nucleic acid extracted and extraction protocol used are all specified, as follows:

$$S[n_1, n_2, \dots]\_\mathrm{T[P]}$$

Where:

- S is the sample series, equal to FR1 and FR2 for the samples from colon resection surgery, M for the samples from the previous KRAS mutation status project (FR and M are abbreviations for sample source)
- n_1, n_2,… is the sequence of progressive sample identifiers in the series (e.g. FR1[3,4] denotes the 3rd and 4th samples in the FR1 series)
- T denotes the nucleic acid type, and may be D or R where D stands for DNA, R stands for RNA.
- P denotes the extraction protocol as shown in **Table 1**.

Protocol	Abbreviation
Qiagen GeneRead DNA FFPE Kit	D[Q-GR]
Qiagen AllPrep DNA/RNA FFPE Kit	D[Q-AP]
Magbio Genomics HighPrep FFPE Tissue DNA Kit	D[MB-HP]
Promega ReliaPrep FFPE gDNA Miniprep System	D[P-RP]
Promega Maxwell RSC DNA FFPE Kit	D[P-MR]
OMEGA bio-tek E.Z.N.A. FFPE DNA Kit	D[OB-E]
OMEGA bio-tek Mag-Bind FFPE DNA Kit	D[OB-MB]
Macherey-Nagel NucleoSpin DNA FFPE XS	D[MN-NSX]
BioChain FFPE Tissue DNA Extraction Kit – Column	D[BC-FC]
Analytik Jena blackPREP FFPE DNA Kit	D[AJ-BP]
Qiagen RNeasy FFPE Kit	R[Q-RE]
OMEGA bio-tek E.Z.N.A. FFPE RNA Kit	R[OB-E]
Macherry-Nagel NucleoSpin totalRNA FFPE XS	R[MN-NSX]
OMEGA bio-tek Mag-Bind FFPE RNA Kit	R[OB-MB]

Table 1: Abbreviation used to identify extraction protocol in sample IDs.

4.4 NUCLEIC ACID EXTRACTION KITS

In the diagnostic and therapeutic fields, standardized pre-analytical methods are essential in order to guarantee reproducibility and robustness of each results obtained.

In this case, genomic DNA and RNA obtained from FFPE tumor tissues are the starting points for NGS library preparation and NGS sequencing.

The isolation of high quantities of pure, double stranded, highly concentrated, not contaminated genomic DNA and RNA molecules is prerequisite for successful and reliable large scale genotyping and expression analysis. For these reasons, it is necessary to find a FFPE nucleic acid extraction kit having characteristics compatible with these requirements.

4.4.1 COMMERCIAL NUCLEIC ACID EXTRACTION KIT TECHNOLOGY

The FFPE nucleic acid extraction techniques considered are shown in the following sections.

4.4.1.1 SPIN COLUMN-BASED NUCLEIC ACID PURIFICATION TECHNOLOGY

This method allows purifying the nucleic acids more quickly than conventional in-house methods (for example, guanidinium thiocyanate – phenol - chloroform extraction).

The common steps involved in all commercial column based extraction kits used in the thesis work are:

1. Deparaffinization: the paraffin removal from biological sample occurs with mechanisms that differ according to the kits used (by heating, dissolution, etc). In general, during this phase biological sample is isolated and purified from paraffin.
2. Column conditioning: a buffer at a particular composition converts the surface or functional groups on the column into a particular chemical form useful for the following purification step.
3. Tissue lysis: chemical and enzymatic treatments allow the nucleic acid release from biological membranes, proteins and other biomolecules.
4. Nucleic acids purification: contaminants like proteins, lipids are removed from nucleic acids that are bound to the column during several washing step by using washing buffer.
5. Elution: a proprietary buffer is used to release the nucleic acid from the column, so that it can be collected in a purified state.

4.4.1.2 MAGNETIC BEAD BASED NUCLEIC ACID PURIFICATION TECHNOLOGY

This separation process allows an efficient nucleic acid extraction from FFPE tissues. Magnetic bead based extraction kit can be used in both manual and high throughput application if coupled with liquid handler automation instruments.

The common steps involved in commercial bead based extraction kits are similar to column based kits:

1. Deparaffinization: the paraffin removal from biological sample occurs with mechanisms that differ according to the kits used (by heating, dissolution, etc). In general, during this phase biological sample is isolated and purified from paraffin.
2. Tissue lysis: mechanical, chemical or enzymatic treatments allow nucleic acid release from biological membranes and other biomolecules.
3. Nucleic acids purification: contaminants like proteins, lipids are removed from nucleic acids that are bound to the beads during several washing step by using washing buffer, while the beads are held in place by using a magnet.
4. Elution: proprietary buffer is used to release the nucleic acid, so that it can be collected in a purified state, while the beads are held in place by using a magnet.

4.4.2 NUCLEIC ACID KITS LIST

The genomic DNA and RNA extraction kits analyzed in this Ph.D. work are reported in **Table 2**.

Product Name	Manifacturer
blackPREP FFPE DNA Kit	**Analytik-jena**
innuPREP FFPE DNA kit IPC 16	**Analytik-jena**
AxyPrep MAG FFPE (DNA-RNA-miRNA) Kit	**Axygen**
Formapure DNA Extraction Kit	**Beckman Coulter**
FFPE Tissue DNA Extraction Kit	**BioChain**
FFPE Tissue DNA Extraction Kit – Columns	**BioChain**
FFPE Tissue DNA Extraction Kit - Magnetic Beads	**BioChain**
Clean FFPE DNA	**CleanNa**
truXTRAC FFPE DNA kit	**Covaris**
FFPE DNA Extraction kit (Bioruptor Plus & Standard)	**Diagenode**
NucleoMag DNA FFPE	**Macherey-Nagel**
NucleoSpin 96 DNA FFPE	**Macherey-Nagel**
NucleoSpin DNA FFPE XS	**Macherey-Nagel**
HighPrep FFPE Tissue DNA Kit	**MagBio Genomics**
FFPE RNA/DNA Purification Plus Kit	**Norgen Biotek**
FFPE DNA Purification Kit	**Norgen Biotek**
E.Z.N.A. FFPE DNA Kit	**OMEGA bio-tek**
Mag-Bind FFPE DNA	**OMEGA bio-tek**
Prepito FFPE Kit	**Perkin Elmer**
Chemagic FFPE DNA Kit special	**Perkin Elmer**
Promega Maxwell RSC DNA FFPE Kit	**Promega**

ReliaPrep FFPE gDNA Miniprep System	**Promega**
QIAamp DNA FFPE Tissue Kit	**Qiagen**
GeneRead DNA FFPE Kit	**Qiagen**
EZ1 DNA Tissue kit	**Qiagen**
AllPrep DNA/RNA FFPE Kit	**Qiagen**
AllPrep DNA/RNA/miRNA	**Qiagen**
MagCore Genomic DNA FFPE One-step kit	**RBC Bioscience**
MagCore Genomic DNA Tissue kit	**RBC Bioscience**
Invisorb Genomic DNA Kit II	**Stratec Molecular**
InviMag FFPE DNA Kit/IG	**Stratec Molecular**
GeneJET FFPE DNA Purification Kit	**Thermo Scientific**
MagMAX FFPE DNA Isolation Kit	**Thermo Scientific**
Ion AmpliSeq Direct FFPE DNA Kit	**Thermo Scientific**
RecoverAll Total Nucleic Acid Isolation Kit for FFPE	**Thermo Scientific**
RecoverAll Total Nucleic Acid Isolation Kit	**Thermo Scientific**
299008 - DNA Extraction Kit, FFPE Tissue, BioAssay	**United State Biological**
D3878-01F - DNA Extraction Kit, FFPE Tissue, BioAssay	**United State Biological**
MagPurix FFPE DNA Extraction Kit	**ZINEXTS**

Table 2: DNA extraction kits and respective company producers.

The RNA extraction kits analyzed in this work are reported in **Table 3**.

Product Name	Manufacturer
NucleoSpin totalRNA FFPE XS	Macherey-Nagel
E.Z.N.A. FFPE RNA Kit	OMEGA bio-tek
Mag-Bind FFPE RNA kit	OMEGA bio-tek
RNeasy FFPE Kit	Qiagen

Table 3: RNA extraction kits and respective company producers.

4.5 NUCLEIC ACID EXTRACTION KITS SCREENING CRITERIA

The purpose of this screening was the pre-selection of suitable DNA extraction protocols to be tested for nucleic acid purification from both types of samples described above.

The parameters and features considered to evaluate and find potential candidates for genomic DNA extraction kit are:

- NGS technology, for the purpose of discarding kits which are not compatible with Illumina sequencing technology;
- Instrument restrictions: it is used for excluding kits which require dedicated instrumentation and/or binding technical support;
- Extraction technology (spin column-based or magnetic bead based);
- Deparaffinization method which may be heating, proprietary deparaffinization solution, mineral oil, sonication;
- Operator safety risk for the purpose of excluding kits having toxic reagents for health professionals;
- Cross-linking correction for the purpose of excluding protocols that do not include steps for repairing cross-linking alterations induced by formalin;
- Citosine deamination repair: this feature determines if a kit is able to convert uracil in cytosine;
- Availability of a manual protocol for familiarization before porting to automation (for magnetic beads based protocols only);
- Type of automatic system on which the protocol is intended to run: general purpose robotic platforms or dedicated (typically benchtop) instrument (for magnetic beads based protocols only);

- NGS compliance: this parameter refers to investigate if the kit can be used for downstream NGS analysis;
- Deparaffinization without centrifugation: this parameter is important for easier and less expensive fully walk-away automation (for magnetic beads based protocols only);
- CE IVD marking: this parameter is important for clinical and pharmacological applications;
- FFPE input specifications: number, thickness, surface area of FFPE sections, or maximum amount of FFPE input and embedded tissue input;
- Elution volume;
- Quality of supplier collaboration and acceptance of OEM contractual terms to be applied to critical suppliers in the supply chain for production of CE IVD marked products.

4.5.1 NUCLEIC ACID EXTRACTION PROTOCOLS

Extraction reactions of the kits obtained after the previous pre-selection study were carried out according to the manufacturers protocol. Different conditions in term of number of input FFPE tissue sections and final elution volume were tested.

The aim of these experiments was to find nucleic acids extraction kits with adequate performances for inclusion as a component of a CE IVD marked workflow for NGS-based diagnostic screening for tumor-associated mutations.

4.6 NUCLEIC ACID QUANTIFICATION

The quantification of nucleic acid yields, obtained using the selected extraction kits with the two samples typologies, was performed by:

- the absorbance method using the ThermoFisher Scientific NanoDrop 2000 spectrophotometer.
- the fluorescence method using the ThermoFisher Scientific Qubit 3 Fluorimeter with the Qubit dsDNA HS Assay Kit and Qubit RNA HS Assay Kit.

4.6.1 ABSORBANCE: THERMOFISHER SCIENTIFIC NANODROP 2000

The ThermoFisher Scientific NanoDrop 2000 Spectrophotometer is a benchtop instrument that can be used for nucleic acid and protein quantification using 0.5-2 µl sample volumes.

The procedure used for measuring the absorbance and the concentration of nucleic acids requires that the instrument is firstly set for nucleic acid molecules analysis, then the blank (the buffer where the

molecule of interest is dissolved in) and the sample solutions are respectively measured. NanoDrop 2000 spectrophotometer records also both blank and sample spectra.

The sample intensities along with the blank intensities are used to calculate the sample absorbance according to the following equation:

$$Absorbance = -\log\left[\frac{Intensity_{sample}}{Intensity_{blank}}\right]$$

The measured light intensity of both the sample and the blank are required to calculate the absorbance at a given wavelength.

The Beer-Lambert equation is used to correlate the calculated absorbance with concentration:

$$A = \varepsilon \cdot b \cdot C$$

For nucleic acid quantification, the Beer-Lambert equation is modified to use a factor with units of ng-cm/microliter.

$$C = \frac{(A \cdot \varepsilon)}{b}$$

Where:

- A is the absorbance represented in absorbance units.
- ε is the wavelength-dependent molar absorptivity coefficient (or extinction coefficient) with units of liter/mol-cm.
- The generally accepted extinction coefficients for nucleic acids are:
- Double-stranded DNA: 50 ng-cm/µL
- Single-stranded DNA: 33 ng-cm/µL
- RNA: 40 ng-cm/µL
- b is the optical pathlength in cm.
- C is the analyte concentration in moles/liter or molarity.

The nucleic acid absorbance values are normalized to a 1.0 cm (10.0 mm) path for all measurements. Although it is generally recognized that quantification based on optical density leads to inaccurate results [63], collecting absorption spectra as a visual "fingerprint" of sample purity could be useful. As numerical parameters of sample purity, the commonly used A260/A280 and A260/A230 absorbance ratios were included in the instrument report.

In general, an A260/A280 ratio of 1.8 is accepted for "pure" DNA, while a ratio value about 2.0 is generally accepted for "pure" RNA.

The 260/230 values for a "pure" nucleic acid are often higher than the respective 260/280 values and are commonly in the range of 1.8-2.3.

4.6.2 FLUORESCENCE: THERMOFISHER SCIENTIFIC QUBIT 3 FLUOROMETER

The ThermoFisher Scientific Qubit 3.0 fluorometer is a benchtop fluorometer that can be used for the quantification of DNA, RNA, microRNA, and protein molecules using specific fluorescence-based Qubit quantitation assays.

The DNA and RNA concentrations of the thesis samples were obtained using ThermoFisher Scientific Qubit 3.0 fluorometer with Qubit dsDNA HS Assay Kit and Qubit RNA HS Assay Kit respectively.

The procedure used for measuring the concentration of nucleic acids requires that the fluorescent staining solution is firstly prepared. Secondly, the fluorescent staining solution is added to the standard solutions and the nucleic acid sample solutions. Finally, after the setting of the instrument and the creation of calibration curve using the standards solutions, the sample solutions are respectively measured.

The Qubit 3.0 fluorometer provides nucleic acid values in ng/mL. This value corresponds to the concentration after sample dilution into the assay tube. The sample concentration can be calculated with the following equation:

$$C = QF\ value \cdot \frac{200}{x}$$

Where:

- QF value is the value given by the Qubit 3.0 Fluorimeter.
- x is the number of microliters of sample added to the assay tube.

4.7 FRAGMENT SIZE DISTRIBUTION: AGILENT 2200 TAPESTATION SYSTEM

The Agilent 2200 TapeStation system carries out electrophoretic separation of nucleic acids and proteins. In particular, it is able to detect fluorescently stained double stranded DNA, including genomic DNA, total RNA and fluorescently labelled proteins.

For DNA samples, fragment size distribution was determined using the Agilent 2200 TapeStation System with the D1000 ScreenTape Assay, High Sensitivity D1000 ScreenTape Assay, D5000 ScreenTape Assay, High Sensitivity D5000 ScreenTape Assay, and Genomic DNA ScreenTape Assay. The ScreenTape Assay are dedicated cards having channels preloaded with the gel medium used to obtain different fragment size ranges: 35 -1,000 bp (D1000 assays), 100 – 5,000 bp (D5000 assays), 200 - >60,000 bp (Genomic DNA assay).

The instrument automatically performs gel electrophoresis separation of fluorescently stained samples in channels preloaded with the gel medium contained in ScreenTapes. Electropherograms (EPGs) are collected as curves of fluorescence intensity vs distance along the channel after distance conversion to fragment size expressed in base pairs (bp) by calibration using a reference ladder sample. The instrument provides also quantification using reference peaks from markers added to the sample, both across the full fragment size range of the specific ScreenTape used, or across selected regions in the EPG.

4.8 SAMPLE QUALITY CONTROLS FOR NGS LIBRARY PREPARATION: QIASEQ DNA QUANTIMIZE SYSTEM

The QIAseq DNA QuantiMIZE System utilizes a high performance SYBR Green I Real Time PCR-based approach to determine the quantity and quality of sample DNA that is amenable to PCR-based targeted enrichment prior to next-generation sequencing (NGS).

In detail, the QIAseq DNA QuantiMIZE System uses two Real Time PCR assays (Assay 100 and Assay 200, generating sequences around 100 bp and 200 bp, respectively) to query more than 40 discrete genomic loci that are randomly distributed in the genome for the quantification and qualification of DNA extracted from FFPE samples and provide guidance for the appropriate amount of input DNA for NGS libraries preparation reactions. The use of the included high-quality reference gDNA allows quantification of amplifiable molecules in gDNA samples, without the need for serial dilutions.

The procedure used for the QIAseq DNA QuantiMIZE Assay involves mainly four steps: the sample preparation, PCR master mixes preparation, Real Time PCR sample preparation and analysis.

In the sample preparation step, the DNA extracts having concentrations higher than 50 ng/µl must be diluted to 20 ng/µl. Samples with concentrations lower than 50 ng/µl are used without dilution step.

In the PCR mixes preparation step, two different master mixes must be prepared that differ for the primer solutions used (Primer Assay100 and Primer Assay 200 respectively). The common elements are the RNase-/DNase-free water and the QIAseq qPCR SYBR Green Mastermix Low ROX. Both Primer Assay solutions and QIAseq qPCR SYBR Green Mastermix Low ROX are included in the kit. The quantities required for each master mix are calculated for 16 reactions: 14 samples, the control gDNA and the negative control.

In the last two steps, the DNA samples and the two controls, after they have been diluted again with RNase-/DNase-free water, are individually analyzed with six reactions. In detail, three reactions (triplicate) with Assay 100 PCR master mix and three reactions (triplicate) with the Assay 200 PCR

master mix are carried out. Each volume reaction is 25 µl, 20 µl of master mix and 5 µl of sample/control.

The runs were performed with Applied Biosystem 7500 Fast following the Real Time PCR cycling program showed below:

Step	Cycles	Duration	Temperature (°C)
Initial denaturation	**1**	**10 min**	**95**
Denaturation	**40**	**15 s**	**95**
Annealing/Extension/ Data acquisition		**2 min**	**60**

The name, the concentration analyzed by ThermoFisher Scientific Qubit 3.0 fluorometer and the Ct values measured by the Applied Biosystem 7500 Fast software of each sample were inserted into the Qiagen Microsoft Excel spreadsheet according the manufacturer's instructions for the calculations.

The parameters for determining the quality of each DNA sample (quality control, QC) and the amount required for the library preparation were automatically calculated by the Qiagen excel file.

The parameters are shown and explained below:

- ΔCt100 = Ct100 sample – Ct100 Control
- ΔCt200 = Ct200 sample –Ct200 Control
- Slope = qPCR QC Score. QC Score provides an indication of DNA sample damage or fragmentation. The value is obtained using excel SLOPE function with ΔCt100 and ΔCt200
- qPCR QC Call = Samples with a QC Score <= 0.04 are labeled with "High", meaning high quality and low degree of damage or fragmentation. Samples with a QC Score > 0.04 are labeled with "Low", meaning low quality and a high degree damage or fragmentation.
- "Sample Volume (µL) Recommended " and "Sample Amount (ng) Recommended "= these value were provided by the QIAseq DNA QuantiMIZE Assay Handbook and Qiagen excel file.

4.9 NGS LIBRARY GENES PANELS

4.9.1 ILLUMINA TRUSEQ CUSTOM AMPLICON LOW INPUT

Illumina DesignStudio website is the bioinformatics tool used for defining the target genes of the Illumina library panel kit for high grade serous ovarian cancer. This free web-based software tool provides a simple and powerful method for evaluating loci and creating the custom genotyping assays. Metrics returned by DesignStudio provides prediction information, validation status, and minor allele frequencies from published studies.

The panel contains twenty-four different genes that were chosen based on "state of art" about mutated genes associated with high grade serous ovarian cancer: TP53, BRCA1, BRCA2, CSMD3, NF1, CDK12, FAT3, GABRA6, RB1, BRAF, PIK3CA, KRAS, NRAS, ATM, BARD1, BRIP1, CHEK1, CHEK2, FAM175A, MRE11A, NBN, PALB2, RAD51C, RAD51D [15] [73].

In detail, parameters used for designing the high grade serous ovarian cancer genes panel are:

1. Assay type: DNA. This parameter determines which is the nucleic acid target to sequencing;
2. Assay technology: the library preparation technology (amplicon or target enrichment technology), in this case TruSeq Custom Amplicon Low Input is used (for applications with ≥ 10ng input. Amplicon sizes include 150 bp, 175 bp, and 250 bp. Read length up to 2x150);
3. Species: this parameter indicates the species and source that DesignStudio uses as a reference when designing probes. In this study, Human is the specie of target genes (hg19 is the genes database used for the design);
4. Design name: genes panel name;
5. Description: here it is possible to annotate with information regarding panel design;
6. Instrument: here it is necessary to select the Illumina sequencing machine. MiSeq is the instrument used for this project;
7. Assay version: this parameter determines how much nucleic acid input (ng) is required. In this study, the use of low input DNA for FFPE applications is included. Dual pools design is an option applied on this panel design. Dual pool designs add a second, mirrored set of complementary amplicons to target both DNA strands at all loci, which is especially helpful in FFPE samples. Dual pool design avoids loss of nucleotide information (for example a nucleotide excision on a strand belonging to double helix) resulting from DNA fragmentations induced by formalin fixation, and using the intact DNA strand to obtain the exact nucleotides sequence.

8. Variant Source: database used for single nucleotide polymorphism identification. 1,000 Genomes is the source used for this Ph.D. work because it is one of the most detailed catalogue of human genetic variation.
9. Population: specific subpopulations based on geographic origin.
10. Amplicon Size: the target length of amplicons (in bp) to be used in the design. In this case, the amplicon size required is 175 bp.
11. Add target: this section allows to input genes names or the coordinates of target sequences on chromosome. In this work, genes names are used to identify and design the nucleic acid sequences panel.
12. Targeting options: a target region comprises a pair of coordinates that define a contiguous sequence in the genome
13. Padding per exon: the number of bases (from 0–25) added upstream and downstream of each exon within the target region. In this design, the value is zero.

4.9.2 QIASEQ TARGETED DNA CUSTOM PANEL

The QIAseq Targeted DNA Panel solution is constituted both by proprietary design informatics tool (QIAseq Targeted DNA Custom Panel Builder) and all chemical reagents that allow:

- The targeted library preparation of each FFPE DNA sample.
- The following digital DNA sequencing by utilizing unique molecular indices (UMI they identify each original starting DNA molecule) and sample barcodes (they identify each sample DNA).

The Digital DNA sequencing is a unique approach to detect low-frequency variants with high confidence by overcoming the issues of PCR duplicates, false positives and library bias.

The QIAseq Targeted DNA Custom Panel Builder is the informatics tool used for generating the QIAseq Targeted DNA Panel –COLORECTAL (catalog identifier: CDHS-13060Z-63).

The panel allows the NGS library preparation for three genes correlated to human colorectal cancer KRAS, NRAS and BRAF.

The Qiagen software, unlike Illumina DesignStudio, employs the following parameters to create the library panel:

1. Gene: this section allows to input genes names or the coordinates of target sequences on chromosome. In this work, genes names are used to identify and design the nucleic acid sequences panel considering only exons.

2. Species: this parameter indicates the species and source that Illumina DesignStudio uses as a reference when designing probes. In this study, Human is the specie of target genes (hg19 is the genes database used for the design).
3. NGS platform: the samples libraries were analyzed with Illumina MiSeq instrument.

The QIAseq Targeted DNA Panel –COLORECTAL uses the GRCh37 reference genome to design 63 different primers that can amplify the regions of interest (ROI) with 100% coverage and a size of 4,007 bp.

The principle (**Figure 16**) and the protocol used for the library preparation with QIAseq Targeted DNA Custom Panel kit is described below:

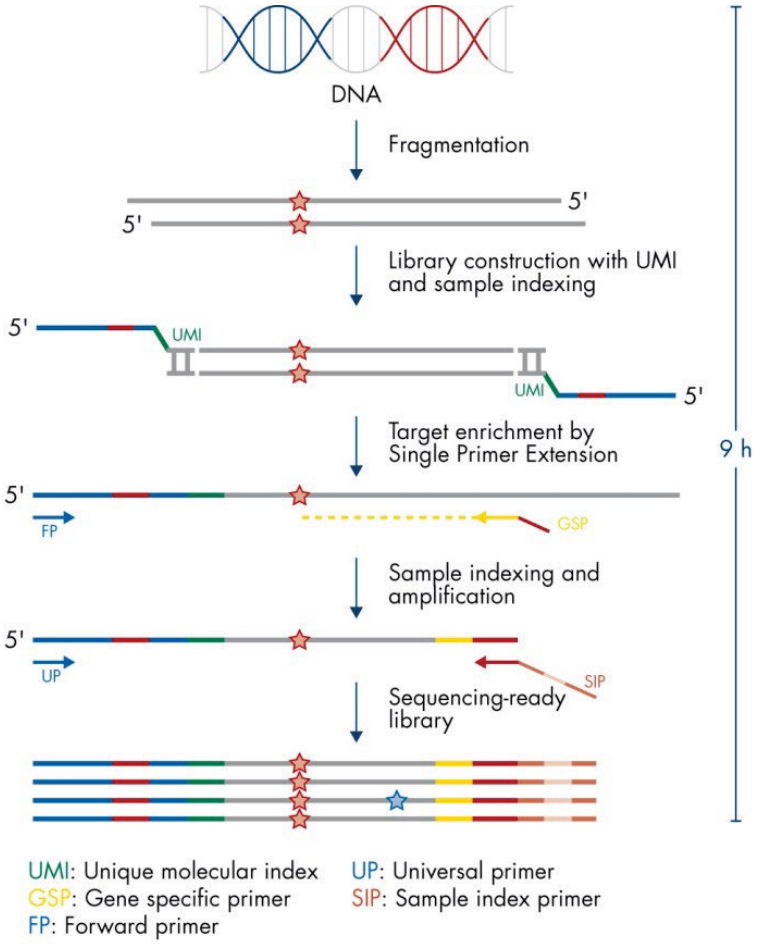


Figure 16: "QIAseq Targeted DNA Panel Handbook" 2017. https://www.qiagen.com/kr/resources/download.aspx?id=8907edbe-a462-4883-ae1b-2759657e7fd0&lang=en pp: 1-76.

1. Fragmentation

40 ng of each high quality sample FFPE DNA or 120 ng of low quality sample DNA (amounts obtained from QIAseq DNA QuantiMIZE System) are first fragmented, end repaired and A-tailed within a single, controlled multi-enzyme reaction. The prepared DNA fragments, obtained using

proprietary enzymes mix, are then ligated at their 5' ends with a sequencing platform-specific adapter containing UMIs and sample index.

2. UMI assignment

Prior to target enrichment and library amplification, each sample DNA molecule is tagged with a unique 12-base fully random sequence or index, commonly referred to as a UMI. Statistically, this process provides 4^{12} possible indices per adapter, and each DNA molecule in the sample receives a unique UMI sequence. In addition, this ligated adapter also contains the first sample index.

3. Target enrichment and final library construction

Target enrichment is performed post-UMI assignment to ensure that DNA molecules containing UMIs are sufficiently enriched in each sequenced sample library. For enrichment, ligated DNA molecules are subject to several cycles of targeted PCR using one region-specific primer and one universal primer complementary to the adapter. A Universal PCR is ultimately carried out to amplify each sample library and add Illumina MiSeq platform specific adapter sequences and additional sample indices.

The sample DNA extracts that will be analyzed with Illumina NGS technology are labeled with the following indexes in the library preparation process (**Table 4**):

Sample ID	Sample Barcode
M[34574]_D[BC-FC]	IL-N711
M[07775]_D[BC-FC]	IL-N712
M[21830]_D[BC-FC]	IL-N714
M[32892]_D[BC-FC]	IL-N715
M[31317]_D[P-RP]	IL-N705
M[13218]_D[P-RP]	IL-N706
M[28108]_D[P-RP]	IL-N707
M[34295]_D[P-RP]	IL-N710
M[36110]_D[Q-GR]	IL-N701
M[01773]_D[Q-GR]	IL-N702
M[30673]_D[Q-GR]	IL-N703
M[31271]_D[Q-GR]	IL-N704

Table 4: Sample DNA extracts and respective sample barcoeds.

The manufacturer does not provide the nucleic acid sequences of primers, UMIs, adapters and samples barcodes.

4.10 LIBRARIES QUANTIFICATIONS FOR NGS ANALYSIS: KAPA LIBRARY QUANTIFICATION KIT – ILLUMINA PLATFORMS

KAPA Library Quantification Kit for Illumina platforms utilizes a SYBR Green I Real Time PCR-based approach to quantify the sample libraries (or pool libraries) flanked by the P5 and P7 flow cell oligo sequences for Illumina instruments.

The library quantification is performed by amplifying the set of six pre-diluted DNA Standards and diluted library samples by qPCR, using the KAPA SYBR FAST qPCR Master Mix and primers targeting the Illumina P5 and P7 flow cell oligo sequences. The average Cq (quantitation cycle) score for each DNA Standard is plotted against $\log_{10}$ (concentration in pM) to generate a standard curve. The concentrations of diluted library samples are then calculated against the standard curve, allowing absolute quantification.

The six-library quantification DNA Standards are 10-fold dilution series (20, 2, 0.2, 0.02, 0.002 and 0.0002 pM respectively) of linear 452 bp templates.

The library quantification primers have the following nucleotide sequences:

- Primer 1: 5'-AAT GAT ACG GCG ACC ACC GA-3'
- Primer 2: 5'-CAA GCA GAA GAC GGC ATA CGA-3'

The procedure used for the KAPA Library Quantification Kit involves mainly three steps: the sample preparation, Real Time PCR master mix preparation and analysis.

In the sample preparation step, each sample library must be diluted 1:200,000 using DNA dilution buffer (10 mM Tris-HCl, pH 8.0 – 8.5 (25°C) + 0.05% Tween 20) as recommended in the KAPA Library Quantification Kit user manual. Then, the six library quantification DNA Standards have to be prepared using the same DNA dilution buffer.

In the Real Time PCR master mix preparation step, the master mix containing the dedicated KAPA SYBR® FAST qPCR Master Mix (2X) solution, Primer Premix (10X) solution and ROX Low (50X) solution, has to be prepared for the analysis.

In the last step, the sample libraries and the library quantification DNA Standards are individually analyzed by triplicate. Each volume reaction is 10.2 µl, 6.2 µL Real Time PCR Mix and 4 µl of sample/standard.

The runs were performed with an Applied Biosystem 7500 Fast DX instrument following the Real Time PCR cycling program showed below:

Step	Cycles	Duration	Temperature (°C)
Initial denaturation	**1**	**5 min**	**95**
Denaturation	**35**	**30 s**	**95**
Annealing/Extension/ Data acquisition		**45 s**	**60**

The concentrations and the Cq values obtained by the Applied Biosystem 7500 Fast DX software of each library quantification DNA Standard were used to generate a standard curve (average Cq vs $\log_{10}$ [pM]) for calculating the concentration of each library or libraries pool solution.

The standard curve has to ensure the following criteria:

- The average ΔCq value between DNA Standards is in the range of 3.1 – 3.6.
- $R^2 \geq 0.99$

The 1:200,000 libraries concentrations were obtained interpolating the library or libraries pool average Cq into the equation derived from the standard curve.

The average size-adjusted concentration (in nM) for each dilution of every library and control were obtained using the following formula:

$$\frac{\boldsymbol{Size\ of\ DNA\ Standard\ (452bp)}}{\boldsymbol{Average\ library\ fragment\ length\ (375bp)}} * \left(\boldsymbol{library\ conc\ (pM)_{1:200,000}}\right) * \mathbf{200,000} * \mathbf{10^{-3}} \boldsymbol{nMpM^{-1}}$$

4.11NGS ANALYSIS: ILLUMINA MISEQ

The NGS analysis is performed using Illumina MiSeq instrument and the MiSeq Reagent Micro Kit v2 300 cycles (150x2) flow-cell.

The parameters used for setting the run are shown below:

- Category: Other
- Select Application: FASTQ Only
- Sample Prep Kit: Nextera XT v2
- Index Reads: 2
- Read Type: Paired End Read
- Cycles for both Read 1 and 2: 151
- FASTQ Only Workflow Specific Setting: Custom Primer for Read 1, Use Adapter Trimming

4.11.1 LIBRARIES POOL LOADING SOLUTION PREPARATION

5 µL of 4nM libraries pool were mixed with 5 µL 0.2M NaOH solution for the denaturation step. After 5 minutes of RT incubation, 990 µL HT1 buffer (hybridization buffer) were added (final concentration is 20pM) to the mix.

The libraries pool loading solution is constituted by 360 µL of 20pM denatured libraries pool, 30 µL of 20pM PhiX Control and 210 µL HT1 buffer (PhiX > or = 5% Spike-in, for low diversity library). The solution was loaded into Illumina Miseq cartridge Micro V2. The final concentration of libraries pool in MiSeq instrument is 12 pM.

4.11.2 QIASEQ A READ 1 PRIMER I PREPARATION

597 µl HT1 (Hybridization Buffer) were added to dilute 3 µl QIAseq A Read 1 Custom Primer I (provided by the QIAseq Targeted DNA Panel kit) to obtain a final concentration of 0.5 µM. The solution was loaded to position 18 of the MiSeq reagent cartridge.

5 RESULTS AND DISCUSSION

The results shown in this thesis chapter reflect the logical steps that must be taken into account when a NGS work- -flow for CE-IVD applications has to be studied.

The starting point involves a study phase where it is necessary to understand the "state of art" about the target disease and the products able to detect it which are available on the market, including an assessment of the technology of interest for the future clinical application (in this thesis the next generation sequencing).

Then, it is necessary to extend the study phase for identifying the requirements, in terms of methods and kits, that allow to obtain a biological material with suitable features for the following analysis.

Finally, when all requirements (both kits and samples) are determined, it is possible to proceed with the experimental phases where the starting biological sample is processed for the final analysis, which allow reaching the info desired.

The experimental steps involve also the quality control tests required for guaranteeing that the sample has the characteristics required for obtaining the correct analysis.

The 5.1 "PATENT SEARCH ANALYSIS" paragraph shows the patent search carried out to check the state of the art on the products and the technologies currently available on the market for detecting the high grade serous ovarian cancer.

The 5.2 "NGS LIBRARY PREPARATION" paragraph shows the Illumina genes panel design based on genes having potential diagnostic, prognostic and therapeutic implications.

The 5.3 "NUCLEIC ACID EXTRACTION KITS SCREENING" paragraph shows the results derived from the screening for the selection of FFPE nucleic acid extraction kits having features potentially compatible with the integration into a clinical-diagnostic-therapeutic process.

The paragraphs from 5.4 to 5.8 shows the results obtained for studying the experimental steps necessary to sequence FFPE samples on the Illumina platform using amplicon technology with well-defined target genes.

In detail:

- The 5.4 "NUCLEIC ACID EXTRACTION KITS PERFORMANCES ANALYSIS" paragraph shows the extraction quantification and characterization of DNA (4.4.1 section) and RNA (4.4.2 section) isolated from FFPE samples necessary for the preparation of the libraries obtained with the kits selected in the 4.3 paragraph. Here, the nucleic acid extraction kits with the best features for NGS application are identified.
- The 5.5 "DNA EXTRACTION FROM FFPE SAMPLES" paragraph shows the extraction results obtained from clinical FFPE samples using the kits selected in the 4.4 paragraph.

- The 5.6 “QUALITY CONTROLS FOR NGS LIBRARIES PREPARATION” paragraph shows the results for determining the quality and the concentrations of the nucleic acids obtained from each DNA extraction kit necessary for the sample libraries preparation step.
- The 5.7 “NGS LIBRARIES QUANTIFICATION ANALYSIS” paragraph shows the results obtained from sample libraries quantification and the libraries average size measurements necessary for the libraries pool preparation to employ for final sequencing on Illumina MiSeq instrument.
- The 5.8 “NGS ANALYSIS” paragraphs shows the results obtained from the sequencing run.

5.1 PATENT SEARCH ANALISYS

The aim of this work was to evaluate the "state of art" regarding the presence of patents relevant for the HERCULES project.

The resulting twenty-two patents are listed in "SUPPLEMENTARY MATERIALS - APPENDIX A".

5.1.1 DISCUSSION

5.1.1.1 GENERAL CONSIDERATIONS

More than 200 results have been found and individually inspected to discard those not relevant for the HERCULES project. Twenty-two results were selected that are based on potential protocols for isolating ovarian cancer subpopulations employing different biomarker typologies, methods and systems for the prediction of ovarian cancer treatment.

5.1.1.2 OVARIAN CANCER IDENTIFICATION CONSIDERATIONS

Nine patents out of twenty-two (patents number 1, 4, 9, 10, 13, 14, 15, 17 and 20 respectively) show methods able to identify and / or diagnose ovarian cancer.

Two patents (1, 4) describe protocols for isolating the cancer stem cells (CSC) from tumor tissue. The CSC identification is performed by analyzing protein expression and miRNA expression profile.

The patent number 9 describes a protocol that uses the Mullerian Inhibiting Substance (MIS) to separate the cancer stem cells from the other cell component, also in this casa employing tumor matrix.

Six patents describe distinct methods employing the analysis of protein mix biomarkers, mRNA patterns, genetic mutations or epigenetic alterations in order to diagnose ovarian cancer, starting from primary tumors, blood and its derivatives.

In all cases described above, there is no specific selection for HGS-OvCa but as tumors belonging to ovarian cancer family. Two patents (patents 10 and 15 respectively) described protocols that diagnose ovarian cancer as part of a heterogeneous set of tumors (lung, gynecological, etc.).

The technologies reported are the Real Time PCR, protein isolation and chips, with possible NGS applications.

5.1.1.3 OVARIAN CANCER TREATMENT CONSIDERATIONS

Twelve patents out of twenty-two (patents number 2, 3, 5, 6, 7, 8, 14, 17, 18, 19, 21 and 22 respectively) show methods able to predict the effectiveness of ovarian cancer surgery or pharmacological treatments.

The patent number 7 describes a protocol capable of predicting the effectiveness of cytoreduction through a genetic expression profile analysis starting from tumor tissue nucleic acids extracts.

The other patents describe procedures in which the mutational analysis of different genes, the gene expression profile, hyper methylation of promoters, identification of CpG island, or combination of genetic alterations with protein biomarkers (e.g. CA-125 and E2F5 mRNA, patent 13) are used to predict efficacy or susceptibility to numerous drugs such as taxol, cis-platinum (22) or immunoglobulins (21).

5.1.1.4 OVARIAN CANCER PROGNOSIS AND MONITORING CONSIDERATIONS

Eight patents out of twenty-two (patents number 10, 11, 12, 14, 16, 17, 18 and 19 respectively) show methods able to provide prognosis basing on the evolution of the disease and monitor the onset of the tumor after surgical / pharmacological treatments.

There are patents that describe methods, protocols, technologies and biomarkers that are able to analyze every aspect of the disease, hence diagnosis, prognosis, pharmacological treatment and monitoring of the reappearance of the pathology.

The patents number 11, 12 and 16 show protocols and biomarkers (both protein and genetic) able to monitor the presence of tumor cells during or after the chemotherapy cycles.

5.1.1.5 FINAL CONSIDERATION

The patents described above show protocols, methods and technologies able to analyze patterns of different biomarkers, both genetic and protein, isolated from distinct biological matrices.

No specific biomarkers for HGS-OvCa are emerged that could provide detailed information on its diagnosis, prognosis, prediction of therapeutic efficacy and follow-up.

Furthermore, no restrictions have been found on the use of NGS technology, particularly Illumina, to be applied to the HGS-OvCa.

The patent analysis showed no restrictions to the “freedom to operate” (FTO) at the moment for prediction of therapy effectiveness. A more specific FTO analysis, as well as a patentability search, will be carried out as soon as specific biomarkers are determined in the HERCULES project.

5.2 NGS LIBRARY PREPARATION KIT FOR HGS-OvCa

The NGS library preparation kit chosen for the nucleic acids sequencing based on Illumina platform is the TruSeq Custom Amplicon Low Input Kit. This kit is a customized, amplicon-based assay for targeted high grade serous ovaian cancer genes panel resequencing using 10 ng of genomic DNA (gDNA).

The TruSeq Custom Amplicon Low Input Kit was chosen for the following reasons:

1. Low DNA input required: the kit requires very low DNA input amounts, about 10 ng of genomic DNA, and, in particular, it is also compatible with formalin-fixed, paraffin-embedded (FFPE) samples, starting from 10 ng of FFPE tumor genomic DNA (depending on DNA quality).
2. UMI: unique molecular identifiers (UMI) (they are optional, normally they are not included in the Illumina with TruSeq Custom Amplicon Low Input Kit) allow to enhance allelic detection and to increase sensitivity. The unique molecular identifiers allow the removal of PCR biases and the detection of individual amplicon molecules.
3. MiSeq compatibility: TruSeq Custom Amplicon Low Input Library Prep Kit prepares library tagged DNA sequences compatible with MiSeq, which is the Illumina instrument present in the R&D company division and available in CE IVD version.

5.2.1 GENE PANEL DESIGN

The high grade serous ovarian cancer genes panel was designed using the online tool Illumina Design Studio (**Table 5** and **Table 6**).

Assay Type	**DNA**	**Amplicons Selected/Designed**	**1,180 / 1,181**
Assay Technology	**TruSeq Custom Amplicon Low Input**	**Targets Selected/Designed**	**482 / 482**
Species	**Human (hg19)**	**Undesignable Targets**	**1**
Design Type	**Dual Pool Design**	**Number Of Gaps**	**16**
Variant Source	**1,000 Genomes**	**Sum Of Gap Distance**	**626 bp**
Population	**African, American, Asian, European**	**Cumulative Target**	**98,796 bp**
Amplicon Lenght	**175 bp**	**Coverage**	**99%**

Table 5: Genes panel design specifications.

Region Name	Target Type	Target	Chr	Start	End	Length	Coverage%	Amplicon	Gaps	SNPs
ATM	CDS	52	11	108,093,559	108,239,826	146,268 bp	99%	101	2	16 bp
	Full Region	1	11	108,141,791	108,142,133	343 bp	100%	4	0	0 bp
	Full Region	1	11	108,143,259	108,143,579	321 bp	100%	3	0	2 bp
	Full Region	1	11	108,186,550	108,186,840	291 bp	100%	3	0	1 bp

	Full Region	1	11	108,098,352	108,098,615	264 bp	100%	3	0	1 bp
	Full Region	1	11	108,235,809	108,236,232	424 bp	100%	5	0	0 bp
BARD1	CDS	11	2	215,593,275	215,674,428	81,154 bp	100%	27	0	5 bp
BRAF	CDS	18	7	140,433,813	140,624,564	190,752 bp	100%	32	0	6 bp
BRCA1	CDS	19	17	41,196,312	41,277,500	81,189 bp	100%	25	0	5 bp
	Full Region	1	17	41,228,505	41,228,631	127 bp	100%	2	0	1 bp
	Full Region	1	17	41,243,452	41,246,877	3,426 bp	100%	36	0	7 bp
	Full Region	1	17	41,258,473	41,258,550	78 bp	100%	1	0	1 bp
	Full Region	1	17	41,197,698	41,197,819	122 bp	100%	1	0	0 bp
BRCA2	CDS	22	13	32,889,617	32,973,809	84,193 bp	99%	111	2	28 bp
	Full Region	1	13	32,900,238	32,900,419	182 bp	71%	1	1	0 bp
	Full Region	1	13	32,953,887	32,954,282	396 bp	100%	4	0	0 bp
BRIP1	CDS	19	17	59,756,547	59,940,920	184,374 bp	100%	45	0	5 bp
CDK12	CDS	13	17	37,617,739	37,690,800	73,062 bp	100%	40	0	3 bp
	Full Region	1	17	37,686,857	37,687,566	710 bp	100%	6	0	0 bp
CHEK1	CDS	10	11	125,495,031	125,546,150	51,120 bp	100%	19	0	4 bp
	Full Region	1	11	125,499,127	125,499,355	229 bp	100%	3	0	0 bp
CHEK2	CDS	12	22	29,083,731	29,137,822	54,092 bp	87%	15	2	1 bp
	Full Region	1	22	29,120,965	29,121,355	391 bp	100%	4	0	1 bp
	Full Region	1	22	29,115,383	29,115,473	91 bp	73%	1	1	1 bp
CSMD3	CDS	72	8	113,235,159	114,449,242	1,214,084 bp	99%	138	4	24 bp
FAM175A	CDS	9	4	84,382,094	84,406,290	24,197 bp	100%	16	0	5 bp
FAT3	CDS	25	11	92,085,262	92,629,635	544,374 bp	100%	145	0	26 bp
GABRA6	CDS	7	5	161,112,658	161,129,598	16,941 bp	100%	13	0	5 bp
	Full Region	1	5	161,115,955	161,116,342	388 bp	100%	4	0	0 bp
KRAS	CDS	5	12	25,358,180	25,403,854	45,675 bp	100%	8	0	2 bp
MRE11A	CDS	19	11	94,150,469	94,227,040	76,572 bp	99%	31	1	6 bp
NBN	CDS	16	8	90,945,564	90,996,899	51,336 bp	100%	32	0	7 bp
NF1	CDS	47	17	29,421,945	29,704,695	282,751 bp	99%	96	1	18 bp
	Full Region	1	17	29,559,718	29,560,231	514 bp	100%	4	0	0 bp

	Full Region	1	17	29,562,629	29,563,039	411 bp	100%	4	0	0 bp
	Full Region	1	17	29,663,351	29,663,932	582 bp	100%	6	0	1 bp
	Full Region	1	17	29,683,978	29,684,387	410 bp	100%	4	0	0 bp
	Full Region	1	17	29,664,837	29,665,157	321 bp	100%	3	0	0 bp
	Full Region	1	17	29,548,868	29,549,005	138 bp	100%	2	0	0 bp
NRAS	CDS	4	1	115,247,085	115,259,515	12,431 bp	92%	6	1	1 bp
PALB2	CDS	11	16	23,614,483	23,652,678	38,196 bp	100%	38	0	7 bp
	Full Region	1	16	23,649,171	23,649,450	280 bp	100%	3	0	1 bp
PIK3CA	CDS	16	3	178,866,311	178,952,497	86,187 bp	100%	35	0	9 bp
	Full Region	1	3	178,927,974	178,928,353	380 bp	100%	4	0	0 bp
	Full Region	1	3	178,947,792	178,948,164	373 bp	100%	4	0	2 bp
RAD51C	CDS	8	17	56,769,963	56,811,692	41,730 bp	100%	12	0	2 bp
	Full Region	1	17	56,772,292	56,772,551	260 bp	100%	3	0	0 bp
RAD51D	CDS	7	17	33,426,811	33,446,888	20,078 bp	100%	8	0	0 bp
	Full Region	1	17	33,430,273	33,430,563	291 bp	100%	3	0	0 bp
	Full Region	1	17	33,427,975	33,428,384	410 bp	100%	4	0	0 bp
RB1	CDS	23	13	48,877,883	49,056,026	178,144 bp	100%	39	0	4 bp
	Full Region	1	13	48,954,189	48,954,377	189 bp	82%	2	1	0 bp
	Full Region	1	13	49,039,134	49,039,504	371 bp	100%	5	0	1 bp
TP53	CDS	3	17	7,571,720	7,590,868	19,149 bp	100%	3	0	2 bp
	Full Region	1	17	7,576,853	7,577,155	303 bp	100%	3	0	1 bp
	Full Region	1	17	7,578,177	7,578,554	378 bp	100%	4	0	0 bp
	Full Region	1	17	7,579,312	7,579,912	601 bp	100%	5	0	3 bp
	Full Region	1	17	7,576,540	7,576,657	118 bp	100%	2	0	1 bp

Table 6: Correlations between genes with their coordinates in genome and amplicons designed by Illumina DesignStudio free web-based software.

5.2.2 DISCUSSION

5.2.2.1 GENERAL CONSIDERATIONS

The design is based on TruSeq Custom Amplicon Low Input as library preparation technology. Being FFPE DNA samples damaged by strong tissue paraffin inclusion process, this panel uses dual pool

design in which amplicons are designed for both DNA strands in order to eliminates false positives that can arise from deamination events during formalin fixation. The Illumina software designs two panels, one for each strand of DNA double helix. Functionally, the panels can be considered only one.

5.2.2.2 PANEL DETAIL CONSIDERATIONS

The software has calculated 1,180 amplicons (with a length of 175 bp) for analyzing 482 specific sequences belonging to 24 genes, which were chosen from literature.

The coverage, based on the alignment with 1,000 Genomes database (which includes African, American, Asian and European populations), is 99% because an undesignable target is present (i.e. a DNA sequence that the software is not able to design for sequencing). 16 gaps (a gap is any base position in the region of interest, for example exon, that is intended for enrichment, but fails to be theoretically enriched by the probe set) are shown (**Table 5**).

Table 6 shows the details about the genes analyzed, the chromosomal regions used for amplicons design (chromosomes coordinates included), the length of the regions where amplicons are designed, the coverage (the total number of non overlapping bases covered by probes divided by the total number of bases in design), the amplicon number and the single nucleotide polymorphisms (SNP) present in each sequence analyzed.

The amplicons were designed on coding DNA sequences (CDS), but an amount of them were merged in order to maximize the coverage, because some amplicons overlap in parts.

In some cases, the coverage was equal or lower than 99% because the software was not able to design primers. The Illumina DesignStudio has created a panel that includes also non pathological nucleotide variations, in this way the software helps users to identify the relevant mutations instead of normal polymorphisms.

5.2.2.3 FINAL CONSIDERATIONS

This work allowed becoming familiar with the Illumina technology, the requirements specification step, the features and the Illumina informatics tool available. It is also directed at design a gene panel, which was aimed to prepare libraries with the amplicons and the UMIs technologies for Illumina platform.

This genes panel was designed by considering genes correlated to high grade serous ovarian cancer but not include specific and univocal mutations present only in this type of ovarian carcinoma. The purpose of the Company is to design, test and validate a gene panel that contains nucleic acid alterations able to correlate with precise information about the diagnosis, prognosis and the therapy.

The genes panel is subject to possible revision based on the results that come from the HERCULES project.

Illumina MiSeq instrument will employ DNA derived from FFPE high grade serous ovarian cancer samples, which will be provided from University of Trieste, Department of Medical Sciences, as a partner involved in HERCULES project.

5.3 NUCLEIC ACID EXTRACTION KITS SCREENING

An initial screening of 39 DNA extraction protocols was performed looking at the following parameters, relevant for the choice of the protocol as a component of a CE IVD marked workflow, through analysis of the technical documentation and direct contacts with potential suppliers.

The screening activity and the following performances tests are focalized on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

Concerning nucleic acid extraction kits for liquid handler automation, the ideal extraction kit is based on magnetic beads technology. It is necessary to analyze kits that have both manual and automation protocols for determining their performances and identifying the kits which are compatible with a CE-IVD NGS workflow.

About nucleic acid extraction kits for benchtop instruments, the kit can be used only with dedicated instruments.

The preselection phase consists in screening for kits based on various requirements.

Operator safety risk: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

Eight kits were excluded in this work because contain xylene (a toxic reagent [64]) used in deparaffinization step. Actually, xylene is used in anatomo-pathological laboratory for paraffin removal step, but Italian legislative decree imposes to minimize xylene use [65]. For this reason, and for company policy, kits containing xylene were discarded.

FFPE nucleic acid yield: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

This is one of the most important feature analyzed in every kit after the preselection phase. Genomic DNA yield, in terms of quality (considering that DNA derived by FFPE is already damaged from its preparation) and purity (potential library preparation and sequencing reactions inhibitors), depends from sample and kit characteristics, so it is not possible to know a priori the extraction results. The kit user manuals, confirmed by scientific literature, report that yield depends essentially on the characteristics of every sample (for example tissue type), FFPE process used, time-spent from FFPE process, preservation method, etc [66]. The selection based only on technical documentation was not conclusive, thus requiringexperimental measurements.

Quantity of FFPE tissue: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

This parameter depends strictly on kit used. In general, quantity can be expressed in tissue slide dimensions (area approximately 150-200 mm^2 and thickness about 5-50 µm) or weight (range 5-50 mg). After every DNA extraction it is necessary to determine DNA concentration (fluorescence is the method chosen in this project).

Elution volume: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

This characteristic influences the nucleic acid concentration obtained from extraction kits. The elution volume has to be lower than 50 µl, preferably up to 30 µl.

Cross-linking correction: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

Formalin fixation leads to chemical cross-links to proteins, RNA, and other DNA molecules, with concomitant fragmentation of DNA, which inhibits molecular analysis in a length-dependent fashion. In general, molecular analyses of formalin-fixed DNA that require DNA segments shorter than 300 bp are usually successful, while those requiring a length between 300 and 1,000 bp succeed inconsistently, and those requiring a length of greater than 1,000 bp are often unsuccessful [67]. This step is necessary to remove the crosslinks (chemical modification caused by formalin fixation) from the DNA.

Cytosine deamination repair: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

Cytosine deamination has been demonstrated to contribute to background noise for DNA sequencing of ancient and formalin fixed paraffin embedded (FFPE)-treated DNA [68]. It can manifest as base substitution of C to T (C>T), G>A mutations or, specifically in FFPE, cytosine into uracil due to formalin treatment. It is demonstrated that uracil N-glycosylase (UNG) can specifically remove artificially-induced uracils from the DNA obtained from the FFPE sample and reduce background noise for DNA sequencing [69]. Among all analyzed kits, only one contains UNG treatment, GeneRead DNA FFPE Kit.

NGS instrument compatibility: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

Three kits were excluded because both user manual and manufacturer does not warrant the use of nucleic acid extract in downstream NGS assays or Illumina NGS technology. In particular, Thermo Scientific Ion AmpliSeq Direct FFPE DNA Kit, which contains reagents involved not only in genomic nucleic acid extraction but also in initial step of library preparation for ION TORRENT platform (with a chemistry that differ drastically from ILLUMNIA technology).

Extraction technology: This requirement was applied on nucleic extraction kits suitable for liquid handler automation. Magnetic beads extraction kits are the best choice for automation use because they require more simple and liquid handler compatible devices (for example magnetic and thermo-shaking instruments). Another important aspect that reinforces the choice of magnetic beads extraction kits is the problem of the clogging of the column, which results to be difficult to manage in automation.

Deparaffinization methods: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

The selected kits employ different processes for removing paraffin from FFPE tumor tissues. Actually, it is not possible to determine which is the better technique based only on user manual and literature. The selection based only on technical documentation is not enough, it is necessary to proceed with experimental measurements.

Deparaffinization with ultrasound technology: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

Two kits were excluded because they carry out deparaffinization by ultrasound. Ultrasound technology for nucleic acid extraction, since the required instrument is expensive, would be affordable only if the same technology had to be chosen also for library preparation due to distinctive performance over library preparation protocols not requiring the ultrasound DNA fragmentation step. However, ultrasound DNA fragmentation is usually employed in whole genome sequencing and is not usually required for targeted sequencing based on amplicons. Furthermore, using FFPE samples, DNA is already degraded, so that DNA fragmentation is much less useful For these reasons, kits based on ultrasound technology for deparaffinization, and the related instrumentation, were excluded.

Deparaffinization without centrifugation: This requirement was applied on nucleic extraction kits suitable for liquid handler automation.

This aspect is important from the point of view of commercial exploitation because in a full walk away system, the deparaffinization phase should be included. The advantage of having a protocol that does not employ the centrifuge in the system would make the process simpler and less expensive. Very particular centrifuges are used in deparaffinization and these instruments are very expensive. Processes not requiring centrifugation would give a strong competitive advantage, therefore these kits strongly favored.

Instrument restrictions: This requirement was applied on nucleic acid extraction kits for benchtop instruments. Two kits were excluded because their use is restricted only to dedicated instrumentation and technical support is difficult. These aspects are important parameters for studying diagnostic kits because technical and general supports are required in order to solve every aspect in case of problems.

Compatibility with "open" automated instruments: This requirement was applied on nucleic extraction kits suitable for liquid handler automation. Automated platforms, such as liquid handler, would be needed in diagnostic laboratory with high throughput sample analysis. The machines need to be programmed for integrating protocols or aspects of specific and optimized protocols, for this reason, open instruments are required.

CE-IVD mark: This requirement was applied on manual nucleic acid extraction kits (column based and magnetic beads based kit), nucleic acid extraction kits for benchtop instruments and nucleic extraction kits suitable for liquid handler automation.

The IVD (*in vitro* diagnostics) devices are regulated by EU Directive 98/79/EC. *In vitro* diagnostic medical device means any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, equipment, or system, whether used alone or in combination, intended by the manufacturer to be used in vitro for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information:

- concerning a physiological or pathological state, or
- concerning a congenital abnormality, or
- to determine the safety and compatibility with potential recipients, or
- to monitor therapeutic measures [70].

In preparatory studies, nucleic acid extraction kits with CE-IVD mark are considered for their integration in a potential marketable prototype biomarker kit for predicting HGS-OvCa patients response to combinatorial therapeutic modalities. When a kit is CE-IVD marked, it means it has features that guarantee stable and well-defined performances for diagnostic application.

The use of CE-IVD marked kit in a diagnostic/prognostic/therapy predictive workflow is not essential if the manufacturer, in this case DOTT.DINO PALADIN, validates the whole workflow process

certificating the reproducibility, the sensibility, the specificity and the other characteristics required for the CE-IVD labeling.

Other aspects were taken in consideration for this nucleic acid extraction kit screening that are based on the opportunity of choosing a specific supplier. In particular, the quality of supplier collaboration and the acceptance of OEM contractual terms are critical aspects to consider in the supply chain for production of CE IVD marked products (such as acceptable minimum quantity, reserved batch, bulk reagents, additional supporting product conformity information and testing, traceability, product end-of-life management, acceptance of auditing from our side or from Notified Bodies).

Finally, the screening phase led to the selection of 10 kits that are reported in **Table 7**.

Product Name	Company
blackPREP FFPE DNA Kit	Analytik-jena
FFPE Tissue DNA Extraction Kit - Columns	BioChain
NucleoSpin DNA FFPE XS	Macherey-Nagel
HighPrep FFPE Tissue DNA Kit	MagBio Genomics
E.Z.N.A. FFPE DNA Kit	OMEGA bio-tek
Mag-Bind FFPE DNA	OMEGA bio-tek
Promega Maxwell RSC DNA FFPE Kit	Promega
ReliaPrep FFPE gDNA Miniprep System	Promega
GeneRead DNA FFPE Kit	Qiagen
AllPrep DNA/RNA FFPE Kit	Qiagen

Table 7: Nucleic acid extraction kits obtained from initial screening. Commercial names and Company brands are reported.

These kits will be tested to determine their analytical performances for the following Illumina based NGS application.

5.4 NUCLEIC ACID EXTRACTION KITS PERFORMANCES ANALYSIS

Nucleic acid quantification analysis were made for both DNA and RNA biomolecules.

5.4.1 DNA QUANTIFICATION TESTS

DNA molecules are the most important targets for oncology application because it is possible to detect and identify the different mutations that characterize the various tumor subtypes summarized in several diagnostic, prognostic and therapeutic guidelines.

The testing results obtained for the protocols that passed the initial screening are reported in the following sections and the results obtained from each nucleic acid characterization are preceded by a short summary of features specific of each protocol.

5.4.1.1 ANALYTIK JENA - BLACKPREP FFPE DNA KIT

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The processes of deparaffinization, lysis and partial reversal of modifications of nucleic acid due to fixation are all conducted in the same buffer at 65°C, 90°C, and RT respectively under continuous shaking;
- The protocol requires max 50 mg of FFPE section as input material.

The concentrations and the yields of DNA obtained with Analytik Jena blackPREP FFPE DNA Kit are reported in **Table 8**. Concentrations obtained with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		Yield (ng)
			Qubit	Nanodrop	
FR1[48]_D[AJ-BP]	1	30	0.67	4.1	20.04
FR1[58]_D[AJ-BP]	1	30	0.77	4.2	23.22
FR1[25,67]_D[AJ-BP]	2	30	1.61	4.8	48.3
FR1[39,49]_D[AJ-BP]	2	30	1.36	5.5	40.8

Table 8: Concentration of DNA extracted with Analytik Jena blackPREP FFPE DNA Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 17** (samples from the FR1 series, 1 and 2 FFPE sections).

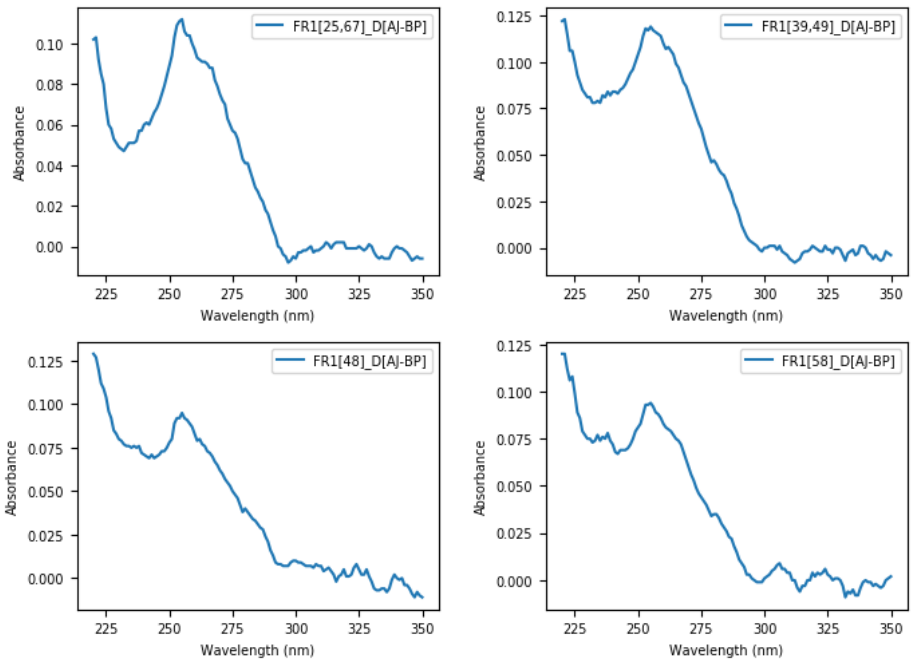


Figure 17: Absorbance plots of DNA extracted with Analytik Jena blackPREP FFPE DNA Kit, 1 and 2 FFPE sections input, elution in 30 µl.

Absorbance ratios are shown in **Table 9**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[48]_D[AJ-BP]	1	30	2.06	1.03
FR1[58]_D[AJ-BP]	1	30	2.36	1.11
FR1[25,67]_D[AJ-BP]	2	30	2.34	1.96
FR1[39,49]_D[AJ-BP]	2	30	2.36	1.36

Table 9: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Analytik Jena blackPREP FFPE DNA Kit

5.4.1.2 BIOCHAIN - FFPE TISSUE DNA EXTRACTION KIT - COLUMN

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The paraffin removal process involves a proprietary deparaffinization solution that operates at 90°C;
- The protocol requires 4-8 FFPE sections as input material with a 5-10 µm thickness.

The concentrations and the yields of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column are shown in **Table 10**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)					Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	TapeStation gDNA	
FR1[15]_D[BC-FC]	1	50	4.8	16.4	0.651	2.98	10.9	238
FR1[34]_D[BC-FC]	1	50	4.7	16.0	0.46	2.96	10.7	233
FR1[1,11]_D[BC-FC]	2	50	14.3	34.7	4.96	12	28.9	715
FR1[2,69]_D[BC-FC]	2	50	10.5	22.3	0.794	4.97	23.6	525

Table 10: Concentration of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 18** (samples from the FR1 series, 1 and 2 FFPE sections).

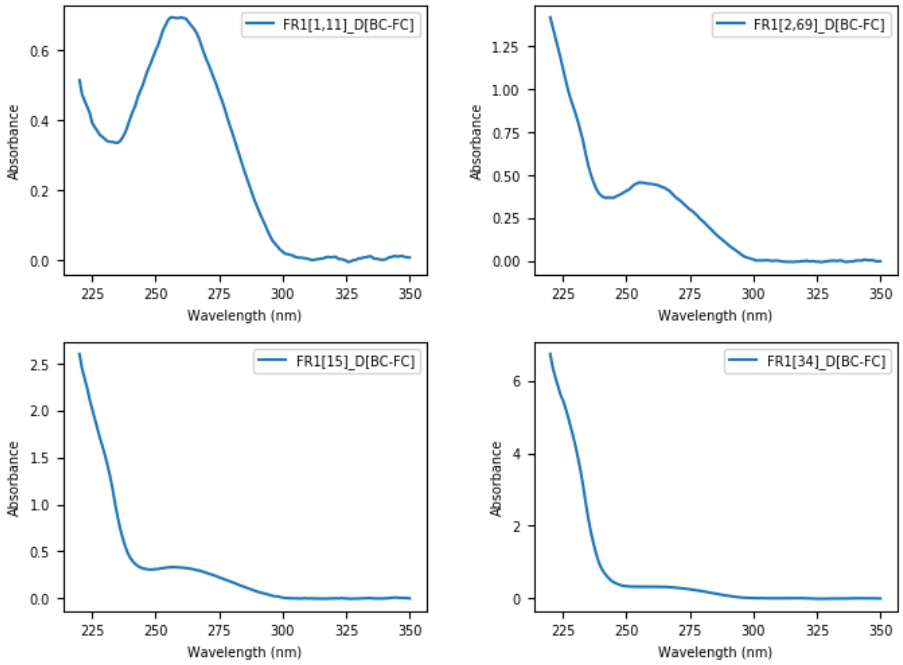


Figure 18: Absorbance plots of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column, 1 and 2 FFPE sections input, elution in 50 µl.

Absorbance ratios are shown in **Table 11**. A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit – Column.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[15]_D[BC-FC]	1	50	1.87	0.21
FR1[34]_D[BC-FC]	1	50	1.64	0.08
FR1[1,11]_D[BC-FC]	2	50	1.89	2.01
FR1[2,69]_D[BC-FC]	2	50	1.93	0.52

Table 11: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column

EPGs are shown in **Figure 19**.

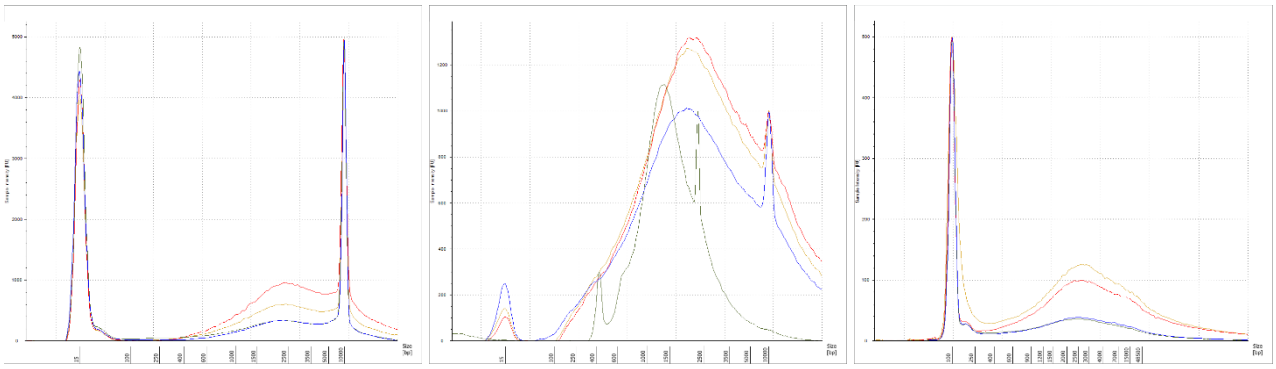


Figure 19: EPGs of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column, 1 and 2 FFPE sections input, elution in 50 µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[15]_D[BC-FC] (blu), FR1[34]_D[BC-FC] (green), FR1[1,11]_D[BC-FC] (red), FR1[2,69]_D[BC-FC] (brown).
Center: HS D5000 ScreenTape Assay, samples FR1[15]_D[BC-FC] (blu), FR1[34]_D[BC-FC] (green), FR1[1,11]_D[BC-FC] (red), FR1[2,69]_D[BC-FC] (brown).
Right: gDNA ScreenTapeAssay, samples FR1[15]_D[BC-FC] (blu), FR1[34]_D[BC-FC] (green), FR1[1,11]_D[BC-FC] (red), FR1[2,69]_D[BC-FC] (brown).

5.4.1.3 MACHEREY-NAGEL - NUCLEOSPIN DNA FFPE XS

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The paraffin removal process involves a proprietary deparaffinization solution that operates at 60°C;
- The processes of partial reversal of modifications of nucleic acid due to fixation is conducted at 90°C;
- The protocol requires up to 5 mg tissue, up to 15 mg paraffin (about 7 sections 10 µm x 250 mm2).

The concentrations and the yields of DNA obtained with Macherey-Nagel NucleoSpin DNA FFPE XS are reported in **Table 12**. Concentration obtained with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)				Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	
FR1[50]_D[MN-NSX]	1	20	0.86	7.5	NA	0.339	6.00
FR1[37]_D[MN-NSX]	1	20	0.30	5.6	NA	0.193	5.72
FR1[36,20]_D[MN-NSX]	2	20	0.19	4.6	NA	0.214	17.2
FR1[19,26]_D[MN-NSX]	2	20	0.30	3.8	0.177	0.090	3.84

Table 12: Concentration of DNA extracted with Macherey-Nagel NucleoSpin DNA FFPE XS.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 20** (samples from the FR1 series, 1 and 2 FFPE sections).

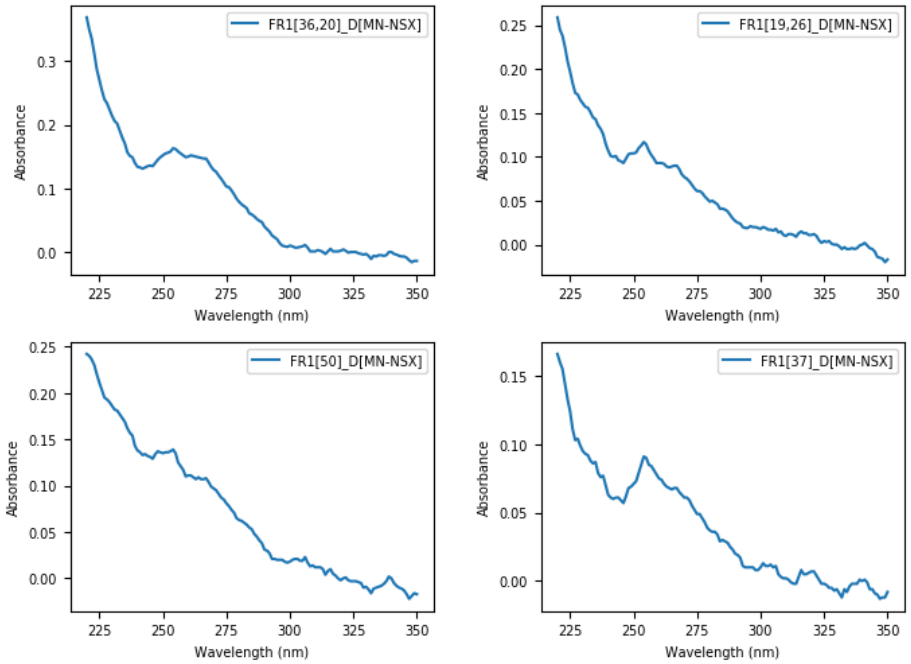


Figure 20: Absorbance plots of DNA extracted with Macherey-Nagel NucleoSpin DNA FFPE XS, 1 and 2 FFPE sections input, elution in 20 µl.

Absorbance ratios are shown in **Table 13**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[50]_D[MN-NSX]	1	20	1.76	0.60
FR1[37]_D[MN-NSX]	1	20	2.04	0.80
FR1[36,20]_D[MN-NSX]	2	20	1.90	0.70
FR1[19,26]_D[MN-NSX]	2	20	1.89	0.58

Table 13: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Macherey-Nagel NucleoSpin DNA FFPE XS

EPGs are shown in **Figure 21**.

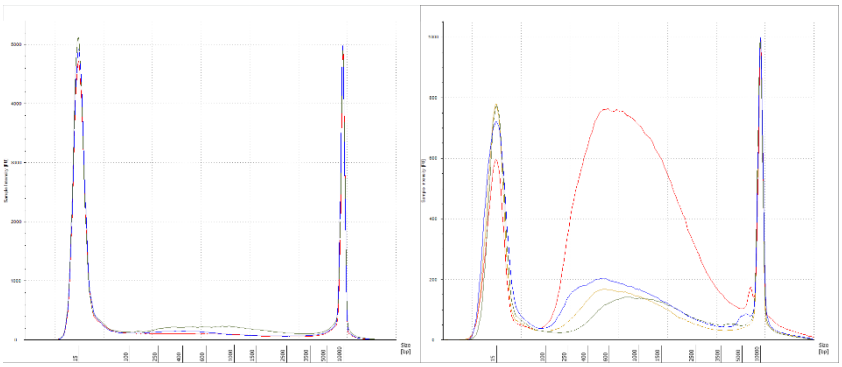


Figure 21: EPGs of DNA extracted with Macherey-Nagel NucleoSpin DNA FFPE XS, 1 and 2 FFPE sections input, elution in 20 µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[37]_D[MN-NSX] (blu), FR1[50]_D[MN-NSX] (green), FR1[36,20]_D[MN-NSX] (red), FR1[19,26]_D[MN-NSX] (brown)
Right: HS D5000 ScreenTape Assay, samples FR1[37]_D[MN-NSX] (blu), FR1[50]_D[MN-NSX] (green), FR1[36,20]_D[MN-NSX] (red), FR1[19,26]_D[MN-NSX] (brown)

5.4.1.4 MAGBIO GENOMICS - HIGHPREP FFPE TISSUE DNA KIT

The main features of this protocol are:

- The kit uses a magnetic beads based extraction method;
- The paraffin removal process involves a proprietary deparaffinization solution that operates at 65°C;
- The protocol requires 2-3 FFPE sections as input material with a 5-10 µm thickness.

The concentrations and the yields of DNA extracted with Magbio Genomics HighPrep FFPE Tissue DNA Kit are shown in **Table 14**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)				Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	
FR1[23]_D[MB-HP]	1	30	3.38	100.5	NA	0.045	101.4
FR1[57]_D[MB-HP]	1	30	4.44	109.8	0.121	NA	133.2
FR1[10,42]_D[MB-HP]	2	30	5.96	134.4	NA	0.031	178.8
FR1[47,66]_D[MB-HP]	2	30	9.70	185.5	NA	0.084	291.0

Table 14: Concentration of DNA extracted with Magbio Genomics HighPrep FFPE Tissue DNA Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 22** (samples from the FR1 series, 1 and 2 FFPE sections)

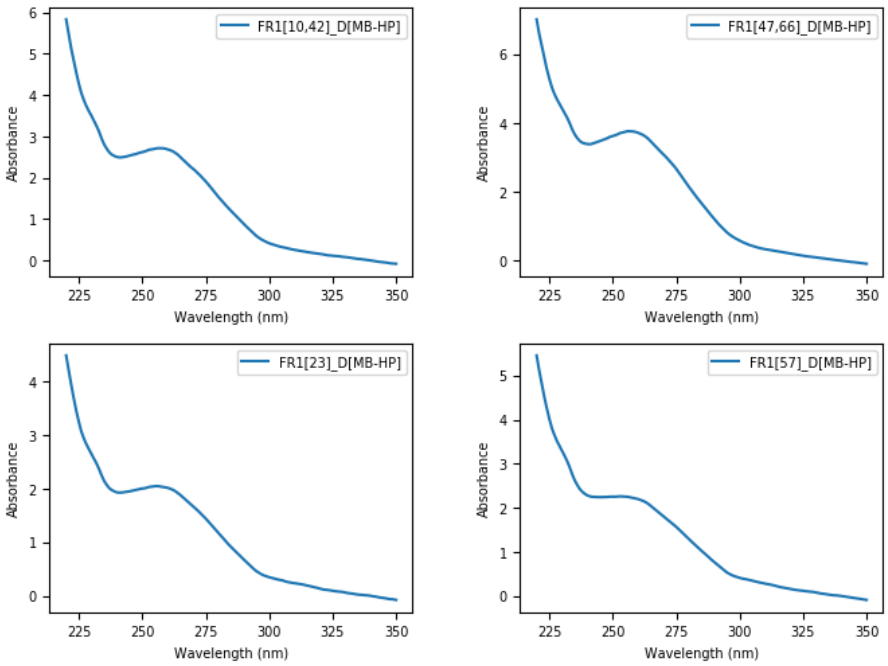


Figure 22: Absorbance plots of DNA extracted with Magbio Genomics HighPrep FFPE Tissue DNA Kit, 1 and 2 FFPE sections input, elution in 30 µl.

Absorbance ratios are shown in **Table 15**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[23]_D[MB-HP]	1	30	1.72	0.76
FR1[57]_D[MB-HP]	1	30	1.70	0.67
FR1[10,42]_D[MB-HP]	2	30	1.75	0.77
FR1[47,66]_D[MB-HP]	2	30	1.73	0.84

Table 15: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Magbio Genomics HighPrep FFPE Tissue DNA Kit

EPGs are shown in **Figure 23**.

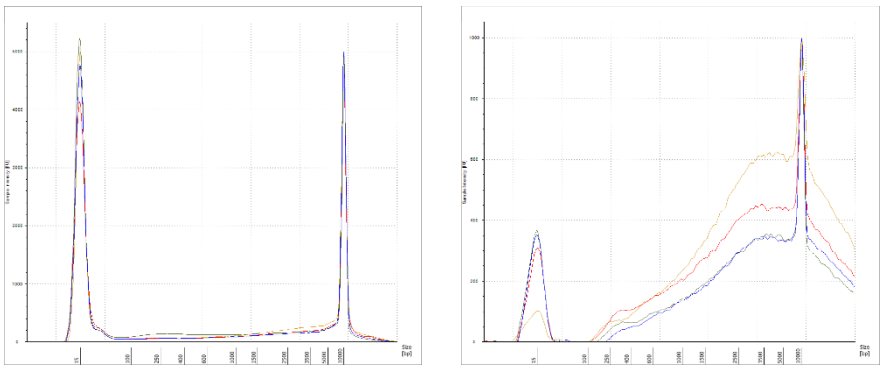


Figure 23: EPGs of DNA extracted with Magbio Genomics HighPrep FFPE Tissue DNA Kit from samples in the FR1 series, 1 and 2 FFPE sections input, elution in 30 µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[23]_D[MB-HP] (blu), FR1[57]_D[MB-HP] (green), FR1[10,42]_D[MB-HP] (red), FR1[47,66]_D[MB-HP] (brown)
Right: HS D5000 ScreenTape Assay, samples FR1[23]_D[MB-HP] (blu), FR1[57]_D[MB-HP] (green), FR1[10,42]_D[MB-HP] (red), FR1[47,66]_D[MB-HP] (brown)

5.4.1.5 OMEGA BIO-TEK - E.Z.N.A. FFPE DNA KIT

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The paraffin removal process involves deparaffinization by heating to 90°C;
- The protocol requires 1-2 FFPE sections as input material with a 5-10 µm thickness.

The concentrations and the yields of DNA obtained with OMEGA bio-tek E.Z.N.A. FFPE DNA Kit are reported in **Table 16**. Concentration obtained with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)				Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	
FR1[35]_D[OB-E]	1	50	0.37	1.5	0.077	0.138	18.4
FR1[70]_D[OB-E]	1	50	0.71	3.1	NA	0.237	35.7
FR1[21,51]_D[OB-E]	2	50	0.93	8.8	0.073	NA	46.3
FR1[22,52]_D[OB-E]	2	50	0.77	2.9	NA	NA	38.7

Table 16: Concentration of DNA extracted with OMEGA bio-tek E.Z.N.A. FFPE DNA Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 24** (samples from the FR1 series, 1 and 2 FFPE sections).

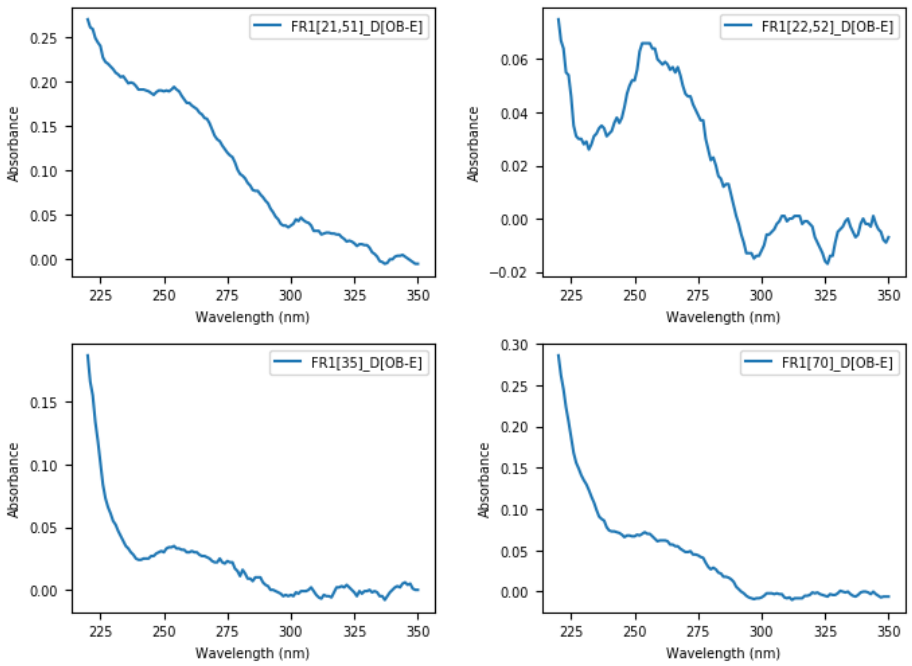


Figure 24: Absorbance plots of DNA extracted with OMEGA bio-tek E.Z.N.A. FFPE DNA Kit, 1 and 2 FFPE sections input, elution in 50 µl.

Absorbance ratios are shown in **Table 17**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[35]_D[OB-E]	1	50	2.73	0.55
FR1[70]_D[OB-E]	1	50	2.31	0.46
FR1[21,51]_D[OB-E]	2	50	1.83	0.82
FR1[22,52]_D[OB-E]	2	50	2.62	2.10

Table 17: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with OMEGA bio-tek E.Z.N.A. FFPE DNA Kit

EPGs are shown in **Figure 25**.

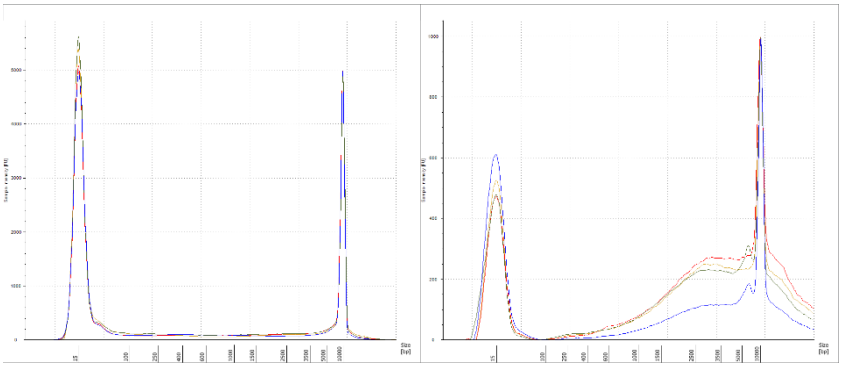


Figure 25: EPGs of DNA extracted with OMEGA bio-tek E.Z.N.A. FFPE DNA Kit, 1 and 2 FFPE sections input, elution in 50 µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[35]_D[OB-E] (blu), FR1[70]_D[OB-E] (green), FR1[21,51]_D[OB-E] (red), FR1[22,52]_D[OB-E] (brown).
Right: HS D5000 ScreenTape Assay, samples FR1[35]_D[OB-E] (blu), FR1[70]_D[OB-E] (green), FR1[21,51]_D[OB-E] (red), FR1[22,52]_D[OB-E] (brown).

5.4.1.6 OMEGA BIO-TEK - MAG-BIND FFPE DNA KIT

The main features of this protocol are:

- The kit uses a magnetic beads based extraction method;
- The paraffin removal process involves a proprietary deparaffinization solution that operates at 80°C;
- The protocol requires 2-3 FFPE sections as input material with a 5-10 μm thickness.

The concentrations and the yields of DNA extracted with OMEGA bio-tek Mag-Bind FFPE DNA Kit are shown in **Table 18**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (μl)	Concentration (ng/μl)				Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	
FR1[33]_D[OB-MB]	1	30	2.92	21.8	NA	NA	87.6
FR1[28]_D[OB-MB]	1	30	2.38	28.4	0.097	0.227	71.4
FR1[14,45]_D[OB-MB]	2	30	5.22	26.0	NA	0.595	156.6
FR1[9,60]_D[OB-MB]	2	30	9.2	36.7	NA	0.048	276

Table 18: Concentration of DNA extracted with OMEGA bio-tek Mag-Bind FFPE DNA Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 26** (samples from the FR1 series, 1 and 2 FFPE sections).

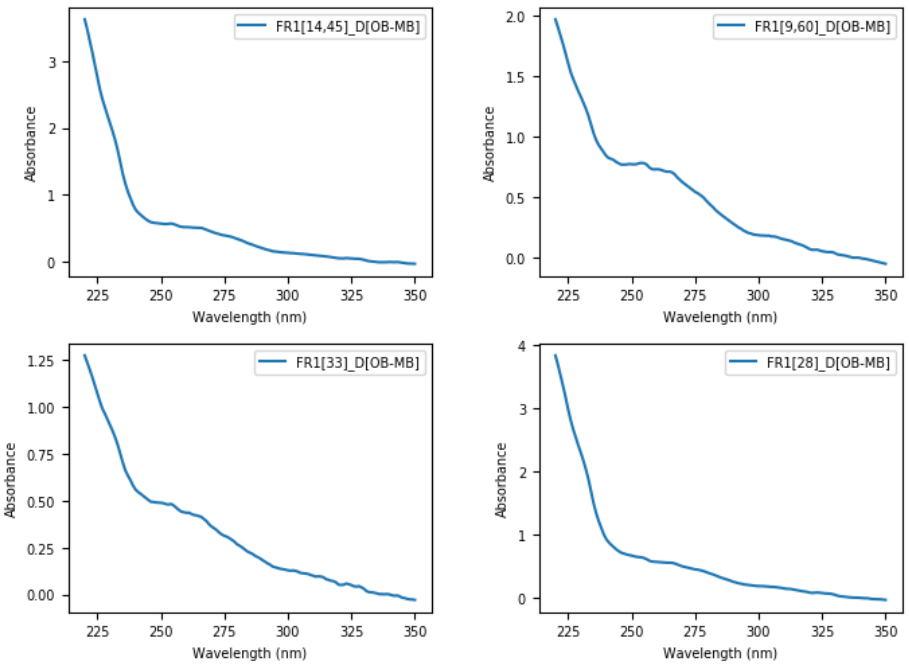


Figure 26: Absorbance plots of DNA extracted with OMEGA bio-tek Mag-Bind FFPE DNA Kit, 1 and 2 FFPE sections input, elution in 30 µl.

Absorbance ratios are shown in **Table 19**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[33]_D[OB-MB]	1	30	1.63	0.48
FR1[28]_D[OB-MB]	1	30	1.43	0.25
FR1[14,45]_D[OB-MB]	2	30	1.51	0.25
FR1[9,60]_D[OB-MB]	2	30	1.59	0.55

Table 19: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with OMEGA bio-tek Mag-Bind FFPE DNA Kit

EPGs are shown in **Figure 27**.

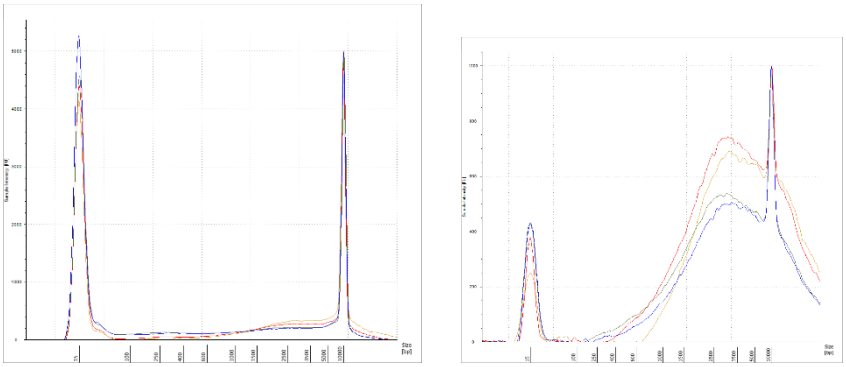


Figure 27: EPGs of DNA extracted with OMEGA bio-tek Mag-Bind FFPE DNA Kit, 1 and 2 FFPE sections input, elution in 30 µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[33]_D[OB-MB] (blu), FR1[28]_D[OB-MB] (green), FR1[14,45]_D[OB-MB] (red), FR1[9,60]_D[OB-MB] (brown)
Right: HS D5000 ScreenTape Assay, samples FR1[33]_D[OB-MB] (blu), FR1[28]_D[OB-MB] (green), FR1[14,45]_D[OB-MB] (red), FR1[9,60]_D[OB-MB] (brown)

5.4.1.7 PROMEGA - MAXWELL RSC DNA FFPE KIT

The main features of this protocol are:

- The kit uses magnetic beads based extraction method;
- The paraffin removal process involves mineral oil that operates at 80°C;
- The kit can be used only with Maxwell RSC Instrument;
- The protocol requires 5–10µm FFPE sections as input material with a size range of 20mm^2 to 200mm^2 for a total of up to 2.0mm^3.

The concentrations and the yields of DNA obtained with Promega Maxwell RSC DNA FFPE Kit are reported in **Table 20**. Concentrations obtained with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		Yield (ng)
			Qubit	Nanodrop	
FR2[13]_D[P-MR]	1	50	10.2	40.4	510
FR2[5]_D[P-MR]	1	50	7.78	56.2	389
FR2[9]_D[P-MR]	1	50	4.58	25.4	229
FR2[16]_D[P-MR]	1	50	5.54	30.2	277

Table 20: Concentration of DNA extracted with Promega Maxwell RSC DNA FFPE Kit.

Absorbance ratios are shown in **Table 21**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR2[13]_D[P-MR]	1	50	1.82	1.45
FR2[5]_D[P-MR]	1	50	1.78	1.5
FR2[9]_D[P-MR]	1	50	1.7	1.02
FR2[16]_D[P-MR]	1	50	1.69	0.98

Table 21: A260/A280 and A260/A230 absorbance ratios from of DNA extracted with Promega Maxwell RSC DNA FFPE Kit.

5.4.1.8 PROMEGA - RELIAPREP FFPE GDNA MINIPREP SYSTEM

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The paraffin removal process involves mineral oil that operates at 80°C;
- The protocol requires 5–50µm FFPE sections as input material, total equivalent of ≤100 µm.

The concentrations and the yields of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System are shown in **Table 22**. Concentration obtained with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method

using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)					Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	TapeStation gDNA	
FR1[4]_D[P-RP]	1	30	9.0	28.3	9.38	8.73	17.2	270
FR1[5]_D[P-RP]	1	30	5.7	15.2	0.877	10.3	9.39	169.8
FR1[12]_D[P-RP]	1	30	6.7	19.0	0.519	2.63	8.90	200.4
FR1[13]_D[P-RP]	1	30	6.2	14.9	0.484	3.47	11.5	186.6
M[31317]_D[P-RP]	5	30	97.2	985.9	NA	NA	NA	2,916
M[37229]_D[P-RP]	5	30	106.0	772.9	NA	NA	NA	3,180
M[13218]_D[P-RP]	5	30	99.4	805.6	NA	NA	NA	2,982
M[28108]_D[P-RP]	5	30	100.0	636.8	NA	NA	NA	3,000
M[20770]_D[P-RP]	5	30	13.0	107.0	NA	NA	NA	390
M[34295]_D[P-RP]	5	30	102.0	905.5	NA	NA	NA	3,060
M[32172]_D[Q-GR]	5	30	67.4	475.1	NA	NA	NA	2,022

Table 22: Concentration of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 28** (samples from the FR1 series, 1 FFPE section) and **Figure 29** (samples from the M series, 5 FFPE sections).

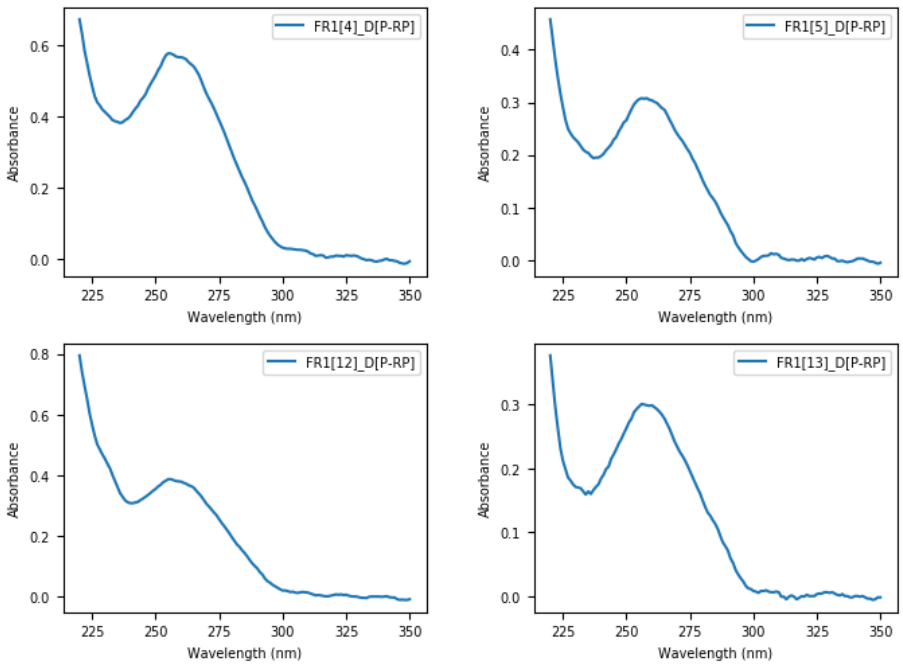


Figure 28: Absorbance plots of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System, 1 FFPE section input, elution in 30 µl.

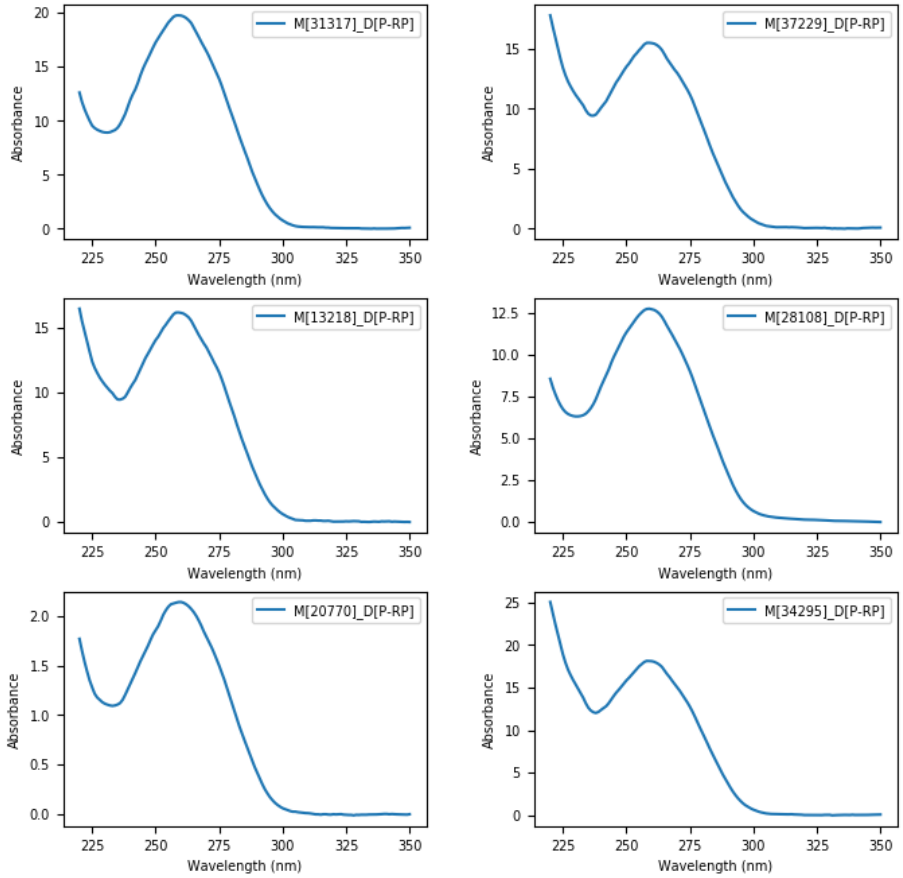


Figure 29: Absorbance plots of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System, 5 FFPE sections input, elution in 30 µl.

Absorbance ratios are shown in **Table 23**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[4]_D[P-RP]	1	30	1.89	1.37
FR1[5]_D[P-RP]	1	30	2.00	1.32
FR1[12]_D[P-RP]	1	30	1.93	0.83
FR1[13]_D[P-RP]	1	30	2.00	1.74
M[31317]_D[P-RP]	5	30	1.87	2.21
M[37229]_D[P-RP]	5	30	1.81	1.39
M[13218]_D[P-RP]	5	30	1.85	1.53
M[28108]_D[P-RP]	5	30	1.85	2.02
M[20770]_D[P-RP]	5	30	1.92	1.92
M[34295]_D[P-RP]	5	30	1.88	1.18

Table 23: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System

EPGs are shown in **Figure 30**.

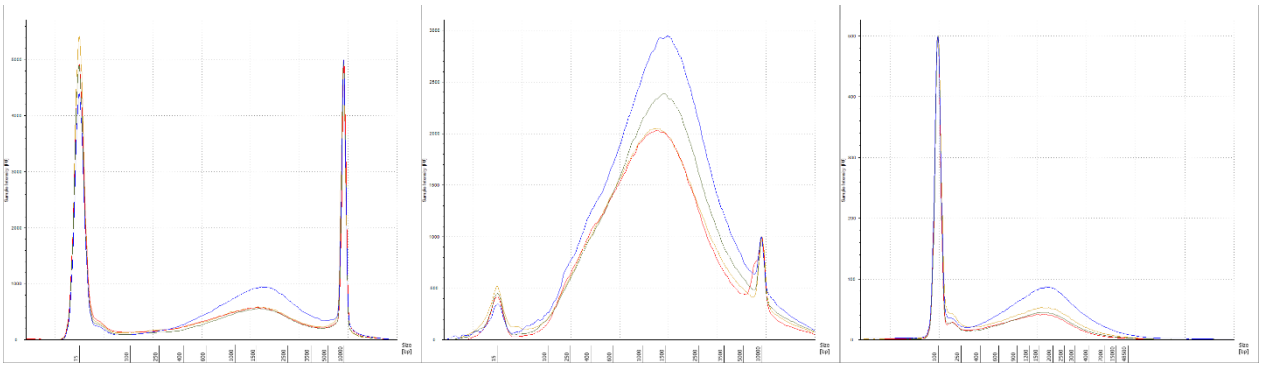


Figure 30: EPGs of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System, 1 FFPE section input, elution in 30µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[4]_D[P-RP] (blu), FR1[5]_D[P-RP] (green), FR1[12]_D[P-RP] (red), FR1[13]_D[P-RP] (brown)
Center: HS D5000 ScreenTape Assay, samples FR1[4]_D[P-RP] (blu), FR1[5]_D[P-RP] (green), FR1[12]_D[P-RP] (red), FR1[13]_D[P-RP] (brown)
Right: gDNA ScreenTapeAssay, samples FR1[4]_D[P-RP] (blu), FR1[5]_D[P-RP] (green), FR1[12]_D[P-RP] (red), FR1[13]_D[P-RP] (brown)

5.4.1.9 QIAGEN - ALLPREP DNA/RNA FFPE KIT

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- Simultaneous purification of genomic DNA and total RNA;
- DNA and RNA are released sequentially by differential solubilization at 90°C and 80°C respectively of the same FFPE sample;
- The paraffin removal process involves a proprietary deparaffinization solution that operates at 56°C;

- The protocol requires 1-4 FFPE sections input with 10 µm thickness and 150 mm^2 surface area or 1-2 FFPE sections input with 20 µm sections of 150 mm^2 surface area.

The concentrations and the yields of DNA obtained with Qiagen GeneRead DNA FFPE Kit are reported in **Table 24**. Concentrations measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		Yield (ng)
			Qubit	Nanodrop	
FR2[10]_D[Q-AP]	1	30	84	23.3	276.6
FR2[3]_D[Q-AP]	1	30	21.8	23.6	265.2
FR2[2,19]_D[Q-AP]	2	30	9.22	171.6	2520
FR2[8,11]_D[Q-AP]	2	30	8.84	48.4	654

Table 24: Concentration of DNA extracted with Qiagen AllPrep DNA/RNA FFPE Kit.

Absorbance ratios are shown in **Table 25**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR2[10]_D[Q-AP]	1	30	1.91	3.92
FR2[3]_D[Q-AP]	1	30	1.91	1.55
FR2[2,19]_D[Q-AP]	2	30	1.83	2.71
FR2[8,11]_D[Q-AP]	2	30	1.9	2.85

Table 25: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Qiagen AllPrep DNA/RNA FFPE Kit.

5.4.1.10 QIAGEN - GENEREAD DNA FFPE KIT

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The paraffin removal process involves a proprietary deparaffinization solution that operates at 56°C;
- The processes of partial reversal of modifications of nucleic acid due to fixation is conducted at 90°C;
- The Removal of deaminated cytosine is operated by UNG at 56°C;
- The protocol requires 1 FFPE section input with a thickness up to 10 µm.

The concentrations and the yields of DNA extracted with Qiagen GeneRead DNA FFPE Kit are shown in **Table 26**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and Agilent 2200 TapeStation instruments are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method

using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)					Yield (ng)
			Qubit	Nanodrop	TapeStation D5000	TapeStation HS D5000	TapeStation gDNA	
FR1[3]_D[Q-GR]	1	20	10.2	25.0	4.2	8.9	20.3	204
FR1[18]_D[Q-GR]	1	20	13.5	33.5	1.5	19.2	26.2	270
FR1[27]_D[Q-GR]	1	20	15.2	28.1	6.6	12.4	27.3	304
FR1[30]_D[Q-GR]	1	20	18.0	43.2	7.3	18.2	32.8	360
M[36110]_D[Q-GR]	5	30	75.6	547.3	10.0	NA	272.0	2,268
M[05088]_D[Q-GR]	5	30	58.0	279.3	12.3	NA	88.1	1,740
M[17773]_D[Q-GR]	5	30	76.0	1,060.5	25.8	NA	236.0	2,280
M[30673]_D[Q-GR]	5	30	44.4	136.3	NA	NA	NA	1,332
M[15725]_D[Q-GR]	5	30	6.2	18.7	NA	NA	NA	187.2
M[31271]_D[Q-GR]	5	30	114.0	1,140.9	NA	NA	NA	3,420
M[32172]_D[Q-GR]	5	30	67.4	475.1	NA	NA	NA	2,022

Table 26: Concentration of DNA extracted with Qiagen GeneRead DNA FFPE Kit.

Absorption spectra obtained with the Nanodrop instrument are shown in **Figure 31** (samples from the FR1 series, 1 FFPE section), **Figure 32** and **Figure 33** (samples from the M series, 5 FFPE sections).

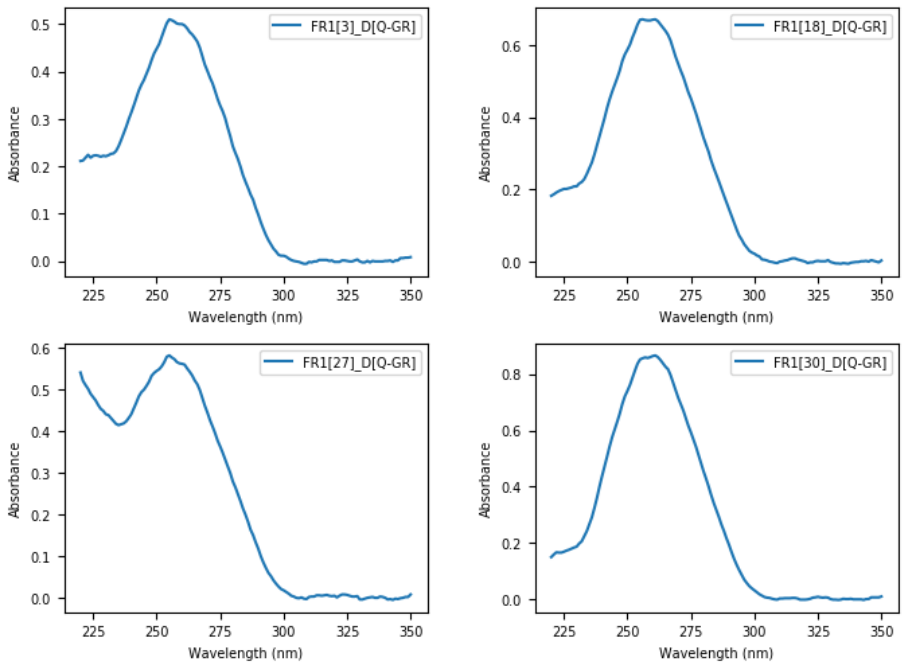


Figure 31. Absorbance plots of DNA extracted with Qiagen GeneRead DNA FFPE Kit, 1 FFPE section input, elution in 20 µl.

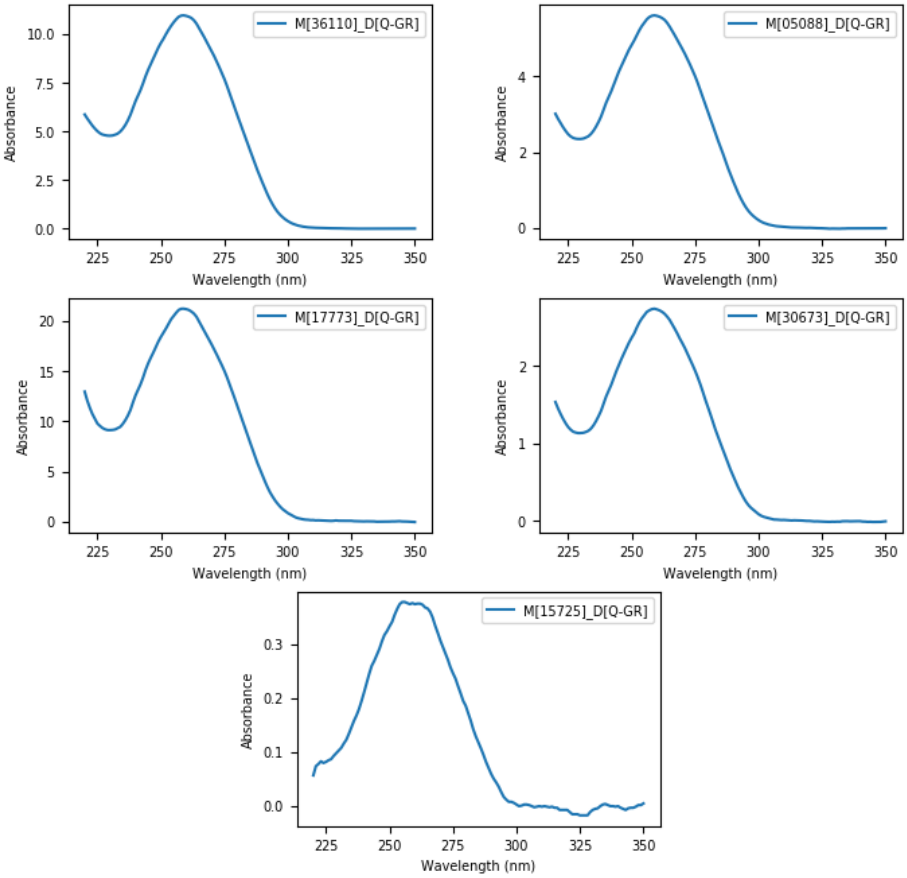


Figure 32: Absorbance plots of DNA extracted with Qiagen GeneRead DNA FFPE Kit, 5 FFPE sections input, elution in 30 µl.

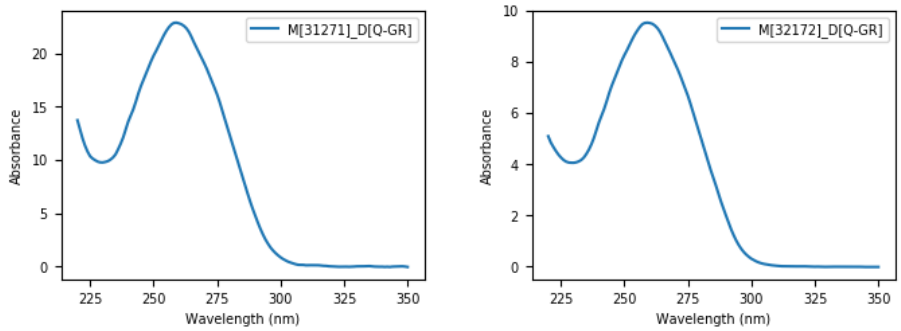


Figure 33: Absorbance plots of DNA extracted with Qiagen GeneRead DNA FFPE Kit, 5 FFPE sections input, elution in 30 µl.

Absorbance ratios are shown in **Table 27**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[3]_D[Q-GR]	1	20	2.07	2.26
FR1[18]_D[Q-GR]	1	20	1.96	3.21
FR1[27]_D[Q-GR]	1	20	2.01	1.27
FR1[30]_D[Q-GR]	1	20	1.90	4.61
M[36110]_D[Q-GR]	5	30	1.86	2.29
M[05088]_D[Q-GR]	5	30	1.82	2.38
M[17773]_D[Q-GR]	5	30	1.83	2.32
M[30673]_D[Q-GR]	5	30	1.85	2.40
M[15725]_D[Q-GR]	5	30	2.01	3.63
M[31271]_D[Q-GR]	5	30	1.85	2.33
M[32172]_D[Q-GR]	5	30	1.88	2.34

Table 27: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Qiagen GeneRead DNA FFPE Kit.

EPGs are shown in **Figure 34** and **Figure 35**. EPGs for R series with HS D5000 are not plotted since markers were not detected and consequently fragment size calibration was not performed by the TapeStation software.

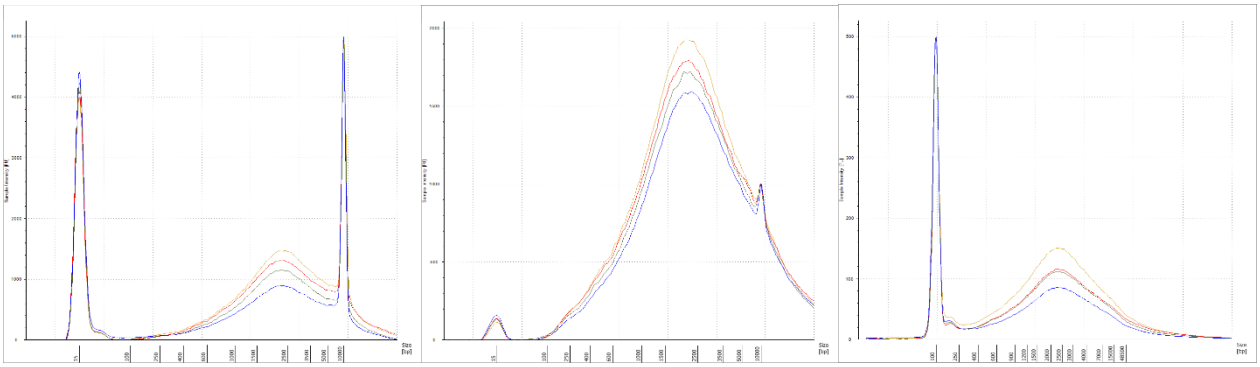


Figure 34: EPGs of DNA extracted with Qiagen GeneRead DNA FFPE Kit from samples in the FR1 series, 1 FFPE section input, elution in 20 µl. The narrow peaks are markers included in the assay reagents.
Left: D5000 ScreenTape Assay, samples FR1[3]_D[Q-GR] (blue), FR1[18]_D[Q-GR] (green), FR1[27]_D[Q-GR] (red), FR1[30]_D[Q-GR] (brown)
Center: HS D5000 ScreenTape Assay, samples FR1[3]_D[Q-GR] (blue), FR1[18]_D[Q-GR] (green), FR1[27]_D[Q-GR] (red), FR1[30]_D[Q-GR] (brown)

Right: gDNA ScreenTapeAssay, samples FR1[3]_D[Q-GR], FR1[18]_D[Q-GR] (green), FR1[27]_D[Q-GR] (red), FR1[30]_D[Q-GR] (brown)

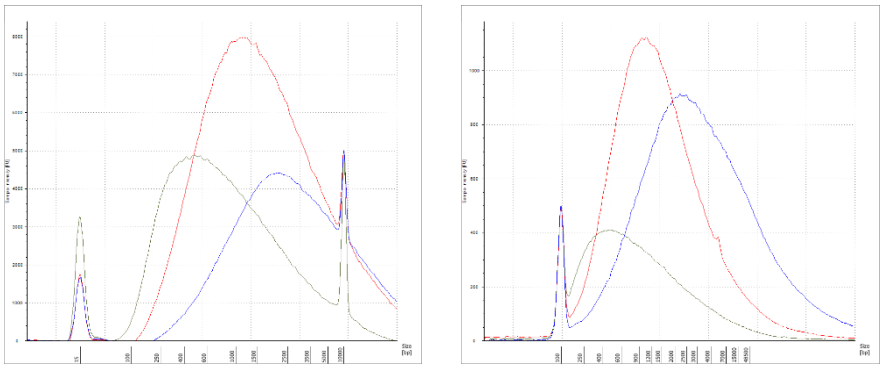


Figure 35: EPGs of DNA extracted with Qiagen GeneRead DNA FFPE Kit from samples in the M series, 5 FFPE section input, elution in 30 µl.
Left: D5000 ScreenTape Assay, M[36110]_D[Q-GR] (blu), M[05088]_D[Q-GR] (green), M[17773]_D[Q-GR] (red)
Right: gDNA ScreenTapeAssay, M[36110]_D[Q-GR] (blu), M[05088]_D[Q-GR] (green), M[17773]_D[Q-GR] (red)

5.4.1.11 PERFORMANCES OVERVIEW

In order to compare the performances obtained from the extraction assays, the concentrations and yields per section of DNA isolated by FFPE FR slices (all derived by the same non-tumoral human sample) are reported in **Table 28**.

Extraction kit	Conc./section (ng/µl)	Yield/section (ng)
Qiagen GeneRead DNA FFPE Kit	14.2 ± 2.8	285 ± 57
Qiagen AllPrep DNA/RNA FFPE Kit	9.0 ± 0.2	271 ± 6
Promega Maxwell RSC DNA FFPE Kit	7.0 ± 2.2	351 ± 108
Promega ReliaPre FFPE gDNA Miniprep System	6.9 ± 1.3	207 ± 38
BioChain FFPE Tissue DNA Extraction Kit - Column	5.5 ± 1.0	273 ± 50
MagBio Genomics HighPrep FFPE Tissue DNA Kit	3.9 ± 0.8	117 ± 23
OMEGA bio-tek Mag-Bind FFPE DNA	3.1 ± 0.9	94 ± 26
Analytik-jena blackPREP FFPE DNA Kit	0.73 ± 0.06	22.0 ± 1.8
OMEGA bio-tek E.Z.N.A. FFPE DNA Kit	0.48 ± 0.14	24.2 ± 6.9
Macherey-Nagel NucleoSpin DNA FFPE XS	0.28 ± 0.12	5.6 ± 2.4

Table 28: Concentrations and yields of DNA per FR sample section obtained from the tested nucleic acid extraction kits.

5.4.1.12 DISCUSSION

5.4.1.12.1 GENERAL CONSIDERATIONS

With the purpose of determining the performances of DNA extraction kits, different quantities of FFPE slices and elution volumes were used.

In detail, concerning the slices number employed in the DNA extraction tests, 1 to 2 slices for the FR series were used; 5 slices of M series samples have been employed only with Qiagen GeneRead DNA FFPE Kit and Promega ReliaPrep FFPE gDNA Miniprep System.

The reasons for using FR series samples are three:

- To study and compare the results obtained by different extraction kits in standard conditions, employing biological samples with the same characteristics from tissue matrix, nucleic acid quantity and FFPE preparation protocol.
- To verify the capability and sensitivity of the kits to extract DNA from a single FFPE slice, which may occur in real clinical routine practice;
- To evaluate the performances of the DNA isolation kits in presence of low input biological material conditions and in the user manuals recommended conditions.

Samples belonging to M series may represent samples having features similar to those observed in samples typically analyzed in diagnostic workflow in terms of DNA fragmentation derived from FFPE process, pathological characteristics and donor origin (human tumor mass).

Taking into account the elution volumes employed, the choice of microliters to use in the experiments (20 µl, 30 µl and 50 µl respectively) depends on the indications reported by each kit user manual. An important aspect of these nucleic acid isolation tests is to employ the lower elution volume possible in order to obtain the highest FFPE DNA concentration.

In fact, a critical aspect in the NGS library preparation is the DNA sample input concentration. Furthermore, it is necessary to have the highest FFPE DNA concentration in small elution volume, considering that the quality of DNA isolated from FFPE samples may be not good due to the fragmentations and the mutations derived from fixation processes.

5.4.1.12.2 DNA CONCENTRATION CONSIDERATIONS

The Qiagen GeneRead DNA FFPE Kit shows the highest DNA concentration value (approximately 14 ng/µl) It is also the only kit among those tested including a step for removal of deaminated cytosine. Other kits that performed markedly above the others are the Promega ReliaPrep FFPE gDNA Miniprep System (6.9 ng/µl) and the BioChain FFPE Tissue DNA Extraction Kit – Column (5.5 ng/µl).

The Qiagen AllPrep DNA/RNA FFPE Kit shows apparently good performances (9.0 ng/µl), furthermore it is the only kit able to isolate both DNA and RNA molecules. Since only a few data are available after outliers elimination, the kit requires further tests for statistic studies and considerations.

The magnetic beads based extraction kits used in these tests can be used for both manual and automatic applications based on the experimental protocol used for the extraction procedure.

The only exception is Promega Maxwell RSC DNA FFPE Kit because it requires a dedicated instrument.

Here, MagBio Genomics HighPrep FFPE Tissue DNA Kit, OMEGA bio-tek Mag-Bind FFPE DNA kit were tested with the manual protocol, instead Promega Maxwell RSC DNA FFPE Kit was tested Maxwell RSC Instrument.

The magnetic beads based protocols have worse performance than the best spin column based protocols in terms of nucleic acid concentration obtained, in fact the concentrations observed are approximately between 51 % (Promega Maxwell RSC DNA FFPE Kit, 7.0 ng/µl) and 78 % (OMEGA bio-tek Mag-Bind FFPE DNA, 3.1 ng/µl) lower than Qiagen GeneRead DNA FFPE Kit.

Among the magnetic beads based extraction kits, Promega Maxwell RSC DNA FFPE Kit shows the best performances (7.0 ng/µl), but, compared with the others which have both manual and automatic protocols, this kit can be used only with its own extractor Maxwell RSC Instrument.

The MagBio Genomics HighPrep FFPE Tissue DNA Kit performances are a little bit higher than the OMEGA bio-tek Mag-Bind FFPE DNA Kit (3.9 ng/µl and 3.1 ng/µl respectively). The advantage of MagBio Genomics HighPrep FFPE Tissue DNA Kit compared to the other magnetic beads based kit, is that it does not require the use of a centrifuge during the deparaffinization step. For this reason, MagBio Genomics HighPrep FFPE Tissue DNA Kit should be kept as a candidate for protocol porting to fully walk-away (i.e. including deparaffinization) robotic high throughput systems, where the presence of a centrifuge increases considerably product cost and system complexity for sample loading and retrieving.

According to its specifications and the results of the testing, the Qiagen GeneRead DNA FFPE Kit protocol is optimized for one single FFPE section or low input tissue amount.

In situations where multiple FFPE sections are available for DNA extraction, the Promega and Biochain have the potential to give a higher nucleic acid concentration by extracting from more tissue in the same elution volume.

In the tests, an indication that the Qiagen GeneRead DNA FFPE Kit protocol processing capability is saturated by higher amounts of tissue is the inability to obtain meaningful EPGs with the HS D5000 ScreenTape Assay on M series samples.

The DNA concentrations values obtained using ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are approximately from 2 to 30 times higher than the concentrations measured with ThermoFisher Scientific Qubit 3.0 fluorometer.

5.4.1.12.3 ABSORBANCE CONSIDERATIONS

The 260/280 absorbance ratios for all kits tested, excluding Qiagen GeneRead DNA FFPE Kit and Promega ReliaPrep FFPE gDNA Miniprep System, are affected by both low and high nucleic acid absorbance so that no conclusion can be drawn on sample purity.

Nevertheless during the DNA extraction phases, some traces of chemical contaminants (in particular for low 260/280 absorbance ratios) or residual RNA (for high 260/280 absorbance values) were isolated with DNA molecules.

These considerations are also confirmed by the absorption spectra obtained for each sample of the FR series, where the shapes and peaks at 260 nm deviate or, as in the cases of the OMEGA bio-tek Mag-Bind FFPE DNA Kit, OMEGA bio-tek E.Z.N.A. FFPE DNA Kit and Macherey-Nagel NucleoSpin DNA FFPE XS, are not very intense.

About Qiagen and the Promega protocols (and to a lesser extent for the Biochain protocol), the A260/A280 ratios are around the expected value for DNA (1.8-1.9) while the A260/A230 ratio reflects the presence of contaminants which are intrinsic of nucleic acid extraction protocols from FFPE.

The user manuals for commercial NGS Library Preparation kits generally recommend fluorescence-based quantification and advice against UV-spectrometer methods, which can overestimate DNA concentrations due to the presence of RNA and other contaminants (e.g. QIAseq Targeted DNA Panel Handbook, and Illumina TruSeq Custom Amplicon Low Input Kit Reference Guide).

Considering the scientific literature, we have found two somewhat conflicting evidences that describe the use of absorbance values for determining the DNA quantification for NGS library preparation.

Some authors demonstrated that Nanodrop readings generally are two to 10 times higher than the Qubit readings, attributed to residual RNA, denatured single-stranded DNA (ssDNA) and other contaminants in the sample, preferring Qubit (but Nanodrop is not excluded) because of very good correlation with ddPCR [71]. On the other hand, other authors found that in 50% of samples the highest reading are not from Nanodrop and extremely high differences are not observed, so fluorescence-based quantification methods, Nanodrop and qPCR are all suitable for NGS downstream applications [72].

Basing on literature and nucleic acid concentrations obtained with the absorbance and fluorescent measurements, high ratios of Nanodrop-to-Qubit estimations were found, for these reasons absorbance spectra are considered useful only as a rough visual qualitative assessment of sample purity and absence of major issues in the extraction process.

5.4.1.12.4 FRAGMENTS SIZE DISTRIBUTION CONSIDERATIONS

The EPGs measured using the D1000 assays are not shown since the only feature present, being limited to low fragment sizes, is the initial rise of the fragment size distribution, which is visible also using other ScreenTape assays. No further peaks or structures, specific of the low fragment sizing range of the D1000 assays can be seen in the D1000 EPGs.

Very large inaccuracies result also from ScreenTape assays quantifications, notably with large differences between estimations even from EPGs which look similar. The internals of the automatic analysis performed by the supplied software needs to be further elucidated in order to understand the reason of this behavior. However, the ScreenTape assays are useful because of the capability of detecting differences in fragment size distribution introduced by the different protocols. In fact, looking at the EPGs of the FR1 series, which are all taken from the same FFPE block, the center position of the broad fragment size distributions is relatively consistent among EPGs of DNA extracted with the same extraction protocol. On the other hand, EPGs measured using DNA extracted with different extraction protocols differ, ranging from around 600 bp for Macherey-Nagel NucleoSpin DNA FFPE XS to 4,000 bp for Magbio Genomics HighPrep FFPE Tissue DNA Kit. Correlation with NGS quality metrics and results will clarify if the TapeStation or similar instruments may have a role as a component of an IVD workflow.

The fragment size distribution of M series biological samples, performed with Agilent 2200 TapeStation system, were made only with Qiagen GeneRead DNA FFPE Kit and Promega ReliaPrep FFPE gDNA Miniprep System because these kits have shown the best performances compared to the other kits.

From the EPGs of M series samples shown in **Figure 35**, in detail focusing on the width of the peaks, it is possible to observe that the DNA fragmentation patterns are wide. This may depend on the nature of each sample, and the FFPE processes, mainly from the nucleic acid degradation and chemical alterations caused by formalin fixation step. Both High Sensitivity D5000 ScreenTape Assay and Genomic DNA ScreenTape Assay results (they differ among themselves for the fragment size ranges, 100 – 5,000 bp and 200 – 60,000 bp respectively) give approximately the same fragment distribution values, namely M[36110]_D[Q-GR] equal to 500 bp, M[05088]_D[Q-GR] equal to 1,100 bp and M[17773]_D[Q-GR] equal to 2,300 bp.

These observations confirm that there were no errors in the identification of fragment size distribution.

5.4.1.12.5 FINAL CONSIDERATIONS

Considering the results obtained from the extraction tests, the DNA isolation kits which will be used for the following NGS library preparation are shown in in **Table 29**.

Product Name	Company
GeneRead DNA FFPE Kit	Qiagen
ReliaPrep FFPE gDNA Miniprep System	Promega
FFPE Tissue DNA Extraction Kit - Columns	BioChain

Table 29: DNA extraction kits selected from performance analysis. Commercial names and Company brands are reported.

The selected kits are all based on column isolation technology because the magnetic beads based extraction kits resulted less performant in comparison to column extraction kits and, for the first NGS experiment, it is important to become familiar both with manual kits and NGS workflow.
Considering the column based extraction kits, the Qiagen GeneRead DNA FFPE Kit shows the best performances. The Promega ReliaPrep FFPE gDNA Miniprep System kit shows results lower than Qiagen GeneRead DNA FFPE Kit, but it may extract more quantity of DNA since Qiagen GeneRead DNA FFPE Kit requires only one slice of FFPE sample according to the user manual. Although BioChain FFPE Tissue DNA Extraction Kit – Columns is the least performing of the three, it has the potential to be integrated into a fully validated and CE-IVD marked workflow.
Considering the extraction kit for benchtop instrument, Promega Maxwell RSC DNA FFPE Kit is the best because it has performances comparable to those of its equivalent columns, considered among the best in the tests.
Considering the extraction kit for liquid handler instrument, although MagBio Genomics HighPrep FFPE Tissue DNA Kit has lower performance than automation-compatible kits, MagBio Genomics HighPrep FFPE Tissue DNA Kit is comparable with Promega Maxwell RSC DNA FFPE Kit, it has the advantage not requiring the use of centrifuges in deparafinization processes.

5.4.2 RNA QUANTIFICATION TESTS

RNA molecules are becoming increasingly important in oncology field because the study of expression is helping clinicians and scientists to understand the molecular mechanics of many tumors important for early diagnosis, prognosis and to identify new therapeutic targets.
RNA extraction protocols were selected from the same preselected DNA extraction suppliers.

5.4.2.1 MACHEREY-NAGEL - NUCLEOSPIN TOTALRNA FFPE XS

The main features of this protocol are:

- The kit uses a spin column based extraction method;

- The paraffin removal process involves a proprietary deparaffinization solution that operates at 56°C;
- The protocol requires up to 10 FFPE input sections material, 1–20 µm thickness, up to 5mg embedded tissue.

The concentrations of RNA obtained with Macherey-Nagel NucleoSpin totalRNA FFPE XS are reported in **Table 30**. For this protocol the fluorescence from the nucleic acid absorption peak was too low to allow the calculation of concentration with the ThermoFisher Scientific Qubit 3.0 fluorometer instrument. Taking into consideration the results from the other protocols, concentrations obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer should be considered only for the purpose of qualitative evaluation relative to similar absorbance curves.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)	
			Qubit	Nanodrop
FR1[24]_R[MN-NSX]	1	14	Too low	14.4
FR1[38]_R[MN-NSX]	1	14	Too low	15.4
FR1[59]_R[MN-NSX]	1	14	Too low	51.1
FR1[68]_R[MN-NSX]	1	14	Too low	20.5

Table 30: Concentration of RNA extracted with Macherey-Nagel NucleoSpin totalRNA FFPE XS Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 36** (samples from the FR1 series, 1 FFPE section).

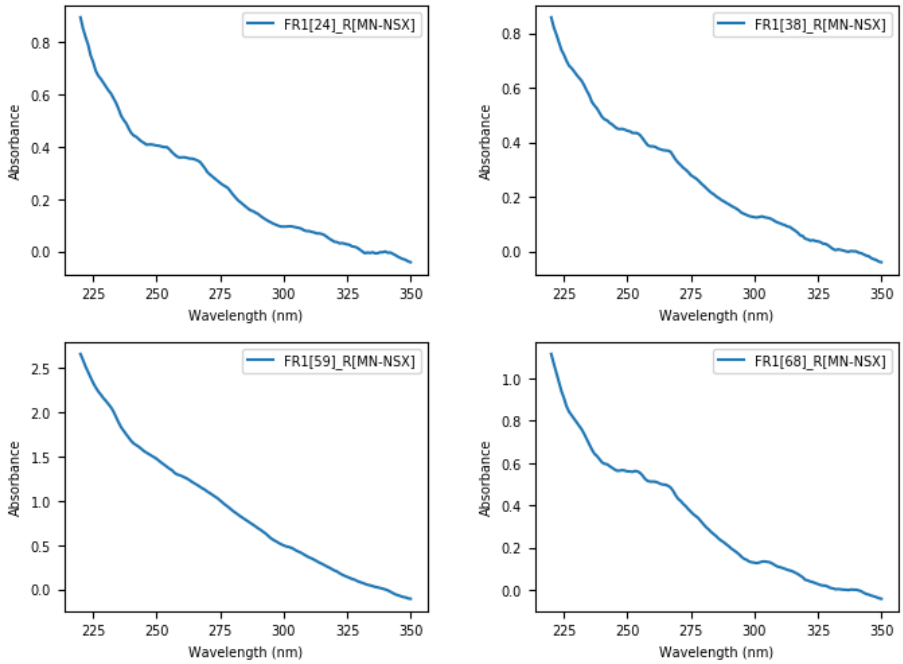


Figure 36: Absorbance plots of RNA extracted with Macherey-Nagel NucleoSpin totalRNA FFPE XS, 1 FFPE section input, elution in 14 µl.

Absorbance ratios are shown in **Table 31**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[24]_R[MN-NSX]	1	14	1.65	0.57
FR1[38]_R[MN-NSX]	1	14	1.60	0.60
FR1[59]_R[MN-NSX]	1	14	1.44	0.60
FR1[68]_R[MN-NSX]	1	14	1.66	0.65

Table 31: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with Macherey-Nagel NucleoSpin totalRNA FFPE XS.

5.4.2.2 OMEGA BIO-TEK - E.Z.N.A. FFPE RNA KIT

The main features of this protocol are:

- The kit uses a spin column based extraction method;
- The paraffin removal process involves deparaffinization by heating to 80°C;
- The partial reversal of modifications of nucleic acid due to fixation at 80°C;
- The protocol requires 3-4 FFPE sections as input material with 5-10 µm thickness.

The concentrations and the yields of RNA extracted with OMEGA bio-tek E.Z.N.A. FFPE RNA Kit are shown in **Table 32**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		Yield (ng)
			Qubit	Nanodrop	
FR1[7]_R[OB-E]	1	14	24.4	47.9	341.6
FR1[31]_R[OB-E]	1	14	47.6	73.8	666.4
FR1[17,32,62]_R[OB-E]	3	14	80.2	177.4	1,122.8
FR1[6,16,63]_R[OB-E]	3	14	126	207.9	1764

Table 32: Concentration of RNA extracted with OMEGA bio-tek E.Z.N.A. FFPE RNA Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 37** (samples from the FR1 series, 1 and 3 FFPE sections).

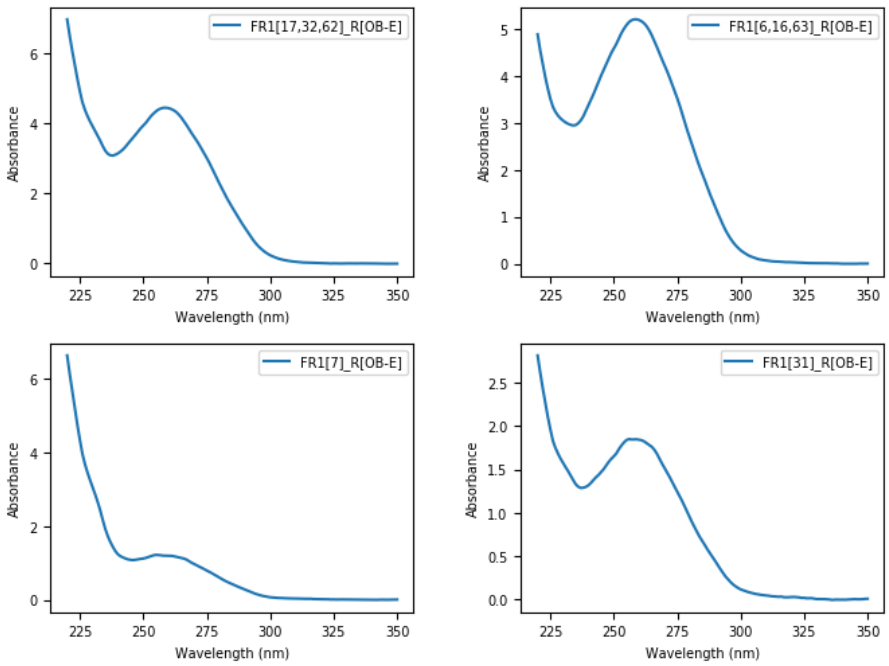


Figure 37: Absorbance plots of RNA extracted with OMEGA bio-tek E.Z.N.A. FFPE RNA Kit, 1 and 3 FFPE sections input, elution in 14 µl.

Absorbance ratios are shown in **Table 33**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[7]_R[OB-E]	1	14	1.99	0.39
FR1[31]_R[OB-E]	1	14	1.96	1.17
FR1[17,32,62]_R[OB-E]	3	14	1.94	1.13
FR1[6,16,63]_R[OB-E]	3	14	1.94	1.70

Table 33: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with OMEGA bio-tek E.Z.N.A. FFPE RNA Kit.

5.4.2.3 OMEGA BIO-TEK - MAG-BIND FFPE RNA KIT

The main features of this protocol are:

- The kit uses a magnetic beads based extraction method;
- The paraffin removal process involves deparaffinization by heating to 80°C;
- The partial reversal of modifications of nucleic acid due to fixation at 80°C;
- The protocol requires 2-5 FFPE sections as input material with 5-10 µm thickness.

The concentrations of RNA obtained with OMEGA bio-tek Mag-Bind FFPE RNA Kit are reported in **Table 34**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		Yield (ng)
			Qubit	Nanodrop	
FR1[44]_R[OB-MB]	1	30	4.0	8.8	120
FR1[64]_R[OB-MB]	1	30	Too low	7.6	NA
FR1[40,45,56]_R[OB-MB]	3	30	13.8	22.9	414
FR1[41,43,65]_R[OB-MB]	3	30	10.4	18.2	312

Table 34: Concentration of RNA extracted with OMEGA bio-tek Mag-Bind FFPE RNA Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 38** (samples from the FR1 series, 1 and 3 FFPE sections).

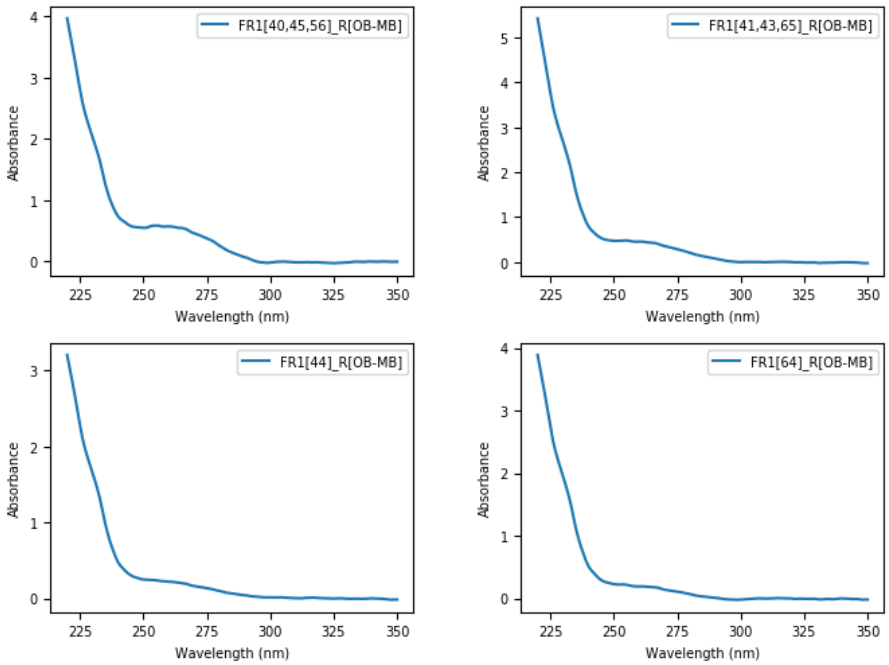


Figure 38: Absorbance plots of RNA extracted with OMEGA bio-tek Mag-Bind FFPE RNA Kit, 1 and 3 FFPE sections input, elution in 30 µl.

Absorbance ratios are shown in **Table 35**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[44]_R[OB-MB]	1	30	2.24	0.13
FR1[64]_R[OB-MB]	1	30	2.79	0.10
FR1[40,45,56]_R[OB-MB]	3	30	2.21	0.28
FR1[41,43,65]_R[OB-MB]	3	30	2.20	0.17

Table 35: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with OMEGA bio-tek Mag-Bind FFPE RNA Kit.

5.4.2.4 QIAGEN - RNEASY FFPE KIT

The main features of this protocol are:

- The kit uses a spin column based extraction method;

- The paraffin removal process involves a proprietary deparaffinization solution that operates at 56°C;
- The partial reversal of modifications of nucleic acid due to fixation at 80°C;
- The protocol requires up to 40 µm total sum of input FFPE section with a surface area of up to 250 mm^2.

The concentrations and the yields of RNA extracted with Qiagen RNeasy FFPE Kit are shown in **Table 36**. Concentration measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are reported only for the purpose of giving an overview on the deviation from the concentration obtained with the fluorimetric method using ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		Yield (ng)
			Qubit	Nanodrop	
FR1[8]_R[Q-RE]	1	14	7.68	54.5	107.52
FR1[29]_R[Q-RE]	1	14	10.3	24.9	144.2
FR1[53]_R[Q-RE]	1	14	8.8	23.8	123.2
FR1[61]_R[Q-RE]	1	14	5.38	38.3	75.32

Table 36: Concentration of RNA extracted with QIAGEN RNeasy FFPE Kit.

Absorption spectra obtained with the ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are shown in **Figure 39** (samples from the FR1 series, 1 FFPE section).

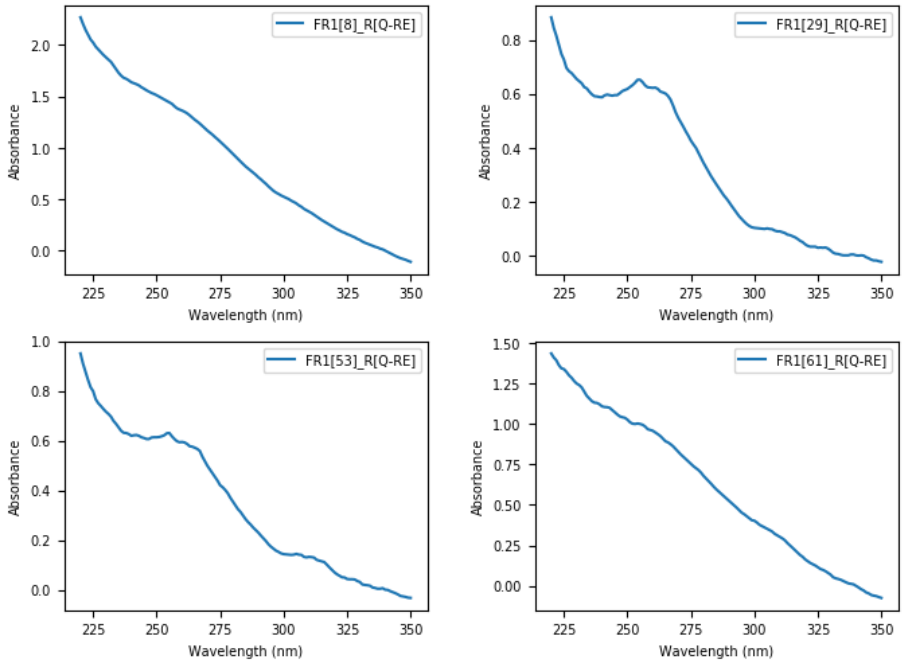


Figure 39: Absorbance plots of RNA extracted with QIAGEN RNeasy FFPE Kit, 1 FFPE section input, elution in 14 µl.

Absorbance ratios are shown in **Table 37**.

Sample ID	Number of FFPE sections	Elution volume (µl)	A260/A280	A260/A230
FR1[8]_R[Q-RE]	1	14	1.45	0.72
FR1[29]_R[Q-RE]	1	14	1.82	0.95
FR1[53]_R[Q-RE]	1	14	1.67	0.83
FR1[61]_R[Q-RE]	1	14	1.41	0.77

Table 37: A260/A280 and A260/A230 absorbance ratios from absorbance curves of DNA extracted with QIAGEN RNeasy FFPE Kit.

5.4.2.5 PERFORMANCES OVERVIEW

The results of the RNA concentrations and yields obtained from FFPE FR slices are summarized in **Table 38**.

Extraction kit	Conc./section (ng/µl)	Yield/section (ng)
OMEGA bio-tek E.Z.N.A. FFPE RNA Kit	35.2 ± 9.9	493 ± 138
Qiagen RNeasy FFPE	8.0 ± 1.8	113 ± 25
OMEGA bio-tek Mag-Bind FFPE RNA	3.0 ± 1.8	91 ± 54
Macherey-Nagel NucleoSpin totalRNA FFPE XS	Too low	Too low

Table 38: Concentrations and yields of RNA per FR 1 sample section obtained from the tested nucleic acid extraction kits.

5.4.2.6 DISCUSSION

5.4.2.6.1 GENERAL CONSIDERATIONS

With the purpose to determine the performances of RNA extraction kits, different quantities of FFPE slices and elution volumes were used.

In detail, concerning the slices number, 1 to 3 slices for the FR 1 series were used, while for elution volume values, 14 µl and 30 µl have been used.

The reasons are similar to those explained for DNA isolation assays, namely to obtain information about the ability and sensitivity of the kits to extract RNA from a single FFPE slice and in the manuals recommended conditions.

5.4.2.6.2 RNA CONCENTRATION CONSIDERATIONS

The OMEGA bio-tek E.Z.N.A. FFPE RNA Kit shows the best RNA extraction performance among the tested kits (about 35 ng/µl), four time higher then Qiagen RNeasy FFPE (8 ng/µl).

The Macherey-Nagel NucleoSpin totalRNA FFPE XS results to be the product with the worse performance, compared to the other kits employing 1 FFPE slice. In fact the ThermoFisher Scientific Qubit 3.0 fluorometer was not able to measure the RNA concentrations in elution solutions obtained with this kit. Tests using 3 FFPE slices are not done due to the very low performance recorded with just 1 slice.

The OMEGA bio-tek Mag-Bind FFPE RNA, the only magnetic beads based protocol used in these experiments, is markedly inferior than the best spin column based protocol OMEGA bio-tek E.Z.N.A. FFPE RNA Kit.

Although they are based on different technologies and produced by the same manufacturer, the RNA concentrations reached with OMEGA bio-tek Mag-Bind FFPE RNA are approximately 90 % lower than OMEGA bio-tek E.Z.N.A. FFPE RNA Kit (35 ng/µl versus 3 ng/µl respectively).

The RNA concentrations values obtained using ThermoFisher Scientific NanoDrop 2000 Spectrophotometer are approximately from 2 to 7 times higher than the concentrations measured with ThermoFisher Scientific Qubit 3.0 fluorometer,. These observations are aligned with those made for DNA extraction tests results.

5.4.2.6.3 ABSORBANCE CONSIDERATIONS

The 260/280 absorbance ratios for Qiagen RNeasy FFPE and Macherey-Nagel NucleoSpin totalRNA FFPE XS are affected by low nucleic acid absorbance, this aspect may indicates presence of possible chemical contaminants in the eluates.

These considerations are also confirmed by the absorption spectra and the low values 260/280 absorbance ratios obtained for each sample analisys.

On the other hand, OMEGA bio-tek Mag-Bind FFPE RNA Kit and OMEGA bio-tek E.Z.N.A. FFPE RNA Kit show value close to 2.0, but also in these cases the absorption spectra and the low A260/A230 ratio values indicate that the RNA molecules in eluates are not pure.

5.4.2.6.4 FINAL CONSIDERATIONS

These are data collected by preliminary experiments, further two kits will be tested in order to find the best RNA extraction kit for NGS application based on RNA molecules, in detail they are Promega ReliaPrep FFPE Total RNA Miniprep System and Qiagen AllPrep DNA/RNA FFPE.

To sum up, the RNA quantification assays made by fluorimetry with the ThermoFisher Scientific Qubit 3.0 fluorometer instruments identifies OMEGA bio-tek E.Z.N.A. FFPE RNA Kit as the kit with the optimal protocol, while the optical absorption method suffers from the same limitations affecting the measurement on DNA described in previous section.

5.5 DNA EXTRACTION FROM FFPE SAMPLES

In order to become familiar with NGS workflow based on Illumina technology, twelve samples belonging to M series were selected for preparing the NGS libraries which will be pool together for the following sequencing step.

Samples details are summarized in **Table 39**.

Sample ID	Characterization	Mutation	Kit	Company
M[34574]_D[BC-FC]	PCR k-KRAS	12 Val	FFPE Tissue DNA Extraction Kit - Column	BioChain
M[07775]_D[BC-FC]	PCR k-KRAS	13 Asp	FFPE Tissue DNA Extraction Kit - Column	BioChain
M[21830]_D[BC-FC]	PCR k-KRAS	WT	FFPE Tissue DNA Extraction Kit - Column	BioChain
M[32892]_D[BC-FC]	PCR k-KRAS	WT	FFPE Tissue DNA Extraction Kit - Column	BioChain
M[31317]_D[P-RP]	PCR k-KRAS	12 Asp	ReliaPrep™ FFPE gDNA Miniprep System	Promega
M[13218]_D[P-RP]	PCR k-KRAS	13 Cys	ReliaPrep™ FFPE gDNA Miniprep System	Promega
M[28108]_D[P-RP]	PCR k-KRAS	WT	ReliaPrep™ FFPE gDNA Miniprep System	Promega
M[34295]_D[P-RP]	PCR k-KRAS	WT	ReliaPrep™ FFPE gDNA Miniprep System	Promega
M[36110]_D[Q-GR]	PCR k-KRAS	12 Asp	Qiagen GeneRead DNA FFPE	Qiagen
M[01773]_D[Q-GR]	PCR k-KRAS	12 Asp	Qiagen GeneRead DNA FFPE	Qiagen
M[30673]_D[Q-GR]	PCR k-KRAS	WT	Qiagen GeneRead DNA FFPE	Qiagen
M[31271]_D[Q-GR]	PCR k-KRAS	WT	Qiagen GeneRead DNA FFPE	Qiagen

Table 39: Association between M series samples and FFPE extraction kits for NGS analysis. K-RAS Mutations characterized in previous company development project are reported.

For each FFPE DNA extraction kit chosen from previous experiments and shown in **Table 29**, two M series FFPE samples with wild type KRAS gene (WT) and two with KRAS mutations (12 Asp, 12 Val, 13 Asp and 13 Cys respectively) have been selected.

The M series samples are used because they derived from FFPE biopsied human tissues characterized by KRAS pathological mutations.

The testing results obtained from the four protocols are reported in the following sections.

5.5.1 BIOCHAIN - FFPE TISSUE DNA EXTRACTION KIT - COLUMN

The concentrations and the yields of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column are shown in **Table 40**. Concentrations were analyzed with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and the ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method for nucleic acid concentration measurements.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		A260 / A280	A260 / A230	Yield (ng)
			Qubit	Nanodrop			
M[34574]_D[BC-FC]	5	50	64.6	437	1.9	2.29	3,230
M[07775]_D[BC-FC]	5	50	5.12	23.6	2.01	2.24	256
M[21830]_D[BC-FC]	5	50	46.4	177.8	1.93	2.2	2,320
M[32892]_D[BC-FC]	5	50	9.68	32.4	2.08	1.67	484

Table 40: Concentrations, Yields and absorbance ratios of DNA extracted with BioChain FFPE Tissue DNA Extraction Kit - Column.

5.5.2 PROMEGA - RELIAPREP FFPE GDNA MINIPREP SYSTEM

The concentrations and the yields of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System are reported in **Table 41.** Concentrations were analyzed with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and the ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method for nucleic acid concentration measurements.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		A260 / A280	A260 / A230	Yield (ng)
			Qubit	Nanodrop			
M[31317]_D[P-RP]	5	30	97.2	985.9	1.87	2.21	2,916
M[13218]_D[P-RP]	5	30	99.4	805.6	1.85	1.53	2,982
M[28108]_D[P-RP]	5	30	100	636.8	1.85	2.02	3,000
M[34295]_D[P-RP]	5	30	102	905.5	1.88	1.18	3,060

Table 41: Concentrations, Yields and absorbance ratios of DNA extracted with Promega ReliaPrep FFPE gDNA Miniprep System.

5.5.3 QIAGEN - GENEREAD DNA FFPE KIT

The concentrations and the yields of DNA extracted with Qiagen GeneRead DNA FFPE Kit are shown in **Table 42**. Concentrations were analyzed with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and the ThermoFisher Scientific Qubit 3.0 fluorometer instrument, which is considered the reference method for nucleic acid concentration measurements.

Sample ID	Number of FFPE sections	Elution volume (µl)	Concentration (ng/µl)		A260 / A280	A260 / A230	Yield (ng)
			Qubit	Nanodrop			
M[36110]_D[Q-GR]	5	30	75.6	547.3	1.86	2.29	2,268
M[01773]_D[Q-GR]	5	30	76	1,060.5	1.83	2.32	2,280
M[30673]_D[Q-GR]	5	30	44.4	136.3	1.85	2.4	1,332
M[31271]_D[Q-GR]	5	30	114	1,140.9	1.85	2.33	3,420

Table 42: Concentrations, Yields and absorbance ratios of DNA extracted with Qiagen GeneRead DNA FFPE Kit.

5.5.4 EXTRACTIONS OVERVIEW

The data recorded from the DNA extraction reactions are recapitulated in **Table 43**.

Sample ID	k-RAS Mutation	Number of FFPE slices	Elution volume (µl)	Qubit Concentration (ng/µl)	Nanodrop Concentration (ng/µl)	A260 / A280	A260 / A230	Yield (ng)
M[34574]_D[BC-FC]	12 Asp	5	50	64.6	437	1.9	2.29	3,230
M[07775]_D[BC-FC]	12 Asp	5	50	5.12	23.6	2.01	2.24	256
M[21830]_D[BC-FC]	WT	5	50	46.4	177.8	1.93	2.2	2,320
M[32892]_D[BC-FC]	WT	5	50	9.68	32.4	2.08	1.67	484
M[31317]_D[P-RP]	12 Asp	5	30	97.2	985.9	1.87	2.21	2,916
M[13218]_D[P-RP]	13 Cys	5	30	99.4	805.6	1.85	1.53	2,982
M[28108]_D[P-RP]	WT	5	30	100	636.8	1.85	2.02	3,000
M[34295]_D[P-RP]	WT	5	30	102	905.5	1.88	1.18	3,060
M[36110]_D[Q-GR]	12 Val	5	30	75.6	547.3	1.86	2.29	2,268
M[01773]_D[Q-GR]	13 Asp	5	30	76	1,060.5	1.83	2.32	2,280
M[30673]_D[Q-GR]	WT	5	30	44.4	1,36.3	1.85	2.4	1,332
M[31271]_D[Q-GR]	WT	5	30	114	1,140.9	1.85	2.33	3,420

Table 43: Association between M series mutated samples and the respective details and results.

5.5.5 DISCUSSION

5.5.5.1 GENERAL CONSIDERATIONS

The starting material has been five FFPE slices for each sample used by the DNA extraction kits, in this way all kits employed the same amount of FFPE slices for the tests.

The elution volume used by BioChain - FFPE Tissue DNA Extraction Kit was 50 µl compared to Promega ReliaPrep FFPE gDNA Miniprep System and Qiagen GeneRead DNA FFPE Kit that employed only 30 µl. Although BioChain - FFPE Tissue DNA Extraction Kit works with elution volumes higher than the other two kits, it has been chosen because it demonstrated good performances in preceding DNA quantification tests.

5.5.5.2 DNA CONCENTRATION CONSIDERATIONS

Considerations about DNA extraction performances cannot be done since the FFPE slices analyzed are different among themselves, as a matter of fact they derive from different human biological samples.

The concentrations values recorded using ThermoFisher Scientific NanoDrop 2000 Spectrophotometer confirm the same observations done in the previous DNA and RNA isolation experiments, in fact the quantifications obtained with NanoDrop 2000 Spectrophotometer are approximately from 3 to 13 times higher than the concentrations measured with ThermoFisher

Scientific Qubit 3.0 fluorometer. Considering the past and the present observations, the concentrations measured with NanoDrop 2000 Spectrophotometer cannot be used for quantifying DNA for the following NGS libraries preparation step.

5.5.5.3 ABSORBANCE CONSIDERATIONS

The A260/A280 and A260/A230 values recorded with NanoDrop 2000 Spectrophotometer are useful parameters for quality analysis.

In general, A260/A280 absorption ratios obtained from DNA samples are good. In detail, the A260/A280 ratios values obtained from DNA Promega ReliaPrep FFPE gDNA Miniprep System (included between 1.85 and 1.88) and Qiagen GeneRead DNA FFPE Kit (included between 1.83 and 1.86) are close to 1.8 value (it is normally used to assess the purity of nucleic acids). The A260/A280 amounts acquired form DNA extracted with BioChain - FFPE Tissue DNA Extraction Kit are a little bit higher than the others (included between 1.9 and 2.08), probably the BioChain kits have isolated some traces of residual RNA or chemical contaminants during the nucleic acid extraction procedures.

Considering the A260/A230 absorbance ratios, almost all the DNA samples show values included between 1.8 and 2.3. The M[34295]_D[P-RP], M[13218]_D[P-RP] and M[32892]_D[BC-FC] samples have value lower than 1.8 (1.18, 1.53 and 1.67 respectively), this may indicate the presence of co-purified contaminants in the eluates.

In three cases, more precisely M[01773]_D[Q-GR], M[31271]_D[Q-GR] and M[30673]_D[Q-GR] show A260/A230 absorbance ratios higher then 2.3, this may indicate, for example, problems during the reading phases.

5.5.5.4 FINAL CONSIDERATIONS

These results indicate that, in general, the DNA extractions from FFPE samples were done properly. In fact, these observations are supported by the A260/A280 and A260/A230 values measured with ThermoFisher Scientific NanoDrop 2000 Spectrophotometer.

The concentrations obtained with ThermoFisher Scientific Qubit 3.0 fluorometer have to be used with for the next quality control analysis for obtaining the parameter useful for the libraries preparation reaction.

5.6 QUALITY CONTROLS FOR NGS LIBRARIES PREPARATION

The NGS libraries preparation step must take place with DNA extracts having precise information about their quality and quantity. These aspects are essential because contaminants or other characteristics incompatible with the subsequent phases of NGS do not have to be there and because it is fundamental to know precisely each DNA concentration to calculate exactly the quantities of samples in all subsequent steps up to the final sequencing experiment.

For these reasons, it is necessary to perform tests that are able to analyze the quality and quantity of the nucleic acids solution for all the subsequent steps of the NGS workflow.

5.6.1 REAL TIME PCR QUANTIFICATION ANALYSIS

With the purpose to determine the quality and the quantity of amplifiable DNA belonging to M series FFPE samples for the following NGS libraries preparation step, the QIAseq DNA QuantiMIZE System kit based on Real Time PCR technology was used.

The **Table 44** shows the results useful for the quality control interpretation.

Sample ID	DNA Concentration (ng/µl)	Ct 100	Ct 200	ΔCt 100	ΔCt 200	Slope
M[34574]_D[BC-FC]	20.00	19.66±0.042	23.62±0.059	1.08	5.60	0.05
M[07775]_D[BC-FC]	5.12	20.60±0.115	23.36±0.135	2.03	5.35	0.03
M[21830]_D[BC-FC]	46.40	17.62±0.034	19.57±0.052	-0.96	1.55	0.03
M[32892]_D[BC-FC]	9.68	18.49±0.085	20.48±0.075	-0.09	2.46	0.03
M[31317]_D[P-RP]	20.00	16.41±0.031	18.04±0.161	-2.17	0.02	0.02
M[13218]_D[P-RP]	20.00	16.85±0.024	19.31±0.133	-1.72	1.30	0.03
M[28108]_D[P-RP]	20.00	16.97±0.166	18.69±0.097	-1.61	0.67	0.02
M[34295]_D[P-RP]	20.00	16.27±0.037	17.83±0.030	-2.30	-0.19	0.02
M[36110]_D[Q-GR]	75.60	30.41±0.416	22.91±0.385	12.55	4.32	-0.08
M[01773]_D[Q-GR]	20.00	17.22±0.030	19.27±0.117	-1.35	1.25	0.03
M[30673]_D[Q-GR]	44.40	19.78±0.126	23.11±0.027	1.20	5.09	0.04
M[31271]_D[Q-GR]	20.00	16.53±0.019	19.07±0.094	-2.05	1.05	0.03
NCT	NA	NA	NA	NA	NA	NA
Control	NA	18.58±0.105	18.02±0.054	0.00	0.00	0.00

Table 44: qPCR results obtained testing M series mutated samples DNA extracts with QIAseq DNA QuantiMIZE System kit on Applied Biosystems 7500 Fast instrument.

Taking into account the results described in the **Table 44** and the indications reported in the QIAseq DNA QuantiMIZE System user manual, the **Table 45** shows the qualitative and quantitative outputs interpretations and the amounts of FFPE M series samples DNA, in terms of microliters and nanograms, required for the libraries preparation reactions.

Sample ID	qPCR QC Score	qPCR QC Call	Sample Volume (µL) Recommended	Sample Amount (ng) Recommended	DNA Amount used (ng)	Sample Volume (µL) used
M[34574]_D[BC-FC]	0.045	Low	6-13	120-250	120	6
M[07775]_D[BC-FC]	0.033	High	8-20	40-100	40	8
M[21830]_D[BC-FC]	0.025	High	1-2	40-100	40	1
M[32892]_D[BC-FC]	0.026	High	4-10	40-100	40	4
M[31317]_D[P-RP]	0.022	High	2-5	40-100	40	2
M[13218]_D[P-RP]	0.030	High	2-5	40-100	40	2
M[28108]_D[P-RP]	0.023	High	2-5	40-100	40	2
M[34295]_D[P-RP]	0.021	High	2-5	40-100	40	2
M[36110]_D[Q-GR]	-0.082	High	1-1	40-100	40	1
M[01773]_D[Q-GR]	0.026	High	2-5	40-100	40	2
M[30673]_D[Q-GR]	0.039	High	1-2	40-100	40	1
M[31271]_D[Q-GR]	0.031	High	2-5	40-100	40	2

Table 45: Association between qualitative and quantitative outputs, QIAseq DNA QuantiMIZE System user manual recommendations and samples amounts for NGS libraries preparation phase. The output values are obtained using excel data analysis file provided by the Qiagen Company.

5.6.1.1 DISCUSSION

5.6.1.1.1 GENERAL CONSIDERATIONS

Before the qPCR analysis, DNA extracts having concentrations higher than 50 ng/µl were diluted to 20 ng/µl, as suggested by the user manual protocol, as show in **Table 44**.

All the calculations were made inserting Ct 100, Ct 200 and concentration values in the excel data analysis file provided by the kit manufacturer.

5.6.1.1.2 Ct CONSIDERATIONS

Almost all the Ct 100 and Ct 200 values obtained for each sample using two different master mixes (Assay 100 or Assay 200, generating amplicon sizes around 100 bp and 200 bp, respectively) differ a little bit among them, approximately from 1.5 to 3.5 Ct values. Only the Ct values measured with M[36110]_D[Q-GR] sample diverge significantly, in fact Ct 100 is 30.41 and Ct 200 is 22.91 with a Ct value difference of 7.5 (**Table 44**).

5.6.1.1.3 QUALITY CONTROL OUTPUT CONSIDERATIONS

ΔCt 100 and 200 values (derived from the difference between ΔCt sample and ΔCt control numbers) are parameters for calculating the "slope" value, which is the qPCR quality control score required to interpreter the results (see the section "Material and Methods" for the details).

Almost all the DNA samples have shown quality control scores that are lower than 0.04 (included between -0.082 and 0.039) (**Table 45**), this evidence means that the DNA samples have high quality

and low degree of damage or fragmentation, for this reason they have the qPCR quality control call labeled as “High”.

Only the M[34574]_D[BC-FC] sample has the quality control score value higher than 0.04, in fact it is 0.045. Basing on the kit user manual indications, the value means DNA quality is low and it shows a high degree damage or fragmentation (the qPCR M[34574]_D[BC-FC] quality control call is reported as“Low”).

5.6.1.1.4 FINAL CONSIDERATIONS

The QIAseq DNA QuantiMIZE System has provided essential qualitative and quantitative data for the following libraries preparation reaction. All DNA extracts obtained with BioChain - FFPE Tissue DNA Extraction Kit, Promega ReliaPrep FFPE gDNA Miniprep System and Qiagen GeneRead DNA FFPE Kit can be used for NGS application.

Combining the results derived from the instrumental experiment, the mathematical elaborations and the recommendations provided by the product user manual, it was possible to identify the correct volume and amount of each DNA exctract sample for employing in the NGS libraries preparation (**Table 45**).

Finally, 40 ng of each high quality DNA sample and 120 ng of M[34574]_D[BC-FC], that are the lower amount suggested by the kit user manual, will be used for the next libraries preparation reactions.

5.7 NGS LIBRARIES QUANTIFICATION ANALYSIS

In order to proceed with the subsequent NGS experiment, the accurate quantification of the libraries obtained for each individual sample with QIAseq Targeted DNA Panel and the measuring of the average sizes of amplicons are necessary.

The precise libraries quantification, both from the concentrations and DNA molecule sizes points of view, is essential for the preparation of libraries pool (in which each library is labeled by an index that identifies the starting sample) that it has to be loaded on the flow-cell of the Illumina MySeq NGS instrument.

5.7.1 FLUORESCENT-BASED NGS LIBRERIES QUANTIFICATION AND AMPLICON DISTRIBUTION ANALYSIS

With the purpose to measure the concentration of each library derived from the M series FFPE samples, the ThermoFisher Scientific Qubit 3.0 fluorometer and Agilent 2200 TapeStation instruments were used for the measurements.

The results are summarized in **Table 46**.

Sample ID	Sample Barcode	Concentration (ng/µl)	
		Qubit	TapeStation HS D5000
M[34574]_D[BC-FC]	IL-N711	12.4	18.6
M[07775]_D[BC-FC]	IL-N712	16.5	14.6
M[21830]_D[BC-FC]	IL-N714	20.6	15.7
M[32892]_D[BC-FC]	IL-N715	16.4	14.8
M[31317]_D[P-RP]	IL-N705	16.5	14.7
M[13218]_D[P-RP]	IL-N706	18.7	24.7
M[28108]_D[P-RP]	IL-N707	16.9	24
M[34295]_D[P-RP]	IL-N710	16	9.7
M[36110]_D[Q-GR]	IL-N701	15.2	15.2
M[01773]_D[Q-GR]	IL-N702	13	15.2
M[30673]_D[Q-GR]	IL-N703	13.3	16.3
M[31271]_D[Q-GR]	IL-N704	17.5	19.3

Table 46: Libraries concentrations measured with ThermoFisher Scientific Qubit 3.0 fluorometer and Agilent 2200 TapeStation instruments. Each sample library is associated to its specific molecular barcode.

The analysis for determining the amplicons size distribution of each sample library were performed with Agilent 2200 TapeStation and High Sensitivity D5000 ScreenTape Assay. The **Figure 40** shows the virtual electrophoresis gel resulted from the assay.

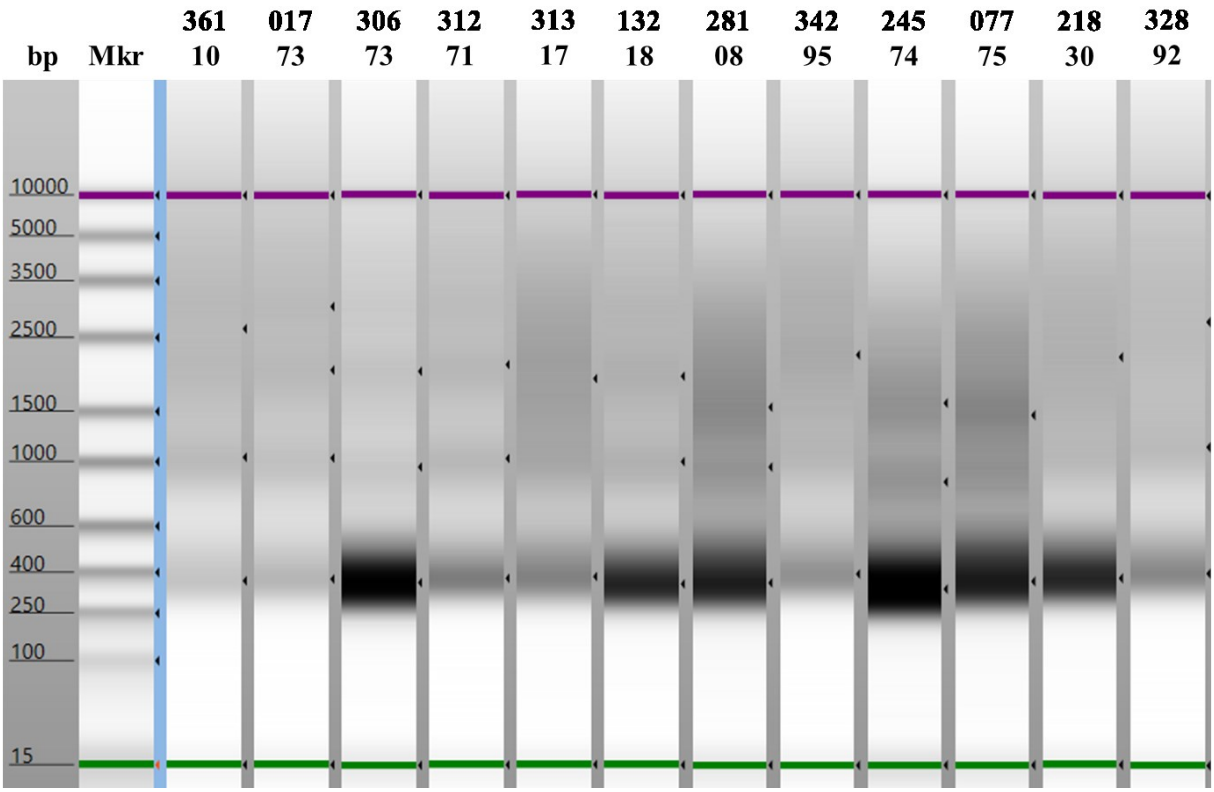


Figure 40: Gel image of M series samples libraries run realized with the Agilent 2200 TapeStation system/High Sensitivity D5000 ScreenTape Assay.

The EPGs belonging to each sample library are shown below. In detail, M[36110]_D[Q-GR] and M[0773]_D[Q-GR] libraries (**Figure 41**), M[30673]_D[Q-GR] and M[31271]_D[Q-GR] libraries (**Figure 42**), M[31317]_D[P-RP] and M[13218]_D[P-RP] libraries (**Figure 43**), M[28108]_D[P-RP] and M[34295]_D[P-RP] libraries (**Figure 44**), M[34574]_D[BC-FC] and M[07775]_D[BC-FC] libraries (**Figure 45**), M[21830]_D[BC-FC] and M[32892]_D[BC-FC] libraries (**Figure 46**).

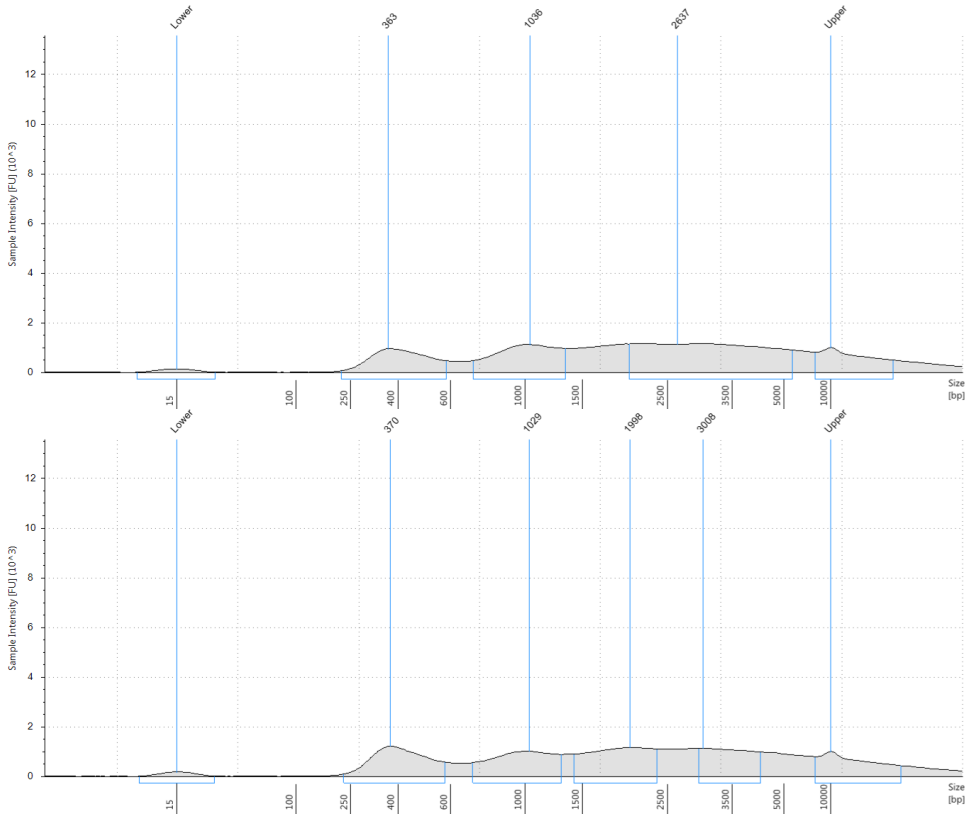


Figure 41: EPGs of DNA Libraries produced with QIAseq Targeted DNA Panel kits from samples in the M series. Up: HS D5000 ScreenTape Assay, samples M[36110]_D[Q-GR]. Down: HS D5000 ScreenTape Assay, samples M[01773]_D[Q-GR].

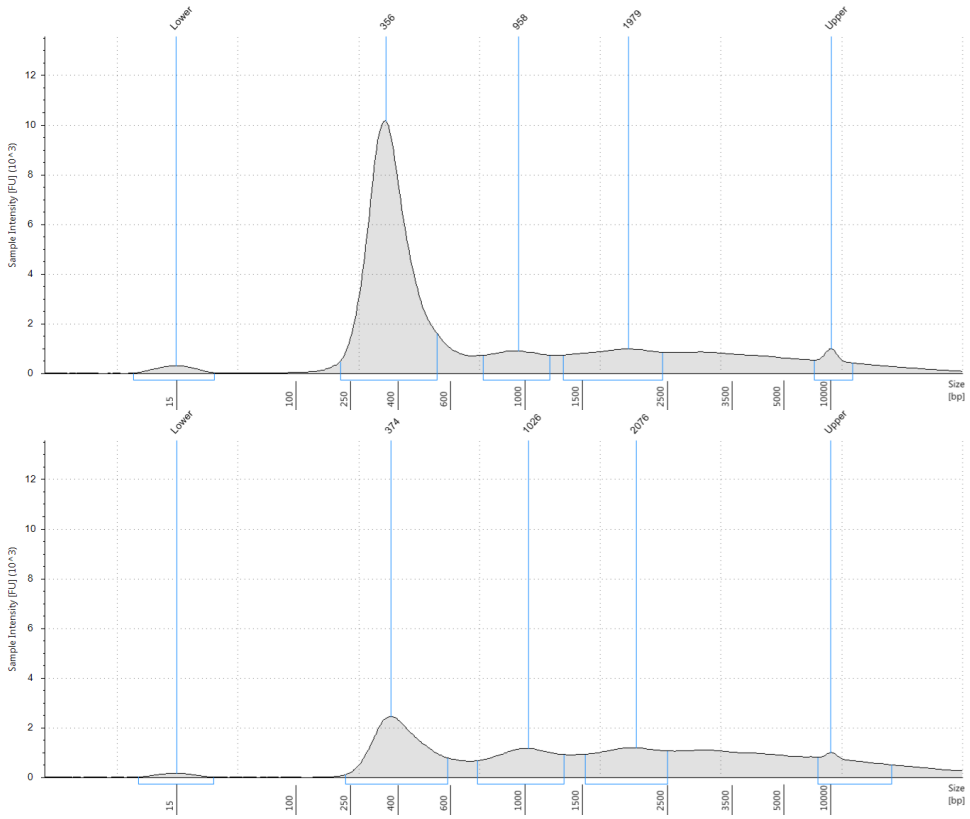


Figure 42: EPGs of DNA Libraries produced with QIAseq Targeted DNA Panel kits from samples in the M series. Up: HS D5000 ScreenTape Assay, samples M[30673]_D[Q-GR]. Down: HS D5000 ScreenTape Assay, samples M[31271]_D[Q-GR].

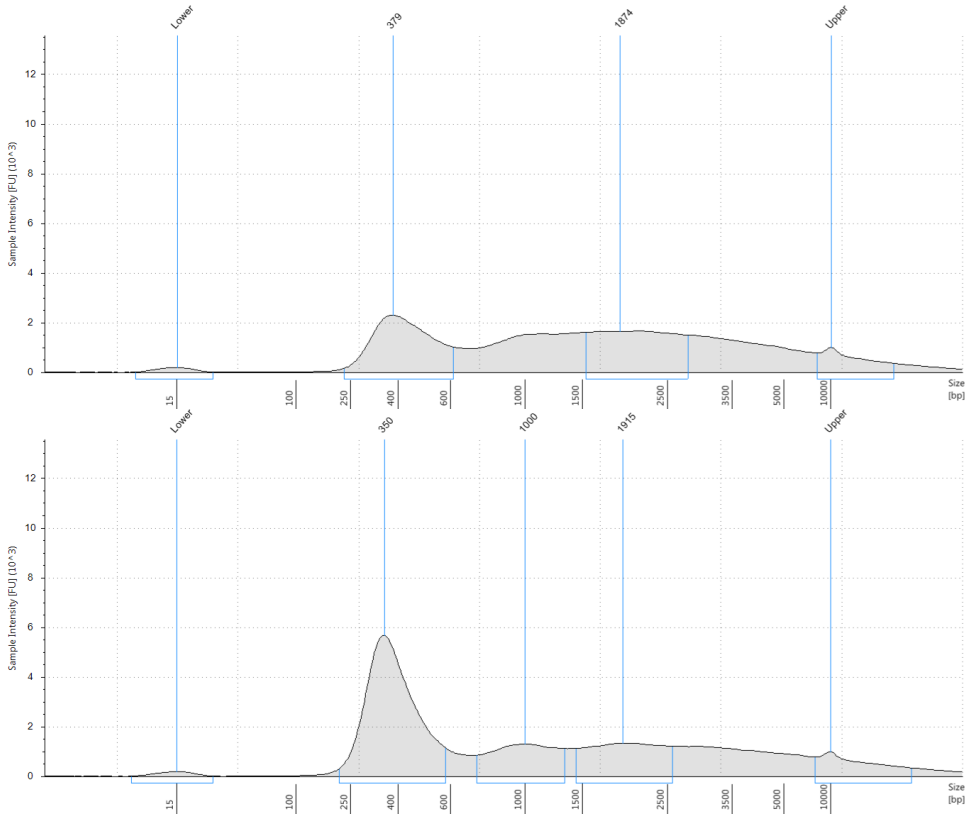


Figure 43: EPGs of DNA Libraries produced with QIAseq Targeted DNA Panel kits from samples in the M series. Up: HS D5000 ScreenTape Assay, samples M[31317]_D[P-RP]. Down: HS D5000 ScreenTape Assay, samples M[13218]_D[P-RP].

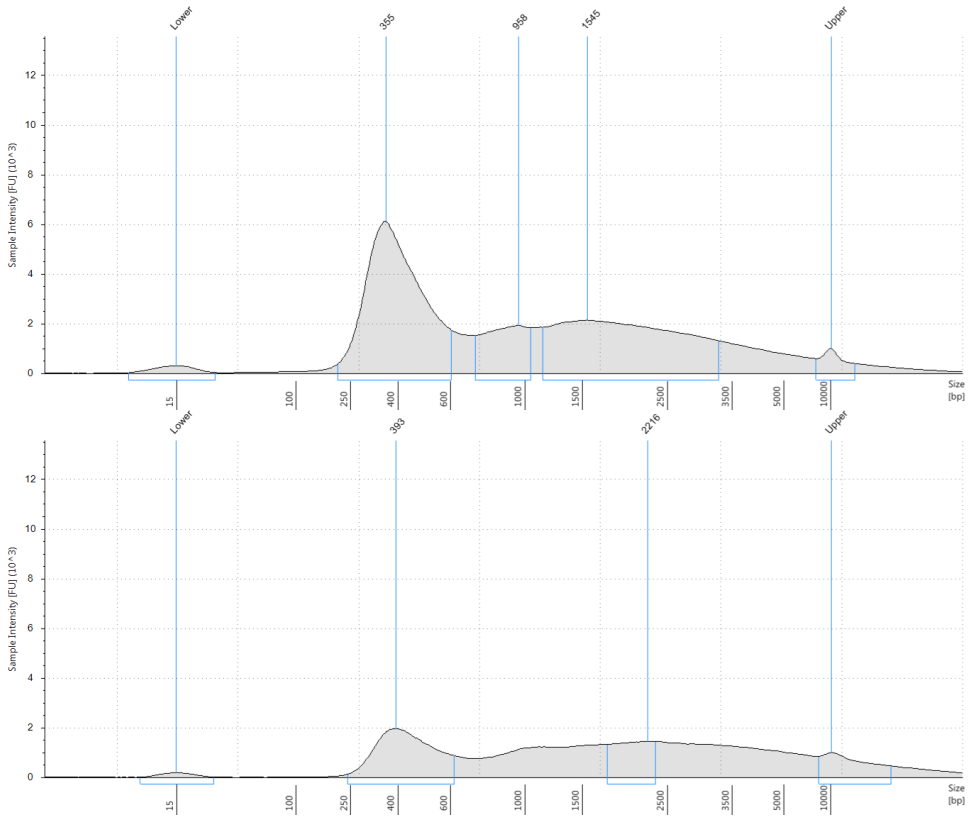


Figure 44: EPGs of DNA Libraries produced with QIAseq Targeted DNA Panel kits from samples in the M series. Up: HS D5000 ScreenTape Assay, samples M[28108]_D[P-RP]. Down: HS D5000 ScreenTape Assay, samples M[34295]_D[P-RP].

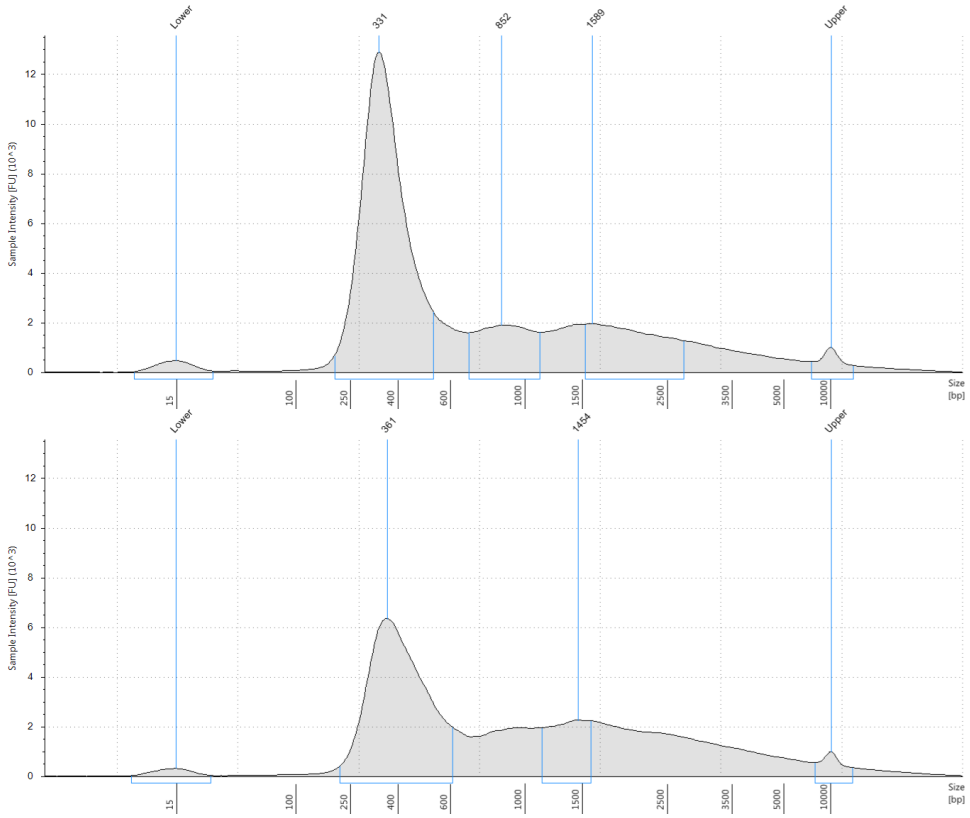


Figure 45: EPGs of DNA Libraries produced with QIAseq Targeted DNA Panel kits from samples in the M series. Up: HS D5000 ScreenTape Assay, samples M[34574]_D[BC-FC]. Down: HS D5000 ScreenTape Assay, samples M[07775]_D[BC-FC].

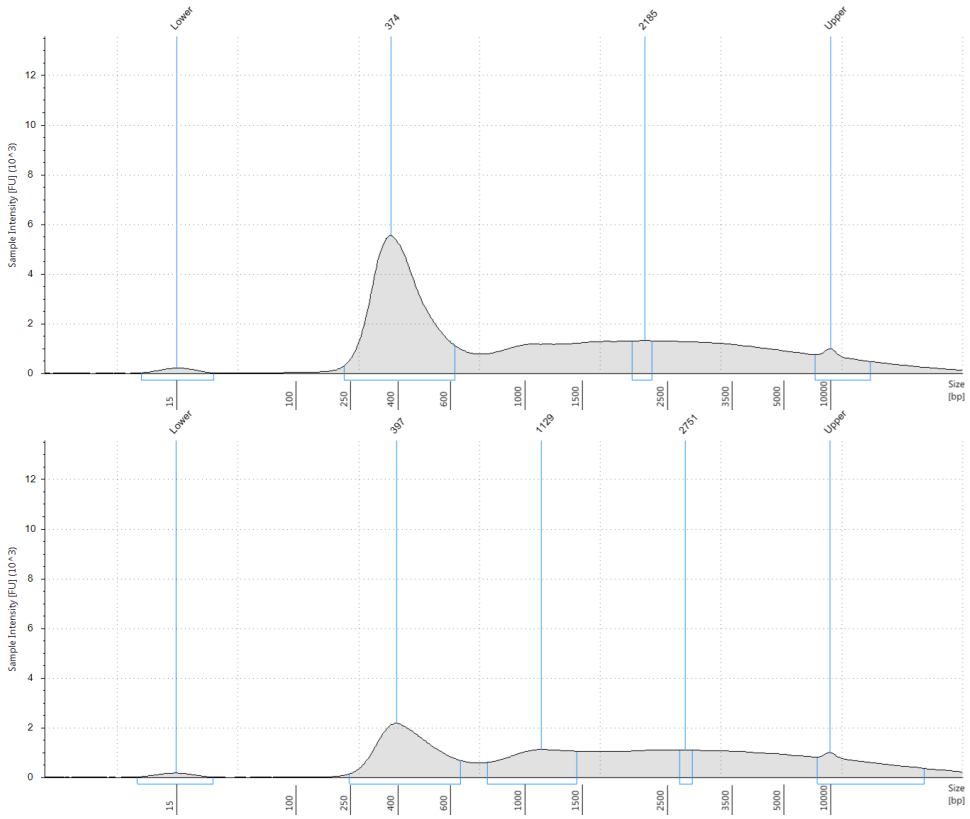


Figure 46: EPGs of DNA Libraries produced with QIAseq Targeted DNA Panel kits from samples in the M series. Up: HS D5000 ScreenTape Assay, samples M[21830]_D[BC-FC]. Down: HS D5000 ScreenTape Assay, samples M[32892]_D[BC-FC].

The data recorded from the ThermoFisher Scientific Qubit 3.0 fluorometer and Agilent 2200 TapeStation measurements are recapitulated in **Table 47**:

Sample ID	Sample Barcode	Concentration (ng/µl)		Average size (bp)
		Qubit	TapeStation HS D5000	
M[34574]_D[BC-FC]	IL-N711	12.4	18.6	375
M[07775]_D[BC-FC]	IL-N712	16.5	14.6	375
M[21830]_D[BC-FC]	IL-N714	20.6	15.7	375
M[32892]_D[BC-FC]	IL-N715	16.4	14.8	375
M[31317]_D[P-RP]	IL-N705	16.5	14.7	375
M[13218]_D[P-RP]	IL-N706	18.7	24.7	375
M[28108]_D[P-RP]	IL-N707	16.9	24	375
M[34295]_D[P-RP]	IL-N710	16	9.7	375
M[36110]_D[Q-GR]	IL-N701	15.2	15.2	375
M[01773]_D[Q-GR]	IL-N702	13	15.2	375
M[30673]_D[Q-GR]	IL-N703	13.3	16.3	375
M[31271]_D[Q-GR]	IL-N704	17.5	19.3	375

Table 47: Libraries concentrations measured with ThermoFisher Scientific Qubit 3.0 fluorometer and Agilent 2200 TapeStation instruments. Each sample library is associated to its specific molecular barcode.

5.7.1.1 DISCUSSION

5.7.1.1.1 GENERAL CONSIDERATIONS

The DNA libraries having a final volume of 28 µl that were used for these experiments derive from the libraries preparation process.

5.7.1.1.2 LIBRARIES CONCENTRATION CONSIDERATIONS

The concentrations measured for each M series sample library using the ThermoFisher Scientific Qubit 3.0 do not differ very much from each other, in fact they are included in a range between 12.4 ng/µl (M [34574] _D [BC -FC]) and 20.6 ng/µl (M [21830] _D [BC-FC]).

Comparing the concentrations obtained for each sample library employing the ThermoFisher Scientific Qubit 3.0 fluorometer and the Agilent 2200 TapeStation instruments, general variations were observed, that in some cases, such as M [34295] _D [P-RP] M [28108] _D [P-RP] M [34574] _D [BC-FC], differ significantly. These aspects reflect the same observations obtained during initial DNA extraction kits performance experiments.

5.7.1.1.3 AMPLICONS SIZE DISTRIBUTION CONSIDERATIONS

The software of Agilent 2200 TapeStation instrument has provided for each sample library the average amplicon size distribution equal to 375 bp.

Taking into account the EPGs resulted for each sample library, the highest peaks (which represent the main molecule size distribution) have shown values close to 375 bp (included between 331 and

397). In general, amplicon size distributions having values higher than 375 bp were obtained, in particular for M[31317]_D[P-RP] (**Figure 43**), M[28108]_D[P-RP], M[34295]_D[P-RP] (both shown in **Figure 44**), M[34574]_D[BC-FC] and M[07775]_D[BC-FC] libraries (both shown in **Figure 45**). In these cases, DNA molecules with size distribution included between 1000 and 2500 bp are shown (**Figure 40**).

5.7.1.1.4 FINAL CONSIDERATIONS

The ThermoFisher Scientific Qubit 3.0 fluorometer has provided quantitative data for each sample library. The concentrations measured during these assays are useful for the following NGS experiments.

Furthermore, the measurements obtained with the Agilent 2200 TapeStation and High Sensitivity D5000 ScreenTape Assay have shown the average amplicon size for each M series sample library which will be analyzed during the Illumina sequencing experiment.

5.7.2 REAL TIME PCR QUANTIFICATION ANALYSIS

5.7.2.1 LIBRARIES QUANTIFICATION

In order to quantify accurately each sample library for preparing the pool libraries solution for the next Illumina sequencing run, the KAPA Library Quantification Kit for Illumina Platforms based on Real Time PCR technology was used.

The **Table 48** shows the conversion, from ng/µl to nM, of libraries concentrations obtained with ThermoFisher Scientific Qubit 3.0 fluorometer.

Sample ID	Sample Barcode	Library Concentration(ng/µl)	Library Concentration (nM)
M[34574]_D[BC-FC]	IL-N711	12.4	50.101
M[07775]_D[BC-FC]	IL-N712	16.5	66.667
M[21830]_D[BC-FC]	IL-N714	20.6	83.232
M[32892]_D[BC-FC]	IL-N715	16.4	66.263
M[31317]_D[P-RP]	IL-N705	16.5	66.667
M[13218]_D[P-RP]	IL-N706	18.7	75.556
M[28108]_D[P-RP]	IL-N707	16.9	68.283
M[34295]_D[P-RP]	IL-N710	16	64.646
M[36110]_D[Q-GR]	IL-N701	15.2	61.414
M[01773]_D[Q-GR]	IL-N702	13	52.525
M[30673]_D[Q-GR]	IL-N703	13.3	53.737
M[31271]_D[Q-GR]	IL-N704	17.5	70.707

Table 48: Libraries concentrations expressed on ng/µl to nM. Each sample library is associated to its specific molecular barcode.

The **Table 49** shows the results useful for the standard curve calculation. It is fundamental for the accurate libraries quantification (**Figure 47**).

Sample ID	Standard Concentration (pM)	qPCR Cq
Standard 1	20	8.02 ± 0.06
Standard 2	2	11.49 ± 0.03
Standard 3	0.2	15.01 ± 0.05
Standard 4	0.02	18.49 ± 0.03
Standard 5	0.002	22.04 ± 0.03
Standard 6	0.0002	25.71 ± 0.02
NTC	0.0	NA

Table 49: Association between standards and samples barcoded with Cq values respectively obtained using KAPA Biosystems Library Quantification Kits.

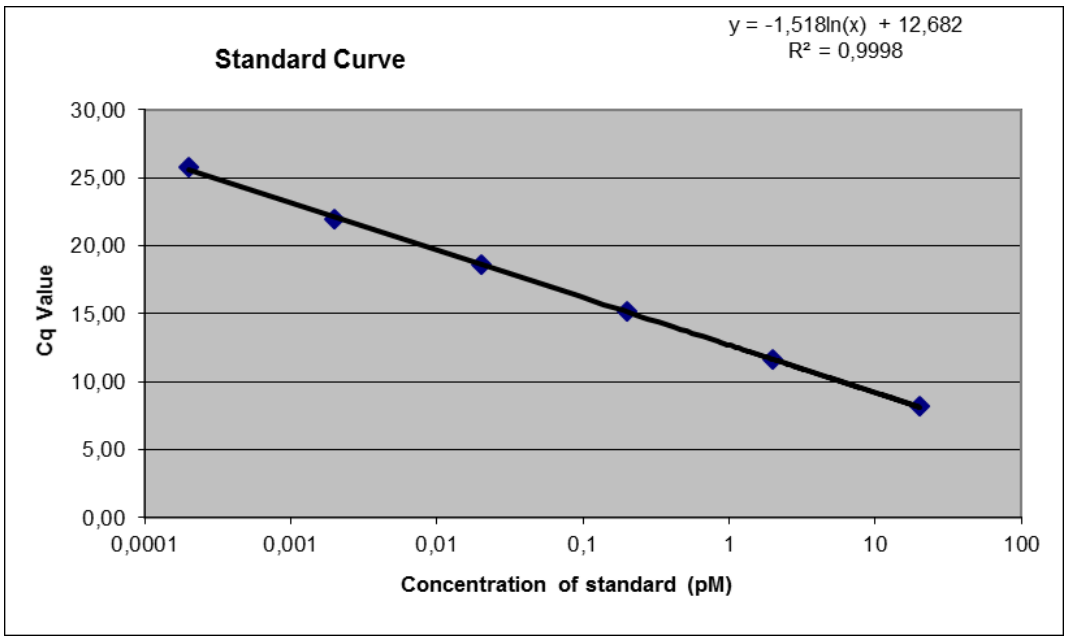


Figure 47: Standard curve generated with 6 different DNA Standards concentrations (20, 2, 0.2, 0.02, 0.002 and 0.0002 pM respectively) which are included in KAPA Library Quantification kit.

The results obtained from the Real Time PCR assay and the mathematical calculations are summarized in **Table 50**.

Sample ID	qPCR Cq	Dilution Factor	Per sample volume (µl)	Dilution Corrected
			Concentration (pM)	Concentration (pM)
Standard 1	8.02 ± 0.06	NA	19.329 ± 0.70	19.329
Standard 2	11.49 ± 0.03	NA	2.005 ± 0.04	2.005
Standard 3	15.01 ± 0.05	NA	0.203 ± 0.01	0.203
Standard 4	18.49 ± 0.03	NA	0.021 ± 0.00	0.021
Standard 5	22.04 ± 0.03	NA	0.002 ± 0.00	0.002
Standard 6	25.71 ± 0.02	NA	0.001 ± 0.00	0.001
NTC	NA	NA	NA	NA
M[34574]_D[BC-FC]	14.69 ± 0.08	1:200,000	0.249 ± 0.01	49,873
M[07775]_D[BC-FC]	13.92 ± 0.02	1:200,000	0.411 ± 0.00	82,244
M[21830]_D[BC-FC]	13.82 ± 0.11	1:200,000	0.440 ± 0.03	87,940
M[32892]_D[BC-FC]	13.81 ± 0.24	1:200,000	0.448 ± 0.07	89,609
M[31317]_D[P-RP]	13.78 ± 0.21	1:200,000	0.455 ± 0.06	91,027
M[13218]_D[P-RP]	13.89 ± 0.06	1:200,000	0.420 ± 0.02	84,076
M[28108]_D[P-RP]	13.75 ± 0.22	1:200,000	0.463 ± 0.07	92,557
M[34295]_D[P-RP]	13.91 ± 0.07	1:200,000	0.414 ± 0.02	82,824
M[36110]_D[Q-GR]	13.62 ± 0.08	1:200,000	0.501 ± 0.03	100,183
M[01773]_D[Q-GR]	14.05 ± 0.08	1:200,000	0.378 ± 0.02	75,563
M[30673]_D[Q-GR]	14.73 ± 0.17	1:200,000	0.244 ± 0.03	48,783
M[31271]_D[Q-GR]	13.54 ± 0.47	1:200,000	0.544 ± 0.18	108,827

Table 50: Samples quantification measured with graphic interpolation obtained using KAPA Biosystems Library Quantification Kit user manual indications. Concentrations expressed in pM are reported.

The **Table 51** the comparison between concentrations measured using ThermoFisher Scientific Qubit 3.0 fluorometer and the Real Time PCR based KAPA Library Quantification Kit.

Sample ID	Sample Barcode	Qubit Library Concentration (nM)	KAPA Library Concentration (nM)
M[34574]_D[BC-FC]	IL-N711	50.101	60.114
M[07775]_D[BC-FC]	IL-N712	66.667	99.131
M[21830]_D[BC-FC]	IL-N714	83.232	105.997
M[32892]_D[BC-FC]	IL-N715	66.263	108.009
M[31317]_D[P-RP]	IL-N705	66.667	109.718
M[13218]_D[P-RP]	IL-N706	75.556	101.339
M[28108]_D[P-RP]	IL-N707	68.283	111.562
M[34295]_D[P-RP]	IL-N710	64.646	99.830
M[36110]_D[Q-GR]	IL-N701	61.414	120.754
M[01773]_D[Q-GR]	IL-N702	52.525	91.079
M[30673]_D[Q-GR]	IL-N703	53.737	58.799
M[31271]_D[Q-GR]	IL-N704	70.707	131.173

Table 51: Libraries concentrations (nM) measured with ThermoFisher Scientific Qubit 3.0 fluorometer and KAPA Biosystems Library Quantification Kit. Each sample library is associated to its specific molecular barcode.

The Real Time PCR data were used for calculating the volume of each sample library necessary for obtaining an equimolar sample libraries pool (**Table 52**). The libraries pool solution contains equal amplicon concentration, expressed in nM, belonging to each sample library.

Sample ID	qPCR Cq	KAPA Library Concentration (nM)	Library volume required (µl)
M[34574]_D[BC-FC]	14.69 ± 0.08	60.114	9.78
M[07775]_D[BC-FC]	13.92 ± 0.02	99.131	5.93
M[21830]_D[BC-FC]	13.82 ± 0.11	105.997	5.55
M[32892]_D[BC-FC]	13.81 ± 0.24	108.009	5.44
M[31317]_D[P-RP]	13.78 ± 0.21	109.718	5.36
M[13218]_D[P-RP]	13.89 ± 0.06	101.339	5.80
M[28108]_D[P-RP]	13.75 ± 0.22	111.562	5.27
M[34295]_D[P-RP]	13.91 ± 0.07	99.830	5.89
M[36110]_D[Q-GR]	13.62 ± 0.08	120.754	4.87
M[01773]_D[Q-GR]	14.05 ± 0.08	91.079	6.46
M[30673]_D[Q-GR]	14.73 ± 0.17	58.799	10.00
M[31271]_D[Q-GR]	13.54 ± 0.47	131.173	4.48

Table 52: Volume calculation for samples libraries preparation having the same amplicon concentration, equal to the lower sample library concentration value M[30673]_D[Q-GR], 58.799 nM). The libraries pool volume is 74.83 µl.

The libraries pool constituted by each library described above will quantify using the KAPA Library Quantification Kit.

5.7.2.1.1 DISCUSSION

5.7.2.1.1.1 GENERAL CONSIDERATIONS

The Real Time experiments were made using sample libraries diluted 1:200,000 in 10 µl qPCR final reaction volume as recommended from the KAPA Library Quantification Kit.

The KAPA Library Quantification Kit allows quantifying each sample library both pM concentration and nucleic acid fragment number points of view using the standard curve obtained with six standard sample (20 pM – 0.2 fM) having template length of 452 bp. The standard curve and the quantification results were obtained using the excel-based data analysis file provided by the kit manufacturer.

5.7.2.1.1.2 QUANTIFICATION CONSIDERATIONS

All the samples libraries dilutions are included in the standard concentrations range used for calculation the standard curves useful for the mathematical quantification analysis.

Comparing the quantification results obtained using the ThermoFisher Scientific Qubit 3.0 instrument and the KAPA Library Quantification Kit, the data measured with the first one are lower than those calculated with the second ones. The qPCR results are from 9% to 97% higher than those observed with ThermoFisher Scientific Qubit 3.0 fluorometer. The quantification values obtained using KAPA Library Quantification Kit were used for calculating the volume of libraries pool because the real time PCR analysis is an analytical method more accurate than fluorescent method used by ThermoFisher Scientific Qubit 3.0 fluorometer.

5.7.2.1.1.3 FINAL CONSIDERATIONS

The results were used for calculating the volume required of each sample library for preparing a libraries pool solution that contain the same amplicon concentration of each library (M[30673]_D[Q-GR]], 58.799 nM), a fundamental requirement for analyzing correctly a set of samples in NGS.

5.7.2.2 POOL QUANTIFICATION

In order to quantify the libraries pool for the following Illumina sequencing run, the Real Time PCR analysis was performed one more time using the KAPA Library Quantification Kit

All the results obtained with the assay are recapitulated in **Figure 48** and **Table 53**.

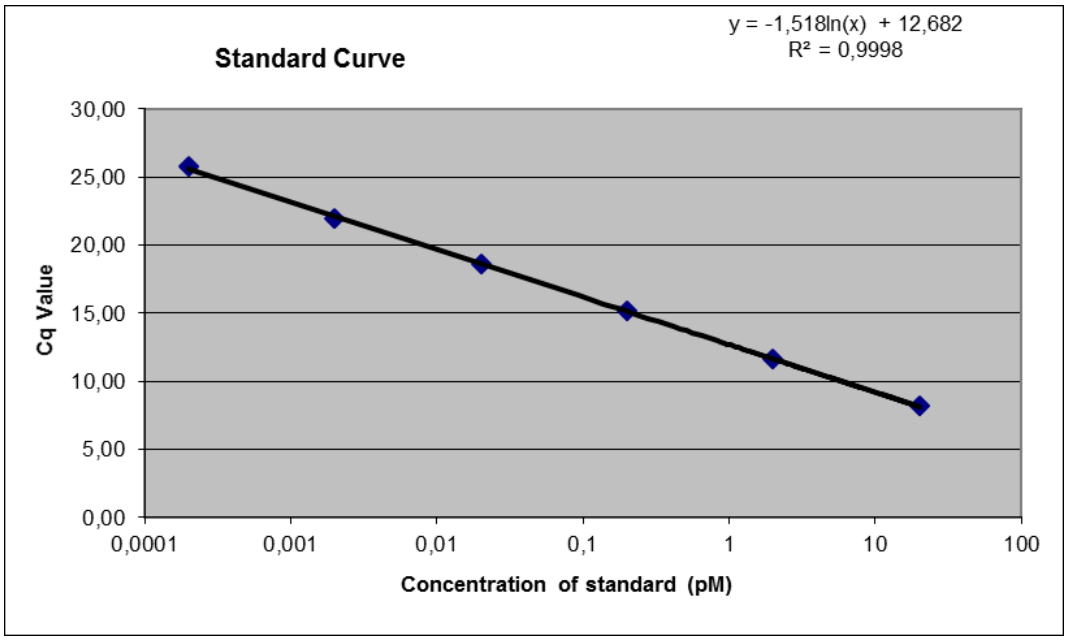


Figure 48: Standard curve generated with 6 different DNA Standards concentrations (20, 2, 0.2, 0.02, 0.002 and 0.0002 pM respectively) which are included in KAPA Library Quantification kit.

Sample ID	qPCR Cq	Dilution Factor	Per sample volume (µl)	Dilution Corrected
			Concentration (pM)	Concentration (pM)
Standard 1	8.18 ± 0.04	NA	19.406 ± 0.46	19.406
Standard 2	11.62 ± 0.03	NA	2.010 ± 0.04	2.010
Standard 3	15.13 ± 0.10	NA	0.199 ± 0.01	0.199
Standard 4	18.59 ± 0.01	NA	0.020 ± 0.00	0.020
Standard 5	21.96 ± 0.21	NA	0.002 ± 0.00	0.002
Standard 6	25.76 ± 0.09	NA	0.002 ± 0.00	0.002
NTC	NA	NA	NA	NA
Libraries pool	13.45 ± 0.24	1:200,000	0.609 ± 0.100	121,851

Table 53: Libraries pool quantification measured with graphic interpolation obtained using KAPA Biosystems Library Quantification Kit user manual indications. Concentrations expressed in pM are reported.

Using the libraries pool concentration expressed nM shown in **Table 54** and making the dilution 1:10, the libraries pool volume required for obtaining a 4nM concentration for the NGS analysis based on Illumina platform is shown in **Table 54**.

Sample ID	qPCR Cq	Library Pool Concentration (nM)	1:10 Library Pool Concentration (nM)	For 4nM in 5µl (µl) of HT1 buffer
Libraries Pool	13.45 ± 0.24	146.87	14.687	1.36

Table 54: Libraries pool volume required for the NGS assay is reported. Amplicon number per µl, concentrations measured using KAPA Biosystems Library Quantification Kit user manual indications are reported.

5.7.2.2.1 DISCUSSION

5.7.2.2.1.1 GENERAL CONSIDERATIONS

The Real Time experiment was made at the same conditions used in previous experiments. The libraries pool diluted 1:200,000 was used for a 10 µl qPCR final reaction volume as recommended from the KAPA Library Quantification Kit.

5.7.2.2.1.2 QUANTIFICATION CONSIDERATIONS

The samples libraries pool dilution is included in the standard concentrations range used for calculation the standard curves useful for the mathematical quantification analysis.

The libraries pool concentration has been calculated using the qPCR concentration per µl x (452 bp - size of KAPA DNA Standard/375 bp - average amplicon length) x the dilution factor.

5.7.2.2.1.3 FINAL CONSIDERATIONS

The result were used for calculating the libraries pool volume required for the last sample preparation which precedes the Illumina flow-cell loading for NGS run.

5.8 NGS ANALYSIS

The results obtained from the sequencing run are shown in **Table 55**.

Sample M Series	M[07775]_D[BC-FC]	M[21830]_D[BC-FC]	M[32892]_D[BC-FC]	M[36110]_D[Q-GR]	M[01773]_D[Q-GR]	M[30673]_D[Q-GR]	M[31271]_D[Q-GR]	M[31317]_D[P-RP]	M[13218]_D[P-RP]	M[28108]_D[P-RP]	M[34295]_D[P-RP]	M[34574]_D[BC-FC]
Total Read Fragments	396,591	345,038	331,924	326,582	389,201	394,278	289,644	347,648	330,579	326,701	363,830	426,266
Read Fragments Dropped	27,860	27,867	22,807	30,366	44,048	32,246	27,826	43,321	43,649	37,819	45,564	59,087
UMIs	29,739	34,969	40,420	88,533	108,839	15,241	51,791	84,174	27,921	38,150	70,746	27,891
Read Fragments Analyzed	368,731	317,171	309,117	296,216	345,153	362,032	261,818	304,327	286,930	288,882	318,266	367,179
Average Read Fragments per UMI	12.40	9.10	7.60	3.30	3.20	23.80	5.10	3.60	10.30	7.60	4.50	13.20
Primers Number	63	63	63	63	63	63	63	63	63	63	63	63
Average UMIs per Primer	472.05	555.06	641.59	1,405.29	1,727.60	241.92	822.08	1,336.10	443.19	605.56	1,122.95	442.71
Average Reads per Primer	5,852.87	5,034.46	4,906.62	4,701.84	5,478.62	5,746.54	4,155.84	4,830.59	4,554.44	4,585.43	5,051.84	5,828.24
Target Bases Number	4,007	4,007	4,007	4,007	4,007	4,007	4,007	4,007	4,007	4,007	4,007	4,007
Average UMIs per base	553.49	711.99	837.45	1,687.62	2,030.35	243.80	971.41	1,703.97	517.60	721.84	1,397.75	427.04
Average Read Depth	14,269.85	12,428.57	11,954.54	11,036.86	12,812.18	13,871.39	9,833.62	11,678.96	11,039.21	10,728.11	12,096.16	13,449.44
% of bases >= 5% of Average Read Depth	100	100	100	100	100	100	100	100	100	100	100	100
% of bases >= 10% of Average Read Depth	100	100	100	100	100	100	100	100	100	100	100	100
% of bases >= 20% of Average Read Depth	99.25	99.38	99.38	99.38	99.38	99.38	99.38	99.38	98.93	99.05	99.38	97.53
% of bases >= 30% of Average Read Depth	96.73	95.66	97.95	99.20	94.83	98.80	94.53	93.01	94.56	93.81	93.16	91.71

Table 55: The results obtained from the sequencing of the libraries pool are reported. The NGS analysis was performed using Illumina MiSeq instrument and the MiSeq Reagent Micro Kit v2 300 cycles (150x2) flow-cell.

5.8.1 DISCUSSION

5.8.1.1 GENERAL CONSIDERATIONS

The libraries pool sequencing analysis was performed properly, this outcome highlights that the previous steps of the NGS workflow have been accomplished correctly.

Each sample library (designed for KRAS, NRAS and BRAF genes analysis) included in the pool has been sequenced, obtaining reads values useful for the interpretation phase of between 260,000 and 370,000.

All library samples have a fraction of fragment reads were excluded automatically (fragment reads dropped) from the software for the next base call identification phase because they are off-target, namely they are not included into the regions of interest. This condition occurs normally in all NGS analysis.

The values of UMI detected for each library in general is high, this is an important for revealing PCR issues and mutations expressed at very low levels.

The results reveal high coverage of regions of interest. In fact, high values of read depth (the number of times that a given nucleotide in the reference genome has been read in an experiment) were observed in all sample libraries, the values are in the range from 9,800X to 14,300X.

Other observations have shown that the sensitivity of this experiment was approximately 92%, detecting the molecular characterizations obtained through the standard goal method of 11 samples out of 12 (data not shown).

Furthermore, the sequencing test showed the presence of numerous SNPs both on mutated and wild type samples.

5.8.1.2 FINAL CONSIDERATIONS

The fact that the NGS experiment has generated results demonstrates that previous workflow procedures have been successful.

The numerical outputs are interesting because, using 63 different primers for amplifying the 4,007 target bases of three genes involved in colorectal cancer, the average read depths resulted higher than 9,800, namely each target base is read 9,800 times with the fragment reads.

The observations show that the NGS experiment has identified the molecular features of 11 samples out of 12, with a sensitivity higher than 90%, but further investigations are required to compare the data obtained through sequencing with the goal standard method.

6 CONCLUSIONS

The intent of this Ph.D. research activity was to investigate the processes required for developing an in vitro diagnostic (IVD) workflow finalized to a targeted next generation sequencing analysis for oncological application.

This Ph.D. research work is an integral part of the HERCULES project (Compre**HE**nsive cha**R**acterisation and effe**C**tive combinatorial targeting of high-grade sero**U**s ovarian cancer via sing**LE**-cell analysi**S**), a Horizon 2020 project where AB ANALITICA, in collaboration with the DOTT. DINO PALADIN Company, has to develop and validate a marketable prototype biomarker kit, based on NGS technology, for predicting HGS-OvCa patient response to combinatorial therapeutic modalities.

During the three Ph.D. years, two distinct activities, with parallel implications, were performed:

- The patent search analysis for evaluating the “state of art” about the presence of patents significant for the HERCULES project followed by the gene panel design based on genes involved in HGS-OvCa. These works are specific for the HERCULES project.
- The study and the selection of FFPE DNA and RNA extraction kits for NGS application, followed by the examinations of all processes necessary to sequence human FFPE samples on the Illumina platform using amplicon technology on well-defined target genes panel. These works are not specific for the HERCULES project, but also for the introduction of NGS technology in the Company and the acquisition of skills and know-how enabling the development of IVD devices based on NGS technology.

The aim of the patent search analysis was to evaluate the “state of art“ about the presence of patents or patent applications relevant for the commercial exploitation of the results expected from the HERCULES project. More than 200 results have been analyzed and 22 were selected because they had contents related with HGS-OvCa.

The patent analysis showed no restrictions on the "freedom to operate" at the moment of the analysis for the prediction of the effectiveness of the therapy, allowing to proceed with the study and the development of a panel of genes dedicated to HGS-OvCa.

The purposes of the NGS library preparation kit and the gene panel design were to identify the requirements and the features for designing a gene panel focused on HGS-OvCa genetic alterations.

The TruSeq Custom Amplicon Low Input Kit is the kit selected to generate a NGS library based on amplicons able to amplify the coding regions of 24 genes (TP53, BRCA1, BRCA2, CSMD3, NF1,

CDK12, FAT3, GABRA6, RB1, BRAF, PIK3CA, KRAS, NRAS, ATM, BARD1, BRIP1, CHEK1, CHEK2, FAM175A, MRE11A, NBN, PALB2, RAD51C, RAD51D) involved in HGS-OvCa.
The advantages of this product are:

- It requires as low as 10 ng of FFPE tumor genomic DNA.
- Unique molecular indexes (UMIs) and sample barcodes can be used for enhancing the allelic detection and creating a library pool derived from different samples respectively.
- It is compatible with the Illumina MiSeq instrument.

The NGS tests employing this HGS-OvCa gene panel were not done due to the lack of clinically validated gene mutations from prognostic and therapeutic points of view specific for high grade serous ovarian cancer.

The aim of the nucleic acid extraction kits screening was to select the DNA and RNA extraction kits, specific for FFPE samples, having features compatible with the integration into a CE-IVD NGS workflow, for both manual and automatic applications.
Numerous aspects were considered for selection, which was based on the study of user manuals and feedbacks from the manufacturers. The screening phase led to the selection of 10 DNA extraction kits (both column and magnetic beads based protocols), out of 39 kits, for the following performance tests.

The purpose of nucleic acid extraction kits performances analysis was to identify the DNA and RNA extraction kits derived from the previous screening phase having the best performances, in terms of quantity and quality of nucleic acid extracts, to be employed in the following NGS workflow development.
The kits were tested using FFPE sample slices of FR series, which derived from a non-tumoral tissue isolated from a human colorectal cancer resection. These biological samples were employed in order to study and compare the results obtained by different extraction kits in standard conditions of matrix, nucleic acid quantity and FFPE preparation protocol.
The assays were made using both absorbance and fluorescent methods (ThermoFisher Scientific NanoDrop 2000 Spectrophotometer and ThermoFisher Scientific Qubit 3.0 fluorometer respectively), in particular the last one is considered the reference method for the DNA/RNA quantification.
The results obtained from the DNA extraction performances tests have demonstrated that:

- for manual application, the column based extraction kits performs better than magnetic beads based kits.

- the kits with the best performances for NGS application are Qiagen GeneRead DNA FFPE Kit, Promega ReliaPrep FFPE gDNA Miniprep System kit and BioChain FFPE Tissue DNA Extraction Kit – Columns.
- for automatic extraction based on benchtop instrument, Promega Maxwell RSC DNA FFPE Kit shows performances similar to Promega ReliaPrep FFPE gDNA Miniprep System kit, so it could be integrated into a NGS workflow with its dedicated instrument.
- for automatic extraction based on liquid handler instrument, MagBio Genomics HighPrep FFPE Tissue DNA Kit has the advantage of not requiring the use of centrifuges in deparafinization processes, although the performance are worse than Qiagen GeneRead DNA FFPE Kit, Promega ReliaPrep FFPE gDNA Miniprep System kit and BioChain FFPE Tissue DNA Extraction Kit – Columns.

The results obtained from the RNA extraction performances tests have demonstrated that:

- column based extraction kits show performances higher than magnetic beads based kit.
- OMEGA bio-tek E.Z.N.A. FFPE RNA Kit shows the best performances, in fact extracting at least 4 times more RNA than the other kits tested. So, it could be taken into consideration as the best candidate for a NGS workflow based on RNA molecules.

Twelve FFPE samples from the M series, which are part of a previous company development project on KRAS mutation detection by real time PCR, were used for studying the NGS workflow with the Qiagen GeneRead DNA FFPE Kit, the Promega ReliaPrep FFPE gDNA Miniprep System kit and the BioChain FFPE Tissue DNA Extraction Kit – Columns.

The aim of the Real Time PCR analysis performed with QIAseq DNA QuantiMIZE System kit was to determine the quality of amplifiable DNA belonging to M series FFPE samples and the quantity necessary for generating the NGS libraries.

The QIAseq DNA QuantiMIZE System kit is a commercial product specific for NGS applications; it uses two qPCR assays that interrogate 20 genomic loci to determine the amounts of amplifiable DNA in biological samples and its quality.

The assay demonstrated that all DNA extracts obtained with Qiagen GeneRead DNA FFPE Kit, Promega ReliaPrep FFPE gDNA Miniprep System kit and BioChain FFPE Tissue DNA Extraction Kit – Columns have a quality compatible with NGS application.

For this reason, they were used for generating the sample libraries based on amplicon technology coupled with UMI and sample molecular tags, providing the amounts required of DNA extracts.

The purpose of libraries quantification assays was to obtain accurate values about the amplicons size and the libraries concentration for generating the libraries pool for the final sequencing analysis.
The electrophoretic separation experiment performed with the Agilent 2200 TapeStation and High Sensitivity D5000 ScreenTape Assay demonstrated that each sample library show an average amplicon size distribution equal to 375 bp, compatible with the Illumina sequencing technology.
The libraries quantification was performed with ThermoFisher Scientific Qubit 3.0 fluorometer and the KAPA Library Quantification Kit for Illumina Platforms based on Real Time PCR technology.
The KAPA Library Quantification Kit for Illumina Platforms is a commercial product specific for NGS applications; it utilizes a SYBR Green I Real Time PCR-based approach to generate a calibration curve with six standard samples for quantifying the sample libraries and the libraries pool.
These experiments demonstrated that the ThermoFisher Scientific Qubit 3.0 fluorometer tends to underestimate the libraries concentrations compared to KAPA Library Quantification Kit. Considering that the qPCR employs a calibration curve generate from standard controls, the libraries pool for the NGS run was produced based on the results obtained from the KAPA Library Quantification Kit.

The aims of the sequencing analysis were to observe if the previous steps of the NGS workflow were successful and also to compare the M series samples mutations obtained by sequencing run with the mutations identified by employing the gold standard method.
The results show that the library pool has been sequenced, demonstrating that the NGS workflow is reliable, in fact all previous NGS workflow steps have been performed correctly.
Each library has incorporated high values of UMIs and presents high depth coverage values. These observations demonstrate that the gene panel designed for library preparation using UMI technology allows the detection of low allelic frequency nucleotide variations and the possible false positives that may derive from sequencing reactions.
Finally, these results show that the NGS experiment is in agreement with the gold standard in 11 samples out of 12, demonstrating a sensitivity higher than 90%.
Further investigations are needed to compare the data obtained through sequencing with the gold standard method in order to understand if this NGS workflow has the potentiality to be improved and integrated into a CE-IVD NGS workflow for oncology applications.

7 REFERENCES

[1] International Agency for Research on Cancer, World Health Organization (WHO). GLOBOCAN 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.

[2] International Agency for Research on Cancer, World Health Organization (WHO). http://globocan.iarc.fr/old/burden.asp?selection_pop=207840&Text-p=United+States+of+America&selection_cancer=290&Text-c=All+cancers+excl.+non-melanoma+skin+cancer&pYear=23&type=0&window=1&submit=%C2%A0Execute and http://globocan.iarc.fr/old/burden.asp?selection_pop=63990&Text-p=European+Union+%28EU-28%29&selection_cancer=290&Text-c=All+cancers+excl.+non-melanoma+skin+cancer&pYear=23&type=0&window=1&submit=%C2%A0Execute.

[3] World Health Organization (WHO). Media centre, Cancer Fact sheet February 2017. http://www.who.int/mediacentre/factsheets/fs297/en/.

[4] Mariotto A.B., et al. "Projections of the Cost of Cancer Care in the U.S.: 2010-2020". J Natl Cancer Inst. 2011. doi: 10.1016/j.amepre.2011.04.008, pp: 117-127.

[5] Cooper G.M. et al. "The Development and Causes of Cancer. In The Cell: A Molecular Approach". 2nd Edition, Sunderland: Sinauer Associates. 2000. https://www.ncbi.nlm.nih.gov/books/NBK9963/.

[6] Hanahan D., et al. "The Hallmarks of Cancer Review." Cell. 2000. https://doi.org/10.1016/S0092-8674(00)81683-9. pp: 57-68.

[7] Hanahan D., Weinberg R. A. "Hallmarks of cancer: the next generation." Cell. 2011. https://doi.org/10.1016/j.cell.2011.02.013, pp: 646-669.

[8] Meinhold-Heerlein I. et al. "The heterogeneity of ovarian cancer". Arch Gynecol Obstet. 2014. https://doi.org/10.1007/s00404-013-3114-3, pp: 237–239.

[9] Prat J. "Staging classification for cancer of the ovary, fallopian tube, and Peritoneum". International Journal of Gynecology and Obstetrics. 2014. https://doi.org/10.1016/j.ijgo.2013.10.001, pp: 1-5.

[10] Kurman R.J. et al. "WHO Classification of Tumours of Female Reproductive Organs". WHO Classification of Tumours. 4th Edition. 2014. Aufl. Lyon: WHO Press.

[11] National Cancer Institute, Cancer Staging. https://www.cancer.gov/about-cancer/diagnosis-staging/staging.

[12] Testa U.et al. "Ovarian Cancers: Genetic Abnormalities, Tumor Heterogeneity and Progression, Clonal Evolution and Cancer Stem Cells". Medicines (Basel). 2018. https://doi.org/10.3390/medicines5010016, pp: 1-74.

[13] Kurman R.J. et al. “Origin and molecular pathogenesis of ovarian high-grade serous carcinoma”. Ann Oncol. 2013. https://doi.org/10.1093/annonc/mdt463, pp: 16-21.

[14] Badgwell D. “Early detection of ovarian cancer”. Dis Markers. 2007. https://dx.doi.org/10.2217%2F17520363.2.3.291, pp: 397-410.

[15] “The Cancer Genome Atlas Research Network. Integrated genomic analysis of ovarian carcinoma”. Nature. 2011. https://doi.org/10.1038/nature10166, pp: 609–615.

[16] Bowtell D.D. et al. “The genesis and evolution of high-grade serous ovarian cancer”. Nat Rev Cancer. 2010. https://doi.org/10.1038/nrc2946, pp: 803–808.

[17] Mavaddat N. et al. “Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA)”. Cancer Epidemiol Biomarkers Prev. 2012. https://doi.org/10.1158/1055-9965.EPI-11-0775, pp: 134-147.

[18] Bookman M.A. “First-line chemotherapy in epithelial ovarian cancer”. Clin Obstet Gynecol. 2012. https://dx.doi.org/10.1080%2F15548627.2015.1009781, pp: 96-113.

[19] Eisenkop S.M. et al. “In Pursuit of optimal cytoreduction for advanced epithelial ovarian cancer: a commentary”. Gynecol Oncol. 2006. https://www.ncbi.nlm.nih.gov/pubmed/16876853, pp: 329-335.

[20] Leffers N. et al.”Prognostic significance of tumor-infiltrating T-lymphocytes in primary and metastatic lesions of advanced stage ovarian cancer”. Cancer Immunol Immunother. 2009. https://doi.org/10.1007/s00262-008-0583-5, pp: 449-459.

[21] Ledermann J.A. et al. “eUpdate – Ovarian Cancer Treatment Recommendations” ESMO Guidelines Committee. 2016. https://www.esmo.org/Guidelines/Gynaecological-Cancers/Newly-Diagnosed-and-Relapsed-Epithelial-Ovarian-Carcinoma/eUpdate-Treatment-Recommendations, pp: 1-2.

[22] Goff B.A., et al. “Frequency of symptoms of ovarian cancer in women presenting to primary care clinics”. JAMA. 2004. https://doi.org/10.1001/jama.291.22.2705, pp: 2705-2712.

[23] Conte P.F. at al. “Linee guida tumori dell’ovaio” AIOM 2015. www.aiom.it

[24] Bookman M.A. et al. “Optimal primary therapy of ovarian cancer”. Annals of Oncology. 2016. https://academic.oup.com/annonc/article/27/suppl_1/i58/1786141, pp: i58–i62.

[25] Bell J. et al. ”Randomized phase III trial of three versus six cycles of adjuvant carboplatin and paclitaxel in early stage epithelial ovarian carcinoma: a Gynecologic Oncology Group study”. Gynecol Oncol. 2006. https://doi.org/10.1016/j.ygyno.2006.06.013, pp: 432–439.

[26] Pignata S. et al. ”Carboplatin plus paclitaxel versus carboplatin plus pegylated liposomal doxorubicin as first-line treatment for patients with ovarian cancer: the MITO-2 randomized phase III trial”. J Clin Oncol. 2011. https://doi.org/10.1200/JCO.2010.33.8566, pp: 3628-3635.

[27] Perren T.J. et al. ”A phase 3 trial of bevacizumab in ovarian cancer”. N Engl J Med. 2011. https://doi.org/10.1056/NEJMoa1103799, pp: 2484-2496.

[28] Burger R.A. et al.”Incorporation of bevacizumab in the primary treatment of ovarian cancer”. N Engl J Med. 2011. https://doi.org/10.1056/NEJMoa1104390, pp: 2473-2483.

[29] De Picciotto N. et al. “Ovarian cancer: Status of homologous recombination pathway as a predictor of drug response”. Critical Reviews in Oncology/Hematology. 2016. https://www.croh-online.com/article/S1040-8428(16)30038-5/abstract, pp 50–59.

[30] Evans T. et al.“PARP inhibitors in ovarian cancer: evidence, experience and clinical potential”. Therapeutic Advances in Medical Oncology. 2017. https://dx.doi.org/10.1177%2F1758834016687254, pp: 253–267

[31] Alison J. et al. “Poly(ADP-ribose) polymerase (PARP) inhibitors for the treatment of ovarian cancer”. Cochrane Database of Systematic R, eviews. 2015. https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD007929.pub3/epdf/full, pp:1-50.

[32] “An overview of Lynparza and why it is authorised in the EU” EMA. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Summary_for_the_public/human/003726/WC500180153.pdf

[33] National Comprehensive Cancer Network (NCCN). Guidelines Epithelial Ovarian Cancer (including Fallopian Tube Cancer and Primary Peritoneal Cancer). Version 2.2018. https://www.nccn.org/professionals/physician_gls/pdf/ovarian_blocks.pdf, pp: 1-130.

[34] Kurman R.J. et al. “The dualistic model of ovarian carcinogenesis revisited, revised, and Expanded”. American Journal of Pathology. 2016. https://doi.org/10.1016/j.ajpath.2015.11.011, pp: 733-747.

[35] de Leng W.W.J. et al. “Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material”. PLoS ONE.2016. doi: 10.1371/journal.pone.0149405, pp: 1-18.

[36] Levy S.E. et al. “Advancements in Next-Generation Sequencing”. Annu. Rev. Genom. Hum. Genet. 2016. https://doi.org/10.1146/annurev-genom-083115-022413, pp: 95-115.

[37] Yohe S. et al. “Review of Clinical Next-Generation Sequencing”. Arch Pathol Lab Med. 2017. https://doi.org/10.5858/arpa.2016-0501-RA, pp: 1544-1557.

[38] Naidoo N. et al. “Human genetics and genomics a decade after the release of the draft sequence of the human genome”. Hum Genomics. 2011. https://dx.doi.org/10.1186%2F1479-7364-5-6-577, pp: 577–622.

[39] Kulski .K. “Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications”, IntechOpen, 2016. http://dx.doi.org/10.5772/61964, pp: 3-60.

[40] Shendure J. et al. “Next-generation DNA sequencing”. Nat Biotechnol. 2008. https://doi.org/10.1038/nbt1486, pp: 1135-1145.

[41] “An Introduction to Next-Generation Sequencing Technology” 2017. https://www.illumina.com/Documents/products/Illumina_Sequencing_Introduction.pdf, pp: 1-16.

[42] Buermans H.P. et al. “Next generation sequencing technology: Advances and applications”. Biochimica et Biophysica Acta. 2014. https://doi.org/10.1016/j.bbadis.2014.06.015, pp: 1932–1941

[43] Froyen G. et al. “Validation and Application of a Custom-Designed Targeted Next-Generation Sequencing Panel for the Diagnostic Mutational Profiling of Solid Tumors”. PLoS ONE.2016. https://doi.org/10.1371/journal.pone.0154038, pp: 1-16.

[44] Kerick M. et al., “Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity”. BMC Med Genomics. 2011. https://doi.org/10.1186/1755-8794-4-68, pp: 4-68.

[45] Thomas A. et al. “From targets to targeted therapies and molecular profiling in non-small cell lung carcinoma”. Ann Oncol. 2013. https://doi.org/10.1093/annonc/mds478, pp: 577-585.

[46] Luthra R. et al.”Next-Generation Sequencing in Clinical Molecular Diagnostics of Cancer: Advantages and Challenges”. Cancers (Basel). 2015. https://doi.org/10.3390/cancers7040874, pp: 2023–2036.

[47] Jennings L.J. et al. “Guidelines for Validation of Next-Generation SequencingeBased Oncology Panels”. The Journal of Molecular Diagnostics, 2017. http://dx.doi.org/10.1016/j.jmoldx.2017.01.011, pp: 341-365.

[48] Spencer D.H. et al. ”Performance of common analysis methods for detecting low frequency single nucleotide variants in targeted next generation sequence data”. J Mol Diagn. 2014. https://doi.org/10.1016/j.jmoldx.2013.09.003, pp:75–88.

[49] van Dijk E.L. et al. “Library preparation methods for next-generation sequencing: Tone down the bias”,Exp Cell Res. 2014. http://dx.doi.org/10.1016/j.yexcr.2014.01.008, pp: 1-9.

[50] Wayne P.A. “ Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline”. CLSI document MM09-A2. ed 2. 2014. https://clsi.org/media/1481/mm09a2_sample.pdf, pp:1-122.

[51] Spencer D.H. et al. “Comparison of Clinical Targeted Next-Generation Sequence Data from Formalin-Fixed and Fresh-Frozen Tissue Specimens”. J Mol Diagn. 2013. https://doi.org/10.1016/j.jmoldx.2013.05.004, pp: 623-633.

[52] Auerbach C. et al. “Genetic and cytogenetical effects of formaldehyde and related compounds”. Mutat Res. 1977. https://www.ncbi.nlm.nih.gov/pubmed/331091, pp. 317-361.

[53] Inadome Y. et al. “Selection of higher molecular weight genomic DNA for molecular diagnosis from formalin-fixed material”. Diagn Mol Pathol. 2003. https://www.ncbi.nlm.nih.gov/pubmed/14639109, pp. 231-236.

[54] Chen G. et al. “Cytosine deamination is a major cause of baseline noise in next-generation sequencing”. Mol Diagn Ther. 2014. https://doi.org/10.1007/s40291-014-0115-2, pp: 587-593.

[55] Cline J. Et al. ”Pcr fidelity of pfu dna polymerase and other thermostable dna polymerases”. Nucleic Acids Res. 1996. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146123/, pp: 3546-3551.

[56] Yan B. et al. “Coverage analysis in a targeted amplicon-based next-generation sequencing panel for myeloid neoplasms”. J Clin Pathol. 2016. https://doi.org/10.1136/jclinpath-2015-203580, pp: 801-804.

[57] Blomquist T.M. et al. “Targeted RNA-sequencing with competitive multiplex-PCR amplicon Libraries”. PLoS One. 2013. https://doi.org/10.1371/journal.pone.0079120, pp: 1-14.

[58] Kukita Y. et al. “High-fidelity target sequencing of individual molecules identified using barcode sequences: de novo detection and absolute quantitation of mutations in plasma cell-free dna from cancer patients”. DNA Res. 2015. https://doi.org/10.1093/dnares/dsv010, pp: 269–277.

[59] Xu C. et al. “Detecting very low allele fraction variants using targeted DNA sequencing and a novel molecular barcode-aware variant caller”. BMC Genomics. 2017. https://doi.org/10.1186/s12864-016-3425-4, pp: 1-11.

[60] Roy S. et al. “Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment”. Arch Pathol Lab Med. 2016. https://doi.org/10.5858/arpa.2015-0507-RA, pp: 958-975.

[61] Roy S. et al. “Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines”. J Mol Diagn. 2018. https://doi.org/10.1016/j.jmoldx.2017.11.003, pp: 4-27.

[62] Cock P.J.A. et al. “The Sanger FASTQ file format for sequences with quality scores, and the Solexa/ Illumina FASTQ variants”. Nucleic Acids Res. 2009. https://dx.doi.org/10.1093%2Fnar%2Fgkp1137, pp: 1767-1771.

[63] "QIAseq Targeted DNA Panel Handbook" 2017. https://www.qiagen.com/kr/resources/download.aspx?id=8907edbe-a462-4883-ae1b-2759657e7fd0&lang=en, pp: 1-76.

[64] https://echa.europa.eu/registration-dossier/-/registered-dossier/15482/2/1.

[65] https://www.ispettorato.gov.it/it-it/Documenti-Norme/Documents/Dlgs-81-08-Integrato.pdf.

[66] N.I. Lindeman et al. "Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors. Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology" Arch Pathol Lab Med. 2013. https://doi.org/10.1097/JTO.0b013e318290868f, pp: 823-859.

[67] Srinivasan M. at al, "Effect of fixatives and tissue processing on the content and integrity of nucleic acids" Am J Pathol. 2002. https://doi.org/10.1016/S0002-9440(10)64472-0, pp: 1961-1971.

[68] Do H. et al. "Dramatic reduction of sequence artefacts from DNA isolated from formalinfixed cancer biopsies by treatment with uracil-DNA glycosylase" Oncotarget. 2012. https://dx.doi.org/10.18632%2Foncotarget.503, pp: 546–558.

[69] Guoli C. at al, "Cytosine Deamination is a Major Cause of Baseline Noise in Next Generation Sequencing" Mol Diagn Ther. 2014. https://dx.doi.org/10.1007%2Fs40291-014-0115-2, pp: 587–593.

[70] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31998L0079.

[71] Arreaza G. et al, "Pre-Analytical Considerations for Successful Next-Generation Sequencing (NGS): Challenges and Opportunities for Formalin-Fixed and Paraffin-Embedded Tumor Tissue (FFPE) Samples". Int. J. Mol. Sci. 2016. https://doi.org/10.3390/ijms17091579, pp: 1–8.

[72] Heydt C. et al, "Comparison of Pre-Analytical FFPE Sample Preparation Methods and Their Impact on Massively Parallel Sequencing in Routine Diagnostics" PLoS ONE. 2014. https://doi.org/10.1371/journal.pone.0104566, pp: 1–11.

[73] Moschetta M. et al. "BRCA somatic mutations and epigenetic BRCA modifications in serous ovarian cancer". Annals of Oncology. 2016. https://doi.org/10.1093/annonc/mdw142, pp: 1449–1455.

8 ACKNOWLEDGEMENTS

Firstly, I would like to express sincere gratitude to Dr. Dino Paladin and Prof. Sara De Martin who have allowed me to carry out the Ph.D. course in pharmacological sciences.

I would like to thank sincerely Dr. Diego Boscarino for his precious support, without him I would not have been able to conduct this research activity.

I thank all my colleagues in DOTT. DINO PALADIN, AB ANALITICA and BIOFILED INNOVATION companies for the great collaboration in numerous activities carried out in these years.

I would also like to thank my family and all my friends for the wonderful moments spent with me since and for the patience that they had in the final part of writing the thesis.

Finally, I must express my profound gratitude to Sonia, my wife-to-be, for giving me unfailing support and continuous encouragement in all these years, particularly in these years of Ph.D. and thesis writing. This result would not have been possible without you!

Thank You!

9 SUPPLEMENTARY MATERIALS

9.1 APPENDIX A

1. **ID: US20150030583(A1)**

Title: Methods of Treating Serosal Cancer.

Priority date: 01/04/ 2011

Publication date: 29/01/2015

Description: The application describes methods to isolate (from serosal cavity), prepare, characterize, grow a clonally pure population of serosal cancer stem cells (CSCs) which leads to the new insights for treating serosal cancer. The present invention is directed to methods of treating serosal cancer in a patient undergoing chemotherapy by administering a hyaluronan synthase inhibitor, a hyaluronidase, a collagenase, or a combination thereof, for a time and in an amount to augment the chemotherapy, to improve patient quality of life, to increase patient survival time and/or to cause remission of symptoms, wherein the chemotherapy comprises administering to the patient a therapeutically-effective amount of a compound selected from the group consisting LBH-589 (Panobinostat), NVP-AUY922, LAQ824 (NVP-LAQ824, Dacinostat), colchicine, brefeldin A and diphenyleneiodonium chloride, or a combination thereof. These treatment methods of the invention can also be used in combination with radiation treatment or other conventional cancer therapy.

Relevance to the HERCULES project: Methods are included to isolate CSC using miRNAs and specific proteins, for the purpose of screening test compounds for anti-proliferative effects on cell cultures.

Inventors: Malcolm A.S. Moore, Server A. Ertem.

Owner: Sloan Kettering Institute For Cancer Research

Also published as: WO2012135588 (A2), WO2012135588 (A3).

INPADOC patent family: US2015030583 (A1), WO2012135588 (A2), WO2012135588 (A3).

Status US2015030583 (A1): 25.02.2016 Requirement for Restriction/Election

Status WO2012135588 (A3): 01.10.2013 Non- entry into the national phase in DE, 01.10.2013EP: PCT APP. NOT ENT. EUROP. PHASE

2. **ID: US20140045853(A1)**

Title: Materials, methods, and systems for treating cancer

Priority date: 06/08/ 2012

Publication date: 13/02/2014

Description: Aspects of the invention relate to methods of treatment, and to kits and systems for the same including materials for determining that an individual is susceptible to developing cancer (in particular ovarian cancer) and if warranted treating the patient for cancer or initiating a monitoring strategy and/or taking a preventive action.

The invention is focused on susceptibility but it include also method of treatment, the method comprising determining that a patient has cancer, assaying a sample comprising a RAD51D-encoding nucleic acid molecule, or a complement thereof, from the patient for the presence of a mutation in said nucleic acid, and, where a mutation is found in said nucleic acid, treating the patient using a DNA damaging agent or Topoisomerase I (TOPO I) inhibitor.

Relevance to the HERCULES project: Methods for determining the susceptibility of people to develop cancer (in particular ovarian cancer) and methods of treatment, based on specific combinations of gene mutations.

Inventors: Sabera Nazneen Rahman.

Owner: The Institute Of Cancer Research: Royal Cancer Hospital.

Also published as: No further publications.

INPADOC patent family: US2014045853 (A1)

Status: 05.02.2016 final rejection

3. **ID: US2005032724(A1)**

Title: Methods for improved treatment of cancer with irinotecan based on mrp1.

Priority date: 23/07/ 2001

Publication date: 10/02/2005

Description: The present invention relates to the use of irinotecan or derivative thereof for the preparation of a pharmaceutical composition for treating cancer, especially, colorectal cancer, cervical cancer, gastric cancer, lung cancer, malignant glioma, ovarian cancer, and pancreatic cancer in a patient having a genotype with a variant allele which comprises a polynucleotide in accordance with the present invention. The present invention relates to a method for selecting a suitable therapy for a subject suffering from several carcers, including also the ovarian cancer.

Relevance to the HERCULES project: Method for selecting a suitable therapy for a subject suffering from several cancers, including also the ovarian cancer, using MRP1 gene and protein mutations as biomarkers.

Inventors: Gunther Heinrich, Reinhold Kerb.

Owner: Gunther Heinrich, Reinhold Kerb.

Also published as: CA2454627A1, CA2454637A1, CA2454640A1, CA2454643A1, CA2454648A1, WO2003013533A2, WO2003013533A3, WO2003013533A9, WO2003013534A2, WO2003013534A3, WO2003013534A9, WO2003013535A2, WO2003013535A3, WO2003013535A9, WO2003013536A2, WO2003013536A3, WO2003013536A9, WO2003013537A2, WO2003013537A3, WO2003013537A9-

INPADOC patent family: US2005032724 (A1), AU2002328945(A1), AU2002328950(A1), AU2002328952(A1), AU2002328953(A1), AU2002331290(A1), CA2454627(A1), CA2454637 (A1), CA2454640(A1), CA2454643(A1), CA2454648(A1), EP1408972(A2), EP1408973(A2), EP1408974(A2), EP1408975 A2), EP1438050(A2), JP2005501840(A), JP2005504759(A), JP2005505526(A), JP2005506971(A), JP2005508312(A), WO03013533(A2), WO03013533(A3), WO03013533(A9), WO03013534(A2), WO03013534(A3), WO03013534(A9), WO03013535(A2), WO03013535(A3), WO03013535(A9), WO03013536(A2), WO03013536(A3), WO03013536(A9), WO03013537(A2), WO03013537(A3), WO03013537(A9).

Status WO03013533 (A2), WO03013534 (A2), WO03013535 (A2), WO03013536 (A2): 20.02.2003 DESIGNATED STATES AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZM ZW, 20.02.2003 DESIGNATED COUNTRIES FOR REGIONAL PATENTS GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW AM AZ BY KG KZ MD RU TJ TM AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LU MC NL PT SE SK TR BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG

Status US2005032724 (A1): 01.04.2009 abandonment

Status AU2002328945 (A1), AU2002328950 (A1), AU2002328952 (A1), AU2002328953 (A1), AU2002331290 (A1): 08.04.2004 PCT APPLIC. NOT ENTERING NATIONAL PHASE

Status CA2454627 (A1), CA2454637 (A1), CA2454640 (A1), CA2454643 (A1), CA2454648 (A1): 23.07.2010 DEAD

Status EP1408972 (A2): 02.06.2010 REFUSED

Status EP1408973 (A2), EP1408974 (A2), EP1408975 (A2), EP1438050 (A2): 18.08.2010 DEEMED TO BE WITHDRAWN

Status JP2005501840 (A), JP2005504759 (A), JP2005508312 (A) : 21.04.2010 DECISION OF REFUSAL

Status JP2005505526 (A): 24.03.2010 DECISION OF REFUSAL

Status JP2005506971 (A): 28.02.2009 WRITTEN WITHDRAWAL OF APPLICATION

4. **ID: US20140194366(A1)**

Title: Ovarian cancer stem cells and methods of isolation and uses thereof.

Priority date: 03/06/ 2001

Publication date: 10/07/2014

Description: The present invention relates to methods, assays, compositions and kits related to the CD44+/CD24+/EpCam+/ECad− subpopulation of ovarian cancer stem cells which are selected or enriched by chemotherapeutic agents and inhibited by MIS (Mullerian Inhibiting Substance) and MIS mimetics.

The inventors also discovered that the 3+E-Cadherin negative population (e.g., CD44+, CD24+, EpCam+, ECad−, here referred to as "3+/Ecad−") have increased growth in vitro and in vivo and resistance to chemotherapeutic agents, and their proliferation and/or growth can be inhibited by Mullerian Inhibiting Substance (MIS) or a small molecule MIS mimetic, SP600125, or analogues or variants thereof.

Relevance to the HERCULES project: Use of specific biomarkers to determine the resistance to chemotherapeutic agents and find new therapeutic approaches.

Inventors: Patricia K. Donahoe, David T. MacLaughlin, Jose Teixeira.

Owner: The General Hospital Corporation.

Also published as: WO2012167101A1.

INPADOC patent family: US2014194366(A1), WO2012167101(A1).

Status US2014194366 (A1): 01.02.2016 Non-Final Rejection

Status WO2012167101 (A1): 25.06.2014 EP: PCT APP. NOT ENT. EUROP. PHASE

5. **ID: AU2014275006(A1)**

Title: Assays, methods and kits for analyzing sensitivity and resistance to anti-cancer drugs, predicting a cancer patient's prognosis, and personalized treatment strategies.

Priority date: 04/06/ 2013

Publication date: 11/12/2014

Description: The application describes assays, methods and kits for analyzing sensitivity of a subject's cancerous tumor to a drug, predicting responses of cancerous tumors to drugs, determining the prognosis of a subject having a cancerous tumor, and developing a personalized therapy or treatment strategy for the subject. The assays, methods and kits involve analyzing gene and protein expression signatures or profiles of a subject's cancerous tumor, testing candidate drugs in cancerous cells from the subject's cancerous tumor, and classifying a subject's cancerous tumor based on ovarian cell and fallopian tube cell cell-of-origin gene expression signatures. Using these methods, a suitable drug (or

drugs) is identified, the subject can be treated with that drug, and a personalized therapy can be developed for the subject.

Relevance to the HERCULES project: Use gene and protein expression signatures to develop a personalized therapy or treatment strategy for cancer affected patients.

Inventors: INCE TAN A.

Owner: University Of Miami.

Also published as: WO2014197543 (A1), CA2914026 (A1).

INPADOC patent family: AU2014275006(A1), CA2914026(A1), WO2014197543(A1).

Status AU2014275006 (A1): not available in the Espacenet database

Status WO2014197543 (A1): 24.12.2015 ENTRY INTO THE NATIONAL PHASE IN AU, 04.12.2015 NON-ENTRY INTO THE NATIONAL PHASE IN DE, 27.11.2015 ENTRY INTO THE NATIONAL PHASE IN CA

Status CA2914026 (A1): not available in the Espacenet database

6. **ID: US20090298068(A1)**

Title: Method and test kit for the diagnosis and/or making predictions about and/or for the assessment of the efficacy of therapeutic agents for the treatment of ovarian cancer and method of planning a regimen for the treatment of ovarian cancer

Priority date: 16/03/2006

Publication date: 03/12/2009

Description: The invention relates to a method and a test kit for diagnosing ovarian cancer and/or making predictions in case of ovarian cancer as well as a method for estimating the effectiveness of therapeutic agents during the treatment of ovarian cancer, the promoter hypermethylation of the TUSC3 marker in a biological sample that is to be analyzed, preferably in a tissue sample or biological liquid that is to be analyzed, being measured. The result of said method can be used for planning an ovarian cancer treatment.

Relevance to the HERCULES project: method to correlate ovarian cancer biomarker to estimate the effectiveness of therapeutic agents during the treatment of ovarian cancer.

Inventors: Michael Krainer, Dietmar Pils, Robert Zeillinger.

Owner: Michael Krainer, Dietmar Pils, Robert Zeillinger

Also published as: EP1994173A2, EP1994173B1, WO2007104571A2, WO2007104571A3.

INPADOC patent family: US2009298068(A1), AT503411(A1), AT503411(B1), EP1994173(A2), EP1994173(B1), WO2007104571(A2), WO2007104571(A3).

Status US2009298068 (A1): 23.02.2012 Abandonment

Status WO2007104571 (A2), WO2007104571 (A3): 17.09.2008 NON-ENTRY INTO THE NATIONAL PHASE IN DE

Status EP1994173 (A2), EP1994173B1: (VARYING DATES) LU HU TR AT LI DE GB FR CH MT IE MC BE BG RO SK IT PL NL ES EE DK CZ PT SI CY LV SE GR IS CY FI LT IS LAPSED

Status AT503411 (A1), AT503411 (B1): 15.11.2013 LAPSE BECAUSE OF NOT PAYING ANNUAL FEES

7. **ID: US20160047000(A1)**

Title: Methods and systems for treatment of ovarian cancer.

Priority date: 21/03/2013

Publication date: 18/02/2016

Description: The inventors have identified gene signatures (In some embodiments, the marker gene(s) is selected from the group consisting of MMP2, TIMP3, ADAMTS1, VCL, TGFB1, SPARC, CYR61; EGR1, SMADs; GLIB, VCAN, CNY61, LOX, TAFs, ACTA2, POSTN, CXCL14, CCL13, FAP, NUAK1, PTCH1, TGFBR2; and TNFAIP6. In some embodiments, the marker gene(s) is selected from the group consisting of POSTN, CXCL14, CCL13, FAP, NUAK1, PTCH1, TGFBR2; and TNFAIP6. In some embodiments, the marker is the level of phosphorylated SMAD2 and/or SMAD3. In some embodiments, the assays, methods, and systems described herein are directed to determination of the expression level of a gene product of at least two genes in a biological sample of a subject, e.g. at least two genes, or at least three genes, or at least four genes, or, e.g. all of the following genes: POSTN, CXCL14, CCL13, FAP, NUAK1, PTCH1, TGFBR2; and TNFAIP6). which permit the identification of patients who will benefit from (e.g. have optimal outcomes) cytoreductive surgery as treatment for ovarian cancer. Accordingly, provided herein are methods of treatment, assays, and systems relating to ovarian cancer and the administration of cytoreductive surgery. In one aspect, the technology described herein relates to a method of treatment comprising, detecting, in a sample obtained from a subject in need of treatment for ovarian cancer, the level of activation of at least one pathway, and administering cytoreductive surgery to the subject if the level of activation is not increased relative to a reference level.

Relevance to the HERCULES project: Use of specific gene signatures which permit the identification of patients who will benefit from cytoreductive surgery as treatment for ovarian cancer.

Inventors: Michael Birrer, Giovanni Luigi PARMIGIANI, Markus Riester, Wei Wei.

Owner: General Hospital Corporation, Dana-Farber Cancer Institute, Inc.

Also published as: WO2014153442A2, WO2014153442A3.

INPADOC patent family: US2016047000(A1), WO2014153442(A2), WO2014153442(A3).

Status US2016047000 (A1): 15.01.2015 ASSIGNEMENT

Status WO2014153442 (A2) WO2014153442(A3): 22.10.2015 EP Application deemed to be withdrawn

8. **ID: WO2012019000(A2)**

Title: Biomarkers for the identification monitoring and treatment of ovarian cancer.

Priority date: 04/10/2010

Publication date: 09/02/2012

Description: The present invention relates:

A) the discovery that certain biological markers (referred to herein as "DNARMARKERS"), such as proteins, nucleic acids, polymorphisms, metabolites, protein modifications, nucleic acid modifications, chromosomes, and other analytes, as well as certain physiological conditions and states, present or altered in subjects with an increased risk of developing a recurrent ovarian cancer.

B) Method with a predetermined level of predictability for monitoring the effectiveness of treatment or selecting a treatment regimen for ovarian cancer by detecting the level of an effective amount of DNARMARKERS in a first sample from the subject at a first period of time and optionally detecting the level of an effective amount of DNARMARKERS in a second sample from the subject at a second period of time. The level of the effective amount of DNARMARKERS detected at the first period of time is compared to the level detected at the second period of time or alternatively a reference value. Effectiveness of treatment is monitored by a change in the level of the effective amount of DNARMARKERS from the subject. Treatment is for example, chemotherapy and /or radiotherapy. Chemotherapeutic agents include a platinating agent (e.g., carpoplatin or oxiplatin), taxane, a PARP inhibitor and any combination of these.

Relevance to the HERCULES project: Methods of treating ovarian cancer and methods of accessing/monitoring the responsiveness of a cancer cell to a therapeutic compound.

Inventors: William E. Pierceall, Kam Marie Sprott, David T. Weaver.

Owner: On-Q-ity.

Also published as: WO2012019000A3.

INPADOC patent family: WO2012019000(A2), WO2012019000(A3).

Status WO2012019000 (A2), WO2012019000 (A3): 28.08.2013 EP PCT APP. NOT ENT. EUROP. PHASE

9. **ID: EP2676678(A1)**

Title: Methods to identify and enrich populations of ovarian cancer stem cells and somatic stem cells and uses thereof.

Priority date: 17/07/2007

Publication date: 25/12/2013

Description: The present invention provides methods for treating and/or preventing ovarian cancer in a subject by administering to the subject an effective amount of Mullerian Inhibiting substance and/or an effective amount of an agent that inhibits BCRP1. The present invention further provides methods to identify and/or enrich for populations of ovarian cancer stem cells and populations of somatic ovarian stem cells, in particular, enrichment for populations of coelomic somatic ovarian stem cells, subcoelomic/stromal somatic ovarian stem cells and periphilar medullary somatic ovarian stem cells. The present invention also provides somatic ovarian stem cell markers and ovarian cancer stem cell markers, as well as methods to identify agents which selectively inhibit the proliferation of ovarian cancer stem cells as compared to somatic ovarian stem cells.

Relevance to the HERCULES project: Methods for treating and/or preventing ovarian cancer in a subject by administering to the subject an effective amount of Mullerian Inhibiting substance and/or an effective amount of an agent that inhibits BCRP1

Inventors: Patrick Donahoe, Paul Szotek, David Maclaughlin, Rafael Pieretti-Vanmarcke, David Dombkowski, Frederic Preffer.

Owner: The General Hospital Corporation.

Also published as: CA2693144A1, EP2190481A2, EP2190481A4, EP2190481B1, EP2764874A1, US20100273160, WO2009012357A2, WO2009012357A3.

INPADOC patent family: EP2676678(A1), CA2693144(A1), EP2190481(A2), EP2190481(A4), EP2190481(B1), EP2764874(A1), US2010273160(A1), US9289492(B2), WO2009012357(A2), WO2009012357(A3).

Status EP2676678 (A1): 18.06.2014 DEEMED TO BE WITHDRAWN

Status WO2009012357 (A2) WO2009012357 (A3): 15.01.2010 CA ENTRY INTO NATIONAL PHASE, 19.01.2010 NON-ENTRY INTO THE NATIONAL PHASE IN DE, 07.07.2010 US ENTRY INTO NATIONAL PHASE

Status US2010273160 (A1), US9289492 (B2): 04.02.2016 ASSIGNEMENT

Status EP2764874 (A1): 06.05.2015 FIRST EXAMINATION REPORT

Status EP2190481 (A2), EP2190481 (A4), EP2190481 (B1): 29.02.2016 CH PATENT CEASED, (VARYING DATES) SI MC IT DK PL AT IS RO ES EE CZ SK NL HR SE GR LV LT FI NO LAPSED, 02.12.2015 NO OPPOSITION FILED, 30.11.2015 FR POSTGRANT ANNUAL FEES PAID TO NATIONAL OFFICE, 30.10.2015 GB POSTGRANT ANNUAL FEES PAID TO

NATIONAL OFFICE, 30.10.2015 DE POSTGRANT ANNUAL FEES PAID TO NATIONAL OFFICE

Status CA2693144 (A1): 24.07.2013 EXAMINATION REQUEST

10. ID: EP2156187(A1)

Title: Methods and compositions for diagnosis and/or prognosis of ovarian cancer and lung cancer.

Priority date: 15/06/2007

Publication date: 24/02/2010

Description: Methods and compositions for diagnosis, prognosis and monitoring of ovarian cancer and lung cancer are provided. Assays that detect NHERF-I (or one or more markers related thereto) and NHERF-I -containing complexes are used to assign a diagnosis to a subject being assessed for the presence of ovarian or lung cancer; assign a prognostic risk to a subject suffering from ovarian or lung cancer; or monitor the course of ovarian or lung cancer treatment in a subject.

Relevance to the HERCULES project: Methods relating to the identification and use of markers (NHERF-I) for the diagnosis of ovarian cancer, for stratification of risk in ovarian cancer patients, and for monitoring therapy in ovarian cancer patients.

Inventors: Kelline M. Rodems, David W. Oelschlager, Uday Kumar Veeramallu, Joseph Buechler.

Owner: Biosite Incorporated.

Also published as: EP2156187A4, US20100272635, WO2008157383A1.

INPADOC patent family: EP2156187(A1), EP2156187(A4), US2010272635(A1), WO2008157383(A1).

Status EP2156187 (A1) EP2156187 (A4): 25.08.2010 WITHDRAWN

Status US2010272635 (A1): 13.03.2013 Abandonment

Status WO2008157383 (A1): 06.07.2010 US ENTRY INTO NATIONAL PHASE, 16.12.2009 NON-ENTRY INTO THE NATIONAL PHASE IN DE

11. ID: WO2013188391(A3)

Title: Biomarkers for ovarian cancer.

Priority date: 11/06/2012

Publication date: 19/12/2013

Description: Biomarkers are provided that are useful for the detection or diagnosis of ovarian cancer. The biomarkers are also useful for determining whether the ovarian cancer is active, is in remission, or is recurring. One embodiment provides a method for assessing therapeutic outcome of a treatment for ovarian cancer by determining the amount of one or more proteins in a blood sample from a

subject in ovarian cancer remission, wherein the one or more proteins are selected from the group consisting of sICAM, sVCAMl, sTNFR-II, sgpl30, MMP2, and combinations thereof, and wherein elevated serum amounts of the one or more proteins relative to a control indicates that the subject has poor overall survival relative to subjects in remission for ovarian cancer having lower serum amounts of the one or more serum proteins. Typically groups of 3 to 5 of these markers are assayed.
Relevance to the HERCULES project: Use of specific biomarkers to determine if ovarian cancer is active, is in remission, or is recurring and to provide a method for assessing therapeutic outcome of a treatment for ovarian cancer.
Inventors: Jin-Xiong She, Ashok SHARMA.
Owner: Georgia Regents University.
Also published as: US20150276747, WO2013188391A2.
INPADOC patent family: WO2013188391(A2), WO2013188391(A3), US2015276747(A1).
Status WO2013188391 (A2) WO2013188391 (A3): 01.07.2015 EP PCT APP. NOT ENT. EUROP. PHASE
Status US2015276747 (A1): 30.12.2014 ASSIGNMENT

12. ID: WO2013166480(A1)

Title: Detectors of serum biomarkers for predicting ovarian cancer recurrence.
Priority date: 04/05/2012
Publication date: 07/11/2013
Description: The present invention provides polypeptide marker antigens (ex CA125 and a pull of specific nucleotide sequences) for detecting the presence of autoantibody biomarkers associated with a risk of ovarian cancer recurrence in serum. The present invention also provides an antibody binding assay for detecting the presence of autoantibody biomarkers associated with a risk of ovarian cancer recurrence. The present invention also provides a method for treating ovarian cancer recurrence in an ovarian cancer patient on the basis of the determined risk of recurrence.
Relevance to the HERCULES project: Use of polypeptide marker antigens for detecting the presence of autoantibody biomarkers and providing a method for treating ovarian cancer recurrence.
Inventors: Michael Tainsky, Madhumita Chatterjee, Gregory DYSON, Nancy Levin
Owner: Wayne State University.
Also published as: CA2911257A1, US20150139980.
INPADOC patent family: WO2013166480(A1), CA2911257(A1), US2015139980(A1).

Status WO2013166480 (A1): 04.11.2015 ENTRY INTO THE NATIONAL PHASE IN CA, 27.05.2015 EP PCT APP. NOT ENT. EUROP. PHASE , 04.11.2014 US ENTRY INTO NATIONAL PHASE

Status US2015139980 (A1): 25.02.2016 Requirement for Restriction/Election

13. ID: EP2683835(A2)

Title: Molecular biomarker set for early detection of ovarian cancer.

Priority date: 08/03/2011

Publication date: 15/01/2014

Description: Embodiments of the present invention concern methods and compositions related to detection of ovarian cancer, including detection of the stage of ovarian cancer, in some cases. In particular, the invention encompasses use of expression of TFAP2A and in some embodiments CA125 and/or E2F5 to identify ovarian cancer, including detecting mRNA and/or protein levels of the respective gene products. The levels of TFAP2A and optionally CA125, and/or E2F5 are utilized for determination of a treatment regimen for ovarian cancer. In some embodiments of the invention, the methods are employed to determine if an individual should be given a particular cancer therapy, for example whether an individual would be resistant to a particular cancer drug.

Relevance to the HERCULES project: Use of CA125 and/or E2F5 mRNA and/or protein levels to identify ovarian cancer and to determine a particular ovarian cancer therapy.

Inventors: Vladimir BAJIC, Mandeep KAUR.

Owner: King Abdullah University Of Science And Technology.

Also published as: CA2829300A1, US9057107, US20120231090, US20150284809, WO2012120377A2, WO2012120377A3, WO2012120377A9.

INPADOC patent family: EP2683835(A2), AU2012226530(A1), CA2829300(A1), US2012231090(A1), US9057107(B2), US2015284809(A1), WO2012120377(A2), WO2012120377(A9), WO2012120377(A3).

Status EP2683835 (A2): 29.10.2014 FIRST EXAMINATION REPORT

Status AU2012226530 (A1): not available in the Espacenet database

Status CA2829300 (A1): not available in the Espacenet database

Status US2012231090 (A1) US9057107 (B2): 18.09.2013 ASSIGNMENT

Status US2015284809 (A1): 09.06.2015 ASSIGNMENT

Status WO2012120377 (A2) WO2012120377 (A9) WO2012120377 (A3): 26.09.2013 ENTRY INTO THE NATIONAL PHASE IN AU, 06.09.2013 ENTRY INTO THE NATIONAL PHASE IN CA

14. ID: WO2013096661(A1)

Title: Methylation biomarkers for ovarian cancer .

Priority date: 22/12/2011

Publication date: 27/06/2013

Description: The present disclosure provides differentially methylated genomic CpG dinucleotide sequences associated with cancer. In particular, differentially methylated genomic CpG nucleotides and their use in diagnostic and prognostic methods for ovarian cancer are disclosed. Biomarkers and methods of their use can also be used to assess an individual prior to, during and/or subsequent to a treatment course for eradicating ovarian cancer (e.g., chemotherapy, radiation therapy, drug therapy, etc.) for determining the effectiveness of a treatment course.

Relevance to the HERCULES project: Use of differentially methylated genomic CpG nucleotides in diagnostic and prognostic methods for ovarian cancer and provide a treatment to eradicate ovarian cancer.

Inventors: Jian-Bing Fan, Marina Bibikova, Thomas Royce, Jeremy R. Chien, Lynn Hartmann, Vijayalakshmi Shridhar.

Owner: Illumina, Inc., Mayo Foundation For Medical Education And Research.

Also published as: No further publications.

INPADOC patent family: WO2013096661(A1).

Status WO2013096661 (A1): 14.01.2015 EP PCT APP. NOT ENT. EUROP. PHASE

15. ID: US20130045947(A1)

Title: Pla2activity as a marker for ovarian and other gynecologic cancers.

Priority date: 11/02/2010

Publication date: 21/02/2013

Description: Materials and Methods are provided for the diagnosis, monitoring, and personalized treatments of gynecological cancers. The methods comprise determining levels of PLA2 activity in sample of tissue or fluid recovered from patient; elevated levels of PLA2 activity are consistent with epithelial ovarian cancer (EOC). These methods include assaying for PLA2 activity within tissue, ascites, blood, and other tissue forms by exposing the patient sample to a fluorogenic compound such as DBPC. The methods disclosed herein further include correlating the fluorogenic detection with a disease state in the patient, including diseases such as gynecological cancers, such as EOC. The methods comprise determining levels of total PLA2 activity, and of specific isoforms of PLA2 such as iPLA2, iPLA2β, cPLA2, among other isoforms.

Relevance to the HERCULES project: Methods are provided for the diagnosis, monitoring, and personalized treatments of gynecological cancers using PLA2 activity as biomarker.

Inventors: Yan Xu

Owner: Indiana University Research And Technology Corporation.

Also published as: WO2011100639A2, WO2011100639A3.

INPADOC patent family: US2013045947(A1), WO2011100639(A2), WO2011100639(A3).

Status US2013045947 (A1): 10.03.2014 Abandonment

Status WO2011100639 (A2) WO2011100639 (A3): 06.03.2013 EP PCT APP. NOT ENT. EUROP. PHASE

16. ID: CA2691980(A1)

Title: Predictive markers for ovarian cancer.

Priority date: 29/06/2007

Publication date: 08/01/2009

Description: The present invention generally relates to cancer biomarkers and particularly to biomarkers associated with ovarian cancer. It provides methods to predict, valuate diagnose, and monitor cancer, particularly ovarian cancer, by measuring certain biomarkers, and further provides a set or array of reagents to evaluate the expression levels of biomarkers that are associated with ovarian cancer. A preferred set of biomarkers provides a detectable molecular signature of ovarian cancer in a subject. The invention provides a predictive or diagnostic test for ovarian cancer, particularly for epithelial ovarian cancer and more particularly for early-stage ovarian cancer. The present invention also provides methods to assess the therapeutic efficacy of existing and candidate chemotherapeutic agents and other types of cancer treatments.

Relevance to the HERCULES project: Use of ovarian cancer biomarkers panel (containing for example C reactive protein, FGF basic protein and CA 19-9) to predict, valuate diagnose, monitor and assess the therapeutic efficacy on ovarian cancer.

Inventors: Brian C. Mansfield, Ping Yip, Suraj Amonkar, Greg P. Bertenshaw.

Owner: Correlogic Systems, Inc., Brian C. Mansfield, Ping Yip, Suraj Amonkar, Greg P. Bertenshaw, Vermillion, Inc.

Also published as: CN101855553A, CN101855553B, EP2171453A1, EP2171453A4, EP2637020A2, EP2637020A3, US8664358, US9274118, US20090004687, US20140221240, WO2009006439A1.

INPADOC patent family: CA2691980(A1), CN101855553(A), CN101855553(B), EP2171453(A1), EP2171453(A4), EP2637020(A2), EP2637020(A3), JP2010532484(A), KR20100062996(A),

KR101262202(B1), KR20120087885(A), MY150234(A), SG182976(A1), US2009004687(A1), US8664358(B2), US2014221240(A1), US9274118(B2), WO2009006439(A1).

Status CA2691980 (A1): 15.03.2013 EXAMINATION REQUEST

Status CN101855553 (A) CN101855553 (B): 11.06.2014 GRANT OF PATENT OR UTILITY MODEL

Status EP2171453 (A1) EP2171453 (A4): 30.05.2012 DEEMED TO BE WITHDRAWN, 01.05.2013 DATE OF RECEIPT OF REQUEST FOR RE-ESTABLISHMENT OF RIGHTS, 30.03.2016 DEEMED TO BE WITHDRAWN

Status EP2637020 (A2) EP2637020 (A3): 18.02.2015 FIRST EXAMINATION REPORT

Status JP2010532484 (A): 07.10.2013 REMOVAL OF RECONSIDERATION BY EXAMINER BEFORE APPEAL

Status KR20100062996 (A) KR101262202 (B1): 02.05.2013 WRITTEN DECISION TO GRANT

Status KR20120087885 (A): 13.02.2012 WITHDRAWAL DUE TO NO REQUEST FOR EXAMINATION

Status MY150234 (A): not available in the Espacenet database

Status SG182976 (A1): not available in the Espacenet database

Status US2009004687 (A1) US8664358 (B2): 24.03.2016 ASSIGNMENT

Status US2014221240 (A1) US9274118 (B2): 24.03.2016 ASSIGNMENT

Status WO2009006439 (A1): 20.12.2013 Request for re-establishment of rights, 30.03.2016 The application is deemed to be withdrawn

17. ID: WO2015042115(A1)

Title: Biomarkers for ovarian cancer.

Priority date: 17/09/2013

Publication date: 26/03/2015

Description: Biomarkers and biomarker panels are provided for making ovarian cancer assessments, for example, diagnosing an ovarian cancer, predicting responsiveness of an ovarian cancer to an ovarian cancer therapy, and monitoring an ovarian cancer. A report may be provided to the patient of the assessment. Also provided are methods, reagents, devices and kits for the use of these biomarkers in making ovarian cancer assessments. Patients can further be treated with in accordance with the assessment of responsiveness.

The inventors have identified two proteins, Prkdc and Rad54L, that are represented at elevated levels in blood samples of subtypes of ovarian cancers, and thus, that find use as biomarkers in providing an ovarian cancer assessment, e.g. diagnosing an ovarian cancer, prognosing an ovarian cancer,

determining a treatment for a subject affected with ovarian cancer, monitoring a subject with ovarian cancer, and the like. The PRKDC gene, also known as "protein kinase, DNA-activated, catalytic polypeptide", DNA-PKcs, HYRC, p350, DNAPK, DNPK1 , HYRC1 , and XRCC7, encodes the catalytic subunit of the DNA-dependent protein kinase (DNA-PK). It functions with the Ku70/Ku80 heterodimer protein in DNA double strand break repair and recombination. The protein encoded is a member of the PI3/PI4-kinase family. The cDNA and protein sequences for PRKDC may be found at Genbank Accession No. NM_006904.6. The RAD54L gene, also known as "RAD54-like", HR54, hHR54, RAD54A, and hRAD54, encodes a protein that belongs to the DEAD-like helicase superfamily, and shares similarity with Saccharomyces cerevisiae Rad54, a protein known to be involved in the homologous recombination and repair of DNA, including DNA double strand break repair. The binding of this protein to double-strand DNA induces a DNA topological change, which is thought to facilitate homologous DNA paring, and stimulate DNA recombination. The cDNA and protein sequences for RAD54L may be found at Genbank Accession No. NM_003579.3.

Relevance to the HERCULES project: Use of Biomarkers and biomarker panels based on gene sequences and proteins for diagnosing an ovarian cancer, predicting responsiveness of ovarian cancer to an ovarian cancer therapy and monitoring an ovarian cancer.

Inventors: Atul J. Butte, Linda Anne SZABO, Purvesh Khatri, Bruce Xuefeng Ling.

Owner: The Board Of Trustees Of The Leland Stanford Junior University.

Also published as: No further publications.

INPADOC patent family: WO2015042115(A1).

Status WO2015042115 (A1): 07.03.2016 EP National basic fee paid, Search fee paid, Designation fee(s) paid, Examination fee paid

18. ID: EP2906724(A1)

Title: Method of prognosis and stratification of ovarian cancer.

Priority date: 12/10/2012

Publication date: 19/08/2015

Description: The present invention proposes, in general terms, methods, systems and kits for providing a prognosis of overall survival or prediction of therapeutic outcome (for example, chemotherapeutic outcome) for a patient suffering from epithelial ovarian cancer, in which expression of let-7b and/or miRNAs with which it is associated and/or genes within which it is associated are used to provide the prognosis and/or prediction of the therapeutic outcome. In another aspect the invention proposes methods and systems for identifying miRNA and/or gene signatures for use in a prognosis or and/or prediction of the therapeutic outcome. Embodiments use the biologically and

clinically relevant 36-mRNA prognostic signature as a high-confidence prognostic tool to significantly stratify HG-EOC patients into three survival- significant, molecularly different and clinically distinct subclasses, which can improve patient risk assessment, management and counseling, as well as provide a solution for the optimization of personalized medicine strategy of treating human ovarian cancers in a clinical setting. Embodiments relate to a method of prognosis and outcome prediction of high-grade epithelial ovarian cancer (HG-EOC) based on the measurements of microRNA let-7b, the 21 let-7b associated miRNAs and the 36 let-7b associated mRNAs in the patient tumor samples. Embodiments relate to the methods of identification and use of the resulting gene or microRNA signatures.

Relevance to the HERCULES project: Use of miRNA panels(for example let-7b) and gene sequences for providing a prognosis of overall survival or prediction of therapeutic outcome in patients with epithelial ovarian cancer.

Inventors: Vladimir Andreevich KUZNETSOV, Zhiqun Tang, Ghim Siong OW, Anna Vladimirovna IVSHINA.

Owner: Agency For Science, Technology And Research.

Also published as: CN104854247A, US20150267259, WO2014058394A1.

INPADOC patent family: EP2906724(A1), CN104854247(A), SG11201502778T(A), US2015267259(A1), WO2014058394(A1).

Status EP2906724 (A1): 30.03.2016 Request for examination was made

Status CN104854247 (A): 16.09.2015 DECISION MADE BY SIPO TO INITIATE SUBSTANTIVE EXAMINATION

Status SG11201502778T (A) SG11201502778T (A): not available in the Espacenet database

Status US2015267259 (A1): 07.03.2016 Non-Final Rejection

Status WO2014058394 (A1): 30.03.2016 EP Request for examination was made

19. ID: WO2014062978(A1)

Title: Molecular signatures of ovarian cancer.

Priority date: 17/19/2012

Publication date: 24/04/2014

Description: Described herein is a method of determining a prognosis of cancer in an individual, including determining the presence or absence of a high level of expression in the individual relative to a normal baseline standard for a single prognostic panel of the following markers, ACTA2, ADAM 12, AEBPl, COLl lAl, COL3A1, COL5A1, COL6A2, CYR61, DCN, FN1, GREM1, LOX, LUM, POSTN, SNAI2, SPARC, TAGLN, THBS2, TIMP3, VCAN, and/or VIM, and prognosing a case of

cancer if the individual demonstrates the presence of a high level of expression relative to a normal baseline standard of at least one of the markers. In other embodiments, the individual demonstrates the presence of a high level of expression relative to a normal baseline standard of at least two, three, four, or five of the markers. In other embodiments, the individual demonstrates the presence of a high level of expression relative to a normal baseline standard of at least six, seven, eight, nine, ten or more of the markers. In other embodiments, the cancer is ovarian cancer. In other embodiments, the prognosis provides a therapeutic selection for the prognosed individual, selected from the group consisting of: chemotherapy, radiotherapy, surgery, and combinations thereof.

Relevance to the HERCULES project: Use of ovarian cancer biomarkers for determining a prognosis of cancer and personalized treatment strategies focused on molecular subtypes of gynecological cancers.

Inventors: Sandra ORSULIC, Beth Y. KARLAN, Xiaojian CUI, Mourad TIGHIOUART, Dong-Joo CHEON.

Owner: Cedars-Sinai Medical Center.

Also published as: CA2886607A1, EP2908913A1, US20150322530.

INPADOC patent family: WO2014062978(A1), AU2013331154(A1), CA2886607(A1), EP2908913(A1), US2015322530(A1).

Status WO2014062978 (A1): 03.12.2015 EP Amendment by applicant (claims and/or description)

Status AU2013331154 (A1): not available in the Espacenet database

Status CA2886607 (A1): not available in the Espacenet database

Status EP2908913 (A1): 30.03.2016 EP Request for examination was made

Status US2015322530 (A1): 04.05.2015 ASSIGNMENT

20. ID: EP2721417(A2)

Title: New markers for early diagnosis of ovarian cancer, monitoring during therapy, and new therapy options during and after chematherapy .

Priority date: 17/06/2011

Publication date: 23/04/2014

Description: The inventors have identified several proteases and a protease inhibitor (CA125, TADG14, TADG15, TADG12, SCCE, MMP-7, ALP, KLK6, and hepsin) that are overexpressed in ovarian cancer tumors. They have developed monoclonal antibodies against the proteins and shown that they can be detected in serum and the levels of the proteins in serum fluctuate during cancer treatment. They have shown that serum assays for the proteases and protease inhibitor can be used for early detection of ovarian cancer, and for monitoring cancer treatment.

Relevance to the HERCULES project: Use of biomarkers for early detection of ovarian cancer, and for monitoring cancer treatment.

Inventors: Timothy J. O'brien, John Beard, Wilbur C. HITT.

Owner: The Board of Trustees of The University of Arkansas.

Also published as: EP2721417A4, US20140193426, WO2012174569A2, WO2012174569A3.

INPADOC patent family: EP2721417(A2), EP2721417(A4), US2014193426(A1), WO2012174569(A2), WO2012174569(A3).

Status EP2721417 (A2) EP2721417 (A4): 06.05.2015 SUPPLEMENTARY SEARCH REPORT

Status US2014193426 (A1): 10.02.2016 Non-Final Rejection

Status WO2012174569 (A2) WO2012174569 (A3): 15.12.2015 EP Application deemed to be withdrawn: despatch of communication + time limit, 30.03.2016 Request for examination was made

21. ID: US8715665(B2)

Title: Methods for treating cancer resistant to erbb therapeutics.

Priority date: 13/04/2007

Publication date: 06/05/2014

Description: Provided herein are methods for treating cancer (ex: lung cancer, brain cancer, breast cancer, head and neck cancer, colon cancer, ovarian cancer, gastric cancer, or pancreatic cancer) that is resistant to treatment with an anti-ErbB therapeutic agent and which is associated with an activating MET gene mutation or a MET gene amplification. The methods involve administering to a subject a combination of an anti-ErbB therapeutic and an anti-MET therapeutic. Also provided are methods for reducing ErbB mediated signaling or PI3 kinase mediated signaling in a cancer cell.

Relevance to the HERCULES project: methods for treating cancer (including ovarian cancer) using ErbB as biomarker.

Inventors: Pasi A. Janne, Jeffrey Engelman, Lewis C. Cantley.

Owner: The General Hospital Corporation, Dana Farber Cancer Institute, Beth Israel Deaconess Medical Center.

Also published as: CA2683559A1, EP2076289A2, EP2076289B1, EP2851091A1, US20120225870, US20140296100, WO2008127710A2, WO2008127710A3.

INPADOC patent family: US2012225870(A1), US8715665(B2), AU2008239594(A1), AU2008239594(B2), CA2683559(A1), DK2076289(T3), EP2076289(A2), EP2076289(B1), EP2851091(A1), ES2529790(T3), JP2010523680(A), JP2014141530(A), US2014296100(A1), WO2008127710(A2), WO2008127710 (A3).

Status US2012225870 (A1) US8715665 (B2): 29.09.2015 CERTIFICATE OF CORRECTION
US2014296100 (A1): 11.09.2015 Final Rejection
Status AU2008239594 (A1) AU2008239594 (B2): 27.02.2014 LETTERS PATENT SEALED OR GRANTED (STANDARD PATENT)
Status AU2013203111 (A1) AU2013203111 (B2): 04.02.2016 LETTERS PATENT SEALED OR GRANTED (STANDARD PATENT)
Status CA2683559 (A1): 22.04.2013 EXAMINATION REQUEST
Status DK2076289 (T3): not available in the Espacenet database
Status EP2076289 (A2) EP2076289 (B1): (VARYING DATES) SI IE LU MC CZ EE SK RO LV GR PL AT HR SE CY PT LT IS NL FI NO LAPSED, (VARYING DATES) IT FR DE GB CH DK ES POSTGRANT: ANNUAL FEES PAID TO NATIONAL OFFICE,
Status EP2851091 (A1): 07.10.2015 Amendment by applicant
Status ES2529790 (T3): not available in the Espacenet database
Status JP2010523680 (A): 01.03.2016 WRITTEN AMENDMENT
Status JP2014141530 (A): 29.01.2016 DECISION OF REFUSAL
Status US2014296100 (A1): 11.09.2015 Final Rejection
Status WO2008127710 (A2) WO2008127710 (A3): 30.03.2016 EP No opposition filed within time limit

22. ID: WO2012151277(A1)

Title: Kits and methods for selecting a treatment for ovarian cancer
Priority date: 02/05/2011
Publication date: 08/11/2012
Description: The present invention relates to methods for the selecting a treatment for a subject having ovarian cancer, in particular methods to distinguish between ovarian cancer patients who will respond to the taxol/platinum chemotherapy and survive longer than seven years versus those who will succumb to the disease within three years.
Relevance to the HERCULES project: Use levels of gene expression of a specific genes as biomarkers for selecting a treatment of ovarian cancer.
Inventors: Kristin Louise Murgic BOYLAN, Walter Cheney LOW, Jason Basil NIKAS, Amy Patrice SKUBITZ
Owner: Applied Informatic Solutions, Inc.
Also published as: No further publications.
INPADOC patent family: WO2012151277(A1).

Status WO2012151277 (A1): 28.05.2014 EP PCT APP. NOT ENT. EUROP. PHASE