UNIVERSITY OF INSUBRIA
EXPERIMENTAL AND APPLIED MEDICINE PhD
Cycle XXIX

METAGENOMICS METHODS FOR THE ANALYSIS OF HUMAN MICROBIOTA TOWARDS A CLINICAL APPLICATION.

Tutor: Prof. Mauro FASANO
Correlator: Prof. Lorenzo DRAGO

PhD Thesis:
Roberta DE GRANDI
Matricola n. 789218

Academic Year 2016- 2017

Contents

Abstract

Several microbial communities colonize different body areas constituting the human microbiota. In recent years, several researches suggested the involvement of these bacteria in maintaining homeostasis of the human body, highlighting as a microbial dysbiosis may be implicated in numerous inflammatory conditions and human diseases. Although "dysbiosis" is referred to an alteration of the "normal microbiota community structure", to date there is not a specific microbial composition, definable as "normobioma", that can characterize and identify a physiological microbiome. The development of high-throughput sequencing technologies and their application to the study of human microbiome allowed the achievement of important results about the interactions between microorganisms and the host. Moreover, metagenomics studies extended scientists' knowledge about the human microbiota and its implications in numerous human diseases. Since the 16S rRNA gene sequencing was shown to be an effective method to investigate microbial communities, recently our laboratory adopted this molecular approach to characterize the interactions between bacteria and host. Therefore, the present PhD work describes our experience in samples collection, experimental procedures and data analysis in studying the human microbiota by means of a metagenomics approach. Indeed, our studies allowed us to understand that first of all, the study population should be well characterized using specific and very strict inclusion and exclusion criteria. In particular, to achieve a complete overview of microbial communities involved in a specific fisio-pathological condition, a standardization of experimental protocols and data analysis should be adopted to carry out metagenomics surveys. Moreover, the taxonomic evaluations through 16S rRNA gene sequencing are just a first step for elucidating the roles of bacteria in the host's health. Indeed, the study and analysis of microbial networks should be usually adopted to clarify and better understand the interactions between microorganisms.

Finally, integrating the metagenomics data which those obtained by metatranscriptomic, metaproteomic, and metametabolomic represents a promising approach to develop new noninvasive bacterial-related diagnostic tools.

Chapter 1

AIM OF THE STUDY

The main objective of the present PhD study was the characterization of human microbiota by means of a Next Generation Sequencing approach. In particular, the research activities were focused on the application and study of different analytic methods useful to investigate and better characterize the human microbiota. The final goal was to improve and deepen our skills and knowledge about interactions between bacteria and host and to evaluate a potential application of metagenomics in clinical microbiology.

Consequently, the first part of this thesis gives an introduction about human microbiota and the methodology used to sequence biological samples. Results and manuscripts realized during the PhD period are included in the present work, while general considerations and conclusions about the characterization and analysis of human microbiota are discussed in the conclusive part.

Chapter 2

INTRODUCTION

2.1 Human microbiota

The human microbiota consists of an incredibly rich and diverse set of microbial populations living in symbiosis with the host organism and colonizing numerous districts of human body including the gut, skin, urogenital tract, oral and nasopharynx cavity. In the last years, numerous studies highlighted its importance for the overall human health. Indeed, today human microbiota, and in particular intestinal and cutaneous ones, is considered the "health organ" since it is involved in different protective, structural and metabolic processes essential for the maintenance of human health. For instance, intestinal bacteria allow the exploitation of numerous food substances that otherwise human enzymes would not be able to digest. In addition, these microorganisms stimulate host's immune system contributing to its defense against potentially pathogenic bacteria.

It has been suggested that each individual acquires his own microbiota already during the gestational period, when maternal microorganisms colonize the fetus, reaching his intestinal environment [1]. Subsequently, after birth, newborn's microbiota is influenced by the different mode of delivery, environmental microorganisms and breastfeeding [2]. Several evidences highlighted that the microbial community contained in breast milk may carry out important functions in modulating both metabolic and immunological activities of newborn. Indeed, breast milk is considered essential not only to fulfil nutritional demands of the baby and enable his rapid growth, but especially to stimulate his immune system and transfer beneficial bacteria able to confer protection against potential pathogens. Therefore, a week after birth, newborn acquires a certain number of bacteria which constitute his first microbiota; this latter, will be strengthened and diversified during life. Consequently, bacterial colonization occurring in the first years of infant's life influences not only the composition of his transient microbiota but also the microbiota stability and composition during adulthood [3]. **Figure 1** shows the possible immunological and microbial changes that characterize the dynamic development of human microbiota.

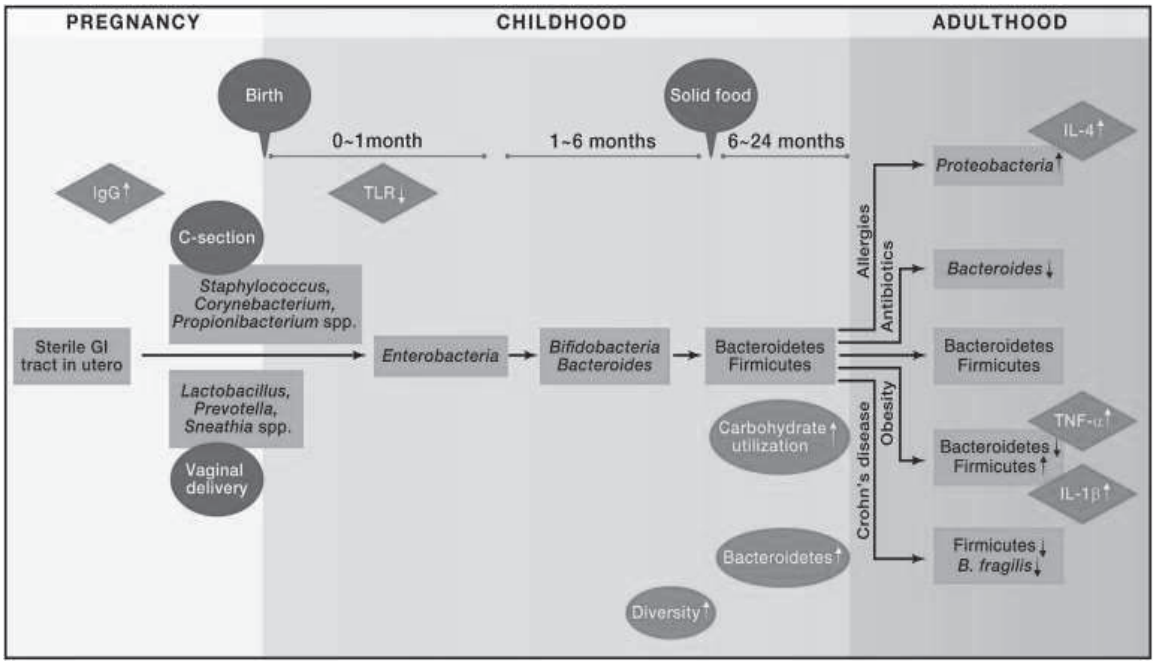

***Figure 1**-Development process of the human microbiota* ***Figure source** from article publish by Clemente JC et al. in Cell. 2012 Mar 16;148(6):1258-70.*

Although the human microbiota is characterized by a very high level of biodiversity, we can distinguish a small number of species in common to all individuals which represent the phylogenetic core of human microbiota. The remaining portion, instead, is variable and it differs both between different populations and individuals belonging to the same ethnic group; this is due to different eating habits, genetic constitution, lifestyle, environmental and geographical conditions and at least, to the age of individuals. Just this variable component ensures that each subject has his own and unique microbiome [4]. For this reason, some researchers dare to say that in the future we may be identified no longer through fingerprints but through our microbiota [5]. Therefore, as human microbiota consists in a vast "universe", today numerous scientific fields are investing on its study and deep characterization. Today, most studies concerning the microbiota are focused on the bacterial community profiling; however, this approach results few useful to elucidate the mechanisms involved in the interaction and cross-talk with the host. The major challenge in evaluating host–microbiome metabolic interactions is understanding how this can affect the individual's health status. In this context, metaproteomics and metabolomics approaches may extend our knowledge on the functional role of the human microbiota and its interactions with

the host, by using different methodologies such as shotgun proteomic, spectroscopy and mass spectrometry or liquid chromatography-mass.

2.1.1 Gut microbiota

Human gut harbors a huge amount of microorganisms, approximately 10^{19} bacteria. The main bacterial phyla are *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria* and *Verrucomicrobia*. Nevertheless, the concentration of different bacterial genera and species

composing the gut microbiota changes along the digestive tract, increasing exponentially from stomach to the descending colon [6]. These microorganisms carry out mainly a metabolic, trophic and protective activity [7]. Together, these three functions guarantee a correct absorption of nutrients and maintain the physiological gut permeability.

Moreover, a correct balance within microorganisms composing the gut microbiota is essential to maintain the correct intestinal pH and control proliferation of potential pathogenic bacteria. Indeed, changes in its microbial composition may lead to an imbalance between the abundance of beneficial and harmful microorganisms. This condition, defined as dysbiosis, may favor host susceptibility to different diseases including metabolic disorders, such as obesity and diabetes, inflammatory bowel disease (IBD) and atopic diseases.

Opportunistic bacteria, such as *Escherichia coli*, *Clostridium* spp. and *Salmonella* spp. can take advantage from aforementioned imbalance condition and cause severe pathologies. Conversely, other intestinal microorganisms are particularly beneficial for the host' health. Among these, bacteria belonging to *Lactobacillus* and *Bifidobacterium* genera are considered the main probiotic microorganisms. They are able both to modulate intestinal microbiota composition and acidify intestinal environment thus reducing the incidence of infections [8]. Furthermore, intestinal probiotic bacteria act not only at gastrointestinal level but also on host's immune system; indeed, they are able to stimulate production of anti-inflammatory cytokines, reducing the onset of respiratory infections and allergic diseases [9], [10]. Therefore, today it is clear that the human gut microbiota plays a crucial role in both the maintenance of gastrointestinal homeostasis and the overall host's health.

2.1.2 Skin microbiota

Although skin is a continuous tissue covering our whole body, it is characterized by a wide variety of habitats including invaginations and specialized niches with different ecology [11]. Each cutaneous district is controlled by different conditions of humidity, temperature, pH and sebum content that, depending on its position on human body, influence cutaneous bacterial colonization.

Firmicutes, Bacteroidetes, Actinobacteria and *Proteobacteria* are the main bacterial phyla constituting the skin microbiota. The epidermis harbors mainly aerobic bacteria; among these, staphylococci (phylum *Firmicutes*), in particular *Staphylococcus epidermidis*, seem to be the main skin colonizers. Differently, a massive presence of facultative anaerobic bacteria belonging to *Propionibacterium* genus (phylum *Proteobacteria*) have been detected both in derma and regions with a high amount of sebum [12].

Skin surface, interposing itself as a barrier between the organism and external environment, protects host from the action of potential pathogens and harmful agents. However, the skin is not only the first line of defense but, thanks to its resident bacteria which constitute the skin microbiota, it is considered a dynamic organ whose structure, composition and function constantly changes during host's life. Considering that, the physiology of infant and adult's skin is widely different.
Especially in the first years of life, baby's skin is more hydrated than that of an adult, having alkaline pH, increased desquamation, high cellular turnover and hair follicles secreting immature sebum. These features favor the colonization of numerous commensal bacterial species which contribute to both skin homeostasis and correct modulation of local immune responses. Moreover, it was demonstrated that bacterial species belonging to *Staphylococcus* genus are dominant in infant's skin microbiota; nevertheless, their abundance decreases towards the end of child's first year of life who, already at 12-18 months old, shows a skin microbiota composition similar to that observed in adults [13].

The first year of life of each individual is characterized by an intense development of the whole organism that involves mainly the maturation of immune system. However, newborn's skin physiology is still evolving and for this reason, if compared to adult skin, it may be

underperforming and more vulnerable to microbial infections and pathological conditions, such as atopic dermatitis and psoriasis [14]. Consequently, it is fundamental that a specific balance between skin barrier and cutaneous microorganisms is established to guarantee the maintenance of host's health.

2.1.3 Breast milk microbiota

Human breast milk contains different bioactive molecules, necessary not only to feed the baby but also to protect him against infections or inflammations. Many of these factors, such as antimicrobial substances, hormones (cortisol) and smaller proteins (as epidermal growth factor, nerve growth factor, insulin like growth factor), may protect lactating mammary gland and strengthen newborn's intestinal environment [17]. Moreover, recent studies showed that breast milk is not sterile as traditionally thought, but it represents a great source of commensal bacteria, including probiotic microorganisms, which may have a positive impact on newborn's health, also contributing to the development of his immune system [9].

Human milk approximately contains $10^3/10^4$ colony forming units (CFU)/ml of bacteria [15]. Every day, the baby takes about 800 ml of milk, ingesting about $1x10^5$ - $1x10^7$CFU of microorganisms responsible for the formation of his initial gut microbiota. Bacteria contained in colostrum and mature milk play a pivotal role on infant's health up to several months after his birth [9]. These microorganisms, indeed, may reduce the incidence and severity of several infections by competitive exclusion mechanisms and the production of antimicrobial compounds, such as bacteriocins [16]. Moreover, breast milk microbiota facilitates digestive processes favoring the metabolization of sugars and proteins [7, 16].

Interestingly, the presence of numerous *Lactobacillus* and *Bifidobacterium* strains in human milk is positively associated with beneficial effects on the host's organism. These bacteria arouse great interest in the scientific community as they are the main probiotic microorganisms used worldwide [17]. The administration of probiotic supplements, for instance, may help to restore the milk microbiota balance after an intense use of antibiotics which can lead to bacterial dysbiosis [9]. Moreover, a decrease of lactobacilli in breast milk has been associated with increased

susceptibility to allergies and gastrointestinal disorders in newborns [18]. Therefore, microorganisms contained in human breast milk are essential to maintain a correct bacterial homeostasis and colonize infants' gut [19]. In particular, the first gut microbiota colonization is influenced by both baby's genetic features and the mode of delivery. Indeed, it is known that vaginal and fecal bacteria, such as bacteroides, bifidobacteria and lactobacilli, are acquired during the passage through the birth canal. Differently, the gut of infants born by Caesarean section is mainly colonized by environmental bacteria such as *Clostridium difficile* and *Escherichia coli* [20]. Also the nutritional characteristics of the human breast milk play a key role in the selection of the gut microbiota during early infancy. Indeed, the human breast milk has a dynamic composition that, in according with the time of lactation, changes during its transition from colostrum to mature milk [21]. Colostrum is produced immediately after the birth until the fifth postpartum day. It is characterized by a large amount of immunoproteins, such as lactoferrin, vitamins and immunostimulating components, as immunoglobulins A (IgA) and leukocytes, which facilitate both the development and maturation of the newborn's immune system [22]. Differently, the mature milk is secreted at 10^{th} day postpartum and if compared to the colostrum, it is characterized by a lower protein content but a higher abundance of fatty acids; oligosaccharides, in particular, represent their main component. The mature milk is also rich in glycans, including glycoproteins, mucins, glycosaminoglycans, and glycolipids. These bioactive compound promotes gut colonization by *Bifidobacterium bifidum* and contribute to the maturation of infant's intestinal microbiota inhibiting the colonization by potential pathogens [23].

Therefore, it has been suggested that the breast milk microbiota composition may modulate a balanced initial immune response and the development of immune-mediated diseases (asthma, inflammatory bowel disease, type 1 diabetes) during both infancy and adulthood [24]. For these reasons, the researchers nowadays agree that breastfeeding leads to several health benefits.

2.2 Next generation sequencing technologies

Sequencing technologies are relatively recent and they have a common "forefather", the Sanger method, which represents the first generation of genomic sequencing; these technologies led to a great improvement of molecular biology, allowing the study of different types of nucleic acids

(DNA, RNA, cDNA, etc.). The high resolution, accuracy, and quality of data obtained with sequencing methods make them ideal to study the whole genomes, gene variants and also to

identify numerous microbial pathogens [25]. The dynamic and rapid evolution of sequencing technologies allowed today their application in research and diagnostic laboratories. In medical field, for instance, they allow diagnosis of diseases related to genomic structure, highlighting a vision of "personalized medicine" thanks to the application of therapeutic strategies close to patients' needs [26]. Nowadays, second-generation sequencing technologies, also defined as "Next

Generation Sequencing" (NGS), consist in *high-throughput* sequencing systems. They allow to generate, simultaneously and in short time, millions of gene sequences constituting the genomic" library " [27]. Consequently, NGS technologies made possible to reduce drastically both time and costs of experimental setting. Moreover, the possibility to perform multiple analysis allows obtaining a huge amount of data in short time; consequently, it is necessary using innovative bioinformatics systems to manage all this new information [28]. In microbiology, the characterization of multiple bacterial DNA in a single biological sample allowed the development of a new study approach defined as Metagenomics.

2.2.1 Metagenomics

Metagenomics is an approach based on genome sequencing of all microorganisms colonizing the same habitat; accordingly, this method allows to bypass cultural isolation of bacteria assessing, instead, the microbial complexity in their natural environment [29].
To standardize the analytical method, metagenomics uses the amplification of 16S rRNA gene, which codes for a particular rRNA contained within the ribosomal small subunit.

The 16S rRNA gene is selected mainly for two reasons:

- it contains both stable and variables regions, these latter are species-specific (**Figure 2**);
- 16S rDNA structure changes very slowly over the time and it is composed by 1500 highly conserved base pairs. Although, there are other genes coding for rRNA, these are not adequate to perform a metagenomics analysis. Indeed, 5S rRNA consists of 120 bp and

the information is thus limited; otherwise, the 23S rRNA consists of 2900 bp, thus it results too complex.

Universal primers are used to recognize the highly conserved regions and allow the duplication and amplification of adjacent hypervariable regions useful for bacterial identification. Consequently, hypervariable regions can be considered as the "identity card" of bacteria.

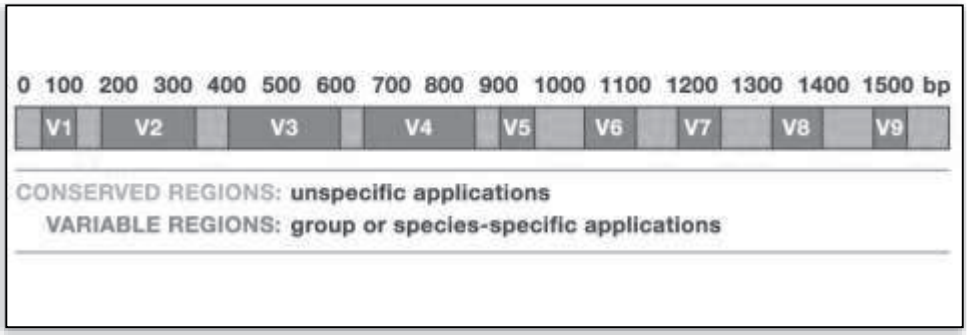

Figure 2 *-16S rRNA gene illustrating the conserved (green) and variable (red) regions.* ***Figure source*** *from article publish by S.* Chakvavorty et al in *MolBiol(Mosk).2015 Sep-Oct;49(5):749-59.*

2.2.2 Ion Torrent ™ PGM sequencing

Nowadays, numerous sequencing technologies allow the parallel analysis of several DNA samples. Among the different NGS platforms we can find the personal Genome Machines (PGMs), which are described as "benchtop" sequencing machines with all advantages of NGS technologies (speed and reduced costs of analysis) [30]. In particular, the Ion Torrent ™ PMG (*Life Technologies, Italy*) platform is based on a semiconductor technology, which exploits *in vitro* the principle occurring in natural biological systems. Indeed, when a single nucleotide is incorporated in the new double-stranded DNA chain by the DNA polymerase enzyme, a hydrogen ion is released as byproduct [31].

The Ion Torrent technology is based on a sophisticated high-density array, with micro-wells inside of which the DNA synthesis and the consequent release of protons is cyclically repeated. Working as a pH meter, the machine records all pH changes generated during the nucleotides annealing, by the release of hydrogen ions (**Figure 3**). Therefore, Ion Torrent appears to be an innovative technology since it does not use fluorescent molecules which are commonly exploited by numerous sequencing machines; furthermore, this platform does not need further capture and analysis devices of fluorescence signal, being thus extremely smaller, faster in the timing processing and much less expensive than other NGS platforms [30, 31].

Using the HiSeq platform, for instance, the fluorescent signal decays after a certain number of cycles determining, as consequence, several analysis problems. Differently, signals detected by Ion Torrent ™ PGM tend to remain stable even after numerous cycles, thus realizing highly

accurate analysis. Finally, a single sequencing run with Ion Torrent PGM requires a very small amount of DNA: 10 nanograms of sample are enough to get sequences (called "*reads*") of optimal length between 200 and 400 bp. For these reasons, in our laboratory, Ion Torrent technology is regularly applied to characterize the composition of human microbiota.

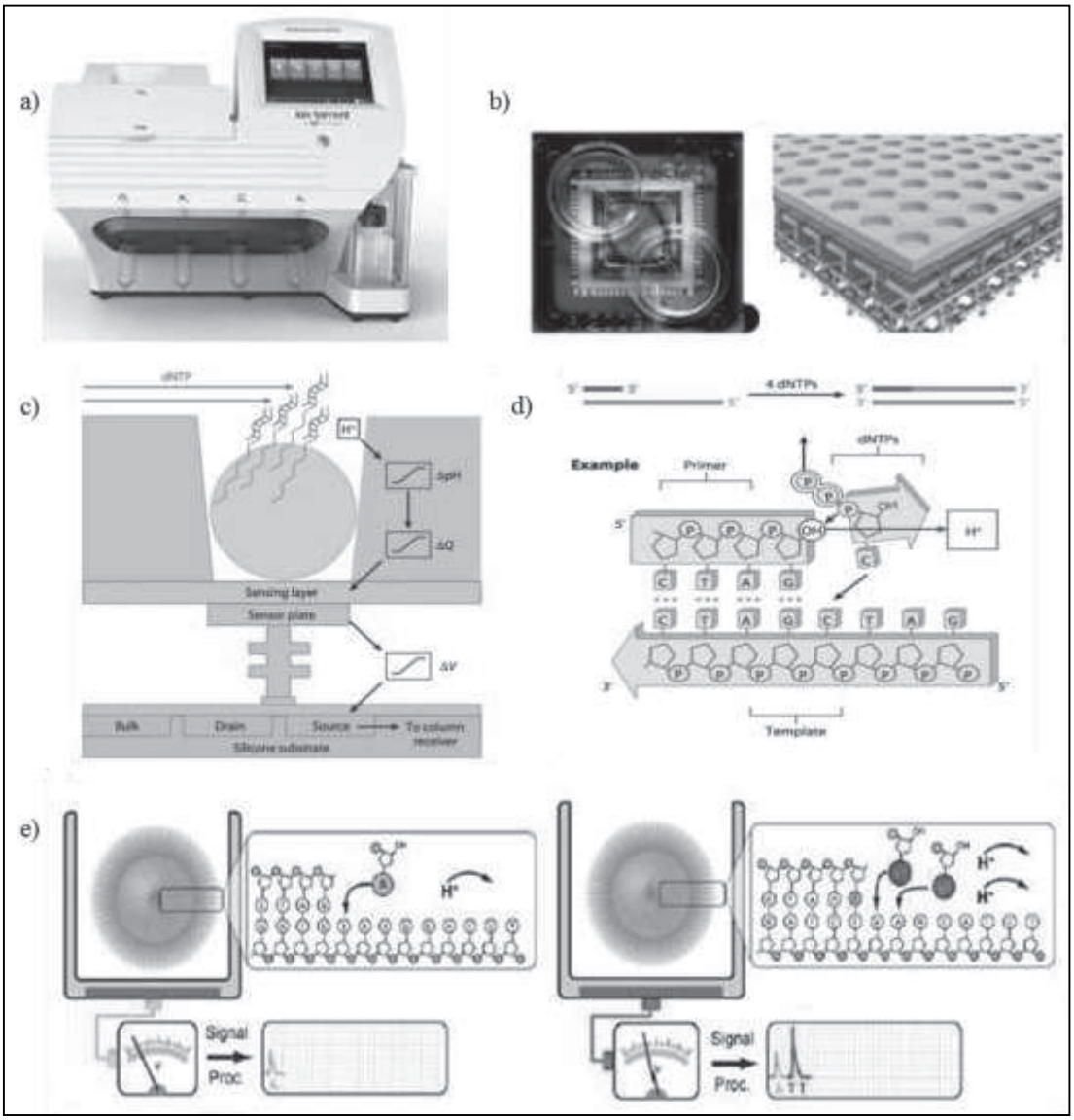

Figure 3*-*Figure reproduced from *N.J. Clifton et al article publishes on 2013;1048:247-84.* ***a)*** *Ion Torrent sequencing;* ***b)*** *Semiconductor packing (on right), a chip section with the wells in evidence (on left);* ***c)****Chemical signal sensor to digital sequence;* ***d)****Pattern of chemical sequencing implemented in Ion Torrent;* ***e)*** *Signal processing by Ion Torrent technology.*

Chapter 3

MATERIALS AND METHODS

3.1 Recruitment of subjects

In vivo studies were carried out on human samples available from different scientific collaborations and they were used to determine the reciprocal influence between host and microbiota.

3.2 Studying the human microbiota

3.2.1 Bacterial DNA extraction and 16S rRNA Gene Amplification

DNA was extracted from biological samples using different DNA Extraction Kits (Norgen, Thorold, Canada and Qiagen, Milano, Italy) following the manufacturer's instructions. They are all column-based DNA extractions kits which are strongly recommended for downstream sequencing [34].

Partial 16S rRNA gene sequences were amplified from extracted DNA using the 16S Metagenomics Kit (Life Technologies, Monza, Italy), designed for rapid analysis of polybacterial samples using the Ion Torrent sequencing technology. Kit includes two primer sets that selectively amplify corresponding hypervariable regions of bacterial 16S RNA gene:

- primer set V2-4-8;
- primer set V3-6, 7-9.

For each sample 2 reaction mix were prepared. For PCR (Polymerase Chain Reaction) reaction was used a SimpliAmp Cycler ™ Thermal Cycler (Applied biosystems, Life Technologies, Monza, Italy) following temperature profile summarized in **Table 3**. A negative control was included to verify the absence of PCR reagents bacterial contamination. Integrity of PCR amplicons were analyzed by electrophoresis on 2% agarose gel.

Component	Sample
2X Environmental Master Mix	15 µl
16S Primer Set (10X)[1]	3 µl
DNA and Negative control (water)	3 µl
Nuclease Free Water	9 µl
Total Volume	**30 µl**

***Table 2**- Quantities of reagents used for the reaction mixture.[1]V2-4-8 or V3-6,7-9*

Stage	Temperature	Time
Holding	95°	10 min
Cycling 28 Cycles	95°	30 sec
	58°	30 sec
	72°	20 sec
Holding	72°	7 min
Holding	4°	∞

***Table 3**- PCR conditions*

3.2.2 Preparing genomic library

PCR products were purified using Agencourt AMPure XP DNA purification beads (Beckman Coulter Genomics, Germany), in order to remove primer dimers according to the supplier's instructions. Amplicons concentration was estimated (ng/µl) by means of Qubit system (Life Technologies, Monza, Italy) and libraries created by means of Ion Plus Fragment Library kit (Life Technologies, Monza, Italy). **Table 4** summarizes the amount of reagents necessary for the purification of amplicons.

Component	Volume	Example
Pooled amplification reaction[1]	X (equal volume of V2-4-8 eV3-6, 7–9 reactions)	40 µl (20 µL each of V2-4-8 and V3-6, 7–9 reactions)
Agencourt® AMPure® XP	1.8X	72 µl

***Table 4**- Reaction mixture for the purification of the amplicon pool.*
[1] Combine equal volume of the 2 reactions

For each sample, bacterial libraries were prepared using an optimal amount of purified amplicons (10-100 ng) quantified by mean of Qubit system (Life Technologies, Monza Italy). The protocol for library preparation provides the use of two different reaction mixtures obtained through the Ion Plus Fragmentary Library Kit (Life Technologies, Monza Italy). **Table 5** reports the initial reaction mixture used to prepare the genomic library.

Component	Volume
Pooled short amplicons, 10-100 ng	79 µl[1]
5X End Repair Buffer	20 µl
End Repair Enzyme	1 µl
Totale	**100 µl**

***Table 5**- Reaction mixture for the initial preparation of the library.*
[1] The 79 µL final consist of nuclease-free water and a quantity of amplified including.

The incubation of template at room temperature with End Repair enzyme allowed the "reconstruction" of amplicon 3' and 5' ends, which is necessary to anchor both adapters and barcode sequences. A further purification, by means of the Agencourt® AMPure® XPbeads, ensured the complete purification of fragments.

Subsequently, barcodes were added to each sample using Ion Xpress Barcode Adapters 1–16 kit (Life Technologies, Italy). **Table 6** shows the amount of reagents necessary to prepare the reaction mixture for the final step of genomic library preparation. Samples were incubated at 25°C for 15 minutes and subsequently, at 72°C for 5 minutes (SimpliAmp thermal cycler, Life Technologies, Monza, Italy) to ligate and nick-repair adapters and barcode to amplicon sequences.

Component	Volum for Barcoded Libraries
DNA	~ 25 µl
10X Ligasi Buffer	10 µl
Ion P1 Adapter	2 µl
Ion Express™ Barcode X[1]	2 µl
dNTP Mix	2 µl
Nuclease-free Water	49 µl
DNA Ligasi	2 µl
Nick Repair Polymerase	8 µl
Total	**100 µl**

***Table 6** - Reaction mixture for the preparation of the final genomic library.*
[1] X= barcode chosen

Finally, library was purified with magnetic beads to obtain a high quality template, free from waste products.

Genomic library was then subjected to an emulsion PCR (emPCR) by means of Ion OneTouch TM 400 Template Kit (Life Technologies, Italy). Each library must be diluted to the optimal concentration of 26 pM, as low concentrations of sample may produce weak signals, thus leading to an inaccurate signal detection; otherwise, an excessive amount of library may generate a huge amount of polyclonal sequences and consequently, not analyzable data. The composition of emPCR reaction mixture is shown in **Table 7**.

Reagent	Volume
Nuclease-free Water	15 µl
Ion PGM™ Hi-Q™ Calibration Standard	10 µl
Ion PGM™ Hi-Q™ Enzima Mix	50 µl
Fresh Diluited Library [1]	25 µl
Ion PGM™ Hi-Q™ ISPs[2]	100 µl
Volume totale	**1000 µl**

***Table 7**–Emulsion PCR Reaction mixture.*
[1] Library final concentration26pM; [2]Ion Sphere™Particles

Subsequently, samples were enriched through a special purification that favored selection of Ion Sphere Particles on which the library amplification reaction occurred. The Ion OneTouch ™ ES (Life Technologies, Monza, Italy) allows the complete automatization of enrichment step. Reagents used for this step and the specific order they were added in the strip are summarized in **Table 8** and **Table 9,** respectively**.**

Order	Component	Volume
1	Tween® Solution	280 µl
2	1 M NaOH	40 µl
-	**Total**	**320 µl**

***Table 8**–Preparation of the Melt-off solution*

Well number	Reagent to dispense in well
1	Entire tamplate-positive ISP sample [~100 µl]
2	130 µl of Dynabeads® MyOne™ Streptavidin C1 Beads
3	300 µl of Ion OneTouch™ Wash Solution
4	300 µl of Ion OneTouch™ Wash Solution
5	300 µl of Ion OneTouch™ Wash Solution
6	Empty
7	300 µl of Melt-Off Solution
8	Empty

***Table 9**–Preparation of the reagents for the template-positive enrichment using Ion OneTouch ™ ES (Life Technologies)*

3.2.3 Ion Torrent PGM sequencing

Library sequencing was carried out on a 316 chip by means of Ion Torrent Personal Genome Machine (PGM) system, using the IonPGM Hi-Q kit (Life Technologies, Monza Italy). Sequencing with Ion Torrent PGM ™ is based on three fundamental steps: Initialization, Chip Check and Sequencing.

Initialization allows the instrument to bring all solutions used in sequencing to the proper pH; this is a critical step since Ion technology is based on the measurement of small changes of pH. Consequently, it is essential that the initial pH is finely controlled. Preparation of each bottle used in the sequencing is reported in **Table 10.** After having reached the optimal pH, 4 tubes containing each one 20 µl of one deoxyribonucleotide (dNTP) (Ion PGM ™ dNTPs Kit, Life Technologies, Italy) were installed on the Ion PGM.

Wash1	350 µl NaOH 100 nM
Wash 2	Ion PGM™ Hi-Q™ Sequencing W2 Solution + 70 µl NaOH 100 nM
Wash3	50mL of Ion PGM™ Hi-Q™ Sequencing W3 Solution

***Table 10**–Composition of Wash1, Wash2 and Wash3 solutions. The Ion PGM ™ Hi-Q ™ Sequencing Solution W2 and W3 are provided with the kit Ion PGM ™ Hi-Q ™ Sequencing Kit (Life Technologies, Italy)*

For the primer annealing, 5µl of Control ISPs and 12 µl of Sequencing Primer (Ion PGM™ Hi-Q™ Sequencing Kit, Life Technologies, Italia) were added directly to the entire volume of enriched-template, for a total volume of 27µl; sample was then subjected to an incubation at 95°C for 2 minutes, followed by incubation at 37°C for 2 minutes. During the annealing step, the Chip Check was performed to test the chip quality and functionality before loading the sample. Following the Chip Check, sample was loaded on chip and sequenced.

At the end of sequencing, reads were filtered by the PGM software to remove low quality and polyclonal sequences. Sequences matching the PGM 3′ adaptor were also automatically trimmed. 16S rRNA sequences were then analyzed by Ion Reporter Software, which comprises a suite of bioinformatics tools that streamline and simplify analysis of semiconductor-based sequencing data. 16S rRNA workflow module in Ion Reporter Soft is able to classify individual reads combining a Basic Local Alignment Search Tool (BLAST) alignment to curated Greengenes database, which contains more than 400,000 records, with a BLAST alignment to the premium curated MicroSEQ ID database, a high quality library of full-length 16S rRNA sequences. The final output of Ion Reporter Software was the identification of microorganisms and their abundance in the sample.

3.3 Analysis methods

Sequencing the human microbiota, we obtain a large amount of data per single subject. Consequently, we need several techniques to reduce data set size, as well as to analyze and

graphically express them, obtaining an immediate visual analysis and a simultaneous interpretation of responses in relation to specific variables.

To date, several methods of analysis are available and divided into:
- sorting methods;
- multivariate analyzes based on hypothesis testing.

Aforementioned methods are used in several scientific fields and on datasets of different nature and structure. In order to carry out our analysis, a particular organization of data was considered. Indeed, our dataset was a matrix where rows were willing the subjects, while in columns OTU count for all bacterial genera identified were arrayed. Therefore, for each subject enrolled in our studies, data were microbial abundances, expressed as relative abundance of OTU.

The sorting methods have descriptive and informative synthesis purpose. They can be divided into two main categories: not bound or bound methods. In the first case, they use only information referring to genera abundances (or OTU), while bound methods take into account other environmental variables associated to dataset (i.e. sex, age and status of the subjects of health, etc...). A bound method that is generally applied to investigate a multivariate dataset is the Principal Component Analysis (PCA), which combines linearly the variables into dataset to explain the maximum variance of initial data using less variables. Differently, multivariate analysis (such as hierarchical cluster analysis and biodiversity indices, including Shannon, Simpson and Chao's indices) based on hypothesis testing allow to assess whether differences between subjects exist, according to genera abundance values, as well as to represent dataset in a diagram according to bacterial similarities. In particular, a hierarchical cluster analysis investigates the associations among variables while, the biodiversity indices allow the characterization of microbial community composition and the better understanding of their structure and dynamics.

The analysis methods applied in the present PhD study were selected from those applied mainly in literature (for instance, biodiversity analysis, PCA and hierarchical clustering), together with a new methodology, the Auto Contractive Map (AutoCM), which is consist in a closed source software based on artificial neural network and may be very useful to investigate the human microbiota.

Chapter 4

LIST OF MANUSCRIPTS

Manuscript 1

Persisting changes of intestinal microbiota after bowel lavage and colonoscopy

Published in European Journal of Gastroenterology & Hepatology

Introduction

The aim of this study was to evaluate the effects of standard bowel lavage on the gut microbiota composition of healthy individuals undergoing to colonoscopy. Fecal bacterial composition was investigated immediately after the bowel cleansing and 1 month after the colonoscopy in order to establish whether this treatment could have long-lasting effects on the intestinal environment.

Main Results

The gut microbiota of individuals subjected to colonoscopy did not completely return to the composition before the colonoscopy. One month after the bowel cleansing, indeed, *Firmicutes* have been observed to be less abundant, while *Proteobacteria* abundance showed a significant increase if compared with samples collected before the examination.

Breakthrough

In this first study, we applied a simply relative abundance analysis to investigate microbiota modifications at phyla, class and family level. Even if several significant changes were observed, we understood how extremely complex metagenomics data are, especially regarding their high variability.