Correlation Between Gut Microbiota and Depression

Ali Naseribafrouei

Master degree in Applied Biotechnology

HEDMARK UNIVERSITY COLLEGE

2013

Table of Contents

ACKNOWLEDGMENT

I hereby present my special thanks to my dedicated supervisors Professor Knut Rudi and Professor Robert Wilson and my statistical advisor Professor Arne Linlokken. I also thank Mr. Oyewumi Opyemi for his kindly help in this regard.

Abstract

Gut microbiota consist of a variety of microorganisms which inhabit the human gut. They play many well-known roles in human gut homeostasis and may be implicated in some pathologic processes especially immunological events. Depression is a chronic syndrome with a pathogenesis linked to various genetic, biological and environmental factors. In recent years inflammatory factors has been mentioned as contributing factors for development and exacerbation of depression. Most of these inflammatory events are associated with disturbances in cortisol secretion and HPA axis. The role of gut microbiota in inflammatory events and co-morbidity of depression and hypercortisolism and other related inflammatory events has long been demonstrated. The aim of this study is to find out if there is a relationship between gut microbiota and depression. In this regard we took fecal samples from 56 people, 37 clinically-depressed patients and 19 control people who were drawn from both outpatient and inpatients at the Sykehuset Inlandet HF hospital, Hamar, Norway. Analysis of gut microbiota was performed by extraction of whole DNA from fecal samples. 16S rRNA gene was then amplified and applied as microbial molecular marker. Direct sequencing on pool of amplified 16SrRNA gene fragment; resolution of the mixed spectra using MCR-ALS multivariate statistics and taxonomic identification of dominant bacterial components using RDP database were then carried out to characterize the present bacteria. Along this, Illumina deep sequencing was performed on the 16SrRNA gene fragments and classification was performed based on the 100 most dominant OTUs. Salivary cortisol measurement (measured in the morning, mid-day and evening) was applied as biological marker for evaluation of baseline unstressed HPA function. Firmicutes and Bacteroidetes were the most dominant phyla in the gut. OTUs belonging to Bacteroidetes phylum collectively were significantly lower among depressed patients in comparison to control people. There were some significant correlations between both bacterial components and operation taxonomic units (OTUs) to depression and cortisol levels, highlighting the involvement of gut microbiota in immunological states and depressive symptoms, yet these correlations are not necessarily cause and effect relationships.

Keywords: depression, gut microbiota, HPA system, 16SrRNA gene, mixed bacteria sequence, salivary cortisol, MCR-ALS and Illumina deep sequencing

ABBREVIATIONS

ACTH – Adrenocorticotropic releasing Hormone

CRH – Corticotropin-releasing Hormone

ddNTP – Dideoxynucleotide triphosphate

DNA – Deoxyribonucleic Acid

dNTP – Deoxynucleotide triphosphate

DST – Dexamethasone Suppression Test

EFA – Evolving Factor Analysis

FISH – Fluorescent in situ Hybridization

GIT – Gastrointestinal Tract

HIV/AIDS – Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome

HPA – Hypothalamus-Pituitary-Adrenal

IBD – Inflammatory bowel disease

MCR-ALS – Multivariate Curve Resolution with Alternating Least Square

mRNA – Messenger Ribonucleic Acid

NGS – Next Generation Sequencing

OTU- Operational Taxonomic Unit

PCR – Polymerase Chain Reaction

RDP – Ribosomal Database Project

RNA – Ribonucleic Acid

rRNA – Ribosomal Ribonucleic Acid

1. Introduction

1.1.1 Definition

Major depressive disorder (MDD) is a mental disorder which presents itself with low mood, low self-esteem, and loss of interest in normally enjoyable activities (Kessler, Nelson, & McGonagle, 1996). It is not amazing to see some of the symptoms of major depressive disorder in other mental disorders such as bipolar mood disorder, schizophrenia or dementia but for a patient to be diagnosed as suffering from "major depressive disorder" all the signs and symptoms must be regarded as a whole and summed up to reach a distinguishing extent and clinically significant and this is up to the clinician to evaluate the presence and the severity of the symptoms (Hirschfeld, 2000). This syndrome has been classified as one of the most disabling and potentially life-threatening syndromes which can affect all aspects of the life including work, education, personal activities and pleasures such as sleeping, eating, sexual intercourse and relationship with family members, colleagues and other people (Haeffel, et al. 2008).

Although scientists and physicians have tried for centuries to know more about all aspects of the disorder and they have made considerable progress in this regard but this disorder is still understood incompletely and many questions remained to be answered in the future (Frank et al., 1991). Among dozens of potential factors to develop the disorder, psychological, psycho-social, hereditary, evolutionary, biologic and iatrogenic factors can be named. Many long lasting physical illnesses such as cancer, autoimmune diseases and diabetes can cause or worsen the depressive disorder both directly or due to side effects of medications (Caspi, et al, 2003).

ICD-10 (World Health Organization's International Statistical Classification of Diseases and Related Health Problems) states that for a patient to be diagnosed as depressed at least two of these three symptoms should present: depressed mood, anhedonia, and reduced energy; while DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders) necessitates the presence of at least one group of symptoms, either depressed mood or anhedonia and these symptoms must persist for at least two weeks for a patient to be diagnosed with major depressive disorder (Almeida and Almeida 1999).

1.1.2 Signs and Symptoms

Major depressive disorder can have deep impact on the patient's life, family relationships, work conditions, personal sense of wellbeing and life satisfaction. This disorder can be ranked among top damaging and burden-imposing diseases such as diabetes, cardiac diseases, renal failure, neurologic disorders and so on (Hasin, 2005). Patients usually complain of very low mood pervading all aspects of life and lack of pleasure in formerly enjoyable activities. They will suffer from preoccupations, thoughts or feelings of worthlessness, inappropriate guilt or regret, helplessness, hopelessness or self-hatred. Besides these symptoms some patients experience psychotic features such as delusions and to lower extent hallucinations (Weissman 1987). Among the most experienced symptoms of major depressive disorder poor concentration and memory (more prominent in those with melancholic or psychotic features); social isolation; loss or decreased libido; thoughts of death and suicide can be underscored (Hahn, et al, 2011). Around 80% of depressed patients are affected of insomnia. Typically depressed patients wake up early in the morning and cannot go back to sleep, which is in contrast to anxiety disorders in which patients have difficulty sleeping. In atypical forms of depression hypersomnia or oversleeping can happen (Hahn, et al, 2011).

In some cases somatic presentations are prominent feature of the disorder including fatigue, headache and digestive problems, and these physical complains can be the main feature of the syndrome. Most depressed patients lose their appetite and weight but in some patients the opposite happens (Goodyer, Tamplin et al. 2000). In elderly patients cognitive symptoms can be more prominent such as forgetfulness and slowness in movements. Since most elderly people have already other physical disorders such as stroke, cardiovascular diseases, Parkinson's disease and chronic obstructive pulmonary diseases the establishment of the depression diagnosis can be challenged and exacerbate the existing physical complains (Patel, et al. 2004).

1.1.3 Etiology

Depression is a multi-factorial disease being caused by biological, psychological and social factors. The diathesis-stress model proposes that depression is caused when stressful life events superimpose on a preexisting vulnerable condition (Uher, & McGuffin, 2010). The preexisting vulnerability can have a genetic basis, due to an interaction between nature and nurture, or schematic, which results from the impact of the world views being learned during childhood (Haeffel, Getchell et al. 2008).

Of around 30 neurotransmitters which so far have been identified, scientists believe three of them, namely serotonin, norepinephrine and dopamine, have the most important role in developing depression in human **(**Hirschfeld, 2000). Antidepressants have their impact on the overall balance of these three neurotransmitters inside structures of the brain that are responsible for emotion, reaction to stress, sleep, appetite, and sexual libido (Ward, Weller et al.1990). Almost all antidepressant agents have their anti-depressive effect through increasing the level of these three main neurotransmitters in the synaptic cleft between neurons in the brain and some of antidepressants directly increase monoamine receptors in the synaptic cleft (Kendler, Gardner , Kenneth, and Charles, 1998). Serotonin is thought to be the regulator of other neurotransmitters; decreased serotonin activity causes this system to act in unusual and erratic ways. Based on this "permissive hypothesis", low levels of serotonin lead in turn to low levels of norepinephrine (Sheline, Gado, & Kraemer, 2003).

Some empirical evidences are in favor of these interactive models. For instance, researchers in New Zealand took a prospective approach for evaluating depression, in order to document how depression emerged in otherwise healthy person in a cohort group. They concluded that the alterations in expression of the serotonin mediator (5-HTT) gene make those under stressful events vulnerable to develop depression later on. This is more prominent in those with one or two short alleles of the 5-HTT gene (Caspi, et al. 2003).

Some evidences are in favor of this hypothesis that depression can be caused partly by an overactive hypothamic-pituitary-adrenal axis (HPA axis) that can mimic the neuro-endocrine response to stress (Anisman, and Hayley, 2012). Investigations found higher levels of hormone cortisol and enlarged pituitary and adrenal glands, suggesting imbalances

in the endocrine system especially cortisol can play a role in some psychological disorders including depression. Oversecretion of corticotrophin-releasing hormone from the hypothalamus is implicated to drive this and developing cognitive and arousal symptoms (Monteleone, 2001).

Recently, Maes reported that increased IgA and IgM-Mediated immune response to lipopolysaccharide (LPS) of gram-negative bacteria in a remarkable number of depressed patients. This shows an increased passage of gram-negative bacteria antigens from the gut into the blood, which is caused by increased gut permeability ("leaky gut") (Maes, 2011) Leaky gut is the result of loosing of the tight junction barrier by inflammatory processes, such as increased production of NFκB and proinflammatory cytokines. This causes, in turn, an increased translocation of otherwise poorly invasive enterobacteria into the blood and the subsequent IgA and IgM-mediated immune responses against LPS of the enterobacteria (Dowlati, Herrmann et al. 2010)

Another recently implicated factor is the potential role of molecules necessary for overall cellular functioning namely cytokines. Cytokines are parts of immune system, causing the depressive symptoms resemble those of sickness behavior. This suggests that depression can result from a maladaptive manifestation of sickness behavior due to abnormalities in circulating cytokines (Dantzer, O'Connor et al. 2008). A meta-analysis of the clinical literature strongly implicates higher blood concentrations of IL-6 and TNF-α in depressed patients compared to controls which can lead to the excessive production of prostaglandin E2 and probably excessive COX-2 expression (Dowlati, Herrmann et al. 2010). Abnormalities in indoleamine 2,3-dioxygenase enzyme activities and metabolism of tryptophan-kynurenine may lead to enhanced production of the neurotoxin quinolinic acid, causing major depression (Müller, Myint & Schwarz, 2011).

Furthermore a Swedish study estimated the heritability rate of depression to be somewhere in the region of 40% for women and 30% for men (Kendler,Gatz, Gardner& Pederson, 2006).

1.1.4 Diagnosis

Diagnosis can be made by a well-trained general practitioner or by a psychiatrist or psychologist by considering patients current symptoms, biographical history, and familial history (Crystal, Sambamoorthi et al. 2003). All the relevant biological, psychological and social factors must be investigated in this regard. The assessment must also include any alcohol or drug abuse, mental state examination including current mood and thought content, especially the presence of the themes of hopelessness or pessimism, self-harm or suicide, and lacking positive thoughts or plan (Mulsant and Ganguli 1999).

In practice around two-thirds of the cases can be missed or unsuccessfully treated by non-psychiatrists due to complication of the symptoms (Cepoiu, McCusker et al. 2008).

It is advisable to carry out a thorough medical examination and selected investigations at the first session to rule out any physical cause. These include blood test measurement of TSH and thyroxine to rule out hypothyroidism; basic electrolytes and calcium for exclusion of metabolic abnormalities; and a full blood count including ESR to exclude any infective or chronic disease (Dale, Sorour, & Milner, 2008). A thorough medical and alcohol consumption history must be taken as well (Brook Dw 2002). Since hypogonadism has been named as of depression causes among men, testosterone levels should be measured to rule out hypogonadism (Orengo, Fullerton & tan 2004). Cognitive symptoms can be observed at the onset of a dementing disorder such as Alzheimer's disease (Reid & Maclullin 2006). Cognitive tests and brain imaging can be useful to distinguish between dementia and depression (Wright & Persad, 2007). If the symptoms are unusual and the course of the disorder is faster than normal, a CT scan is advisable to exclude any brain pathology (Kellner, Rubinow et al. 1983). There is no biological test that can diagnose or confirm major depressive disorder (Zimmerman M 1986). In general, these investigations should be performed only at the first session unless other medical indications persist (Ward, Doerr et al. 1983).

1.1.5 Management

The three most common treatments for depression are psychotherapy, medication, and electroconvulsive therapy. Psychotherapy is the treatment of choice for younger patients under 18, while electroconvulsive therapy is applied only as the last resort (Minter and Mandel 1979). Most depressed patients can be treated outside the hospitals; but if there is a considerable risk for self or others hospitalization must be considered (Thase, 1999).

Antidepressants have no or minimal therapeutic effects during mild or moderate episodes while considerable effect in severe episodes (Fournier et al. 2010). The effects of antidepressants are usually more prominent in comparison to psychotherapy, especially in cases of chronic major depression, although in short-term trials more patients — especially those with milder form of depression —arbitrarily terminate drug consumption, most likely as a result of adverse side effects of the medication and due to patients' preferences for psychological therapies over pharmacological treatments (Cuijpers, van Straten, van Oppen & Andersson, 2008). To achieve the best results and more patient's obedience, antidepressant medication with minimal side-effects and proper dosages should be administered and if necessary, combinations of different classes of antidepressants can be prescribed (Olfson M 2002). Response rates to the first antidepressant administered hover around 50–75%, and it can take at least six to eight weeks from the start of medication to therapeutic effects appear, when the patient can be back to normal life (Karasu , Gelenberg , Merriam & Wang 2000). Antidepressant medication treatment is usually continued for 16 to 20 weeks after remission, to reduce the chance of relapse, and in some cases up to one year of continuation is recommended. People with chronic depression may need to take medication for ever to avoid recurrence (Pampallona, 2004).

Selective serotonin reuptake inhibitors (SSRIs) are usually the first choice of medications prescribed due to their relatively mild side-effects, and their less toxic overdose risk in comparison to other antidepressants and if no desirable therapeutic effect appears after four weeks, patients can be switched to another antidepressant, and this increases the chance of improvement to almost 50% of cases (Whooley & Simon 2000) or switch to the atypical antidepressant Bupropion (Zisook, Rush et al. 2006).

1.2.1 Hypothalamic-Pituitary-Adrenal (HPA) System

The hypothalamic-Pituitary-adrenal axis (HPA) is a neuro-endocrine, stress-response system consisting of the hypothalamus, pituitary gland and adrenal gland. Activation of HPA leads to secretion of Corticotrophin releasing hormone (CRH) from the Para-ventricular nucleus of the hypothalamus. CRH in turn attaches to its receptors in the anterior pituitary gland causing release of adrenocorticotropic hormone (ACTH) which subsequently promotes secretion of glucocorticoids (cortisol in primates and corticosterone in rodents) from the adrenal cortex (Paul, et al. 2009).

Many factors including signals from neural, immune and endocrine system can have an effect on HPA axis (Venkataraman, Munoz, Candido, & Witchel, 2007). Acute and chronic stresses have been mentioned as activators of the HPA system (McEwen, 2000). Cytokines and other inflammatory mediators also can have inhibitory stimulatory effects on HPA axis by altering different components of the system (Path, Bornstein, Ehrhart-Bornstein, & Scherbaum, 1997). HPA system plays an important role in body's physiological protection and homeostasis. Hence, the activation of HPA axis induces only temporal release of cortisol which then comes back to its baseline level through negative feedback system (McEwen, 1998) in which cortisol down-regulates hypothalamic release of CRH which in turn down-regulates ACTH and cortisol secretion (Venkataraman, et al. 2007).

1.2.2 Alterations of HPA system and Hypercortisolemia in depression

Alterations of HPA system has been diagnosed in many mental states including post-traumatic stress disorder (de Kloet et al. 2006), schizophrenia (Walker & Diforio, 1997), social anxiety state (Beaton et al. 2006) and depression (Holtzheimer & Nemeroff, 2006). Hyper-activity and dysfunction of HPA negative feedback system are diagnosed as the cause of hypercortisolism (long-term excessive cortisol production) (Lopez-Duran, Kovacs, & George, 2009) which is a common comorbidity in depressive patients. (Vinberg, Bennike et al. 2008; de Kloet, Joels, & Holsboer, 2005; Gillespie & Nemeroff, 2005; Plotsky, Owens, & Nemeroff, 1998).

Hypercortisolemia disturbs the normal neuro-physiological pathway causing low mood and negative emotions, comprising main symptoms of depression (Nestler, et al. 2002). In healthy individuals, cortisol secretion follows a specific rhythm, having highest levels in the morning and lowest in the evening. (de Weerth, Zijl, & Buitelaar, 2003) However this rhythmic pattern can be impaired in some depressed patients due to exposure to stress conditions. For instance, a meta-analytical study revealed that depressed individuals have a relatively flat and non-responsive pattern of cortisol secretion while non-depressed individuals have a more dynamic and responsive pattern of cortisol secretion in response to stress (Burke, Davis, Otte, & Mohr, 2005).

1.2.3 Analysis of HPA system function

The HPA function can be evaluated by measuring cortisol levels (in urine, blood or saliva); CRH levels (in cerebrospinal fluid), ACTH levels (in blood) and expression level of glucocorticoid/mineralocorticoid receptor (receptors to which cortisol binds to arose physiological response) (Miller, Chen, & Zhou, 2007). Some other methods which can be utilized in analyzing HPA system include dexamethasone suppression test, CRH infusion test, and psychological stress exposure test (Lopez-Duran, et al. 2009).

Cortisol hormone, the primary end-point biochemical product of the HPA system, plays an important regulatory role in metabolic system (regulatory effect on glucose levels), central nervous system (modulates learning, memory and emotion processes); and in the immune system homeostasis (decreases inflammatory responses and lymphocytes migration) (Sapolsky, Romero, & Munck, 2000).

Salivary cortisol measurement has been utilized as indicator for HPA activity in human and animal studies (Gliddon, Darlington, & Smith, 2003; Raff, 2009; van Duijn et al. 2010).

Studies reveal that free cortisol concentration in saliva strikingly reflex both total and free/unbound cortisol concentration in the blood (Kahn, Rubinow, Davis, Kling, & Post, 1988; Perogamvros, Keevil, Ray, & Trainer, 2010).

Measurement of salivary cortisol has some advantages over plasma cortisol measurement including laboratory-independent, stress-free and non-invasive characteristics (Arafah, Nishiyama, Tlaygeh, & Hejal, 2007; Raff, 2009). Cortisol measurement in saliva has also

proved sensitive and reliable method in distinguishing between normocortisolemic and hypercortisolemic patients (Trilck, Flitsch, Ludecke, Jung, & Petersenn, 2005).

1.3.1 Human Gut Microbiome

Human microbiome is a general term being used for the entirety of ecologic community of microorganism, including their collective genomes, present in the human body (Zhu, Wang et al. 2010). There is an estimate that gut micbobiome in total outnumber all human cells by a factor of 10 and also encoding 100-fold more unique genes than human genome (Peterson, et al. 2009). Humans have been defined as "super-organism" since their body has "more microbe than human cells" (Roy, 2001). So the understanding of human biology and physiology has been tightly integrated to the understanding of its complex relationship with the characteristics of resident bacterial species (Davies, 2001)(Figure 1a).

In the dynamic evolutionary process between human being and resident bacteria, three main characteristic forms of relationship have evolved, namely; symbiosis, commensalism and pathogenicity (Bäckhed, Ley, Sonnenburg, Peterson & Gordon, 2005). Symbiosis is a relationship where at least one partner benefits without harming the other; in commensalism, both partners co-exist without harming or any obvious benefit to each other; and in pathogenic relationship the resident bacteria have harmful effect on their host (Bäckhed, et al. 2005; Hooper & Gordon, 2001).

The human microbiome project (HMP), was firstly aimed at characterizing trillions of bacteria cells residing in human body habitats (Legato, 2011). Several different sites on the human body, including nasal passages, oral cavities and oropharynx, saliva, skin, gastrointestinal tract, and urogenital tract were sampled in order to analyze the role of these microbes in human health and disease. The project was in fact "an experimental extension of human genome project with the overall goal of understanding the microbial parts of the human genetics and metabolism" (Peterson et al. 2009).

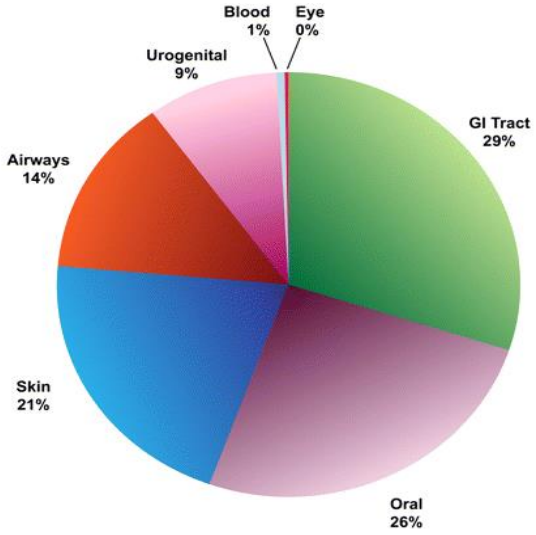


Figure 1a: Distribution pattern of bacteria in different body habitats with their bacteria being sequenced and characterized during the "Human Microbiome Project". The different colored segments represent the percentage of bacteria in that region compared to the total human microbial population. (Peterson et al. 2009)

The distal gut and its associated microbiota is a new field in the attempt for better understanding of human biology and evolution. Gut and intestinal microbiology have experienced a "minirenaissance" in the past 10 years (Marchesi, 2011). In a comprehensive review of the role of the gut microbiota in the health people, Sekirov and colleagues (Sekirov et al., 2012) retrieved the number of publications containing "gut microbiota" and similar terms in ISI Web of Knowledge database for the period from 1990 to 2009. During this period the number of publications increased five-fold in the yearly publication rate. Many reasons can be named as the explaining factors in this regard; the ever-growing recognition of the important role gut microbiota play in both healthy and disease conditions, as well as cross-pollination of ideas from other microbiologists studying other areas of environmental microbiology (Sekirov et al., 2012).

Anatomically, the human gut is divided into six segments, the oral cavity, esophagus, stomach, small intestine (subdivided into the duodenum, jejunum, and ileum), the colon or distal gut (subdivided into the ascending, transverse, and descending colons) and rectum (Marchesi, Eckburg et al)(Figure 1b). While the physiological role of the gut is generally defined as to digest the food we intake, it also provides a niche for colonization by a wide range of microbes. Each gut segment is inhabited by a specific microbial community, which to some extent reflects the role, anatomy and dynamic of that segment. The number and diversity of microbes increases as we move from the stomach to the rectum, where there is one of the most densely populated ecosystems containing between 10^{11}-10^{12} bacteria per gram of luminal bacteria (Whitman et al.1998). Since the distal gut contains one of the densest communities of bacteria [up to 55% of a stool sample comprises of bacteria biomass] it has attracted the majority of attraction (Cummings, Macfarlane & Englyst, 2001). However it is not correct to say that stool sample is the robust representative of the whole colon or small intestine; moreover, the mucosal surface differs significantly in the microbiota composition and density from that found in a stool sample from the same subject (Momozawa et al. 2011).

Due to some challenges encountered the microbiome culture in laboratory, on conventional laboratory media, and directly counting them, the shift has been made from traditional culture-dependent methods towards metagenomic and 16S/18S rRNA gene-based methods to characterize and define the species as well as the function of the species in the specimens. While metatranscriptomic and metaproteomic methods have been implemented, but until present to a lower extent have been applied in comparison to other traditional techniques (Wade, 2002; Echburg et al. 2005).

There has been debate among scientists as the establishment of the climax community in the gastrointestinal system and its perturbations due to interactions between human immune system and the bacterial community. In spite of many other major ecosystems including marine, terrestrial, deep-biosphere and atmosphere where have been colonized for millions to billions of years, the inhabitation of gut occurs immediately after birth and in majority of cases infants are born sterile [Apart from some rare cases where the newborn baby has swallowed amniotic fluid sludge] (Espinoza et al. 2005; Romero et al. 2008). There has

been increasing interest to find out the driving force for the colonization process and the role nature and nurture play in this regard. Specifically considering that findings in this regard can help us understand the potential ramifications for later life health. The climax community appears to be established within the first two years of life and after the first year, it has started to show the generalized adult distal gut inhabitation profile (Palmer et al. 2007). The factors that have been implicated in this regard included: the maternal microbiota (Dominguez-Bello et al. 2010), diet (breast fed versus formula fed; Favier et al. 2003), mode of delivery (normal versus caesarean; Biasucci et al. 2010), full or preterm gestation (Schwiertz et al. 2003; Morowitz et al. 2011), environmental exposure and clinical interventions such as antibiotics (Palmer et al. 2007), or gastrointestinal surgery (Zhang et al. 2009). By using metagenomic DNA (mgDNA) instead of the 16S rRNA gene this pattern can be approved (Koeng et al. 2011). There is a consensus between most scientists that the course of the colonization process is towards a similar outcome, being distal gut microbiota after the age of two predominantly colonized by Firmicutes and Bacteroidetes (Marchesi, 2011).

Apart from the two main phyla presented in almost all samples from the distal gut, the concept of "core microbiota" has not yet been proved to be inclusive and this term has not considered the "micro-eucaryotic" and "viral components" (Marchesi, 2011). The micro-eucaryotic diversity and numbers are several orders of magnitude lower than the bacteria and are leaned towards Candida, Saccharomyces spp. and Blastocystis spp.; while yeasts are rarely observed; only when a dysbiosis in the gut has occurred (Goldman and Huffnagle, 2009). Viral components are assumed at least an order of magnitude higher than the bacterial population in the distal gut; being regarded as drivers of community dynamics by some marine microbiologists (Suttle, 2007).

In addition, there is a significant number of Archaea in the distal gut; of which the most predominant is Euryarchaeota and in particular the Methanobacteriaceae family with *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* the two predominant *Archaea* found (Scanlan et al. 2008).

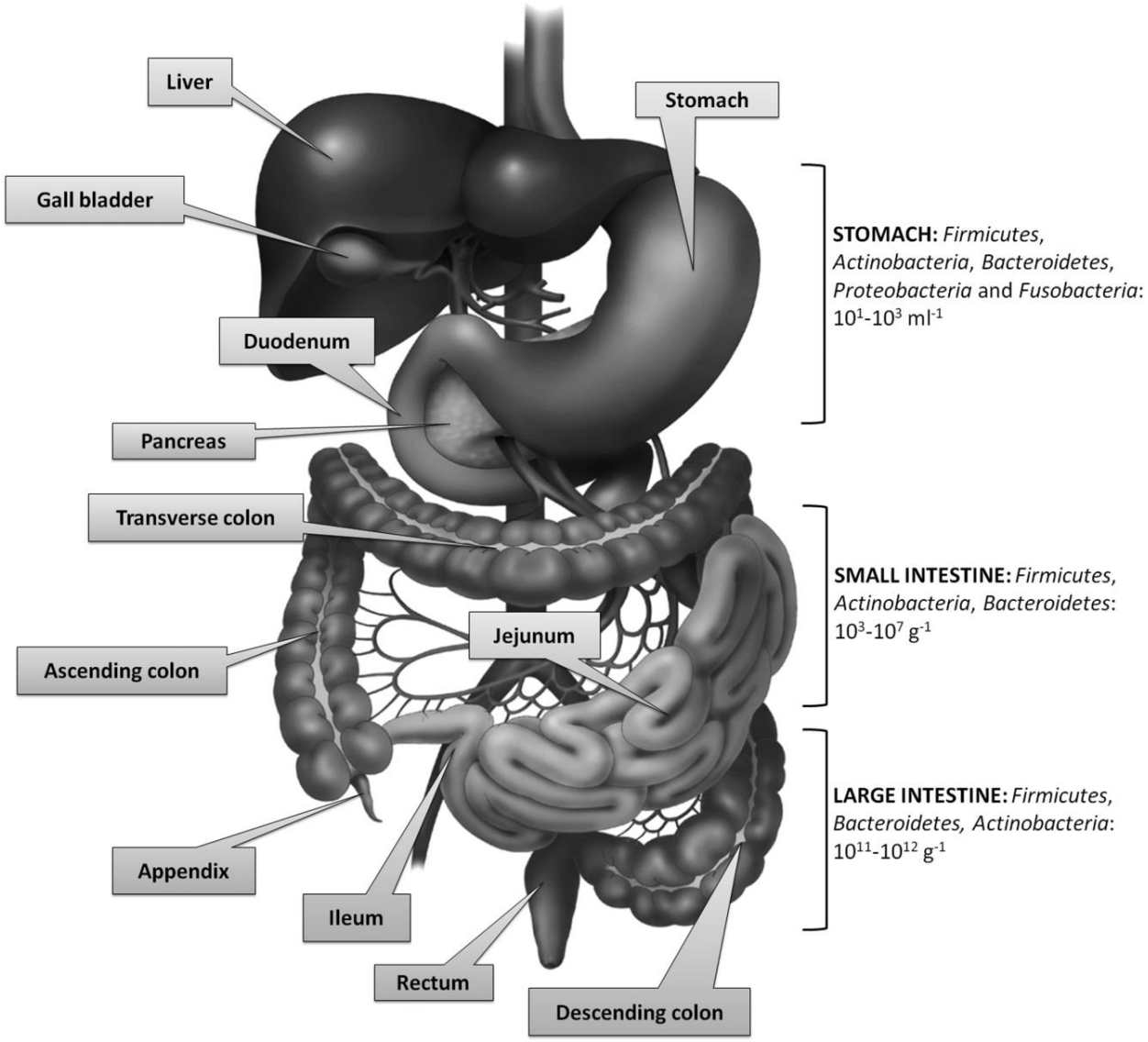


Figure 1b- The anatomy of the gastrointestinal tract, major bacterial phyla and their abundance in each segment. With curtesy to (Marchesi, Eckburg *et al*. 2005; Bik *et al*. 2006; O'Hara and Shanahan, 2006; McConnell *et al*. 2008; van den Bogert *et al*. 2011)

1.3.2 Human Gut Microbiome's contribution to health and disease

The paradigm of the human distal gut microbiome has changed its concept as one that being considered as a source of opportunistic pathogens to a virtual organ with the ability to health status of the host (Marchesi, 2011). In recent years many questions as the role of microbiomes arose; such as if a dysbiosis in the gut microbiota could lead to gastrointestinal diseases. If we consider the gut microbiota as an environmental factor, can they contribute to some diseases like inflammatory bowel disease, colorectal cancer, (Scanlan et al. 2008) irritable bowel syndrome, (Kassinen et al. 2007) Clostridium difficile-associated diarroea (Khoruts et al., 2010) and some general diseases such as cardiovascular disease, (Wang et al. 2011) obesity (Turnbaugh et al. 2008), arthritis, (Wu et al. 2010) and psychiatric diseases (Desbonnet et al. 2008). Many scientists are developing the idea that certain functions and associated microbes are beneficial to the heath of the host (Marchesi, 2011). For both beneficial and harmful effects that microbes can have, some metabolites have benn named among them are: SCFA (short-chain fatty acids), butyrate, acetate, lactate and propionate (Louis, 2010). The two bacterial groups that are mostly responsible for producing butyrate are the F.prauznitzii and Eubacterium rectal/Roseburia groups (Louis, 2010). Butyrate has been implicated in many cases as the contributing factor in controlling apoptosis, cytokine production, energy for colonocytes and mucus synthesis (Guilloteau, 2010). The incidence and prevalence of chronic inflammatory diseases have increased significantly during last decades according to the pattern of increasing urbanization and industrialization. This is particularly prominent for inflammatory disorders that develop in the mucosal tissues of the airway and the gut, such as asthma and inflammatory bowel disease (IBD)(Bach, 2002). The hygiene hypothesis was put forward to explain the increased incidence of these diseases in wealthy societies (Holt et al. 2008). Based on this theory, modern hygiene, and dietary and medical practices change the composition of the gut microbiota and confine exposure of infants to pathogens. This alteration in the microbiota, in combination with genetic and epigenetic factors, affects not only the epithelial mucosal barrier but also perinatal maturation of the immune system, thus increase the disease susceptibility (Renz, Brandtzaeg et al. 2012). Perinatal defects in the induction of "mucosal tolerance" are accompanied with the later development of allergies, autoimmune diseases (such as rheumatoid arthritis, type I

diabetes and systemic lupus erythematosus) and chronic inflammation of the gut and respiratory tract (Tulic et al. 2011).

According to many studies there is an ongoing competition between normal commensal bacteria and pathogen microbes in the gut; and imbalance in their relationship which is in favor of pathogenic bacteria for example due to improper intake of antibiotics can lead to gastrointestinal upset and diarrhea (Young and Schmidt 2004). The normal gut microbiota play this role by occupying attachment sites in mucosal cells, consuming nutritious agents for pathogenic bacteria and producing antibacterial agents such as lactic acid (Sekirov, Russell, Antunes, & Finlay, 2010).

Fermentation of the nutrients is one of the most important duties gut microbiota have. In fact human enzymes are not capable of digesting many nutritious agents which are intaken (Guarner & Malagelada, 2003). Synthesis of many life important agents such as vitamin K and vitamin B group (folate, riboflavin and vitamin B12) is contributed to gut microbiota (LeBlanc et al., 2011). Most of the orally administered drugs are detoxified or get their absorption and bioavailability altered by gut microbiome (Nicholson, Holmes, & Wilson, 2005).

1.4.1 Culture-independent analyzing methods

In spite of culture-dependent methods culture-independent methods do not rely on growing bacteria on specific media for isolation and characterization of them. Currently there are two main types of culture independent methods; namely “non-target fingerprinting technique” and “target specie-specific technique” (Tannock, 2001). Culture-independent molecular methods usually determine target molecular marker (usually bacteria 16SrRNA) using PCR amplification of the pool of target molecular marker; then separation/purification of them into pure components (through cloning, PCR-DG/TTGE, SSCP-PCR or T-RFLP); sequencing and identification through comparing with database reference sequences (Jany & Barbier, 2008). By hybridization of specie-specific signature sequence with labeled oligonucleotide probes (e.g. dot-blots hybridization, FISH and Micro-array); direct identification of gut microbial components can also be performed (Zoetendal, Collier et al. 2004). The nature of the community to be analyzed and the purpose of the experimental

study are two important factors in choosing the molecular technique and target microbial marker (Nocker, et al., 2007). According to recent studies; for increasing the throughput of gut microbial analysis; direct sequencing of the mixed target 16S rRNA followed by mathematical decomposition into pure bacterial spectral components via multivariate statistics (MCR-ALS) can be performed (Zimonja, Rudi, Trosvik, & Næs, 2008).

1.4.2 16S rRNA gene as microbial molecular marker

16S ribosomal RNA (encoded by *16S ribosomal RNA* gene) is a part of 30S small subunit of prokaryotic ribosomes which contains around 1,500 base pairs (Blaut, et al. 2002). This gene has been widely used for phylogenetic identification purposes (Woese, Kandler, & Wheelis, 1990); enabling scientists to distinguish between different species of bacteria in various environments, especially human gut (Blaut, et al. 2002; Favier, et al. 2002). There are many reasons for the widespread use of 16S rRNA as phylogenetic and molecular marker, among them are its presence in all bacteria, its ability to be sequenced directly, and its high degree of functional and evolutionary homology (Woese, 1987). In addition, considering the lack of artifacts of lateral gene transfer, relationships between 16S rRNAs can be applied in evolutionary relationships (Olsen, Lane, Giovannoni, Pace, & Stahl, 1986).

Analysis of *16S rRNA* gene shows nine "highly conserved and interspersed hyper variable regions (V1 – V9)" which include signatures of bacterial phylogenetic groups and species (Baker, Smith, & Cowan, 2003). This variability can be applied for engineering oligonucleotide probes for hybridization in order to distinguish bacteria at different levels of taxonomic hierarchy (Blaut, et al. 2002). Besides this, its conserved nucleotide sequence can be used in designing complimentary primer pairs for PCR amplification of a pool of bacterial *16S rRNA* gene fragment (Baker, et al. 2003). Designing universal primers have an important application in amplifying all the bacterial pool of *16S rRNA* gene (Watanabe, Kodama & Harayama, 2001) or specific primer pairs to identify a particular bacterial species (Rudi, Skulberg, Larsen, & Jakobsen, 1997). After sequencing bacterial *16S rRNA* gene fragment, it can be compared with many expanding 16S rRNA sequence databases in

Ribosome database project in order to identify the bacterial putative taxonomy (Cole et al., 2009a)

1.4.3 DNA Sequencing

DNA sequencing is performed to identify phylogenetic of the community microbial composition. The purpose of DNA sequencing is to determine the sequential order in which nucleotide bases are arranged in the target DNA molecule (Eckburg, et al. 2005). One of the first sequencing methods is "Sanger sequencing" which was first applied to characterize human gut microbial flora using *16S rRNA* gene molecular marker (Sanger, Nicklen et al. 1977)

Sanger sequencing technique comprises of "one-directional synthesis of a single-stranded DNA template by DNA polymerase in the presence of one DNA primer, deoxynucleotidetriphosphates – dNTPs (normal DNA building blocks) and dideoxynucleotidetriphosphates – ddNTPS (lacks 3′ - OH group)". Each of these four ddNTPs is uniquely fluorescently labeled and their subsequent incorporation leads to termination of extension reaction, producing copies of DNA fragments of different lengths (Sanger, Nicklen et al. 1977). Separation of these fragments can be done based on different molecular weight by using capillary electrophoresis while fluorescence light of corresponding wavelengths are read off hence resulting in DNA sequence electropherogram (Sanger, Coulson et al. 1980). Sanger sequencing is considered among first-generation sequencing methods and is considerably rapid and thorough in comparison to more advanced methods (Kircher & Kelso, 2010).

The Next generation sequencing (NGS) also sometimes referred to as high-throughput sequencing methods are more rapid and cost-effective sequencing methods which provide greater efficiency and higher throughput than Sanger sequencing method (Metzker, 2010). Some of the most known next-generation methods are: Roche/454 GS FLX Titanium sequencer, Illumina Genome Analyzer, Applied Bio systems SOLiD and Helicos HeliScope (Kircher & Kelso, 2010). Next generation methods have gained attention because of their potential to process millions of DNA fragments in a single run, which can be applied in

whole genome sequencing of human (Ley et al., 2008) and other model organisms (Hillier et al. 2008). These methods have also been used for profiling gene expression via RNA-sequencing (Nagalakshmi et al. 2008); determination of genomic structural variation (Korbel et al. 2007) and identification of protein-DNA interactions, histone/nucleosome modifications through chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Park, 2009). In addition, rapid molecular-based elucidation of gut microbial flora can be performed more easily by NGS (Rogers & Bruce, 2010).

In comparison to Sanger sequencing, NGS have higher error rates (Table 1a) and for analyzing relatively larger sequence data generated; they need more extensive bioinformatics tools (Kircher & Kelso, 2010; Mardis, 2008). In contrast to Sanger sequencing method in which the error source lies mainly in amplification procedure, polymerase slippage and contamination in the samples, the source of error in NGS includes: amplification step; phasing; mixed beads; neighbor interference and signal decline (Kircher & Kelso, 2010).

The table 1a shows Throughput (measured by amount of megabytes sequenced per day), Length of DNA reads to be sequenced (nucleotides), Quality (measured by average error rate for each nucleotide sequenced) and Costs (denominated in dollar per each megabyte of nucleotide sequenced).The costs amounts can be different according to time. These numbers do not include instrument purchase and maintenance expenses. All the data are based on information gathered by January 2010 (Kircher & Kelso, 2010).

Table 1a- Comparison of Sanger sequencing method with some next generation sequencing technologies (Kircher & Kelso, 2010).

	Throughput	Length	Quality	Costs
Sanger	6Mb/day	800nt	$10^{-4} – 10^{-5}$	~500$/Mb
454/Roche	750Mb/day	400nt	$10^{-3} – 10^{-4}$	~20$/Mb
Illumina	5,000Mb/day	100nt	$10^{-2} – 10^{-3}$	~0.5$/Mb
SOLiD	5,000Mb/day	50nt	$10^{-2} – 10^{-3}$	~0.5$/Mb
Helicos	5,000Mb/day	32nt	10^{-2}	<0.5$/Mb

1.4.4 BigDye Terminator

Applying the recently introduced BigDye™ terminators, large-template DNA can be directly sequenced with custom primers on automated instruments. Cycle sequencing conditions paved the way to sequence DNA samples isolated from a variety of microbial genomes (Heiner, Hunkapiller et al. 1998). Average read lengths of >700 bp from unique primer annealing sites are often adequate to cover final gaps in microbial genome sequencing projects without additional manipulations of template DNA. The technique can also be used to sequence-targeted regions, thereby bypassing laborious subcloning steps (Rosenblum, Lee et al. 1997).

In microbial genome or large-insert clone sequencing projects which apply the predominant random subclone sequencing strategy, progress tends to slow down dramatically at final stages as one confronts gaps. At these sites, DNA could be under-represented or unstable in subclones (Chen et al. 1996; Chissoe et al. 1997). Further sequencing with additional random subclones can be of no help, and it is frequently unavoidable to employ alternative cloning systems or additional methods such as long-range PCR to recover missing DNA (Chen et al. 1996). The variability of performance of these methods and the need for custom-tailored work usually inhibit the late stages of sequencing efforts. In contrast, if one can sequence directly from genomic DNA (or large-insert clones such as BACs or PACs)

with walking primers, tedious task to fill gaps could be fulfilled in a much shorter time (Ogino, Kawasaki et al. 2005).

For instance, in a recent study to sequence the 750-kb genome of *Ureaplasma urealyticum*, assemblage of ~13,000 sequence reads and combinatorial PCR reactions to join contigs remained two gaps. No λ pUC, or M13 subclones were recovered that cover the gaps, nor were PCR products obtained with any of several sets of neighboring primers. The difficulty of cloning these segments can be attributed to repeated sequences in and near the two gaps, but the high sensitivity of the recently introduced BigDye terminator (Rosenblum et al. 1997) allowed direct sequencing of the gap regions on genomic *U. urealyticum* DNA templates. Applying the conditions mentioned in this report, two gaps of 259 and 121 bp were sequenced from both strands with walking primers to complete the project of 751,723 bp (Zakeri, Amparo et al. 1998).

Direct sequencing was successfully applied for larger templates, and reliable results were reproducibly yielded with 1.2-Mb *Mycoplasma fermentans*, 2.3-Mb *Streptococcus pneumoniae*, and 4.6-Mb*Escherichia coli* genomic DNA. Additionally, several difficult gaps in sequencing projects with BAC clones, ranging in size from 140 to 250 kb, have also been covered by this method. Basically the method is of highest applicability whenever 2–3 μg of high-quality large-template DNA is available (Heinar et al. 1998).

1.4.5 Roche/454 GS FLS Titanium Sequencer

This method applies “sequencing by synthesis” based on pyrosequencing. "Sequencing by synthesis" involves taking a single strand of the DNA to be sequenced and then synthesizing its complementary strand enzymatically (Mardis, 2008). Essentially, the method performs sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step (Ronaghi, 2001). The template DNA is immobile, and solutions of A, C, G, and T nucleotides are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence

of the template (Kircher & Kelso, 2010). It differs from Sanger sequencing in that it is based on the detection of pyrophosphate release by nucleotide incorporation, rather than chain termination with dideoxynucleotides (Hamady, Walker et al. 2008).

At first, ssDNA template is hybridized to a sequencing primer and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with the substrates adenosine 5´ phosphosulfate (APS) and luciferin. By adding each of the four deoxynucleoside triphosphates (dNTPs); initiates the second step. DNA polymerase incorporates the correct, complementary dNTPs onto the template. This incorporation releases pyrophosphate (PPi) stoichiometrically. ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5´ phosphosulfate. This ATP allows the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a specific camera and analyzed in a program. Unincorporated nucleotides and ATP are degraded and removed by the apyrase, and the reaction continues with another nucleotide (Ronaghi et al. 1998).

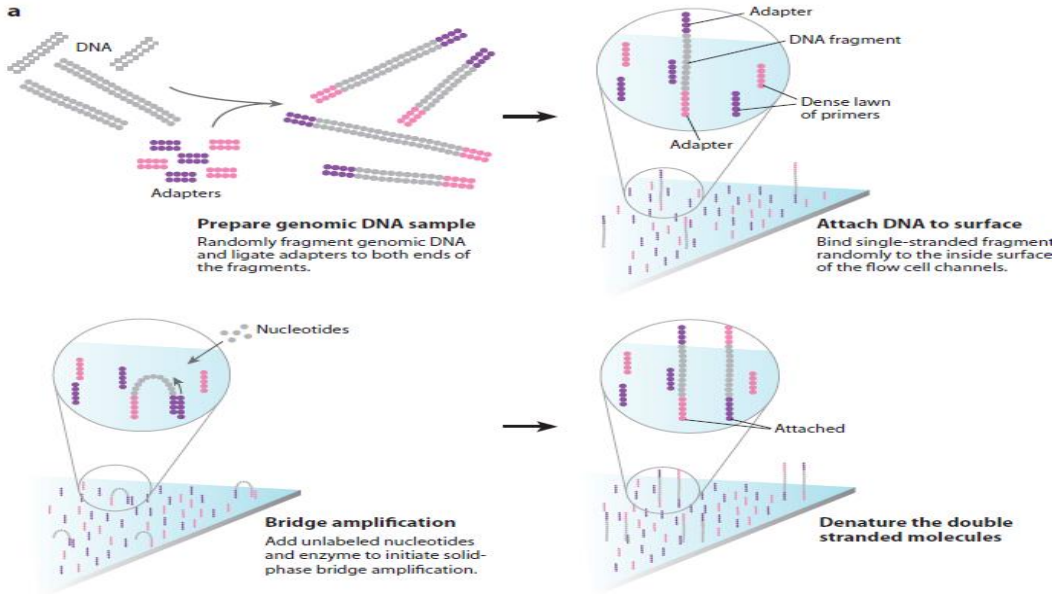
1.4.6 Deep Sequencing

Depth in DNA sequencing means the number of times a nucleotide is read during the sequencing process. Deep sequencing refers to the coverage, or depth, of the process which is many times larger than the length of the sequence under study (Huang and Madan 1999).

The term "deep" has been applied for a wide range of depths (>7x) and the newer term "ultra-deep" has emerged in the scientific literature to refer to even higher coverage (>100x). Although the sequencing accuracy for each individual nucleotide is very high, considering the very large number of nucleotides in the genome means that if an individual genome is only sequenced once, a considerable number of sequencing errors is probable. In addition, rare single-nucleotide polymorphisms (SNPs) are common. To be able to distinguish precisely between sequencing errors and true SNPs, increasing in the sequencing accuracy is more and more necessary; even further by sequencing individual genomes a large number of times (Hampton et al.2011).

1.4.7 Illumina Genome Analyzer

This sequencing method is based on "reversible dye-terminators that enables the identification of single bases as they are introduced into DNA strands" (Dalca & Brudno, 2010). It is often used for sequencing difficult regions, such as homopolymers and repetitive sequences. It can also be applied for whole-genome and region sequencing, transcriptome analysis, small RNA discovery, methylation profiling, and genome-wide protein-nucleic acid interaction analysis **(**Meyer & Kircher 2010). In this method the DNA molecules are first attached to primers on a slide and amplified so that local colonies are formed. Then four types of reversible terminate bases are added, each fluorescently labeled with a different dye and attached with a blocking group. The four bases then compete for binding sites on the template DNA to be sequenced and non-incorporated molecules are washed away. After each synthesis round, a laser is employed causing the removal of the 3' terminal blocking group and the probe. A detectable fluorescent color specific to one of the four bases is then visible, which is recorded through imaging techniques and converted by base-calling algorithms into nucleotide sequences. The process is continued until the full DNA molecule is sequenced. (Figur 1c) (Meyer & Kircher 2010).



Figur 1c- Illumina «flowcell». Genomic DNA hybridize to oligonucleotides via adapters on the DNA fragment. Figur er from Mardis (2008)

1.4.8 Applied Bio systems SOLiD

Applied Biosystems SOLiD (Supported Oligonucleotide Ligation and Detection) technology applies "sequencing by ligation". At first a pool of all possible oligonucleotides of a fixed length are labeled based on the sequenced position. Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences leads to a signal informative of the nucleotide at that position. Before sequencing, the DNA is amplified by emulsion PCR. The resulting beads, each containing single copies of the same DNA molecule, are deposited on a glass slide. The result is sequences of quantities and lengths comparable to Illumina sequencing (Schuster Stephan, 2008; Holt & Jones, 2008).

1.4.9 Helicos Heliscope

The "Helicos Genetic Analysis System" platform was the first commercial NGS implementation to use the principle of single molecule fluorescent sequencing, a method of directly characterization the exact sequence of a piece of DNA (Dalca & Brudno, 2010). The Helicos Genetic Analysis System has the potential to sequence nucleic acids, from several nucleotides to several thousand nucleotides. However, the yield of sequences per unit mass is dependent on the number of 3' end hydroxyl groups, and thus having relatively short templates for sequencing is more efficient than having long templates; optimally about 100-200nt. In this method, there is no need for library preparation or PCR amplification. Input DNA is first fragmented, melted and polyadenylated and at last fluorescently labeled adenine is added (Kircher & Kelso, 2010).

In addition to the NGS mentioned above, there are some newer methods under development expected to have greater efficiency and higher throughput. They include: Pacific Bioscience's Single Molecule Real Time (SMRT) sequencing technology, Oxford Nanopore's BASE technology and IBM'S silicon-based nanopores (Kircher & Kelso, 2010).

1.5.1 Bioinformatics and data analysis

For efficient and fast analysis of the great amount of sequencing data being produced by current molecular-based analytical tools, sophisticated bioinformatics tools are needed (Holt & Jones, 2008). Development of rapid algorithms tools usually results in trade-off between search speed and thoroughness (Smith, 1996). For instance, comparison between BLAST, FASTA and Smith-Waterman reveals that BLAST has highest speed but lowest accuracy while Smith-Waterman has highest accuracy but lowest Speed. (Speed: BLAST>FASTA>Smith-Waterman; Accuracy: Smith-Waterman>FASTA>Smith-Waterman) (Smith, 1996). "BLAST ("Basic Local Alignment Search Tool") and FASTA are heuristic, word-based local sequence alignment algorithm which is based on finding series of short, separate sub-sequences in the query sequences and then matching them with database sequences" (Altschul, Gish, Miller, Myers, & Lipman, 1990; Pearson, 1990). Smith-Waterman, however, applies dynamic programming technique in local sequence alignment and searches for optimal alignments between query and database sequences (Smith & Waterman, 1981). As well as enormous sequence data generated with New Generation Sequencing, the sequences are very short with usually unpaired ends-reads which highlight the importance of development of new bio-informatics tools for alignment purposes and *de novo* assembly of short reads sequences (Pop & Salzberg, 2008). Some examples of alignment algorithms for short read sequences are: Illumina's Eland short-read aligner (Pop & Salzberg, 2008), MAQ – "Mapping and Assembly with Qualities" (Li, Ruan, & Durbin, 2008) Bowtie (Trapnell & Salzberg, 2009) BLAT – "BLAST-Like Alignment Tool",(Kent, 2002) and Burrows-wheeler Transform (Dalca & Brudno, 2010). Of assemblers these can be named: SHARCGS, VCAKE, VELVET, EULER-SR, EDENA and ALLPATHS (Flicek & Birney, 2009).

Query sequence from Small Subunit ribosomal RNA (16S rRNA) gene can be searched for, using Ribosomal Database Project (RDP). RDP is a well-known web-based application (freely available at http://rdp.cme.msu.edu/) which gives data, tools and services related to rRNA sequences (Cole et al., 2005). RDP classifier makes use of naïve Bayesian algorithm to attribute putative taxonomic identity to query sequences. For each query sequence, a subset of eight-base words is randomly selected from the set of separately-created words.

The accumulative probability of finding the words in each subset is calculated for each genus; and sequence annotation at the genus level is determined based on the highest probability while taxa above the genus are assigned according to genus assignment (Cole, Wang, Chai, & Tiedje, 2011; Wang, Garrity, Tiedje, & Cole, 2007). The database is continuously being updated with rRNA sequences from the International Nucleotide Sequence Databases (INSD: GenBank/EMBL/DDBJ). From September 2008 (release 10.3), the Ribosomal Database Project (RDP) contains 33 082 archaeal and 643 916 bacterial small subunit rRNA sequences (Cole et al. 2009b).

One of the 16S rRNA gene sequence databases is Greengenes (available at http://greengenes.lbl.gov) with the potential to screen chimera as well as taxonomic classification (DeSantis et al. 2006).

1.5.2 Analysis of mixed bacterial 16S rRNA gene sequence

In contrary to pure 16S rRNA sequence whose taxonomic classification can be immediately performed through comparison with database reference sequences; mixed sequence spectra generated from direct sequencing of amplified pool of rRNA genes should first be decomposed into pure component. After decomposition stage, the direct sequencing on amplified pool of rRNA genes is easier, faster and less expensive and yields higher throughput than most molecular techniques including cloning (Zimonja et al. 2008).

Another recently developed multivariate technique for analyzing mixed spectral sequence is "Multivariate Curve Resolution with Alternating Least Square (MCR-ALS)" which "separates mixed component spectra into the pure contributions of each component in the sample" (Garrido, Rius, & Larrechi, 2008). MCR-ALS has been promising in many fields and industries. Some of these applications are: "quantitative overlapping component in chromatographic peaks" (Tauler & Barceló, 1993) identifying individual artificial food coloring spectral component (Lachenmeier & Kessler, 2008) and scanning DNA folding pattern using florescence energy transfer (Kumar, Kanchan, Gargallo, & Chowdhury, 2005). Zimonja et al. for the first timed tried to implement the technique in resolving mixed DNA spectra. They carried out a comparison between MCR-ALS and multivariate calibrated

Partial Least Squares (PLS) regression model in a two designed experiments comprising mixtures of three and seven bacteria respectively. The results from both methods were comparable with explained variance of 97.15% (MCR-ALS) versus 97.81 (PLS) and 97.13 (MCR-ALS) versus 97.91% (PLS) for three – and seven – component data set, respectively (Zimonja, et al. 2008). According to these studies, the MCR-ALS was considered as a competent resolution of mixed DNA spectra and can be applied in mixed natural microbial samples (Zimonja, et al. 2008).

The mixed spectral model (Zimonja, et al. 2008) can be formulized as:

D=**CS** + **E**

In which MCR-ALS tries to separate the mixed matrix **D***(IxJ)*, into concentration matrix **C***(Ixk)*, matrix of pure bacteria's spectral profiles **S***(KxJ)* and a residual term **E***(IxJ)*

I = number of mixture samples; J= number of spectral measurements and K = number of constituents in the mixture. Only the initial matrix **D** is known before MCR-ALS analysis. (Zimonja, et al. 2008)

Determination of initial estimates of spectra profiles or concentration matrix is of high importance (Zimonja, et al. 2008). Principal Component Analysis (PCA) can be deployed in this regard. PCA is a multivariate statistical tool with the potential of "data compression and visualization of significant information contained in large data set by converting a number of correlated variables into uncorrelated variables called principal components". This provides a new set of variables that offer linear combinations of the original variables which are uncorrelated and represent the most important structure of the data (Rudi, Zimonja, Trosvik, & Næs, 2007). Evolving Factor Analysis (EFA) can also be utilized in preparing the initial estimates as a starting point in MCR-ALS analysis (Zimonja, et al. 2008).

Multivariate regression models including "Multiple linear Regression" (MLR) and Partial Least Square (PLS); have also been deployed in analyzing mixed bacterial samples. They provide high-throughput and giant quantitative potentials (although PLS is more effective than MLR) in mixed microbial sample analysis. One of the disadvantages of MLR is that it

relies on prior knowledge of unit spectra of genetically coherent units (GCU) while PLS is based on calibration of spectra corresponding to samples of known mixtures ratios (Trosvik Soldani, Tondo & Baldessarini, 2007). These drawbacks are overcome in MCR-ALS because it is a calibration-free regression model with no need for prior knowledge (Zimonja, et al. 2008).

Besides its advantages, there are still some considerations with MCR-ALS method to be attended such as non-negativity (values of pure constituents' spectral profiles and/or their concentrations are non-negative), selectivity and zero-concentration (defined windows of mixture spectra that are descriptive for pure constituents), uni-modality (each constituents has only one peak in its spectrum) and closure (total amount of constituents in each mixture should be constant) for providing effective spectral resolution (Tauler, Smilde, & Kowalski, 1995; Zimonja, et al. 2008)

1.6.1 Significance and purposes of the Project

Depression is a common psychiatric disorder with its etiology has not yet been fully understood. Understanding of contributing and predisposing factors to depression can be of great importance in the prevention, diagnosis and treatment of this burdensome disease. Disturbances in Hypothalamic-Pituitary-Adrenal axis (measurable by salivary cortisol), a neuro-endocrine stress-response system, may have effect in occurring or exacerbation of many psychiatric states including depression. Alterations in the gut microbiome composition have been implicated in causing and complicating depressive symptoms. However, the extent of such potential effects is not fully known. (Figure 1d)

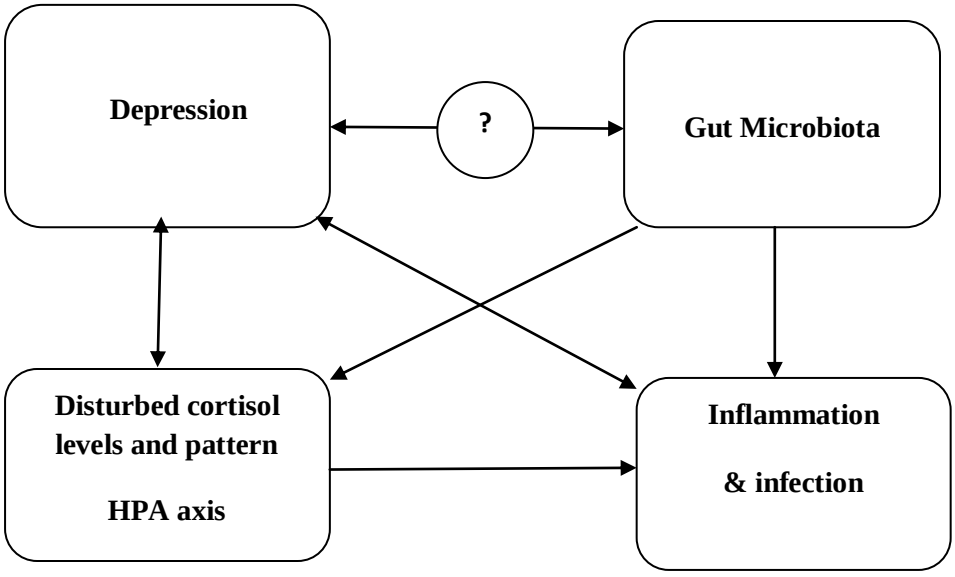


Figure 1d Correlation between depression, cortisol levels and HPA axis, inflammation, infection and gut microbiota. Direct correlation between depression and gut microbiota has not yet been approved by any comprehensive study.

Various studies have also mentioned gut microbiota and their deviations as a prominent feature in depressive patients, although the causative relationship between them is not yet established. It has been well known from long time ago that inflammatory and infectious diseases have a direct and immediate relationship with the level as well as pattern of secretion of cortisol. The role of gut microbiota in provoking immune system especially hommoral immune system has also been understood. Regarding the interest scientist have shown in recent years in finding out more and more about gut microbiota composition and role in maintaining health and normal integrity in the body as well as potential consequences which any deviation and disturbance in gut microbiome can incur on the host; we can find relative small number of studies which directly discuss any relationship between gut microbiota and development and exacerbation of depressive signs and symptoms.

This study aims at finding any direct relationship between gut microbiota and presence or absence of depressive disorder and cortisol levels in sample groups. The gut microbiota composition in depressed patients as well as not-depressed people, here referred to as

control group, will be determined and compared to each other. In order to be more precise in characterizing phyla in gut miribiota we will benefit the next generation techniques namely, Illumia. This study can also provide us with clues about early diagnosis and treatment optimization as well as prognosis of depression disease based on gut microbiome composition. A similar study has been done in the last year by Opyemi O. with 30 samples, 18 depressed patients and 12 control samples; but could not find any significant relationship between depressive disorder and gut microbiota. In this study, we repeated the experiment for both Opyemi's samples and 26 new extra samples, in the hope that by entering more samples in the experiment and using new next generation method, Illumina sequencing, we would find out more about correlation between gut microbiota and its composition in depressed and healthy people and depression in general.

Three main objectives of this project are to:

- Evaluate of correlation between gut microbiota and depression
- Characterize the most common bacterial phyla and species in gut microbiota through stool samples
- Find correlation between baseline HPA function (assessed by salivary cortisol measurement) and depression
- Find any eventual correlation between gut microbiota and cortisol levels

1.6.2 Framework of the Project

This master's thesis is part of a larger clinical survey on depression and immunology led by Prof. Knut Hestad at Sykehuset Inlandet HF, Hamar, Norway.

The outline of scheme is as shown below (figur 1e):

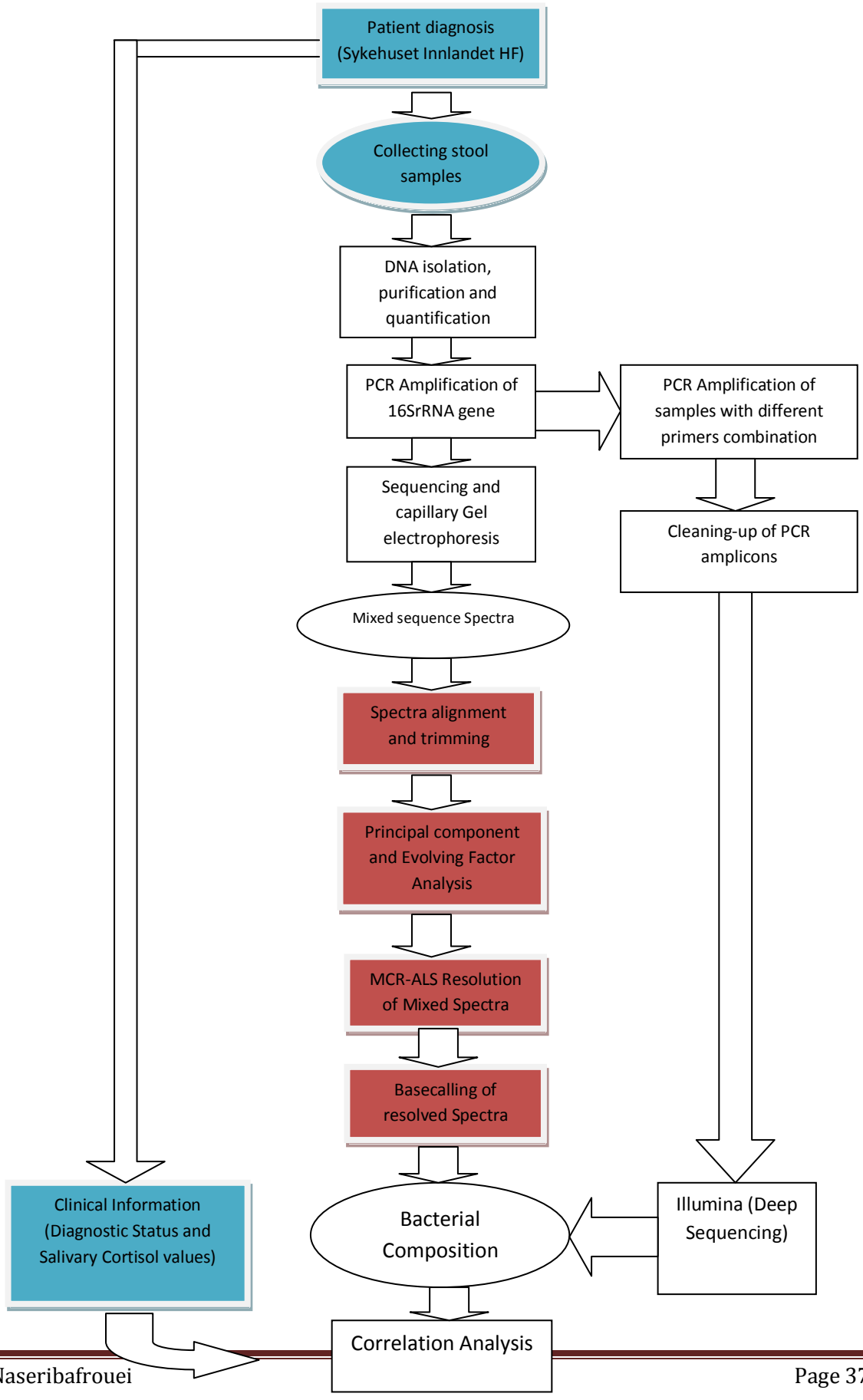

Patient diagnosis (Sykehuset Innlandet HF)
Collecting stool samples
DNA isolation, purification and quantification
PCR Amplification of 16SrRNA gene
PCR Amplification of samples with different primers combination
Sequencing and capillary Gel electrophoresis
Cleaning-up of PCR amplicons
Mixed sequence Spectra
Spectra alignment and trimming
Principal component and Evolving Factor Analysis
MCR-ALS Resolution of Mixed Spectra
Basecalling of resolved Spectra
Clinical Information (Diagnostic Status and Salivary Cortisol values)
Bacterial Composition
Illumina (Deep Sequencing)
Correlation Analysis

Figure 1e-: Schematic outline of the project. Rectangles illustrate procedures while ellipses represent data and materials. Procedures in red background were performed with the help of my main supervisor Prof Knut Rudi (Hedmark University College, Hamar). Procedures, material/data in blue background were fulfilled and gathered by my external supervisor Prof Knut Hestad (Sykehuset Inlandet HF, Hamar).

The experimental group (37samples) includes patients with major depressive disorder as diagnosed under ICD-10:Bipolar affective disorder - current episode severe without psychotic symptoms; Bipolar affective disorder - current episode severe with psychotic symptoms; Severe depressive episode without psychotic symptoms; severe depressive episode with psychotic symptoms; Recurrent depressive disorder - current episode severe without psychotic symptoms; Recurrent depressive disorder – current episode severe with psychotic symptoms. The control group (19samples) includes patients referred to the neurological department for spinal fluid examination without any known neurological or other diseases.

2. Material and Methods

2.1 Stool sample collection and clinical information

Stool samples from fifty six (56) adults were collected from Inland Lillehammer hospital – Sykehuset Inlandet HF. The subjects were chosen from hospitals' inpatients and outpatients. Thirty seven (37) of the samples were from clinically-depressed patients (experimental group); while nineteen (19) were from non-depressed patients (control group).Of these fifty six (56) samples thirty (30) belong to the first round of the experiment which was performed by Opyemi (Opyemi's master thesis, Hedmark University College, 2012) during the last year and the rest twenty six had been chosen randomly for the second round. I repeated the experiment for Opyemi's samples to include more samples in the experiment.

The stool specimens were weighed and then "S.T.A.R" (Stool Transport and Recovery) buffer solution was added to each sample at a ratio of around 1(stool) to 3(S.T.A.R). Samples were gently but thoroughly vortexed to achieve homogenous suspension. Samples were then stored at $-80^{O}C$ for later usage. The very low temperature is of significance to avoid variability in the stool bacterial composition. Deep-freezing has been established to preserve bacterial composition in fecal samples (Dan, Richardson, Miliotis, & Koornhof, 1989; Gorman & Adley, 2004). S.T.A.R. buffer plays its role in preservation of the bacterial composition by binding to inhibitory substances, inactivation of nucleases and stabilizations (S.T.A.R. Buffer, 2004).

The study was designed as a partially-blind experiment in which clinical information including diagnostic status and salivary cortisol values remained unknown until after gut microbial analysis were finalized.

2.2 Salivary Cortisol measurement

Salivary cortisol measurements were fulfilled in the hospital. Three measurements had been done for each patient – during morning and before breakfast (around 08:45 AM), mid-day before lunch (around 13:30 PM) and evening before dinner (around 9:00 PM).

The following things should not be done at least one hour before sampling:

1) Severe physical activity before sample taking, taking shower is also prohibited; but normal morning washing is allowed
2) Brushing the teeth, but washing the mouth with cold water is permitted
3) Using tooth prick was not allowed to avoid any blood in the samples
4) Smoking, eating and drinking

The patients were given a cotton-head swap to chew on in order to take to saliva. These samples were then sent to hormone laboratory in Aker Hospital to be analyzed. None of the patients had any changes in medication during the last month before inclusion and nobody received immunosuppressive drugs. Patients with known or suspected concomitant diseases known to influence inflammatory mediators (such as infections, autoimmune diseases, malignancies, congestive heart disease, and pulmonary disease) were not included in the study. The regional ethic committee approved the study and written informed consent was obtained from each individual.

2.3 DNA Isolation and Purification

The frozen stool samples were first allowed to be thawed on ice pack in the fume cupboard. Autoclaved 2ml tubes containing 250mg glass beads were filled with a suspension volume of 0.5ml of the stool sample. To achieve bacterial cell-lysis; homogenization was performed applying MagNa Lyser (Roche, Switzerland) twice at 2000rpm for 40 seconds and 40 seconds rest between runs. The samples were kept cold during the rest phase in order to avoid DNA degradation due to overheating. This step was followed by centrifugation at 3500rpm for 5minutes. The supernatant lysate solution was then transferred to a new microcentrifuge tube in two replicates for each of the samples. Along with the samples negative control sample (distil water) was included in the experiment to exclude false positive responses which may arise due to DNA contaminants.

50µl supernatant from the tubes was transferred to KingFisher 96- well (KF) plate and after addition of 50µl lysis buffer, 5 µl Proteinase K solution (1.25µl; 20mg/ml) was added to it

to achieve the final concentration of 500µg/ml. The plate was then placed into King Fisher Flex robot and by applying "ProteinaseMagMiniLGC", the samples were incubated at 55^{O}C for 10 minutes. The addition of Proteinase K enzyme was aimed at inactivation of nucleases which may degrade DNA molecule (Lizardi & Engelberg,1979).

When the cell lysis was achieved, 500µl of Binding buffer [5M Guanidinium thiocyanate; 150mM Tris-HCL ph 6.8; 7.5mM EDTA] and 20µl of SiMAG/MP-DNA magnetic Beads (200mg/ml) were added respectively; and then vortexing and incubation was performed for five minutes at room temperature.

The next DNA Purification steps were performed by applying Robotic BioSprint 96 workstation (Qiagen GmbH, Germany) which makes advantage of automatic magnetic—particle based technology in the washing steps and elution of DNA molecule(Qiagen,2012). 105 µl of sample solution with 16 µl Mag particles and 50 µl Ethanolwere added to each well. Three washing steps were respectively performed applying 1ml washing Buffer I [60% Guanidine hydrochloride; 1% Triton-X100; 10mM Tris Hcl pH 8; 1mM EDTA]; 1ml washing Buffer II [70% ethanol] and 1ml washing buffer III [double-distilled water]. At the end, elution reaction was fulfilled with 63ml elution buffer [10mM Tris Hcl pH 8; 1mM EDTA].

Measuring the concentration of purified DNA was carried out by applying Nanodrop spectrophotometer (ND-1000, Solis Biodyne). After DNA isolation, the samples should be appropriately labeled and stored at -4^{O}C for short term and -80^{O}C for long term. In addition, the eluted genomic DNA was evaluated using ethidium bromide-stained 1.5% electrophoretic agarose gel, run at 100voltage for 20 minutes on 1XTAE buffer [40mM Tris; 20mM acetic acid; 1mM EDTA]. A mixture of 5µl of DNA eluate with 1µl of 6X loading buffer [0.25% bromophenol; 0.25% xylene cyanol; 30% glycerol] were loaded into the wells of the solidified gel. 5µl of 1Kb molecular weight DNA ladder was also injected in the first well of the gel. By having such a ladder at the first column, visual evaluation and comparison of DNA presence and size under UV light after the electrophoresis will be possible.

2.4 PCR Amplification of extracted DNA for Sanger sequencing

The Polymerase Chain Reaction amplification was aimed to amplify a conserved region of 16S rRNA gene of about 1,100basepairs size. A PCR master mix at the final concentration of (without DNA template) [1.25U Hotfirepol®DNA polymerase; 1Xhotfirepol®buffer; 2.5mM Magnessium dichloride; 0.2mM ddNTP; 0.2µM Mangala F-1 (forward primer); 0.2µM 16S1015UR (reverse primer)] was at first prepared on ice in DNA-free room of the laboratory. This is mean to prevent any form of DNA contamination. The master-mix was then thoroughly mixed and finally DNA template was added to each respective well. In the first PCR round, the DNA template volume used was one µl. Because by using this amount of template the signals were weak on the agarose gel; this volume was increased to 2µl in the second PCR reaction. In the second PCR-set up; the diluted concentration of template was used. For this reason 2 µl of template was mixed with 8µl of distilled water and of this diluted solution 2µl was used in the PCR .Instead of 17.25µl water, this time 16.25µl water was used to prepare the master mix to reach the total reaction volume (25µl).

The PCR thermocycler was operated at the initial denaturation temperature of 95^{O}C for 15 min. The denaturation step was followed by thirty cycles of denaturation at temperature of 95 ^{O}C for 30 seconds, annealing reaction at 55 ^{O}C for 30 seconds and elongation step at 72 ^{O}C for 1 minute, 20 seconds. Final elongation step was fulfilled at the temperature of 72 ^{O}C for ten minutes to ensure full extension of any remaining single-stranded DNA. The reaction was then completed by a final stage of 10 ^{O}C for indefinite time.

The PCR amplicons were then run on 1.5% agarose gel stained with ethidium-bromide, and after separation of the bands by electrophoresis, UV light was applied to visualize them. The electrophoresis was carried out at 100voltage for 20 minutes.

For Illumina sequencing a similar PCR step was also performed, with the difference of using different combination of forward and reverse primers for each sample. This time the thirty cycle of denaturation, annealing and elongation consisted of 95 ^{O}C for 30 seconds; 50 ^{O}C for one minute and 72 ^{O}C for 45 seconds respectively. 2 µl of undiluted template was also used in this reaction.

2.5 DNA Sequencing and Capillary Gel electrophoresis

Considering that some primers might have not been reacted and eliminated, the PCR amplicons (1µl) were treated with Exonuclease I reaction master mix [ExoI 0.4U/µl (0.2µl); 5X BigDye Sequencing buffer (1µl); DNAse/RNase free water to 5µl] and incubated at 37 ^{O}C for one hour. This incubation stage was followed by treatment at 85^{O}C for 15minutes in order to inactivate the exonuclease I enzyme and at the end the process was finalized at 10 ^{O}C for unlimited time.

The Exo-treated PCR products (5µl) were treated with a master mix of BigDye® Terminator v1.1 (1µl); 5X BigDye sequencing buffer (1µl); 0.32µl of 10µM U515FC30 sequencing primer and DNAse/RNAse-free water up to 10µl volume. The thermocycler was run at initial denaturation stage of 96 ^{O}C for one minute. This stage was followed by twenty five cycles of alternate denaturation (96^{O}C for 30seconds), annealing (50^{O}C for 5seconds) and elongation (60^{O}C for 4minutes). The process ended at 10^{O}C for unlimited time length.

The post sequencing products were cleaned up using the master mix 2 µl EDTA 125mM; 52 µl Ethanol 96^{O} and 2 µl NatriumAcetate 3M. After addition of this master mix to each well, the plate was incubated for 15 minutes in room temperature and then was centrifuged at 4400rpm at 4^{O}C for 30 minutes. Then the plate was upside downed gently to take the supernatant out. After that 70 µl of Ethanol 70 ^{O} was added to each well of the plate and centrifuged again at the 4400 rpm and 4^{O}C for 15 minutes. After centrifuge, the plate was again upside downed to get the supernatant out of wells. To get sure that all the remaining supernatant has got out, the plate was centrifuged upside down for one minute at 150 rpm. The plate was allowed to dry in room temperature for thirty minutes. 30 µl of 0.1mM EDTA was added to each well and the plate was centrifuged again for 5 minutes at 4400rpm at 4^{O}C and then was sent for sequencing.

Purified dye-labelled extension PCR products were then put into the DNA sequencing machine. In the DNA sequencing machine the products were directly injected into the tubes and separated from each other according to size-based capillary electrophoresis on Applied Biosystems 3130xl Genetic Analyser with 36cm-long capillary array and POP-7™ polymer type with injection time of 3seconds. Electrophoretic separation of labeled fragments is

based on the migration of charged fragments under the influence of electric field through a matrix with liquid medium. During this size-based migration, lighter DNA fragments move faster than heavier ones. The last labeled dideoxynucleotide of the varied-sized DNA fragments emits fluorescence signals which are detected by a charge-coupled device camera with four different reading channels for each fluorescence color. Finally, sequences are base-called with Sequencing Analysis v5.3.1 (Applied Biosystems).

2.6 Clean-up of mixed PCR products:

PCR products which were prepared by using different forward and reverse primers so that no template had the same combination of forward and reverse combination, needed to be cleaned up to be sent for Illumina sequencing. For this reason after PCR amplification of all samples, the samples were mixed together; 2µl of strong signaling samples, 4 µl of weak signaling samples and 8 µl of very weak signaling samples were transferred and mixed in a 1.5 ml autoclaved ependorf tube. The cleaning up process was performed applying E.Z.N.A cycle kit manual.(Omega bio-tek, USA,ENZA Cycle pure kit). At first, 250 µl of Buffer CP was added to 50 µl of the mixed PCR products solution in 1.5ml tube and then thoroughly vortexed and then centrifuged for some seconds to collect the drops from the inside of the lid. HiBind® DNA column with 2 ml collection tube was applied and then tubes were centrifuged at 15000 rmp for one minute at room temperature. The HiBind® DNA column was washed with 700 µl of DNA Wash Buffer and then centrifuged again for one minute at room temperature. This step was repeated with 500 µl of DNA Wash Buffer and the collected liquid at the bottom of the collecting tube was discarded. For good DNA yield, the empty HiBind® DNA column was centrifuged again at 15000 rpm for two minutes and then was placed into a clean 1.5ml tube. 30µl of Elution Buffer (10mM Tris,pH 8.5) was added to the column matrix and centrifuged for one minute at 15000 rpm to elude DNA.. The eluted Elution Buffer with DNA was transferred from 1.5 ml tube back onto the column matrix and centrifuged again at 15000 rpm for one minute to increase DNA yield. The pre-clean up and post-clean up samples were run on the agarose gel and taken picture of and

then their concentration was quantified using nanodrop quantification (Quanti-iT^{TM} Assay). During absorbance ratio analysis, the purity of the amplicons after clean-up process is being determined. The primer dimmers should be eliminated during this cleaning-up procedure so that when the before and after clean-up samples are run on the agarose gel and compared under ultraviolet camera, the after clean-up samples should not show primer dimmers in the picture to be suitable for illumine deep sequencing. We used Miseq illumine instrument and were analyzed with Qiime programe.

2.7 Bioinformatics and Statistics

Analytical process of the mixed DNA sequence spectra generated was kindly fulfilled by my supervisor - Professor Knut Rudi. The procedure was carried out under the platform of MATLAB ® R2010a software (The mathWorks Inc., Natick MA, USA), Statistical and Bioinformatics toolboxes for MATLAB®. The statistical analysis of the data was performed with the help and supervision of Professor Arne Linløkken. "R" statistical program was used for data analysis.

2.8 Pre-processing of Raw mixed DNA Sequence

Raw spectral sequences which were gathered from samples must be treated to eliminate some undesirable properties such as having variable start and end points, non-synchronization of peaks' shifts and scaling variation between samples (Zimonja, et al. 2008). In order to improve the data quality, some pre-processing modulations should be performed.

First of all, in order to have sample sequences with same start and end point, we should align them to each other which enable us to trim the sample sequences to a region containg desirable fragment. At the next stage, normalization of sequence data to eliminate scale variability and correlation optimized warping (COW) to correct non-synchoronization of peaks were undertaken (Zimonja, et al. 2008).

2.9 Resolution of processed mixed DNA sequence Spectra using MCR-ALS

Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) was performed on the processed mixed spectra for gathering information on relative concentration of each component in individual sample of the data set. By applying Principal Component Analysis (PCA) and Evolving Factor Analysis (EFA), the number of dominant bacterial spectral component had been determined before. The average concentration of each bacterial component was calculated in all samples including depressed and normal samples.

2.10 Base calling of resolved spectral sequence

In order to assign the order of bases in resolved pure DNA sequences, an in-house base caller script was run in MATLAB® environment. By having different fluorescent peaks with four different color intensities correspond to the four nucleotide-types (T-thymine, C-cytosine, A-Adenine and G-guanine) we would be able to determine the order of bases in the spectrum. The software was designed to look after the intensities measured of each nucleotide for peaks one after another. Finally, the bases are sorted and the best calculated nucleotide sequence is attributed to each spectrum.

2.11 Taxonomic identification of bacterial components

In order to attribute taxonomic identity to the bacterial components, each sequence was compared to annotated reference sequences in Ribosomal Database classifier program which applies Naïve Bayesian algorithm (Q. Wang, et al. 2007). The RDP Naïve Bayesian classifier enables us to thoroughly and effectively attribute 16S rRNA sequences to the bacterial taxonomy down to genus level (Wang et al. 2007). RDP provides us with a bootstrap confidence score which represents the level of confidence of assignment of query sequence to the taxon (Lan, Wang, Cole, & Rosen, 2012). The default bootstrap is set at 80% for almost full-length of 16S rRNA sequence; while 50% cut-off has proved to be satisfactory for correct classification of partial query sequence of length shorter than 250 base pairs but longer than 50 base pair (Claesson et al. 2012). RDP classifier was run

applying a confidence threshold value of 50% because all our read query sequences were within these boundaries.

2.12 PCR amplification of extracted DNA for Illumina

For Illumina sequencing another PCR step was also performed, with the difference of using different combination of forward and reverse primers for each sample so as to none of the samples in our experiment had the same combination of forward and reverse primers. These combinations are presented in the appendix. After performing the PCR, the strength of the signals were visualized under the UV light, so that the amount of the PCR apmlicons to be sent for Illumina deep sequencing was determined according to the severity of signals. For strong signals we mixed 2µl of the PCR amplicons; for moderate signals 4µl of PCR amplicons and for weak signals 8µl of the final PCR amplicons to the mixture we sent for Illumina deep sequencing. This time the thirty cycle of denaturation, annealing and elongation consisted of 95 ^{O}C for 30 seconds; 50 ^{O}C for one minute and 72 ^{O}C for 45 seconds respectively. 2 µl of undiluted template was also used in this reaction. After sequencing the DNA stretch in the 16S rRNA gene, the 100 most dominant operational taxonomic units (OTUs) were provided and classification was performed according to these OTUs. The classifier which was used in this regard was RDP Native Bayesian rRNA classifier version 2.5 and the taxonomical hierarchy was RDP 16S rRNA training set 9.

2.13 Correlation Analysis

2.13.1 Correlation of bacterial components or OUT with depression

Since the distribution of none of the bacterial components is normal we could not use T-test or ANOVA. Instead of these tests we applied non-parametric tests such as Krusskal Wallis and Fisher exact test. The null hypothesis (which claims that there was no difference in components' or OUT concentration between depressed and non-depressed) was rejected if P-value was 0.05 or less. Since we had two replicates of each person in our samplings we used the average levels of each replicate in all our analyses. At first we got sure that there was no significant statistical difference between these two replicates. For Fisher exact test we used median value of the bacterial component and divided the sample values in two,

those above this threshold and those under this threshold and compared them between depressed and non-depressed people.

2.13.2 Correlation of Salivary Cortisol measurements with depression

The average salivary cortisol values (morning, mid-day and evening) measured between both depressed and non-depressed samples is illustrated by Bar-plots graph. Evaluation of any correlation between salivary cortisol values and depression was also performed using Spearman rank test analysis. The null hypothesis (which claims that there is no difference in average salivary cortisol concentration between depressed and control group) can be rejected at P-values lower than 0.05.

2.13.3 Correlations between bacterial components or OTUs and salivary cortisol values

The correlation between bacteria components' proportions and salivary cortisol measurements and OTUs and cortisol measurements was evaluated in the two groups using "Spearmann" statistical test. The aim is to reveal if there is a significant relationship between bacteria components proportions (in mixed data sequencing) or OTUs (in Illumina sequencing) and cortisol concentration in depressed and control subjects. In this regard Microsoft Excel 2010 and "R" statistical software (http://www.r-project.org/) were used. The null hypothesis (which states that there is no significant relation between bacterial component or OTUs and cortisol values) was rejected at significance level of 0.05 or less.

Since the data were not normally distributed in the bacterial components, we used Spearman statistical test.

3. Results

3.1 Part One: Mixed Data Sequencing

3.1.1 Electrophoretic gel analysis of PCR amplicons of 16S rRNA gene fragment

The combined extracted DNA was then amplified using specific primers, Mangala F1 and 16 S u 1510 (10mM), to selectively amplify a conserved DNA region of approximately 1.1 kb par size. The DNA amplicons were then run on agarose gel of 1.5% concentration at 100voltage for thirty minutes to get sure that they were successfully amplified. In each case a negative control was run besides the real samples in order to rule out the possibility of DNA contamination of the samples. In negative controls all the agents in the master mix except the template were added. No band should be visible in the negative controls except the band for primer-dimers. In all samples a band composed of primer-dimers was shown. If the signals were weak, the PCR was repeated this time with higher concentration of template or more PCR cycles. (figure3a)

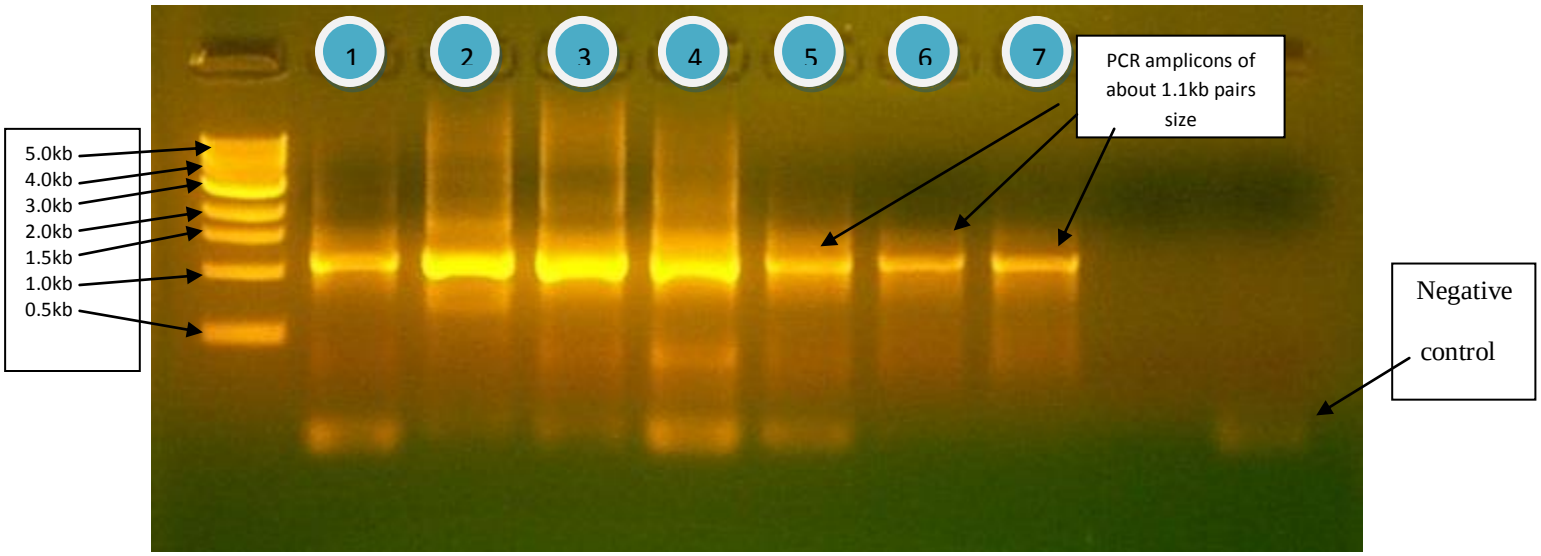


Figure 3a- Representative gel image of PCR amplicons of 16S rRNA gene fragment (about 1.1kb pair size) run along 1kb molecular weight DNA ladder (first lane). The agarose gel of 1.5% concentration was run at 100voltage for 30mins. Columns 2, 3 and 4 show strong signal; while columns 1, 5, 6 and 7 show moderate signal. The last column belongs to the negative control.

3.1.2 Quality of Raw Sequencing Data

The quality of sequenced mixed DNA electropherogram was assessed by viewing by ABI sequence scanner v1.0. One hundred and twelve (120) traces were generated from fifty six samples (two replicates for each sample). If the plate signals were not strong enough, the cleaning and preparation step was repeated to reach a satisfactory signal for most of the plates. Figure 4 is an example of the attained Plate report while different trace score highlighted with different color codes. They were ranked as high quality (score> 20); medium quality (15 < score ≤ 20) and low quality (score≤ 15). However, there were some traces with no readable nucleotide information and zero trace score. Most of the sequences had high/medium trace score (figure 3b). Only mixed sequences of high-quality or medium quality were selected for further analytical procedures. A representative example of raw mixed sequence data as observed by ABI sequence scanner software v1.0 is shown in the appendix.

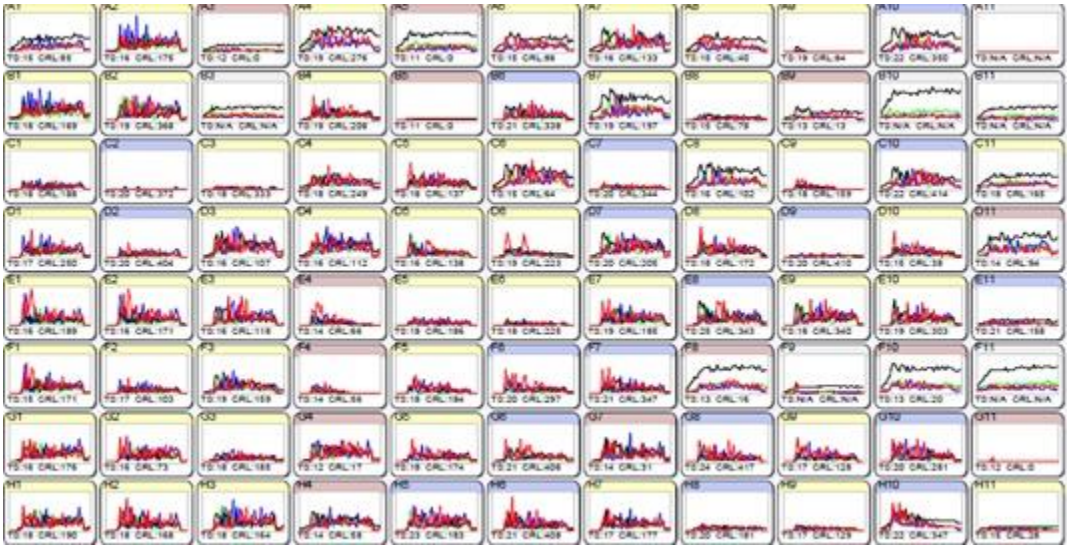


Figure3b- A representative of Plate report of the sequenced samples presenting the trace score value.-TS and Continuous Read Length- CRL (highest uninterrupted stretch of base). Purple color represents low-quality trace score; yellow color medium quality and blue color high-quality. If the trace did not contain readable nucleotide information, no trace score was assigned.

3.1.3 Resolution of Mixed sequence Spectra using MCR-ALS

PCA predicted four dominant bacterial components while EFA predicted twenty dominant bacterial components. MCR-ALS analysis performed on the EFA-predicted twenty components gave the relative concentration of the bacterial components in each mixed sample. Each component has a different concentration in comparison to other components representing a specific bacterial phylum.

3.1.4 Taxonomic classification of the bacteria components (Mixed Data Sequencing)

The taxonomic status of bacterial components at different levels of hierarchical categories (Domain, Phylum, Class, Order, Family and Genus) was determined based on the Ribosomal Database classifier. Reliability of the taxonomic classification of the components, measured by bootstrap threshold, represents a relatively high confidence value in most of the cases. As we move further to the end of the hierarchical classification tree, the confidence of this classification reduces. According to the table 3a, around two thirds of the components (thirteen out of twenty) belong to “firmicutes” phylum; 15% of the components belong to bacteroidetes” phylum; 10% of components belong to “chlamydiae” phylum and only one component (5%) belongs to “chlorophlexi” phylum. None of the phyla presented in this classification have significantly higher or lower abundance between depressed patients in comparison to non-depressed control people.

Table3a. Taxonomic placement of bacterial components at different levels of hierarchy based on mixed data sequencing; determined by using Ribosomal Database classifier tool with confidence value of 50%. The individual components' confidence values, at different hierarchical level are shown in parentheses.

Comp	Domain	Phylum	Class	Order	Family	Genus
Comp1	Bacteria (100%)	Firmicutes (91%)	Clostridia (91%)	Clostridiales (91%)	Ruminococcaceae (82%)	Papillibacter (35%)
Comp2	Bacteria (98%)	Chlamydiae (17%)	Chlamydiae (17%)	Chlamydiales (17%)	Simkaniaceae (8%)	Simkania (8%)
Comp3	Bacteria (100%)	Bacteroidetes (94%)	Bacteroidia (91%)	Bacteroidales (91%)	Prevotellaceae (90%)	Hallella (57%)
Comp4	Bacteria (96%)	Firmicutes (34%)	Clostridia (32%)	Clostridiales (21%)	Lachnospiraceae (12%)	Lactonifactor (2%)
Comp5	Bacteria (99%)	Firmicutes (59%)	Clostridia (51%)	Clostridiales (49%)	Ruminococcaceae (17%)	Anaerotruncus (11%)
Comp6	Bacteria (80%)	Chloroflexi (11%)	Anaerolineae (8%)	Anaerolineales (8%)	Anaerolineaceae (8%)	Longilinea (5%)
Comp7	Bacteria (100%)	Firmicutes (77%)	Clostridia (75%)	Clostridiales (75%)	Ruminococcaceae (67%)	Papillibacter (18%)
Comp8	Bacteria (100%)	Firmicutes (78%)	Clostridia (76%)	Clostridiales (75%)	Clostridiaceae 1 (12%)	Anaerosporobacter (11%)
Comp9	Bacteria (100%)	Firmicutes (100%)	Clostridia (99%)	Clostridiales (99%)	Lachnospiraceae (93%)	Lactonifactor (46%)
Comp10	Bacteria (92%)	Proteobacteria (29%)	Alphaproteobacteria(10%)	Rhodospirillales (5%)	Rhodospirillaceae (5%)	Defluviicoccus (5%)
Comp11	Bacteria (91%)	Firmicutes (33%)	Clostridia (30%)	Clostridiales (30%)	Ruminococcaceae (15%)	Faecalibacterium (9%)
Comp12	Bacteria (99%)	Firmicutes (35%)	Clostridia (28%)	Clostridiales (27%)	Clostridiaceae 4 (6%)	Geosporobacter (5%)
Comp13	Bacteria (95%)	Firmicutes (48%)	Clostridia (44%)	Clostridiales (42%)	Ruminococcaceae (11%)	Hydrogenoanaerobacterium(6%)
Comp14	Bacteria (89%)	Chlamydiae (11%)	Chlamydiae (11%)	Chlamydiales (11%)	Simkaniaceae (10%)	Simkania (10%)
Comp15	Bacteria (86%)	Firmicutes (35%)	Clostridia (30%)	Clostridiales (29%)	Ruminococcaceae (12%)	Fastidiosipila (8%)
Comp16	Bacteria (97%)	Firmicutes (58%)	Clostridia (51%)	Clostridiales (50%)	Eubacteriaceae (21%)	Garciella (19%)
Comp17	Bacteria (100%)	Bacteroidetes (67%)	Bacteroidia (58%)	Bacteroidales (58%)	Bacteroidaceae (31%)	Bacteroides (31%)
Comp18	Bacteria (99%)	Bacteroidetes (82%)	Bacteroidia (72%)	Bacteroidales (72%)	Bacteroidaceae (55%)	Bacteroides (55%)
Comp19	Bacteria (100%)	Firmicutes (100%)	Clostridia (99%)	Clostridiales (99%)	Ruminococcaceae (87%)	Ruminococcus (53%)
Comp20	Bacteria (100%)	Firmicutes (64%)	Clostridia (60%)	Clostridiales (58%)	Lachnospiraceae (49%)	Lachnobacterium (18%)

3.1.5 Evaluation of the similarity between two replicates in each component

Each sample had two replicates, and statistical analysis showed no significant difference between these two replicates. To reject any potential significant difference between the two replicates in each bacterial component, Kruskal-Wallis test was applied. This test is able to find any statistical difference in data which are not normally distributed. Having no significant statistical difference between these replicates assures that the procedures from DNA extraction, purification, PCR proliferation and sequencing were performed thoroughly and no striking deviation and contamination had happened during these different stages (table3b). After getting sure that the two replicates are almost the same, for the rest of statistical analyses the average values of each replicates were used. Using both the replicates separately can lead to different degree of freedom and standard deviation and wrong results. The null hypothesis (which states that there is not any significant relationship between bacterial component and depression) was rejected at significant level of 0.05.

Table3b- Comparison between replicates of each component using Kruskal-Wallis test. None of the components had significant difference between their replicates. (All p-values are higher than significance threshold of 0.05). Null hypothesis states that there is no significant difference between replicate a and replicate b.

componets	pValues
Comp1	0.63
Comp2	0.54
Comp3	0.89
Comp4	0.69
Comp5	0.24
Comp6	0.49
Comp7	0.94
Comp8	0.83
Comp9	0.68
Comp10	0.59
Comp11	0.89
Comp12	0.63
Comp13	0.77
Comp14	0.39
Comp15	0.44
Comp16	0.79
Comp17	0.87
Comp18	0.96
Comp19	0.73
Comp20	0.86

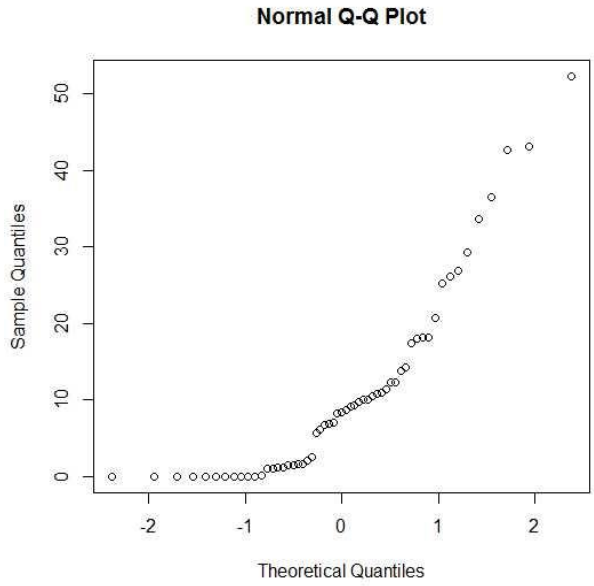


Figure3c- Data distribution pattern for component7; this plot represents distribution pattern of values for bacterial component 7. If the points were dispersed along a relatively straight line from the left lower corner of the plot to the right upper corner of the plot the distribution pattern of the data was regarded as normal. The same distribution pattern is seen for other 19 bacterial components.

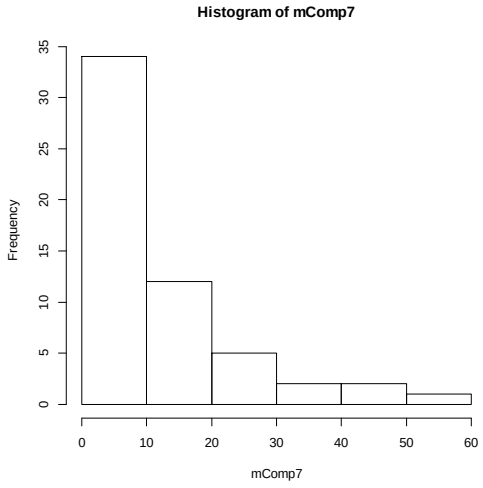


Figure3d- Distribution pattern of bacterial component 7; According to the histogram shows, the values are not normally distributed in component7. The shape is long from the bell shape as in normal distribution pattern.

3.1.6 Correlation between bacterial components and depression

To find any correlation between gut bacterial components and depression state; two different statistical tests were used, namely Fisher exact test (table3c) and Kruskal-Wallis test (table3d). These two tests are non-parametrical tests. Since the values of bacteria components were not normally distributed (figure3c and 3d); unpaired sample T-test analysis and ANOVA test could not be used. In fact in each bacterial component, around 35-40% of the samples had zero value.

Among 20 bacterial components, just component 7 had significant relationship with depression state, which was shown in both statistical tests, Fisher exact test and Kruskal-Wallis test. This component was significantly higher between depressed people than in non-depressed people. This component belongs to Firmicutes, clostridia, clostridiales, Ruminococcaceae, Papillibacter. In Fisher exact test, both mean and median values were used as threshold for analysis which only median gave the significant correlation. The relatively high difference between mean and median arises from non-normal distribution pattern of data in each component. Components 15 (p-value=0.06) and 12 (p-value=0.09) have relatively low p-values although not significant. These two components also belong to Firmicutes, chlostridia phylum and class. This finding is repeated and reinforced in the Kruskal-Wallis test.

Table3c – Fisher exact test results based on mean values and median values. Only component 7 has a significant correlation with depression status. Component 7 belongs to Firmicutes, Clostridium. mcomp refers to the average of replicates (compa and compb) of each component.

	Mean	**p-value**	**Median**	**p-value**
mcomp1	7.61	0.73	2.27	0.79
mcomp2	5.12	0.78	2.28	1
mcomp3	4.73	1	1.27	0.16
mcomp4	11.29	1	1.27	0.16
mcomp5	10.79	0.35	6.08	1
mcomp6	6.88	0.38	3.12	0.26
mcomp7	10.95	0.07	8.38	0.05*
mcomp8	12.39	0.76	7.61	1
mcomm9	10.25	0.50	3.29	0.23
mcomp10	8.82	0.38	4.53	1
mcomp11	12.21	0.23	7.08	1
mcomp12	6.40	0.24	2.99	0.09
mcomp13	8.95	0.14	6.18	0.78

mcomp14	10.84	0.15	4.82	0.40
mcomp15	8.08	0.22	4.58	0.06
mcomp16	7.42	0.50	1.93	0.57
mcomp17	12.14	0.21	1.84	1
mcomp18	17.15	0.78	15.41	0.78
mcomp19	9.85	0.16	3.90	0.08
mcomp20	11.22	0.76	7.74	1

Table3d- Kruskal-Wallis rank sum test; Correlation between each component and depression status. mComp refers to the average of replicates (compa and compb) of each component.

	Chi square	**p-value**
mComp1	0.3593	0.54
mComp 2	0.1504	0.69
mComp 3	2.2101	0.13
mComp 4	0.031	0.86
mComp 5	0.0152	0.90
mComp 6	0.5993	0.43
mComp7	5.081	0.02*
mComp 8	0.6294	0.42
mComp 9	0.3573	0.55
mComp 10	0.1508	0.69
mComp 11	0.9328	0.33
mComp 12	1.1398	0.18
mComp 13	1.0397	0.30
mComp 14	0.2686	0.60
mComp 15	0.3984	0.12
mComp 16	0.5482	0.45
mComp 17	0.0154	0.90
mComp 18	0.0028	0.95
mComp 19	3.5258	0.06
mComp 20	1.188	0.27

3.1.7 Combined correlation effect of components

Along with direct correlation analysis between each component and the depression state, which revealed a significant correlation between component 7 and depression, combined correlation analysis between each two components belonging to the firmicutes phylum and the depression state was carried out by "Spearmann" statistic method. Some components may not have a significant relation with depression state by themselves, but in combination with other components, especially when these two components belong to the same phylum,

may have a significant correlation with the depression. In our experiment, component1 along with component 11 and component 20 had significant correlation with depression state (negative correlation). This was also correct for component 8 which had a significant correlation with depression in combination with component 7 and component 19 (positive correlation).(table3e)

Table3e. Combined correlation effect of components; Spearman statistic analysis was used. Of all the components belonging to Firmicutes phylum, Clostridia class; component1 along with component 11 and component 20 had significant correlation with depression state (negative correlation). This was also correct for component 8 which had a significant correlation with depression in combination with component7 and component19 (positive correlation) Degree of freedom was 54 and p-values less that 0.05 were significant.

Combined correlation	pValue	type of correlation
Comp 1 & Comp 11	0.001	negative correlation
Comp 1 & Comp 20	0.049	negative correlation
Comp 8 & Comp 7	0.044	positive correlation
Comp 8 & Comp 19	0.035	positive correlation

3.1.8 Correlation between cortisol levels and depression

In contrast to distribution pattern of bacterial components which are not normally distributed, the cortisole values are normally distributed. Cortisol was measured in three distinct times of the day, morning, mid-day and evening. Both normal and depressed patients represent a decline in the cortisol levels as the day time goes further and this is clearly shown in the figure 3e. Cortisol has normally its highest secretion level during the morning time and this pattern has long been discovered both in humans and animals. In all three cortisol measurements, the average cortisole levels are higher among depressed patients rather than normal controls; but this difference is only in mid-day measurement significantly higher among depressed patients (p-value=0.01). (table3f)

As it is presented in the figure5, the deviation of cortisol measurements were high both among depressed and normal samples. Cortisol can be influenced by many factors including stress, inflammation, infection, and so on. For finding this relationship between cortisol levels and depression status, Kruskal-Wallis test was used which can be used for both normally and non-normally distributed data.

Table3f- *Kruskal-Wallis rank sum test; correlation between cortisol levels and depression status*. Cortsol was measured in three times during day; morning (between 7-9 AM), mid-day (between 1-2 PM) and evening (between 9-11 PM). Only cortisole 2 (mid-day) is significantly higher between depressed patients.

	Chi square	pValue
Cortisole 1	1.26	0.26
Cortisole 2	5.48	0.01*
Cortisole 3	2.34	0.12

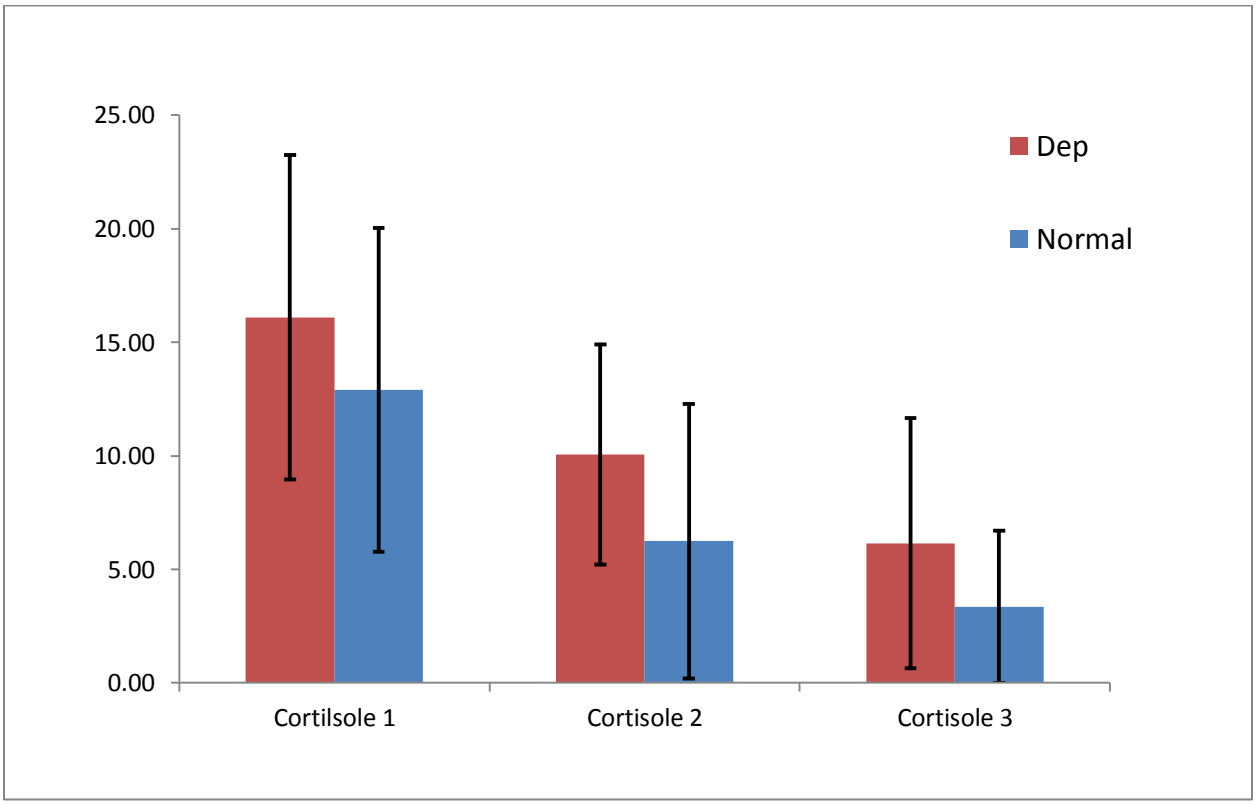


Figure3e. Bar plot of average salivary cortisol values in depressed (red-colored bars) and control (blue-colored bars). Cortsol was measured in three times during day; morning (between 7-9 AM), mid-day (between 1-2 PM) and evening (between 9-11 PM). Salivary cortisol is measured base on ng/ml. The vertical error lines display standard deviations from mean values in both directions.

3.1.9 Correlation between cortisol levels and bacterial components

The summary of statistical analyses of correlation between bacterial components and cortisol values are presented in table3g. It contains information about p-value (statistical significance) calculated based on Spearmann statistic test. This test was applied because the

bacterial components were not normally distributed. As it is highlighted by the asterisk in the table, correlation between mcomponet14 and cortisol levels in the morning and cortisol levels in the mid-day was significant (positive correlation).

Table3g- Correlation between cortisol values and bacterial components; Spearmann statistic test was used, p-value less than 0.05 was considered as significant (positive correlation).

	Cortisol morning	Cortisol mid-day	Cortisol evening
mComp1	0.412	0.092	0.634
mComp2	0.734	0.881	0.752
mComp3	0.123	0.460	0.365
mComp4	0.854	0.247	0.853
mComp5	0.743	0.584	0.265
mComp6	0.574	0.406	0.546
mComp7	0.269	0.116	0.810
mComp8	0.243	0.926	0.632
mComp9	0.345	0.134	0.543
mComp10	0.834	0.084	0.163
mComp11	0.222	0.851	0.734
mComp12	0.867	0.387	0.194
mComp13	8.842	0.181	0.394
mComp14	0.049*	0.012*	0.070
mComp15	0.495	0.847	0.257
mComp16	0.942	0.401	0.277
mComp17	0.264	0.243	0.834
mComp18	0.243	0.410	0.284
mComp19	0.432	0.172	0.537
mComp20	0.745	0.555	0.092

3.2 Part Two: Illumina Deep Sequencing

3.2.1 PCR Proliferation

The PCR was performed according to the protocol mentioned in the material and methods with the different combinations of forward and reverse primers so that none of the samples in our experiment had the same combination of forward and reverse primers. These combinations are presented in the appendix. After performing the PCR, the strength of the signals were visualized under the UV light, so that the amount of the PCR amplicons to be sent for Illumina deep sequencing was determined according to the severity of signals. Figure 7 represents an example of the 1.5% agarose gel being run at 100voltage for 30 minutes in this regard. (figure3f)

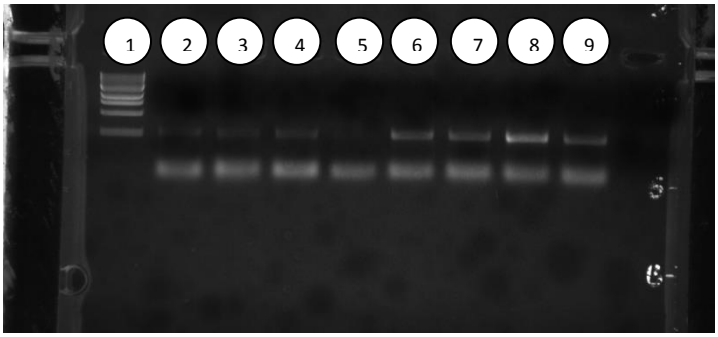


Figure3f- Representative gel image of PCR amplicons for Illumina sequencing. The agarose gel of 1.5% concentration was run at 100voltage for 30mins. The first column is 1kb ladder; Column 8 has strong signal, column 6 moderate signal and column 4 weak signal.

3.2.2 Illumina Deep Sequencing Taxonomic Classification

After mixing the PCR amplicons according to their concentration, the samples were sent for Immumina Sequencing analysis. Table below represents the 100 most dominant operation taxonomic units (OTUs) in the samples. Definition of operational taxonomic units according to NCBI is as follows: "Taxonomic level of sampling selected by the user to be used in a study, such as individuals, populations, species, genera, or bacterial strains." Or it can be regarded as the terminal node in the phylogenetic analysis. Recently some scientist has defined bacterial species as organisms with 16S rRNA gene sequences having at least 97% identity- operational taxonomic units. (Cole et al. 2009b). High resolution approach for OUT assignment was applied so there is possible for one genus to have more than one OUT. Single representative sequences from each OTU were selected and classified using the RDP database. None of the OTUs in our experiment were present in all the samples, neither in depressed patients nor in normal people, rejecting the notion “core microbiota”. The most dominant OUT belongs to Bacteroidetes phylum, Bacteroides genus, presenting in more than 97% of the samples. Among the 10 most dominant OTUs, most of them belong to Bacteroidetes phylum, and among Bacteroidetes phylum, Bacteroides comprises the most frequent genus. Among Firmicutes phylum, Faecalibacterium is the most frequent genus in the gut environment, being present in about 97% of all samples. As it is illustrated, the depth of sequencing in Illumina sequencing is much higher in comparison to previous sequencing method (mixed sequencing analysis) and classification based on Sanger sequencing and many relatively rare species have been sequenced (table3h). The confidence rate for the classification even further down the taxonomic levels is high. As for the previous sequencing results, Bacteroidetes and Firmicutes constitute around 90% of the phyla present in the gut.

Table3h- The table below represents the sequencing and classification of the 100 most dominant OTUs in Illumina sequencing. All of them belong to domain bacteria with 100% confidence. The classifier is RDP Native Bayesian rRNA classifier version 2.5. Taxonomical Hierarchy: RDP 16S rRNA training set 9. The percentage in front of each phylotype shows the confidence rate. If there is not a percentage in front of them, the confidence rate equals 100%.

s	Phylum	Class	Order	Family	Genus
1	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
2	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae(98%)	Bacteroides(98%)
3	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
4	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
5	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia(84%)
6	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
7	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
8	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
9	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
10	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
11	Firmicutes (98%)	Clostridia(98%)	Clostridiales (98%)	Ruminococcaceae (96%)	Ruminococcus (85%)
12	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
13	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides
14	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia(10%)
15	Firmicutes (70%)	Clostridia(69%)	Clostridiales (68%)	Eubacteriaceae (12%)	Anaerofustis(10%)
16	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillibacter(51%)
17	Firmicutes	Negativicutes	Selenomonadals	Veillonellaceae	Dialister
18	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
19	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
20	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
21	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia(96%)
22	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
23	Firmicutes (55%)	Clostridia(52%)	Clostridiales (51%)	Lachnospiraceae (19%)	Cellulosilyticum (13%)
24	Proteobacteria (89%)	Alphaproteobacteria (89%)	Rhizobiales (89%)	Hyphomicrobiaceae (89%)	Gemmiger(89%)
25	Firmicutes (98%)	Clostridia (98%)	Clostridiales (98%)	Ruminococcaceae (98%)	Ruminococcus (90%)
26	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
27	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Coprococcus(94%)
28	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
29	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
30	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae(96%)	Faecalibacterium(96%)
31	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
32	Firmicutes (99%)	Clostridia (99%)	Clostridiales (99%)	Lachnospiraceae (99%)	Pseudobutyrivibrio (40%)
33	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
34	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Barnesiella
35	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
36	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospiracea_incertae_sedis(47%)
37	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
38	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
39	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospiracea_incertae(95%)
40	Bacteroidetes	Bacteroidia (99%)	Bacteroidales (99%)	Rikenellaceae (99%)	Alistipes (92%)
41	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospiracea_incertae(69%)
42	Proteobacteria ia (88%)	Alphaproteobacteria ia (88%)	Rhizobiales (88%)	Hyphomicrobiacea (88%)	Gemmiger (88%)
43	Firmicutes (97%)	Erysipelotrichia (96%)	Erysipelotrichale (96%)	Erysipelotrichacea(96%)	Clostridium XVIII (85%)
44	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae(99%)	Alistipes
45	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillibacter(34%)
46	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
47	Firmicutes (85%)	Clostridia (85%)	Clostridiales (85%)	Ruminococcaceae (85%)	Flavonifractor (53%)

48	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillibacter(91%)
49	Proteobacteria (81%)	Alphaproteobacteria (81%)	Rhizobiales (81%)	Hyphomicrobiaceae (88%)	Gemmiger(81%)
50	Firmicutes (99%)	Clostridia (99%)	Clostridiales (99%)	Ruminococcaceae (90%)	Faecalibacterium(89%)
51	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillibacter(63%)
52	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
53	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
54	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
55	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Oscillibacter(63%)
56	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
57	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae(99%)	Bacteroides(99%)
58	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
59	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
60	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
61	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
62	Bacteroidetes	Bacteroidia (87%)	Bacteroidales (87%)	Rikenellaceae (87%)	Alistipes(87%)
63	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae(84%)	Lachnospiracea_incertae_sedis (69%)
64	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium XlVa(68%)
65	Bacteroidetes	Bacteroidia(96%)	Bacteroidales (96%)	Rikenellaceae(94%)	Alistipes(94%)
66	Proteobacter (89%)	Alphaproteobacter (89%)	Rhizobiales (89%)	Hyphomicrobiaceae (89%)	Gemmiger (89%)
67	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia
68	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Dorea(96%)
69	Proteobacter (74%)	Alphaproteobactea (74%)	Rhizobiales (74%)	Hyphomicrobiaceae (74%)	Gemmiger (74%)
70	Actinobacteria	Actinobacteria	Actinobacteridae	Bifidobacteriales	Bifidobacteriaceae
71	Firmicutes (98%)	Erysipelotrichia (98%)	Erysipelotrichals (98%)	Erysipelotrichacea (98%)	Clostridium XVIII(88%)
72	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Clostridium IV
73	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
74	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospiracea_incertae_sedis(55%)
75	Proteobacteria (86%)	Alphaproteobacterei a (86%)	Rhizobiales (86%)	Hyphomicrobiaceae (86%)	Gemmiger(86%)
76	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
77	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium XlVa(86%)
78	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
79	Firmicutes	Negativicutes (93%)	Selenomonadals (93%)	Acidaminococcaceae (81%)	Acidaminococcus (81%)
80	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae(92%)	Roseburia(66%)
81	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae(99%)	Bacteroides(99%)
82	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium XlVa(84%)
83	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia
84	Firmicutes (77%)	Clostridia (76%)	Clostridiales (76%)	Lachnospiraceae(29%)	Sporobacterium(5%)
85	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
86	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium XlVa(93%)
87	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium
88	Firmicutes (97%)	Erysipelotrichia (97%)	Erysipelotrichals (97%)	Erysipelotrichaceae (97%)	Erysipelotrichaceae_incertae_ sedis (93%)
89	Synergistetes (34%)	Synergistia(34%)	Synergistales (34%)	Synergistaceae(34%)	Pyramidobacter(27%)
90	Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Eubacterium(63%)
91	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
92	Actinobacteria	Actinobacteria	Actinobacteridae	Bifidobacteriales	Bifidobacteriaceae
93	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburi(76%)
94	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium XlVa(74%)
95	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes
96	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia
97	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia(66%)
98	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
99	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia
	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides

Of these 100 most prominent operational taxonomic units (OTUs) there are 42 OTUs which belong to Bacteroidetes phylum. These OTUs were collectively lower in depressed patients in comparison to control people (p-value=0.01). Forty nine OTUs belong to Firmicutes phylum which their abundance had no significant difference between depressed patients and control people (p-value= 0.25). Proteobacteria and Actinobacteria also had no collective correlation with depression state; P-value=0.37 and p-value=0.73 respectively. (Figure3g)

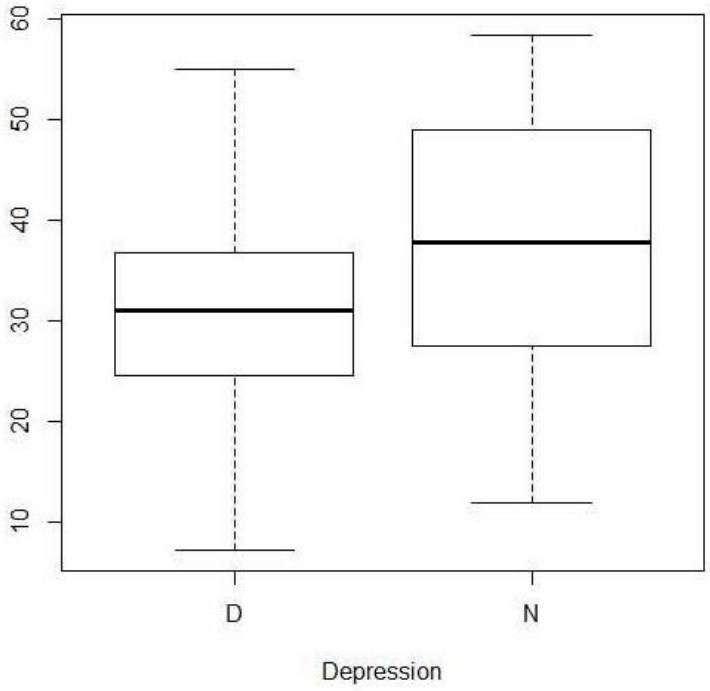


Figure3g- Bacteroidetes phylum related OTUs have collectively lower abundance in depressed patients in relation to control people (p-value=0.01). Kruskal-Wallis test was used in this statistical analysis.

3.2.3 Correlation between OTUs and Depression

Correlation between operational taxonomic units which are representatives of the 16S rRNA gene belonging to the gut microbiota and depression state is analyzed and represented using Kruskal-Wallis statistic test. (table3i). Operational taxonomic units 7, 22, 47, 57, 58, 60, 63, 65, 72, 75, 81, 84, 91, and 96 belonging to Alistipes, Bacteroides, Clostridim IV,

Flavoinfractor, Bacteroides, Prevotella, Lachnospiracea, Gemmiger, Sporobacterium, Blautia were significantly higher between depressed patients. Among these bacteria, Roseburia and Sporobacterium had the lowest p-value and the strongest positive correlation with depression. Operational taxonomic units 5, 14, 34, 46, 54, 59, 97 and 99 belonging to Roseburia, Barnesiella, Bacteroides and Alistipes are significantly lower between depressed patients than normal controls. Of these phylotypes, Roseburia had the lowest p-value and strongest negative correlation with depression.

Table3i- Correlation between Depression and OTUs calculated by Kruskal-Wallis test. In the table below N means normal and D means depressed.

OTUs	**p-value**
1	0.28
2	0.99
3	0.54
4	0.43
5	0.005* D<N
6	0.72
7	0.05*D>N
8	0.52
9	0.85
10	0.64
11	0.16
12	0.16
13	0.41
14	0.02*D<N
15	0.51
16	0.36
17	0.17
18	0.47
19	0.61
20	0.64
21	0.73
22	0.01*D>N
23	0.29
24	0.73
25	0.06
26	0.24
27	0.06
28	0.34
29	0.32
30	0.13
31	0.78
32	0.53
33	0.16
34	0.05*D<N
35	0.25

36	0.44
37	0.49
38	0.08
39	0.35
40	0.12
41	0.69
42	0.08
43	0.26
44	0.26
45	0.85
46	0.01*D<N
47	0.01*D>N
48	0.42
49	0.10
50	0.31
51	0.59
52	0.13
53	0.95
54	0.008*D<N
55	0.22
56	0.25
57	0.02*D>N
58	0.007*D>N
59	0.02*D<N
60	0.001*D>N
61	0.14
62	0.15
63	0.01*D>N
64	0.45
65	0.04*D>N
66	0.86
67	0.94
68	0.07
69	0.87
70	0.78
71	0.93
72	0.02*D>N
73	0.23
74	0.49
75	0.02*D>N
76	0.81
77	0.70
78	0.06
79	0.16
80	0.80
81	0.01*D>N
82	0.19
83	0.10
84	0.009*D>N
85	0.25
86	0.93

87	0.16
88	0.88
89	0.37
90	0.32
91	0.02*D>N
92	0.55
93	0.29
94	0.64
95	0.44
96	0.01*D>N
97	0002*D<N
98	0.87
99	0.002*D<N
100	0.17

3.2.4 Correlation between OTUs and Cortisol Levels

Correlation between operational taxonomic units and cortisol levels was analyzed using Spearman statistical test (table3j). Many of the operational taxonomic units (OTUs) had a significant correlation with one, two or even three cortisol measurements. Around two thirds of these correlations were positive correlations while the rest were negative correlation. Collectively 35 OTUs had significant correlation with cortisol levels in the morning (6 negative and 28 positive), 24 OTUs had significant correlation with cortisol levels in the mid-day (9 negative and 15 positive), and 30 OTUs had significant correlation with evening cortisol levels (12 negative and 18 positive). OTUs 6, 22, 58, 60, 76 and 81 had for example positive correlation with all three cortisol measurements. This correlation was most significant between OUT 81 and cortisol measurement 3 (ρ=+0.44). This OUT belongs to Bacteroidetes phylum and Bacteroides genus. Bacteroides had also positive correlation with cortisol measurements in OTUs 4, 6, 22, 28, 31, 32, 33, 38, 46, 52, 53, 58, 61, 76 and 78. OTUs 25, 83, 86 and 99 had significant negative correlations with all three cortisol measurements. The strongest negative correlation was between OUT 30 and cortisol measurement 3 (ρ=-0.38). This OUT belongs to Firmicutes phylum and Faecalibacterium genus. Faecalibacterium had also negative correlation with cortisol measurements in OTUs 3, 10, 12 and 50. Most of OTUs which have significantly negative correlation with cortisole measurements belong to the Firmicutes phylum, while the majority of OTUs which had significantly positive correlation with cortisol measurements belong to Bacteroidetes phylum. In addition to correlation between each cortisol level and OTus we also analyzed correlation between the difference between cortisol levels in the morning which are usually the highest and cortisol levels in the evening which are usually the lowest, and their corresponding OTU (table 3j).

Table3j- Correlation between cortisol levels and OTUs calculated by Spearman test. In this table each cell shows the p-value of Spearman correlation between that OTU and cortisol level and ρ is representative of positive or negative correlation between two variables. The last column shows the p-value in the Spearman correlation analysis between the difference between cortisol levels in the morning and evening and the corresponding OTU.

OTUs	Cortisol1	Cortisol2	Cortisol3	Difference (Cor1-Cor3)
1	0.73	0.31	0.87	0.65
2	0.17	0.26	0.40	0.23
3	0.01 *ρ=-0.23	0.50	0.06	0.08
4	0.02* ρ=+0.20	0.58	0.21	0.46
5	0.49	0.46	0.41	0.45
6	0.01* ρ=+0.22	0.007* ρ=+0.25	0.01* ρ=+0.23	0.13
7	0.007* ρ=+0.25	0.03* ρ=+0.19	0.63	0.0001* ρ=+0.34
8	0.46	0.60	0.78	0.08
9	0.10	0.73	0.58	0.007* ρ=+0.25
10	0.89	0.008* ρ=-0.24	0.05* ρ=-0.18	0.09
11	0.001* ρ=+0.29	0.48	0.08	0.02* ρ=+0.21
12	0.65	0.001* ρ=-0.29	0.001* ρ=-0.29	0.49
13	0.70	0.29	0.0*2 ρ=+0.20	0.34
14	0.0006* ρ=-0.31	0.01	0.002* ρ=-0.28	0.03* ρ=-0.19
15	0.001* ρ=+0.30	0.94	0.82	0.003* ρ=+0.26
16	0.009* ρ=-0.24	0.30	0.32	0.0003* ρ=+0.32
17	0.65	0.01* ρ=-0.22	0.64	0.25
18	0.31	0.80	0.33	0.45
19	0.41	0.73	0.26	0.46
20	0.93	0.73	0.23	0.95
21	0.15	0.61	0.06	0.008* ρ=+0.24
22	7.517e-09* ρ=+0.51	0.04* ρ=+0.18	2.849e-06* ρ=+0.42	0.005* ρ=+0.25
23	0.05* ρ=+0.18	0.21	0.30	0.0003* ρ=+0.32
24	0.11	0.65	0.65	0.0009* ρ=+0.30
25	0.03* ρ=-0.20	0.02* ρ=-0.21	0.0007* ρ=-0.31	0.86
26	0.33	0.49	0.07	0.16
27	0.72	0.20	0.08	0.14
28	2.941e-06* ρ=+0.42	0.66	0.07	$3.127e^{-07}$* ρ=+0.45
29	0.32	0.07	0.64	0.35
30	0.004* ρ=-0.26	0.96	3.184e-05* ρ=-0.38	0.56
31	0.36	0.41	0.05* ρ=+0.17	0.12
32	0.03* ρ=-0.20	0.58	0.65	0.0003* ρ=-0.32
33	0.05* ρ=+0.17	0.06	0.05* ρ=+0.18	0.01* ρ=+0.22
34	0.80	0.26	0.75	0.39
35	0.52	0.48	0.41	0.85
36	0.75	0.06	0.44	0.64
37	0.32	0.76	0.31	0.07
38	0.04* ρ=+0.19	0.7	0.22	0.04* ρ=-0.19
39	0.02* ρ=+0.20	0.82	0.49	0.0007 ρ=+0.31
40	0.14	0.47	0.19	0.10

41	0.03* ρ=+0.20	0.85	0.004* ρ=+0.26	0.78
42	0.09	0.92	0.0005* ρ=-0.32	0.77
43	0.81	0.94	0.79	0.21
44	0.51	0.57	0.08	0.32
45	0.41	0.92	0.34	0.19
46	0.26	0.80	0.05* ρ=+0.18	0.75
47	0.06	0.02* ρ=+0.21	0.62	0.14
48	0.0001* ρ=-0.35	0.19	0.26	0.001* ρ=+0.29
49	0.001* ρ=+0.29	0.03* ρ=+0.19	0.07	0.0001* ρ=+0.35
50	0.06	0.01* ρ=-0.22	0.99	0.02* ρ=+0.21
51	0.04* ρ=-0.18	0.82	0.89	0.01* ρ=+0.23
52	0.34	0.34	0.04* ρ=+0.18	0.30
53	0.95	0.11	0.03* ρ=+0.20	0.46
54	0.49	0.34	0.34	0.58
55	0.13	0.68	0.13	0.14
56	0.35	0.94	0.15	0.50
57	0.59	0.40	0.68	0.13
58	2.603e-06* ρ=+0.42	0.02* ρ=+0.20	0.003* ρ=+0.28	0.009* ρ=+0.24
59	0.27	0.50	0.99	0.33
60	0.01* ρ=+0.22	0.05* ρ=+0.17	0.02* ρ=+0.21	0.18
61	0.30	0.62	0.004* ρ=+0.26	0.84
62	0.01* ρ=+0.22	0.03* ρ=+0.20	0.07	0.009* ρ=+0.24
63	0.005* ρ=+0.25	0.05* ρ=+0.18	0.84	$5.342e^{-05}$* ρ=+0.37
64	0.72	0.77	0.19	0.70
65	0.86	0.005* ρ=+0.26	0.77	0.51
66	0.27	0.50	0.74	0.004* ρ=+0.26
67	0.73	0.91	0.02* ρ=-0.20	0.60
68	0.68	0.35	0.16	0.95
69	0.14	0.67	0.63	0.24
70	0.42	0.20	0.64	0.60
71	0.21	0.70	0.19	0.89
72	0.72	0.89	0.34	0.03* ρ=+0.20
73	0.63	0.87	0.39	0.42
74	0.11	0.22	0.58	0.07
75	0.83	0.97	0.02* ρ=-0.21	0.02* ρ=+0.20
76	0.004* ρ=+0.26	0.01* ρ=+0.22	0.004* ρ=+0.26	0.07
77	0.25	0.28	0.18	0.40
78	0.01* ρ=+0.22	0.28	0.69	0.02* ρ=+0.21
79	0.02* ρ=+0.20	0.41	0.02* ρ=+0.21	0.43
80	0.05* ρ=+0.17	0.46	0.18	0.05* ρ=+0.17
81	0.0006* ρ=+0.31	0.001* ρ=+0.29	6.042e-07* ρ=+0.44	0.70
82	0.14	0.05* ρ=+0.17	0.43	0.30
83	0.0008* ρ=-0.30	0.03* ρ=-0.19	0.003* ρ=-0.27	0.07
84	0.02*	0.03* ρ=+0.19	0.77	0.01* ρ=+0.23
85	0.06	0.06	0.74	0.002* ρ=+0.28
86	0.05* ρ=-0.17	0.05* ρ=-0.17	0.03* ρ=-0.20	0.17
87	0.61	0.09	0.39	0.68
88	0.10	0.77	0.01* ρ=-0.22	0.42
89	0.26	0.29	0.10	0.21
90	0.22	0.009* ρ=-0.24	0.98	0.01* ρ=+0.23

91	0.64	0.84	0.69	0.98
92	0.88	0.47	0.79	0.62
93	0.21	0.05* ρ=+0.18	0.07	0.95
94	0.16	0.07	0.05* ρ=-0.17	0.35
95	0.01* ρ=-0.23	0.31	0.01* ρ=-0.22	0.57
96	0.0005* ρ=+0.31	0.94	0.05* ρ=+0.18	0.01* ρ=+0.23
97	0.44	0.67	0.13	0.44
98	0.56	0.09	0.08	0.15
99	4.419e-05* ρ=-0.37	0.001* ρ=-0.30	9.747e-05* ρ=-0.35	0.04* ρ=-0.18
100	0.72	0.19	0.26	0.63

4. Discussion

4.1 Correlation between gut microbiota and depression according to mixed data sequencing

In the mixed data sequencing, one of the bacterial components which was independently significant related to depression was component 7 (positive correlation) which is assigned to "Papillibacter" at genus level, a strictly anaerobic, gram-positive, catalase-negative, rod-shaped, motile, non-sporulating, mesophilic bacterium with rounded ends. However, due to the low confidence level (18%) generated by RPD classifier tool, the taxonomic assignment cannot be fully guaranteed.

As well as direct correlation between each component and the depression state, which revealed a significant correlation between component 7 and depression, combined correlation analysis between each two components belonging to the Firmicutes phylum and the depression state was performed by "Spearmann" statistic method. Some bacterial components may in combination with other components have a significant correlation with depression, especially when these two components belong to the same phylum. In our experiment, component1 along with component 11 and component 20 had significant correlation with depression state (negative correlation). This was also correct for component 8 which had a significant correlation with depression in combination with component 7 and component 19 (positive correlation).

In the present study, combination of component 1 and component 11 were significantly lower between depressed patients in relation to non-depressed people. Both components belong to Firmicutes phylum, Clostridia class, but component 1 belongs to Papillibacter genus while component 11 to Faecalibacterium genus. In many other studies, Faecalibacterium has been found to have an immunomodulatory effect on intestinal inflammatory diseases. Harry Sokol and his colleagues demonstrated that a decrease in the abundance and biodiversity of intestinal bacteria inside the dominant phylum Firmicutes was striking in Crohn disease (CD) patients (Sokol, Pigneur et al. 2008). They established that a reduction of a major member of Firmicutes, *Faecalibacterium*, was associated with a

higher risk of postoperative recurrence of ileal CD. According to them, a lower proportion of *Faecalibacterium* on resected ileal Crohn mucosa was responsible for endoscopic recurrence at 6 months. They analyzed the anti-inflammatory effects of *Faecalibacterium* in both *in vitro* and *in vivo* colitis in mice. (Sokol, Pigneur et al. 2008). In another study Sakol and his colleagues, analyzed fecal samples from 22 active Crohn's disease (A-CD) patients, 10 CD patients in remission (R-CD), 13 active ulcerative colitis (A-UC) patients, 4 UC patients in remission (R-UC), 8 infectious colitis (IC) patients, and 27 HS by quantitative real-time polymerase chain reaction (PCR) targeting the 16S rRNA gene (Sokol, Seksik et al. 2009). They found out that bacteria of the phylum Firmicutes (*Clostridium leptum* and *Clostridium coccoides* groups) were less represented in A-IBD patients and IC, compared to healthy people. Also *Faecalibacterium prausnitzii* species (a major representative of the *C. leptum* group) had lower counts in A-IBD and IC patients in comparison to HS. They concluded that the phylum Firmicutes and particularly the species *F. prausnitzii*, were underrepresented in A-IBD patients as well as in IC patients. Since these bacteria could be crucial to gut homeostasis, lower counts of *F. prausnitzii* were firmly associated with a reduced protection of the gut mucosa. (Sokol, Seksik et al. 2009).

In another randomized, double-blind trial, fecal samples were collected from healthy volunteers, and from patients with active CD, ulcerative colitis (UC) and irritable bowel syndrome before and after treatment. The level of *F. prausnitzii* DNA in fecal suspensions was then determined by PCR. The total levels of *F. prausnitzii* in fecal samples from CD patients at presentation were lower than those in the other groups both before and after the treatment. (Jia, Whitehead et al. 2010). All these studies have proved the definitive role of *Faecalibacterium* in dampening inflammatory reaction and they are in accordance with our results.

Another study showed the role of stressors in the composition of gut microbiota in mice. These mice had been exposed to a social stressor called social disruption (SDR), which enhances circulating cytokines and primes the innate immune system for increased reactivity. Stressor exposure significantly altered the community structure of the microbiota. Most strikingly, stressor exposure decreased the relative abundance of bacteria in the genus *Bacteroides*, while increasing the relative abundance of bacteria in the genus *Clostridium*. (Bailey, Dowd et al. 2011). This finding is in favor of the significantly increased levels of

Clostridium in component 7 and combination of components 8 and 7 and 19 between depressed patients in relation to non-depressed people.

4.2 Illumina Sequencing

In the second part of the study, Illumina sequencing, as well as more deeply sequencing and classification of the bacterial taxa present in the gut, their eventual correlation with depression and cortisol levels were studied. In this regard operational taxonomic units (OTUs) were used as the taxonomic index. Deep sequencing allows us to sequence and classify bacteria which are less than 1% of the total bacteria in the gut. We characterized the 100 most dominant OTUs in this regard. Each of this OTUS belongs to one genus while one genus may have more than one OUT. In an study undertaken by Tap and colleagues to analyze the composition of the distal gut microbiota, they generated 10,456 16S rRNA gene sequences from the 17 human fecal DNA samples and analyzed them to determine which sequence were shared and which were unique. They concluded that on average each individual contained 259 operation taxonomic units, although the range was large (159-383) and in total 3180 OTUs were identified by them. Approximately, 79% of these OTUs were only found in one sample and 21% were found at least twice; while no OTUs was found in all 17 distal gut. They showed that just 66 OTUs were present in more than 50% of the samples and proposed these may in some way constitute a "core microbiota" (Tap et al., 2009). Although in this study we characterized just the 100 most dominant OTUs, but none of these OTUs were present in all the samples. The most dominant OTU which belonged to Bacteroidetes phylum and Bacteroides genus were present in about 97% of samples and the least dominant OUT was present in around 75% of samples in our study. Bacteroidetes and Firmicutes constitute about 90% of the OTUs with relatively the same number of OTUs of each phylum; while in our previous mixed sequencing data, Firmicutes was to great extent the most dominant phylum constituting around 65% of the bacterial components followed by Bacteroidetes with 15% of the bacterial components. It is also noteworthy that in our classification both in mixed data sequencing and in high resolution deep sequencing, the lowest degree in taxonomic tree is genus, while this genus further has many species and subspecies which may in turn have different characteristics and properties.

The collective abundance of those OTUs which belong to Bacteroidetes were significantly lower in depressed patients in comparison to control people (p-value=0.01), while there were no significant difference between Firmicutes, Proteobacteria and Actinobacteria phyla in depressed and control people. Firmicutes/Bacteroidetes proportion has been found to be altered in some other physiological and pathological states. Ratio of Firmicutes/Bacteroidetes has been demonstrated to be higher among obese mice and people in comparison to lean ones (Ley, Turnbaugh et al. 2006). In another study this proportion was shown to be evolving during different life stages. For infants, adults and elderly individuals this ratio was measured at 0.4, 10.9 and 0.6, respectively (Mariat et al., 2009).
Some scientists have assigned gut microbiota to one of three enterotypes, driven by *Bacteroides*, *Prevotella* and *Ruminococcus* species (Arumugam et al. 2011). All of these enterotypes were present in our Illumina sequencing, both in depressed and healthy people, and some OTUs belong to one of these enterotypes had positive correlation with cortisol levels.(OTUs 4,6,22, 28,53..belonging to Bacteroides; OTU 60 to Prevotella and OTU 11 belonging to Ruminococcus)

4.3 Correlation between OTUs (gut microbiota) and Depression According to Illumina Deep Sequencing

One gut microbial genus which had significant negative correlation with depression in our analysis was Roseburia from Lachnospiraceae family. In a study on the role of gut microbiota, Claesson and colleagues showed that Roseburia along with Coprococcus had negative correlation with depression and institutionalization due to it (Claesson, Jeffery et al. 2012). In another study carried out by Santos and colleagues it was shown that colitis reduction was associated with higher numbers of mucosal Faecalibacterium, Roseburia, Dialister and Lactobacillus (Santos, Castro et al. 2011).

Operational taxonomic units 7, 22, 47, 57, 58, 60, 63, 65, 72, 75, 81, 84, 91, and 96 belonging to Alistipes, Bacteroides, Flavoinfractor, Bacteroides, Prevotella, Lachnospiracea, Clostridim IV, Gemmiger, Sporobacterium, Blautia have been found to have significantly higher abundance between depressed patients. Among these bacteria, Roseburia and Sporobacterium had the *lowest p-value* and the strongest *positive* correlation with depression. Operational taxonomic units 5, 14, 34, 46, 54, 59, 97 and 99 belonging to

Roseburia, Barnesiella, Bacteroides and Alistipes were found to have significantly lower abundance between depressed patients than normal controls. Of these phylotypes, Roseburia had the lowest p-value and strongest negative correlation with depression. These OTUs belong to Firmicutes, Bacteroidetes and Proteobacteria phyla. It is remarkable that sometimes one OTU belonging to one specific genus has significant positive correlation with depression, while another OTU belonging to the same genus has significantly negative correlation with depression, as this phenomenon is seen in our experiment. It can explain in many ways. First the confidence rate is not always 100%. Second, each Out refers to a specific part of the 16S rRNA gene not entire gene and if organisms resemble each other more than 97% in this relatively short DNA stretch are classified as one operational taxonomic unit and one bacterial species and Thirdly; each genus comprises many species and subspecies which may have different and occasionally contradictory properties. So when one OUT is found to have a significant positive or negative correlation to another variable such as depression it can just be inferred that there is a correlation between this OUT and this variable and the type of this relationship is not necessarily known.

4.4 HPA analysis and its correlation with depression

HPA function can be investigated in three states. These different secretion states include: baseline/non-stressed HPA secretion state (un-stimulated HPA state); stress-reactive HPA secretion state (stimulated HPA activity due to exposure to stressor) and stress-recovery HPA secretion state (post-stimulated HPA activity after stressor exposure)(Bruke et al. 2005). In the present study, the cortisol values show the basal free cortisol secreted into the salivary glands which serves as biological marker for non-stressed, non-stimulated HPA secretion state in both groups during measurement times of the day.

The result shows a decreasing pattern of cortisol secretion with the highest salivary cortisol concentration in the morning and lowest in the evening in both depressed and control groups. This finding is in accordance with previous findings of diurnal/circadian rhythm of cortisol secretion in human where highest levels measured in the morning and lowest in the evening (Burke et al. 2005; deWeerth et al. 2003). Alteration of this rhythmic diurnal pattern of cortisol secretion has previously been documented in depressed patients in

response to stress. Specifically, analysis of stimulated HPA function in the study revealed that depressed individuals represented a relatively flat and unresponsive pattern of cortisol secretion while non-depressed individuals demonstrated a more dynamic and responsive pattern of cortisol secretion in response to stress (Bruke et al. 2005).

Over-activity and/or disruption of HPA regulatory feedback mechanisms can be potential explanatory factors for elevated salivary cortisol measurements in depressed patients. Hypercortisolemia due to alterations of HPA system have been previously observed in depression (Gillespie & Nemeroff, 2005). Basically, hyperactivity of HPA system has been documented in many critically-ill patients, where various biochemical signals from immune, endocrine and nervous system enhance a continual stimulation of HPA system and subsequent prolonged cortisol secretion (Venkataraman, et al., 2007). Besides, disruption of the HPA regulatory feedback mechanism, which controls cortisol secretion, has been observed in depressed patients (Lopez-Duran, et al., 2009). Nonetheless, in order to establish hypercortisolemic diagnosis in depression, further tests like "dexamethasone suppression test" (DST) and "CRH infusion test" are sometimes required (Dinan, 2001).

As well as its usefulness in evaluation of body cortisol levels; "dexamethasone suppression test" (DST) is also considered as a classic test for assessing the integrity of HPA negative feedback mechanism. DST particularly evaluates the integrity of HPA negative feedback mechanism. Prescription of dexamethasone, a synthetic analogue of cortisol, in healthy individuals inaugurate a negative feedback by acting at the level of the anterior pituitary thereby suppressing the secretion of adrenocorticotropic hormone (ACTH) causing decrease in synthesis and release of cortisol from adrenal cortex. Failure of dexamethasone administration in cortisol secretion reduction is assumed as an indicative of dysfunction in the feedback regulatory mechanism. Non-suppression (refractory) of cortisol secretion being challenged with dexamethasone has been found in depressed patients (Gillespie & Nemeroff, 2005). Some studies claimed that gut microbiota play an exciting role in development of HPA axis and its response to stress (Sudo, 2006; Sudo et al., 2004). Although all three cortisol measurements had higher average levels between depressed patients in relation to control people, but only cortisol values being taken in the mid-day were significantly higher in depressed patients and none of the other two measurements

during the day-time were significantly higher. This can to some extent be explained by disturbance in the normal rhythmic secretion pattern of cortisol in depressed patients. Many scientists are interested in finding inflammatory features of the depression state. If cortisol measurements were carried out during several days, it would have been more applicable and informative in HPA axis analysis.

4.5 Correlation between Gut microbiota (bacterial components) and HPA function according to mixed data sequencing

Component 14 had a significant positive correlation with morning and mid-day cortisol levels. This component has been attributed to Chlamydiae phylum and Simkania genus with relatively low certainty (11%). *Simkania* is an intracellular bacterium that was first discovered when rapidly moving particles were observed in cytoplasmic inclusions in laboratory stocks of Vero cells (Friedman, Dvoskin et al. 2003). Transmission electron microscopy of thin sections revealed small, dense forms and larger, homogenous forms, similar in appearance to chlamydial elementary and reticulate bodies. This newly described microorganism, an obligate intracellular, penicillin-resistant microorganism most closely related to the chlamydiae, has been associated with adult community-acquired pneumonia (Kahane, Greenberg et al. 1998). Although this organism tends to inhabit respiratory tract, it has been showed that they can be present in the gastrointestinal tract. Probably concurrent presence of them in both respiratory tract and gastrointestinal tract is explanatory to high cortisol levels among depressed patients who had this microorganism in their intestine. In both two mentioned studies, this organism was associated with high levels of cytokines and serum markers especially IgG antibody. Considering the high potential of spreading from one patient to another, if we find this organism in one patient who has been hospitalized due to a disease such as depression, many other patients who are also hospitalized in the same center demonstrate serological evidence of infection with the same organism (Fasoli, Paldanius et al. 2008). Drinking water supplies are known sources for this organism. Since members of this novel family are commonly observed in free-living amoebae (FLA) as host cells, if a patient is infected with amoebae there is the possibility of showing serological as well as fecal evidence of contamination with *Simkania* (Kahane, Greenberg et al. 2007).

Since the sequencing and taxonomic confidence is quite low, we should cautious with any conclusion regarding this genus.

4.6 Illumina Deep Sequencing; Correlation between OTUs and Cortisol levels

In the results from Illumina sequencing, many of the OTUs had significant correlation either positively or negatively to cortisol levels. In the analysis of those OTUs which had a significant *negative* correlation with cortisol levels, including morning, mid-day or evening cortisol levels, a striking number of them belonged to "Faecalibacterium" genus. As it was discussed before, Faecalibacterium had a well-known effect in reducing inflammatory responses in the GI system (Jia, Whitehead et al. 2010). "Pseudobutyrivibrio" had also a negative correlation with cortisol levels in our Illumina analysis. This phylotype has also been found to possess a potent xylanolytic enzyme system consisting of at least 7 different xylan hydrolases with molar mass 27–145 kDa. In addition to producing butyrate and lactate on different saccharides; these bacteria secrete bacteriocin-like inhibitory substances of $Mz5^T$ which is active against some strains of pathogenic bacteria such as *Salmonella* and *E. coli*. By attachment to the Caco-2 cells, they ensure their successful association with gut epithelial cells. These properties have made some scientists to utilize these bacteria in probiotic products (Čepeljnik, Zorec et al. 2003). Another genus which had a significant *negative* correlation with cortisol measurements in our Illumina analysis was "Eubacterium". This result is in accordance with the results from another study undertaken by Kanauchi and colleagues. They showed that Bifidiobacterium and Eubacterium could increase the level of production of butyrate, which is the primary fuel for mucosa and may play a regulatory role in STAT3 activation and reduction in intestinal inflammation (Kanauchi, Serizawa et al. 2003)

On the other hand, many of those OTUs which had positive correlation with cortisol levels (at all three measurements levels), belonged to "Bacteroides" genus. Indigenous *Bacteroides* strains have been firmly proved to play an undeniable role in the occurrence and exacerbation of ulcerative colitis (UC) as well as some other inflammatory diseases. This pro-inflammatory effect of Bacteroides comes in effect by secretion of some immuno-stimulatory mediators and can be modulated in great extent by the presence of Bifidiobacterium strains (Setoyama, Imaoka et al. 2003). Hollander et al. have demonstated that the intestinal flora of patients with active Crohn's disease and ulcerative colitis is

remarkebly different than that seen in patients with quiescent disease, ulcerative colitis, or normal controls. In the case of active IBD, aerobic bacterial concentrations were elevated, among them Bacteroides fragilis and Bacteroides vulgatus were more prominent (Hollander et al. 1995).

In another experiment on SAMPI/Yit mice, Matsomuto concluded that introduction of gut microbiota to gnobiotic mice could initiate gut inflammation and "Bacteroides" determined the severity of the inflammation; while surprisingly Bacteroides itself had no ability to initiate these inflammatoric reactions (Matsomuto, 2004).

In another study on the effects of introduction of bacteria in the HLA-B27 transgenic rats, it was shown that after colonization of germfree transgenic rats with groups of five to eight bacteria that were either facultative or strictly anaerobic, only transgenic rats colonized with defined bacterial cocktails which contained Bacteroides spp. developed colitis and gastritis (Rath et al. 1996).

"Lachnospiraceae incertae sedis" (OTU63) was shown in our Illumina analysis to have a strong positive correlation with cortisol measurements. This finding was also in common with other studies which showed species belonging to *Lachnospiraceae incertae sedis* are known to express inflammatory flagellin proteins with an important contributing effect in exacerbation of autoimmune diseases (Duck et al. 2007).

Another genus which in our Illumina analysis had a significant *negative* correlation with cortisol levels and thereby inflammation was Oscillibacterium from Firmicutes phylum. Comparable results were reported from some scientists working on diabetes in mice. They showed that development of diabetes and increases inflammatory markers in the serum was accompanied with 28% reduction in Lachnospiraceae and 54% reduction of Oscillibacter genus in diabetic mice versus non-diabetic ones (Springer Healthcare News, 2012).

Many observations have been made regarding the relatively flatter pattern of diurnal salivary cortisol pattern in depressed patients than in non-depressed people (Bruke et al. 2005); but these findings do not necessarily mean that higher differences between morning cortisol levels and evening cortisol levels can be interpreted as an indicator of healthy state, because if the cortisol level in the morning is excessively high or cortisol level in the evening is excessively low, it can result in the high difference between these two levels, while the absolute levels of cortisol are aberrant. In our Illumina study, most of those OTUs

which had higher fluctuations in cortisol levels, had higher morning cortisol levels. Alistipes, Bacterides, Gemmiger, Oscillibacterium, Ruminococcus, Lachnospiraceae and Blautia were among those genera with the highest significant cortisol oscillation rate during day time, while almost all of them had significantly higher morning cortisol levels. Of those OTUs which had significantly higher cortisol difference throughout the daytime without any significant increase in the morning cortisol levels or even decrease in one or more cortisol measurements, were Blautia, Faecalobacterium, Clostridium IV, Eubacterium (Firmicutes) and Gemmiger (Proteobacteria).

Apart from this, we should keep in the mind that there are many other factors which can distort usual salivary cortisol secretion rhythm, such as social stressors like unemployment, one's daily activity and their locations (Ockenfels, Porter et al. 1995); pregnancy and delivery (Allolio, Hoffmann et al. 1990) and so on.

In a study to examine the association between suprachiasmatic nucleus (SCN) control of both awakening and cortisol secretory activity, it was shown that there was a significant correlation between the time of awakening and salivary cortisol secretion pattern and cortisol diurnal profile was negatively correlated with awakening time; meaning that those subjects who awoke earliest had higher levels of cortisol over the 45 minutes following awakening as well as throughout the rest of the day (Edwards, Evans et al. 2001).

In a study aimed at finding correlation btween cortisol secretion pattern and daily experience, 109 healthy employed and unemployed community residents were examined for two days. Of these 109 subjects, Fifty-six (51%) individuals showed typical declines in cortisol during both days, 19 (17%) showed no significant diurnal pattern on both days, and 34 (31%) showed different diurnal patterns on the 2 days. They concluded that although average daily level of cortisol was related to a number of psychosocial and psychiatric factors (e.g. stress and depression), pattern of diurnal cycle was not related to any demographic or psychosocial measures (Smyth, Ockenfels et al. 1997).

In another study performed on severely burnt patients, it was displayed that cortisol levels were higher in burn patients (20.4 ± 1.0 mg/dL) than in control subjects (9.8 ± 1.6mg/dL, $P < .0001$) and the normal circadian decreasing pattern of circulating cortisol levels was markedly blunted in burn patients (Hobson, Havel et al. 2004).

4.7 Integration and Comparison of Mixed Data Sequencing and Illumina Deep Sequencing

In the mixed data sequencing we found 20 bacterial components as the representatives of all bacteria presented in the samples. The confidence rate in this classification was quite low and any conclusion based on this classification must be made with caution. In contrast to this, in Illumina sequencing analysis and classification the confidence rate and the depth of sequencing are much higher. In mixed data sequencing none of the phyla were significantly higher or lower among depressed patients in comparison to control people, while in the Illumina sequencing the collective sum of OTUs belonging to Bacteroidetes phylum were significantly lower among depressed patients in comparison to control people and none of the other phyla had any correlation with depression state. In the first part of the study, the only bacterial component which had independently significant correlation with depression was bacterial component 7 belonging to Firmicutes phylum (77%) and Papillibacter genus (18%) and combination of bacterial components 1& 11 and 1&20 had negative correlation with depression and combination of bacterial components 7&8 and 8&19, all belonging to Firmicutes phylum had positive correlation with depression. Faecalibacterium genus in this analysis had negative correlation with depression and this finding is repeated in Illumina sequencing in which many OTUs belonging to this genus had negative correlation with depression. Bacterial component 7 belonging to Firmicutes phylum and Clostridia class was the only bacterial component with significant *positive* correlation to depression and OUT 72 belonging to Clostridium genus had positive correlation with depression state in Illumina sequencing. Bacterial component 8 belonging to Anaerosporobacter genus (11%) had also positive correlation with depression (combined effect) and OTUs 82 belonging to this genus had positive correlation with depression in Illumina sequencing. Bacterial component 19 belonging to Ruminococcus genus had also combined positive correlation with depression in mixed data analysis and OTU11 belonging to this genus had also positive correlation with depression in Illumina sequencing. Bacterial component 20 belonging to Lachnospiraceae (18%) had combined negative correlation with depression in mixed data analysis, while OUT 63 belonging to this genus had positive correlation with depression.

Of all 20 bacterial components only bacterial component 14 belonging to Chlamydia phylum (11%) and Simkania genus (10%) had positive correlation with morning and mid-

day cortisol measurements, while this genus was not at all among the 100 most dominant OTUs in Illumina sequencing analysis.

4.8 Recommendations for further work

The relatively small number of samples makes the interpretation and conclusion of the study results to some extent mixed with uncertainty; so further experiments are recommended with even more samples so as to enhance the statistical power of the study. By repeating this experiment with a larger sample size, the reproducibility of the results can be evaluated. By measuring the values more accurately and eliminating the disturbing factors and trimming the data, we can avoid noises in the data and have more normally distributed data. During this experiment, we tried to use statistical tests such as "Spearmann" test, "Kendall's test", "Kruskall-Wallis" test, with the potential of overcoming the non-normally distribution pattern of the data. T-test analysis and one-way ANOVA which were primarily designed for normally distributed data, were not applied in statistical analyses of the data; although it was acceptable in statistical analysis of salivary cortisol measurements in depressed and control groups.

In order for more effectively phylogenetic resolution and gut bacteria specie characterization; cloning of amplified bacterial target DNA (16S rRNA gene fragment); followed by sequencing of clone libraries is suggested. Likewise, DNA micro-array containing specie-specific oligonucleotide probes can be utilized for rapid detection and identification.

Three salivary cortisol measurements (morning, mid-day and evening) in this study do not represent a detailed assessment of HPA function in both depressed and control groups. Hence, increasing the frequency of measurements of cortisol level across the 24-hour of the day (like hourly/two-hourly measurements) along with cortisol measurements during different days is recommended in order to have a more thorough perspective of HPA activity. By cortisol measurements in different days, we can avoid some disturbing factors which can alter the cortisol levels just for a limited period of time, such as common cold, menstruation and so on. In this way, we would be able to follow the patients and assess them based on the wax and wane of the depressive symptoms and any potential alterations

in HPA system concurrent with these changes. Also, besides salivary cortisol measurements; other tests like CRH infusion tests and DST tests can be conducted to increase the validity of alterations of HPA system in depression; although they might be more aggressive, time-consuming and expensive.

Although major depressive disorder is not primarily an inflammatory disease, higher levels of cortisol or other peripheral inflammatory biomarkers are seen in 30-50% of depressed patients even in the absence of medical illness. On the other hand, many inflammatory diseases are associated with greater incidence of depressive symptoms (Claes, 2004). Cortisol is just one reactor in the immune system and most of the interactions are mediated through cytokines, growth factors and cellular immune responses. The final and cumulative effect of these elements is enforced by alterations in the neurotransmitters.
Since immunological changes has been documented in depressed patients and the role of gut microbiota in immune response regulation has long been known; it will be of benefit to explore the eventual relationship between gut microbiota and immunological changes in depression making use of cytokines measurements as markers for the latter. This will contribute to answer the question of whether gut bacteria-caused inflammation and depression are linked together or not. Addressing this fundamental question will have major influence on administration of immune modulating medications and targeted handling of the gut microbiota composition in depression treatment.

By entering other explanatory variables such as age, sex, race, medical history with any background disease especially inflammatory diseases, drug history, familial history and so on we would be able to widen our perspective and find out any potential risk factor for developing and exacerbating or alleviating of the depression. Any drug intake especially antibiotics prior to sampling can lead to false and unreliable conclusions.

Since we analyzed the fecal microbiota, our observations may not reflect the bacteria directly interacting with the immune system in the intestine (mucosally associated bacteria MAB) and potentially affect the cortisol levels and depression state. While it is quite simple to collect, it is clear that faeces do not afford a robust proxy for the human gut microbiome in its entirety. In Eckburg and colleagues 2005 culture-independent analysis of the distal gut microbiota (Eckburg et al, 2005) the clearly demonstrated that while the microbiota attached

to the mucosa was similar throughout an individual's large intestine, it was considerably different to the stool sample from the same individual, but whether there is any biological significance in this difference remains to be shown, as the number of luminal bacteria are between 4-6 orders of magnitude less than the mucosally associated bacteria (Zoetendal et al, 2004).

Neither can we determine from our results if the observations were a cause or a consequence of the alterations in the HPA axis or presence or absence of depressive symptoms. Further experimental documentation is therefore required to determine the mechanistic nature of the correlations detected. What we shown, however, is that there is a significant difference in some of the fecal microbiota components between depressed and non-depressed people. Furthermore, we have also shown some deviations in the normal values of the diurnal secretion pattern of cortisol in depressed patients. We believe that with future large-scale studies, several unknown and controversial aspects of the correlation between gut microbiota and depression state can be resolved and a better understanding of the interaction between gut microbiota and immune system and depression can be obtained. Such understanding could lead to early diagnosis of depression and more drastic treatment protocols for both depression and gastrointestinal system diseases.

5. Conclusion

Firmicutes and Bacteroidetes were the most dominant phyla in the gut microbiome. Of the total twenty dominant bacterial phylogroups identified through the MCR/ALS-based mixed sequence resolution, one of them, namely component 7, belonging to Firmicutes phylum, had independently significant correlation with depression (positive correlation). Combination of components 1 &11 and components1&20 were significantly lower in depressed patients and combination of components7&8 and components8&19 were significantly higher in depressed patients in comparison to non-depressed people. All these components belonged to Firmicutes phylum. In the Illumina sequencing analysis, the 100 most dominant operational taxonomic units (OTUs) were characterized. This sequencing method had significantly higher depth in relation to our previous mixed data sequencing method. OTUs belonging to Bacteroidetes phylum were collectively lower in depressed patients than in control people. Of many OTUs which had either positive or negative correlation with depression, OTUs 5, 14, 97 and 99 belonging to Roseburia can be mentioned. Faecalibacterium was one of the genera which was shown to have a dampening effect on inflammation and had significant negative correlation with cortisol measurements. In addition, Psuedobutyrivibrio, Oscillibacter and Eubacterium had reducing effect on cortisol levels in contrast to Lachnospiraceae incertae sedis and Bacteroides which had a significant positive correlation with cortisol measurements. Cortisol values being measured in the salivary secretions in three day times were higher in depressed people in comparison to control group but only cortisol values in the mid-day were significantly higher. Component 14 had a significant correlation with cortisol levels in the morning and mid-day. This component belongs to Chlamydiae phylum and Simkania genus. These finding are strongly in favor of relationship between gut microbiota and depression, but the extent and type of this correlation need to be more investigated and elucidated.

6. References

Allolio, B., et al. (1990). "Diurnal salivary cortisol patterns during pregnancy and after delivery: relationship to plasma corticotrophin-releasing-hormone." Clinical Endocrinology **33**(2): 279-289

Almeida, O. P. and S. A. Almeida (1999). "Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV." International Journal of Geriatric Psychiatry **14**(10): 858-865.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410.

Anisman, H. and Hayley, S. (2012). "Inflammatory Factors Contribute to Depression and Its Comorbid Conditions." Sci. Signal. **5**(244): pe45-.

Arafah, B. M., Nishiyama, F. J., Tlaygeh, H., & Hejal, R. (2007). Measurement of salivary cortisol concentration in the assessment of adrenal function in critically ill subjects: a surrogate marker of the circulating free cortisol. *J Clin Endocrinol Metab, 92*(8), 2965-2971.

Arumugam, M. *et al.*(2011) Enterotypes of the human gut microbiome. *Nature* ;473, 174–180

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host- Bacterial Mutualism in the Human Intestine. *Science, 307*(5717), 1915-1920.

Bach, J. F. (2002).The effect of infections on susceptibility to autoimmune and allergic diseases. *N. Engl. J. Med.* 347, 911–920

Bailey, M. T., et al. (2011). "Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation." Brain, Behavior, and Immunity **25**(3): 397-407

Baker, G. C., Gaffar, S., Cowan, D. A., & Suharto, A. R. (2001). Bacterial community analysis of Indonesian hot springs. *FEMS Microbiology Letters, 200*(1), 103-109. doi: 10.1016/s0378-1097(01)00207-5.

Baker, G. C., Smith, J. J., & Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods, 55*(3), 541-555.

Beaton, E. A., Schmidt, L. A., Ashbaugh, A. R., Santesso, D. L., Antony, M. M., McCabe, R. E., .Schulkin, J. (2006). Low salivary cortisol levels among socially anxious young adults: Preliminary evidence from a selected and a non-selected sample. *Personality and Individual Differences, 41*(7), 1217-1228. doi: 10.1016/j.paid.2006.02.020

Biasucci, G., Rubini, M., Riboni, S., Retetangos, C., Morelli, L., and Bessi, E. (2010) Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev* 86 (Suppl. 1)**:** 13–15.

Bik, E.M., Eckburg, P.B., Gill, S.R., Nelson, K.E., Purdom,E.A., Francois, F., *et al*. (2006) Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA* 103: 732–737.

Brook Dw, B. J. S. Z. C. C. P. W. M. (2002). "DRug use and the risk of major depressive disorder, alcohol dependence, and substance use disorders." Archives of General Psychiatry **59**(11): 1039-1044.

Blaut, M., Collins, M. D., Welling, G. W., Doré, J., van Loo, J., & de Vos, W. (2002). Molecular biological methods for studying the gut microbiota: the EU human gut flora project. *British Journal of Nutrition, 87*(SupplementS2), S203-S211. doi: doi:10.1079/BJN/2002539

Burke, H. M., Davis, M. C., Otte, C., & Mohr, D. C. (2005). Depression and cortisol responses to psychological stress: A meta-analysis. *Psychoneuroendocrinology, 30*(9), 846-856. doi: 10.1016/j.psyneuen.2005.02.010

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., . . . Poulton, R. (2003). Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene. *Science, 301*(5631), 386-389.

Cepoiu, M., et al. (2008). "Recognition of Depression by Non-psychiatric Physicians—A Systematic Literature Review and Meta-analysis." *Journal of General Internal Medicine 23*(1): 25-36.

Chen C.N., Su Y., Baybayan P., Siruno A., Nagaraja R., Mazzarella R., Schlessinger D., Chen E. *(1996) Ordered shotgun sequencing of a 135 kb Xq25 YAC containing ANT2 and 4 possible genes, including three confirmed by EST matches. Nucleic Acids Res. 24:4034–4041.*

Cheryl R. Heiner, Kathryn L. Hunkapiller, Shiaw-Min Chen, John I. Glass,& Ellson Y. Chen(1998) Sequencing Multimegabase-Template DNA with BigDye Terminator Chemistry. *Genome Res. 1998. 8: 557-561*

Chissoe S.L., Marra M.A, Hillier L., Brinkman R., Wilson R.K., Waterston R.H. *(1997) Representation of cloned genomic sequences in two sequence vectors: Correlation of DNA sequence and subclone distribution. Nucleic Acids Res. 25:2960–2966.*

Claes, S. J. (2004). CRH, stress, and major depression: a psychobiological interplay. *Vitam Horm, 69*, 117-150.

Claesson, M. J., et al. (2012). "Gut microbiota composition correlates with diet and health in the elderly." Nature **488**(7410): 178-184.

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam, S. A., McGarrell, D. M., . . . Tiedje, J. M. (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res, 33*(Database issue), D294-296.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., . . . Tiedje, J. M. (2009a). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res, 37*(Database issue), 12.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., . . . Tiedje, J. M. (2009b). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research, 37*(suppl 1), D141-D145. doi: 10.1093/nar/gkn879

Cole, J. R., Wang, Q., Chai, B., & Tiedje, J. M. (2011). The Ribosomal Database Project: Sequences and Software for High-Throughput rRNA Analysis *Handbook of Molecular Microbial Ecology I* (pp. 313-324): John Wiley & Sons, Inc.

Crystal, S., et al. (2003). "Diagnosis and Treatment of Depression in the Elderly Medicare Population: Predictors, Disparities, and Trends." Journal of the American Geriatrics Society **51**(12): 1718-1728.

Cuijpers, P., van Straten, A., van Oppen, P., Andersson, G. (2008). Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *Journal of Clinical Psychiatry*.;69(11):1675–85.

Cummings, J. H., Macfarlane, G.T. & Englyst, H.N. (2001). "Prebiotic digestion and fermentation." The American Journal of Clinical Nutrition **73**(2): 415s-420s.

Čepeljnik, T., et al. (2003). "IsPseudobutyrivibrio xylanivorans strain Mz5T suitable as a probiotic? Anin Vitro study." Folia Microbiologica **48**(3): 339-345.

Dalca, A. V., & Brudno, M. (2010). Genome variation discovery with high-throughput sequencing data. *Briefings in Bioinformatics, 11*(1), 3-14. doi: 10.1093/bib/bbp058

Dale, J., Sorour, E.& Milner, G. (2008). Do psychiatrists perform appropriate physical investigations for their patients? A review of current practices in a general psychiatric inpatient and outpatient setting. *Journal of Mental Health.*;17(3):293–98.

Dan, M., Richardson, J., Miliotis, M. D., & Koornhof, H. J. (1989). Comparison of preservation media and freezing conditions for storage of specimens of faeces. *Journal of Medical Microbiology, 28*(2), 151-154.

Dantzer, R., et al. (2008). "From inflammation to sickness and depression: when the immune system subjugates the brain." Nature Reviews: Neuroscience **9**(1): 46-56.

Davies, J. (2001). In a map for human life, count the microbes, too. *Science, 291*(5512), 2316-2316

Desbonnet, L., Garrett, L., Clarke, G., Bienenstock, J., and Dinan, T.G. (2008) The probiotic Bifidobacteria infantis: an assessment of potential antidepressant properties in the rat. *J Psychiatr Res* 43: 164–174.

de Kloet, C. S., Vermetten, E., Geuze, E., Kavelaars, A., Heijnen, C. J., & Westenberg, H. G. M. (2006). Assessment of HPA-axis function in posttraumatic stress disorder: Pharmacological and non-pharmacological challenge tests, a review. *Journal of Psychiatric Research, 40*(6), 550-567. doi: 10.1016/j.jpsychires.2005.08.002

de Kloet, E. R., Joels, M., & Holsboer, F. (2005). Stress and the brain: from adaptation to disease. [10.1038/nrn1683]. *Nat Rev Neurosci, 6*(6), 463-475.

deSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology, 72*(7), 5069-5072. doi: 10.1128/aem.03006-05

de Weerth, C., Zijl, R. H., & Buitelaar, J. K. (2003). Development of cortisol circadian rhythm in infancy. *Early Hum Dev, 73*(1-2), 39-52.

Dinan, T. (2001). Novel approaches to the treatment of depression by modulating the hypothalamic – pituitary – adrenal axis. *Human Psychopharmacology: Clinical and Experimental, 16*(1), 89-93. doi: 10.1002/hup.188

Docherty, J. P. (1997). "Barriers to the diagnosis of depression in primary care." Journal of Clinical Psychiatry **58**(Suppl 1): 5-10

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 107: 11971–11975.

Dowlati, Y., et al. (2010). "A Meta-Analysis of Cytokines in Major Depression." Biological Psychiatry **67**(5): 446-457.

Duck,L.W.,Walter, M.R., Novak, J. (2007) Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis;*13:1191–201.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., *et al.* (2005) Diversity of the human intestinal microbial flora. *Science* 308: 1635–1638.

Edwards, S., et al. (2001). "Association between time of awakening and diurnal cortisol secretory activity." Psychoneuroendocrinology **26**(6): 613-622.

Espinoza, J., Goncalves, L.F., Romero, R., Nien, J.K., Stites, S., Kim, Y.M., *et al.* (2005) The prevalence and clinical significance of amniotic fluid 'sludge' in patients with preterm labor and intact membranes. *Ultrasound Obstet Gynecol* 25: 346–352.

Fasoli, L., et al. (2008). "Simkania negevensis in community-acquired pneumonia in Italian children." Scandinavian Journal of Infectious Diseases **40**(3): 269-272

Favier, C.F., de Vos, W.M., and Akkermans, A.D.L. (2003) Development of bacterial and bifidobacterial communities in feces of newborn babies. *Anaerobe* 9: 219–229.

Flicek, P., & Birney, E. (2009). Sense from sequence reads: methods for alignment and assembly. [10.1038/nmeth.1376]. *Nat Meth, 6*(11s), S6-S12.

Fournier Jc, D. R. J. H. S. D. and et al. (2010). "Antidepressant drug effects and depression severity: A patient-level meta-analysis." JAMA **303**(1): 47-53.

Frank E, P. R. F. J. R. B. and et al. (1991). "Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence." *Archives of General Psychiatry 48*(9): 851-855.

Friedman, M. G., et al. (2003). "Infections with the chlamydia-like microorganism Simkania negevensis, a possible emerging pathogen." Microbes and Infection **5**(11): 1013-1021.

Garrido, M., Rius, F., & Larrechi, M. (2008). Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Analytical and Bioanalytical Chemistry, 390*(8), 2059-2066. doi: 10.1007/s00216-008-1955-6

Gillespie, C. F. and C. B. Nemeroff (2005). "Hypercortisolemia and Depression." Psychosomatic Medicine **67**(Supplement 1): S26-S28

Gliddon, C. M., Darlington, C. L., & Smith, P. F. (2003). Activation of the hypothalamic–pituitary–adrenal axis following vestibular deafferentation in pigmented guinea pig. *Brain Research; 964*(2): 306-310. doi: 10.1016/s0006-8993(02)04086-6

Goldman, D.L., and Huffnagle, G.B. (2009) Potential contribution of fungal infection and colonization to the development of allergy. *Med Mycol* 47: 445–456.

Gong, J., & Yang, C. (2011). Advances in the methods for studying gut microbiota and their relevance to the research of dietary fiber functions. Food Research International48(2): 916-929.

Goodyer, I. M., et al. (2000). "Recent life events, cortisol, dehydroepiandrosterone and the onset of major depression in high-risk adolescents." The British Journal of Psychiatry **177**(6): 499-504.

Gorman, R., & Adley, C. C. (2004). An evaluation of five preservation techniques and conventional freezing temperatures of −20°C and −85°C for long-term preservation of Campylobacter jejuni. *Letters in Applied Microbiology, 38*(4), 306-310. doi: 10.1111/j.1472-765X.2004.01490.

Guarner, F., & Malagelada, J.-R. (2003). Gut flora in health and disease. *The Lancet, 361*(9356), 512-519. doi: 10.1016/s0140-6736(03)12489-0

Haeffel, G. J., et al. (2008). "Association Between Polymorphisms in the Dopamine Transporter Gene and Depression: Evidence for a Gene-Environment Interaction in a Sample of Juvenile Detainees." Psychological Science **19**(1): 62-69.

Hahn, T., Marquand, A.F., Ehlis, A.C., *et al.* (2010). "Integrating Neurobiological Markers of Depression". *Arch. Gen. Psychiatry 68* (4): 361–368.

Hamady, M., et al. (2008). "Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex." Nat Meth **5**(3): 235-237.

Hasin, D.S. (2005). Epidemiology of major depressive disorder: Results from the national epidemiologic survey on alcoholism and related conditions. *Archives of General Psychiatry 62*(10): 1097-1106.

Heiner, C. R., et al. (1998). "Sequencing Multimegabase-Template DNA with BigDye Terminator Chemistry." Genome Research **8**(5): 557-561.

Hillier, L. W., Marth, G. T., Quinlan, A. R., Dooling, D., Fewell, G., Barnett, D., . . . Mardis, E. R. (2008). Whole-genome sequencing and variant discovery in C. elegans. [10.1038/nmeth.1179]. *Nat Meth, 5*(2), 183-188.

Hirschfeld, R.M. (2000) History and evolution of the monoamine hypothesis of depression. *Journal of Clinical Psychiatry*.;61 (Suppl 6):4–6.

Hobson, K. G., et al. (2004). "Circulating Leptin and Cortisol After Burn Injury: Loss of Diurnal Pattern." Journal of Burn Care & Research **25**(6): 491-499.

Holander, G.A., Simpson, S.J., Misoguchi. E., (1995). Severe colitis in mice with thymic selection. *Immunity; 3*:27-38.

Holt, R. A., & Jones, S. J. M. (2008). The new paradigm of flow cell sequencing. *Genome Research, 18*(6), 839-846. doi: 10.1101/gr.073262.107

Holtzheimer, P. E. & Nemeroff, C. B. (2006). Future prospects in depression research. *Dialogues Clin Neurosci, 8*(2), 175-189.

Hooper, L. V., & Gordon, J. I. (2001). Commensal Host-Bacterial Relationships in the Gut. *Science, 292*(5519), 1115-1118. doi: 10.1126/science.1058709

Huang, X. and A. Madan (1999). "CAP3: A DNA Sequence Assembly Program." Genome Research **9**(9): 868-877.

Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998). Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *Journal of Bacteriology, 180*(18), 4765-4774

Jany, J.-L., & Barbier, G. (2008). Culture-independent methods for identifying microbial communities in cheese. *Food Microbiology, 25*(7), 839-848. doi: 10.1016/j.fm.2008.06.003

Jia, W., et al. (2010). "Is the abundance of Faecalibacterium prausnitzii relevant to Crohn's disease?" FEMS Microbiology Letters **310**(2): 138-144

Kahane, S., et al. (1998). "High Prevalence of "Simkania Z," a Novel Chlamydia-like Bacterium, in Infants with Acute Bronchiolitis." Journal of Infectious Diseases **177**(5): 1425-1429

Kahane, S., et al. (2007). "Domestic water supplies as a possible source of infection with Simkania." The Journal of infection **54**(1): 75-81

Kahn, J.-P., Rubinow, D. R., Davis, C. L., Kling, M., & Post, R. M. (1988). Salivary cortisol: A practical method for evaluation of adrenal function. *Biological Psychiatry, 23*(4), 335-349. doi: 10.1016/0006-3223(88)90284-3

Kanauchi, O., et al. (2003). "Germinated barley foodstuff, a prebiotic product, ameliorates inflammation of colitis through modulation of the enteric environment." Journal of Gastroenterology **38**(2): 134-141

Karasu, T.B., Gelenberg, A., Merriam, A. & Wang, P. (2000). Practice Guideline for the Treatment of Patients With Major Depressive Disorder (Second Edition). *Am J Psychiatry*.;157(4 Suppl):1–45.

Kassinen, A., Krogius-Kurikka, L., Makivuokko, H., Rinttila, T., Paulin, L., Corander, J., *et al*. (2007) The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* **133**: 24–33.

Kellner, C. H., et al. (1983). "Relationship of cortisol hypersecretion to brain CT scan alterations in depressed patients." Psychiatry Research **8**(3): 191-197.

Kendler, K.S., Gardner, C.O. Kenneth, S. and Charles, O. (1998). Boundaries of major depression: An evaluation of DSM-IV criteria. *American Journal of Psychiatry*. *1*,;155(2):172–77.

Kendler, K.S., Gatz, M., Gardner, C.O., Pedersen, N.L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry*.;163(1):109–14.

Kent, W. J. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research, 12*(4), 656-664. doi: 10.1101/gr.229202

Kessler RC, B. P. D. O. and et al. (2003). "The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (ncs-r)." JAMA 289(23): 3095-3105.

Kessler, R.C., Nelson, C. & McGonagle K.A. (1996). Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. *British Journal of Psychiatry*.;*168*(suppl 30):17–30.

Khoruts, A., Dicksved, J., Jansson, J.K., and Sadowsky, M.J. (2010) Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent clostridium difficile-associated diarrhea. *J Clin Gastroenterol* **44**: 354–360.

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing--concepts and limitations. *Bioessays, 32*(6), 524-536.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., *et al*. (2011) Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* 108: 4578–4585.

Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., . . . Snyder, M. (2007). Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science, 318*(5849), 420-426. doi: 10.1126/science.1149504.

Kovalic, D., Kwak, J. H., & Weisblum, B. (1991). General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Res, 19*(16) 4560.

Kumar, P., Kanchan, K., Gargallo, R., & Chowdhury, S. (2005). Application of multivariate curve resolution for the study of folding processes of DNA monitored by fluorescence resonance energy transfer. *Analytica Chimica Acta, 536*(1–2), 135-143. doi: 10.1016/j.aca.2004.12.066

Lachenmeier, D. W., & Kessler, W. (2008). Multivariate Curve Resolution of Spectrophotometric Data for the Determination of Artificial Food Colors. *Journal of Agricultural and Food Chemistry*, **56**(14), 5463-5468. doi: 10.1021/jf800069p

Lan, Y., Wang, Q., Cole, J. R., & Rosen, G. L. (2012). Using the RDP classifier to predict taxonomic novelty and reduce the search space for finding novel organisms. *PLoS ONE, 7*(3), 5.

LeBlanc, J. G., Laiño, J. E., del Valle, M. J., Vannini, V., van Sinderen, D., Taranto, M. P.,
Lederberg, J., McCray, A.T. (2001). 'Ome sweet 'omics: -- A genealogical treasury of words. *The Scientist*, **15**(7), 8.

Legato, M. J. (2011). Worlds Within Worlds: The Human Microbiome. *Gender Medicine*, **8**(5), 339-341. doi: 10.1016/j.genm.2011.08.033

Ley, R. E., et al. (2006). "Microbial ecology: human gut microbes associated with obesity." *Nature* **444**(7122): 1022-1023

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., . . . Wilson, R. K. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. [10.1038/nature07485]. *Nature*, **456**(7218), 66-72.

Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, **18**(11), 1851-1858. doi: 10.1101/gr.078212.108

Liu, W. T., Marsh, T. L., Cheng, H., & Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology*, **63**(11), 4516-4522.

Lizardi, P. M., & Engelberg, A. (1979). Rapid isolation of RNA using proteinase K and sodium perchlorate. *Analytical Biochemistry, 98*(1), 116-122. doi: 10.1016/0003-2697(79)90714-0

Lopez-Duran, N. L., Kovacs, M., & George, C. J. (2009). Hypothalamic–pituitary–adrenal axis dysregulation in depressed children and adolescents: A meta-analysis. *Psychoneuroendocrinology, 34*(9), 1272-1283. doi: 10.1016/j.psyneuen.2009.03.016

Louis, P., Young, P., Holtrop, G., and Flint, H.J. (2010) Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ Microbiol 12*: 304–314.

Ludwig, W., Bauer, S. H., Bauer, M., Held, I., Kirchhof, G., Schulze, R., . . . Schleifer, K. H. (1997). Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiology Letters*, **153**(1), 181-190. doi: 10.1016/s0378-1097(97)00256-5

Maes, M. (2011). "An intriguing and hitherto unexplained co-occurrence: Depression and chronic fatigue syndrome are manifestations of shared inflammatory, oxidative and nitrosative (IO&NS) pathways." Progress in Neuro-Psychopharmacology and Biological Psychiatry **35**(3): 784-794.

Marchesi, J. R. (2011); Human distal gut microbiome. *Environmental Microbiology*.13(12): 3088-3102

Mardis, E. R. (2008). Next-generation DNA sequencing methods *Annual Review of Genomics and Human Genetics* ;**9**:387-402. Palo Alto: Annual Reviews.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9: 387-402.

Mariat, D**.,** Firmesse, O**.,** Levenez**,** F., Guimarães**,**V.D., Sokol**,** H., Doré**,** J., Corthier, G., and Furet, J-P.(2009) The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age.*BMC Microbiology*, 9:123

Matsomuto. S.,(2004) Mucosal Immune Responses to the Introduction of Gut Flora in Mice and the Establishment of a Murine Model of Crohn's Disease. *Bioscience and Microflora.;*23:1-9

McConnell, E.L., Fadda, H.M., and Basit, A.W. (2008) Gut instincts: explorations in intestinal physiology and drug delivery. *Int J Pharm* 364: 213–226.

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *N Engl J Med, 338*(3), 171-179.

McEwen, B. S. (2000). The neurobiology of stress: from serendipity to clinical relevance. *Brain Res, 886*(1-2), 172-189.

Metzker, M. L. (2010). Sequencing technologies [mdash] the next generation. *Nat Rev Genet, 11*(1), 31-46.

Meyer, M. and M. Kircher (2010). "Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing." Cold Spring Harbor Protocols **2010**(6): pdb.prot5448

Miller, G. E., Chen, E., & Zhou, E. S. (2007). If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. *Psychological Bulletin, 133*(1), 25-45. doi: 10.1037/0033-2909.133.1.25

Minter, R. E. and Mandel, M. R. (1979). "The Treatment of Psychotic Major Depressive Disorder with Drugs and Electroconvulsive Therapy." The Journal of Nervous and Mental Disease **167**(12): 726-733.

Momozawa Y., Deffontaine, V., Louis, E., and Medrano, J.F. (2011) Characterization of bacteria in biopass of colon and stools by high throughput sequencing of the V2 region of bacteria 16S rRNA gene in human.*PLos ONE 6*: e16952.

Monteleone, P. (2001). Endocrine disturbances and psychiatric disorders. *Current Opinion in Psychiatry*. ;*14*(6):605–10.

Morowitz, M.J., Denef, V.J., Costello, E.K., Thomas, B.C., Poroyko, V., Relman, D.A., and Banfield, J.F. (2011) Strain resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc Natl Acad Sci USA* **108**: 1128–1133.

Müller, N., Myint, A.M., Schwarz, M.J. (2011). "Inflammatory biomarkers and depression". *Neurotox Res 19* (2): 308–18.

Mulsant, B. H. and Ganguli, M. (1999). "Epidemiology and diagnosis of depression in late life." *The Journal of clinical psychiatry 60 Suppl* 20: 9-15

Muyzer, G., & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek, 73*(1), 127-141. doi: 10.1023/a:1000669317571

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, **320**(5881), 1344-1349. doi: 10.1126/science.1158441.

Nestler, E. J., Barrot, M., DiLeone, R. J., Eisch, A. J., Gold, S. J., & Monteggia, L. M. (2002). Neurobiology of Depression. *Neuron*, **34**(1), 13-25. doi: 10.1016/s0896-6273(02)00653-0

Nicholson, J. K., Holmes, E., & Wilson, I. D. (2005). Gut microorganisms, mammalian metabolism and personalized health care. [10.1038/nrmicro1152]. *Nat Rev Micro*, **3**(5), 431-438.

Nocker, A., Burr, M., & Camper, A. (2007). Genotypic Microbial Community Profiling: "A Critical Technical Review". *Microbial Ecology, 54*(2), 276-289.

Ockenfels, M. C., et al. (1995). "Effect of chronic stress associated with unemployment on salivary cortisol: overall cortisol levels, diurnal rhythm, and acute stress reactivity." Psychosomatic Medicine **57**(5): 460-467.

Ogino, S., et al. (2005). "Sensitive Sequencing Method for KRAS Mutation Detection by Pyrosequencing." The Journal of Molecular Diagnostics **7**(3): 413-421.

O'Hara, A.M., and Shanahan, F. (2006) The gut flora as a forgotten organ. *EMBO Rep* **7:** 688–693.

Olfson M, M. S. C. D. B. E. L. T. T. P. H. (2002). "NAtional trends in the outpatient treatment of depression." JAMA **287**(2): 203-209.

Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., & Stahl, D. A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annual Review of Microbiology, 40*, 337-365.

Orengo, C., Fullerton, G. & Tan, R. (2004) Male depression: A review of gender concerns and testosterone therapy. *Geriatrics.*;*59*(10):24–30.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., & Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences,* **86**(8): 2766-2770.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007) Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.

Pampallona S, B. P. T. G. K. B. M. C. (2004). "Combined pharmacotherapy and psychological treatment for depression: A systematic review." Archives of General Psychiatry **61**(7): 714-719.

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet, 10*(10): 669-680.

Patel, V., et al. (2004). "Editorial: Treating depression in the developing world." Tropical Medicine and International Health **9**(5): 539-541.

Path, G., Bornstein, S. R., Ehrhart-Bornstein, M., & Scherbaum, W. A. (1997). Interleukin-6 and the interleukin-6 receptor in the human adrenal gland: expression and effects on steroidogenesis. *J Clin Endocrinol Metab;* **82**(7): 2343-2349.

Paul J. Lucassen, P. M., A.S. Naylor, A. M. van Dam, A.G. Dayer, B. Czeh, and Charlotte A. *Oomen. (2009). Do depression, stress, sleep diruption, and inflammation alter hippocampal* apoptosis and neurogenesis? In R. M. N. Carmine M. Pariante, David Nutt and Lewis Wolpert (Ed.), *Understanding Depression - A translational approach.* London: Oxford University Press p139

Perogamvros, I., Keevil, B. G., Ray, D. W., & Trainer, P. J. (2010). Salivary Cortisone Is a Potential Biomarker for Serum Free Cortisol. *Journal of Clinical Endocrinology & Metabolism,* **95**(11), 4951-4958. doi: 10.1210/jc.2010-1215

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., . . . Grp, N. H. W. (2009). The NIH Human Microbiome Project. *Genome Research,; 19*(12): 2317-2323. doi: 10.1101/gr.096651.109

Plotsky, P. M., Owens, M. J., & Nemeroff, C. B. (1998). Psychoneuroendocrinology of Pepression: Hypothalamic-Pituitary-Adrenal Axis. *Psychiatric Clinics of North America;* **21**(2): 293-307. doi: 10.1016/s0193-953x(05)70006-x

Pop, M., & Salzberg, S. L. (2008). Bioinformatics challenges of new sequencing technology. *Trends in Genetics, 24*(3), 142-149. doi: 10.1016/j.tig.2007.12.006

Qiagen. (2012). BioSpring 96 Retrieved April, 2012, from http://www.qiagen.com/products/automation/biosprint96.aspx#Tabs=t1

Raff, H. (2009). "Utility of Salivary Cortisol Measurements in Cushing's Syndrome and Adrenal Insufficiency." Journal of Clinical Endocrinology and Metabolism **94**(10): 3647-3655.

Rath, H.C., Herfarth, H.H., Ikeda, J.S., Hamm, T.E., Balish, E., Taurog. J.D., Hammer, R.E., Wilson, K.H., Sartor, R.B. (1996) Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. *Journal of Clinical Investigation* .;**98**(4): 945–953

Renz, H., et al. (2012). "The impact of perinatal immune development on mucosal homeostasis and chronic inflammation." Nature Reviews: Immunology 12(1): 9-23.

Rogers, G. B., & Bruce, K. D. (2010). Next-generation sequencing in the analysis of human microbiota: essential considerations for clinical application. *Mol Diagn Ther; 14*(6): 343-350.

Romero, R., Schaudinn, C., Kusanovic, J.P., Gorur, A., Gotsch, F., Webster, P., *et al.* (2008) Detection of a microbial biofilm in intraamniotic infection. *Am J Obstet Gynecol* **198**: 135.e131–135.e135.

Ronaghi, M. (2001). "Pyrosequencing Sheds Light on DNA Sequencing." Genome Research **11**(1): 3-11.

Ronaghi, M., Uhlén, M. & Nyrén, P (1998). "A sequencing method based on real-time pyrophosphate". *Science* 281 (5375): 363

Rosenblum, B.B., Lee, L.G., Spurgeon, S.L., Khan, S.H., Menchen, S.M., Heiner, C.R., Chen, S.-M. *(1997)* New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res. ;***25***:4500–4504.*

Round, J. L., & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. [10.1038/nri2515]. *Nat Rev Immunol,;* **9**(5): 313-323.

Roy, D. S. (2001). The human superorganism – Of microbes and men. *Medical Hypotheses, ;***74**(2): 214-215.

Rudi, K., Skulberg, O. M., Larsen, F., & Jakobsen, K. S. (1997). Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Applied and Environmental Microbiology,* **63**(7), 2593-2599.

Rudi, K., Zimonja, M., Trosvik, P., & Næs, T. (2007). Use of multivariate statistics for 16S rRNA gene analysis of microbial communities. *International Journal of Food Microbiology,* **120**(1–2), 95-99. doi: 10.1016/j.ijfoodmicro.2007.06.004

S.T.A.R. Buffer. (2004), from http://www.roche-applied-science.com/sis/rtpcr/lightcycler/lightcycler_docs/research_kits/03335208001/3335208a.pdf

Sait, M., Hugenholtz, P., & Janssen, P. H. (2002). Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environmental Microbiology, 4*(11), 654-666. doi: 10.1046/j.1462-2920.2002.00352.x

Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences **74**(12): 5463-5467.

Sanger, F., et al. (1980). "Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing." Journal of Molecular Biology **143**(2): 161-178.

Santos, V. A., et al. (2011). "Physical performance but not muscle function is improved by Infliximab in IBD patients: P-47." Inflammatory Bowel Diseases **17**: S26

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions. *Endocrine Reviews;* **21**(1): 55-89. doi: 10.1210/er.21.1.55

Scanlan, P.D., Shanahan, F., and Marchesi, J.R. (2008a) Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiol;* **8**: 79

Scanlan, P.D., Shanahan, F., Clune, Y., Collins, J.K., O'Sullivan, G.C., O'Riordan, M., *et al*. (2008b) Culture independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ Microbiol* 10: 789–798.

Schuster Stephan C. (January 2008). "Next-generation sequencing transforms today's biology". *Nat. Methods* 5 (1): 16–8.

Schwiertz, A., Gruhl, B., Lobnitz, M., Michel, P., Radke, M., and Blaut, M. (2003). Development of the intestinal bacterial composition in hospitalized preterm infants in comparison with breast-fed, full-term infants. *Pediatr Res* **54**: 393–399.

Sekirov, I., Russell, S. L., Antunes, L. C. M., & Finlay, B. B. (2010). Gut Microbiota in Health and Disease. *Physiological Reviews,* **90**(3), 859-904. doi: 10.1152/physrev.00045.2009

Sekirov, I., Tap, J.,Roudot-Thoraval, F.,Roperch, J.P., Letulle,S., Langella,P., et al. (2011) Microbial dysbiosis in colorectal cancer (CRC)patients Plos ONE 6: e16393.

Sesma, F. (2011). B-Group vitamin production by lactic acid bacteria – current knowledge and potential applications. *Journal of Applied Microbiology ;***111**(6): 1297-1309.

Setoyama, H., et al. (2003). "Prevention of gut inflammation by Bifidobacterium in dextran sulfate-treated gnotobiotic mice associated with Bacteroides strains isolated from ulcerative colitis patients." Microbes and Infection **5**(2): 115-122.

Sheline, Y.I., Gado, M.H., Kraemer, H.C. (2003). Untreated depression and hippocampal volume loss. *American Journal of Psychiatry*.;160(8):1516–18.

Smith, R. F. (1996). Perspectives: sequence data base searching in the era of large-scale genomic sequencing. *Genome Research;* **6**(8): 653-660. doi: 10.1101/gr.6.8.653

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *J Mol Biol;* **147**(1): 195-197.

Smyth, J. M., et al. (1997). "Individual differences in the diurnal cycle of cortisol." Psychoneuroendocrinology **22**(2): 89-105.

Sokol, H., et al. (2008). "Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients." Proceedings of the National Academy of Sciences **105**(43): 16731-16736

Sokol, H., et al. (2009). "Low counts of Faecalibacterium prausnitzii in colitis microbiota." Inflammatory Bowel Diseases **15**(8): 1183-1189.

Springer Healthcare News (2012). "Gut microbiota may dictate vulnerability to diet-induced diabetes." *Springer Healthcare News* **1**(1): 1-2

Suchodolski, J. S. (2011). Companion animals symptomes: Microbes and gastrointestinal health of dogs and cats. *Journal of Animal Science*, **89**(5), 1520-1530. doi: 10.2527/jas.2010-3377

Sudo, N. (2006). Stress and gut microbiota: Does postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response? *International Congress Series, 1287*(0), 350-354. doi: 10.1016/j.ics.2005.12.019

Sudo, N., Chida, Y., Aiba, Y., Sonoda, J., Oyama, N., Yu, X.-N., . . . Koga, Y. (2004). Postnatal microbial colonization programs the hypothalamic–pituitary–adrenal system for stress response in mice. *The Journal of Physiology, 558*(1), 263-275. doi: 10.1113/jphysiol.2004.06338

Suttle, C.A. (2007) Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.

Tannock, G. W. (2001). Molecular assessment of intestinal microflora. *The American Journal of Clinical Nutrition;* **73**(2) :410S-414S.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C,. Furet, JP., et al (2009) Towards the human intestinal microbiota phylogenetic core. Environ Microbiol 11: 2574-2584.

Tauler, R., & Barceló, D. (1993). Multivariate curve resolution applied to liquid chromatography—diode array detection. *TrAC Trends in Analytical Chemistry;* **12**(8): 319-327. doi: 10.1016/0165-9936(93)88015-w

Thase, M.E. (1999). When are psychotherapy and pharmacotherapy combinations the treatment of choice for major depressive disorder?. *Psychiatric Quarterly*.;70(4):333–46.

Trapnell, C., & Salzberg, S. L. (2009). How to map billions of short reads onto genomes. *Nat Biotec; 27*(5): 455-457.

Trilck, M., Flitsch, J., Ludecke, D. K., Jung, R., & Petersenn, S. (2005). Salivary cortisol measurement--a reliable method for the diagnosis of Cushing's syndrome. *Exp Clin Endocrinol Diabetes;* **113**(4): 225-230.

Trosvik, P., Skånseng, B., Jakobsen, K. S., Stenseth, N. C., Næs, T., & Rudi, K. (2007). Multivariate Analysis of Complex DNA Sequence Electropherograms for High-Throughput Quantitative Analysis of Mixed Microbial Populations. *Applied and Environmental Microbiology;* **73**(15): 4975-4983. doi: 10.1128/aem.00128-07

Tsapakis, E.M., Soldani, F., Tondo, L., Baldessarini, R.J. (2008). Efficacy of antidepressants in juvenile depression: meta-analysis. *Br J Psychiatry*.;193(1):10–7.

Tulic, M. K. *et al.*(2011).Differences in innate immune function between allergic and nonallergic children: new insights into immune ontogeny. *J. Allergy Clin. Immunol.* **127**, 470–478

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., *et al*. (2009) A core gut microbiome in obese and lean twins. *Nature* 457: 480–484.

Uher, R. and P. McGuffin (2010). "The moderation by the serotonin transporter gene of environmental adversity in the etiology of depression: 2009 update." Molecular Psychiatry **15**(1): 18-22.

Van Duijn, E., Selis, M. A., Giltay, E. J., Zitman, F. G., Roos, R. A. C., van Pelt, H., & van der Mast, R. C. (2010). Hypothalamic–pituitary–adrenal axis functioning in Huntington's disease mutation carriers compared with mutation-negative first-degree controls. *Brain Research Bulletin;* **83**(5): 232-237. doi: 10.1016/j.brainresbull.2010.08.006

Van Melderen, L. (2010) Toxin-antitoxin systems: why so many, what for? *Curr Opin Microbiol* **13**: 781–785.

Vinberg, M., et al. (2008). "Salivary cortisol in unaffected twins discordant for affective disorder." Psychiatry Research **161**(3): 292-301.

Venkataraman, S., Munoz, R., Candido, C., & Witchel, S. F. (2007). The hypothalamic-pituitary-adrenal axis in critical illness. *Rev Endocr Metab Disor;*, **8**(4): 365-373.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., . . . Smith, H. O. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science;* **304**(5667): 66-74. doi: 10.1126/science.1093857

Wade, W. (2002) unculturable bacteria- the uncharacterized organisms that cause oral infections. J R Soc Med; 95: 81-83

Walker, E. F., & Diforio, D. (1997). Schizophrenia: a neural diathesis-stress model. *Psychol Rev, ;***104**(4): 667-685.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology;***73**(16): 5261-5267. doi: 10.1128/aem.00062-07

Wang, X., Heazlewood, S. P., Krause, D. O., & Florin, T. H. (2003). Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J Appl Microbiol, 95*(3), 508-520.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., DuGar, B., *et al*. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**: 57–63.

Ward, D. M., et al. (1990). "16S rRNA sequences reveal numerous uncultured microorganisms in a natural community." Nature **345**(6270): 63-65.

Ward, N. G., et al. (1983). "Skin conductance: A potentially sensitive test for depression." Psychiatry Research **10**(4): 295-302.

Watanabe, K., Kodama, Y., & Harayama, S. (2001). Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. *Journal of Microbiological Methods;* **44**(3): 253-262. doi: 10.1016/s0167-7012(01)00220-2.

Weissman, M. M. (1987). "Advances in psychiatric epidemiology: rates and risks for major depression." American Journal of Public Health **77**(4): 445-451.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes;the unseen majority. Proc Natl Acad Sci USA 95; 6578-6583.

Whooley, M.A.& Simon, G.E. (2000). Managing Depression in Medical Outpatients. *New England Journal of Medicine*.;343(26):1942–50.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev;* **51**(2): 221-271.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences; 87*(12): 4576-4579.

Wolraich, M. L., et al. (1998). "Examination of DSM-IV criteria for attention deficit hyperactivity disorder in a county-wide sample." Journal of Developmental and Behavioral Pediatrics **19**(3): 162-168

Wright, S.L.& Persad, C. (2007) Distinguishing between depression and dementia in older persons: Neuropsychological and neuropathological correlates. *Journal of geriatric psychiatry and neurology*.;20(4):189–98.

Wu, H.-J., Ivanov, I. I., Darce, J., Hattori, K., Shima, T., Umesaki, Y., . . . Mathis, D. (2010). Gut-Residing Segmented Filamentous Bacteria Drive Autoimmune Arthritis via T Helper 17 Cells. *Immunity;* **32**(6): 815-827.

Young, V. B. and T. M. Schmidt (2004). "Antibiotic-Associated Diarrhea Accompanied by Large-Scale Alterations in the Composition of the Fecal Microbiota." Journal of Clinical Microbiology **42**(3): 1203-1206.

Zakeri, H., et al. (1998). "Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing." Biotechniques **25**(3): 406-410, 412-404.

Zhang, H., Dibaise, J.K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., *et al*. (2009) Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci USA* **106**: 2365–2370.

Zhu, B., et al. (2010). "Human gut microbiome: the second genome of human body." Protein & Cell **1**(8): 718-725.

Zimmerman M, C. W. C. C. W. S. (1986). "A self-report scale to diagnose major depressive disorder." Archives of General Psychiatry **43**(11): 1076-1081.

Zimonja, M., Rudi, K., Trosvik, P., & Næs, T. (2008). Multivariate curve resolution of mixed bacterial DNA sequence spectra: identification and quantification of bacteria in undefined mixture samples. *Journal of Chemometrics;* **22**(5): 309-322. doi: 10.1002/cem.1115.

Zisook, S., et al. (2006). "Use of bupropion in combination with serotonin reuptake inhibitors." Biological Psychiatry **59**(3): 203-210.

Zoetendal, E. G., Collier, C. T., Koike, S., Mackie, R. I., & Gaskins, H. R. (2004). Molecular Ecological Analysis of the Gastrointestinal Microbiota: A Review. *The Journal of Nutrition;* **134**(2): 465-472.

Zoetendal, E. G., Rajilić-Stojanović, M., & de Vos, W. M. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut, 57*(11), 1605-1615. doi: 10.1136/gut.2007.133603

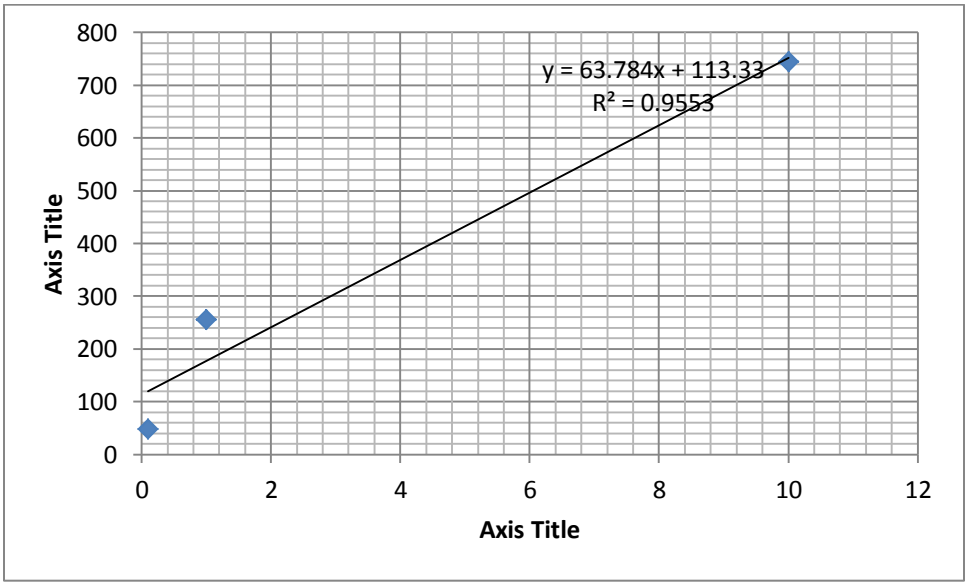
Appendix

Table 7a- DNA quantification at the first DNA extraction phase. The entire DNA in the fecal samples including human and bacterial DNA was extracted at this stage, purified and quantified and if the amount was satisfactory, 16S rRNA gene was selectively amplified later on.

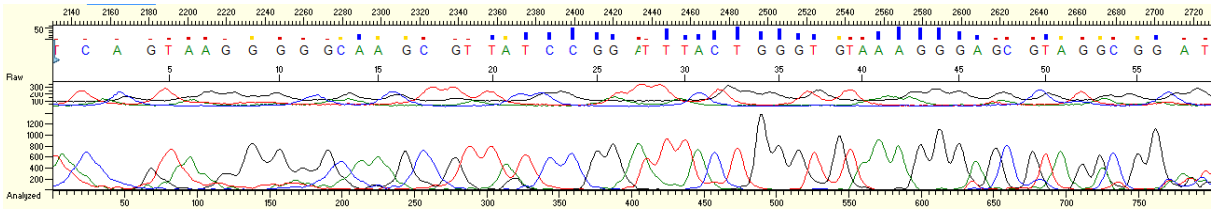
Sample Number	PG	Cal.Concentration	
1a		188	0.2
1b		224	1
2a		178	-0
2b		184	0.1
3a		140	-1
3b		222	1
4a		296	2.8
4b		243	1.5
5a		412	5.5
5b		558	9.1
6a		74	-3
6b		69	-3
7a		197	0.4
7b		151	-1
8a		220	0.9
8b		258	1.8
9a		224	1
9b		410	5.5
10a		91	-2
10b		384	4.9
11a		855	16
11b		902	17
12a		374	4.6
12b		327	3.5
13a		587	9.8
13b		578	9.5
14a		310	3.1
14b		261	1.9
15a		226	1.1
15b		570	9.3
16a		83	-2
16b		54	-3
17a		74	-3
17b		161	-0
18a		339	3.8
18b		416	5.6
19a		623	11
19b		677	12
20a		116	-2
20b		100	-2
21a		128	-1
21b		313	3.2
22a		133	-1
22b		134	-1
23a		341	3.8
23b		198	0.4
24a		1086	22
24bb		375	4.7

25a	256	1.8
25b	276	2.3
26a	90	-2
26b	141	-1
27a	175	-0
27b	204	0.6
28a	236	1.3
28b	204	0.6
29a	137	-1
29b	240	1.4
30a	1136	23
30b	1110	22
34a	287	2.5
34b	168	-0
35a	386	4.9
35b	370	4.5
36a	804	10.8
36b	514	6.28
37a	315	3.16
37b	303	2.97
40a	331	3.41
40b	300	2.93
41a	376	4.12
41b	493	5.95
43a	28	-1.3
43b	19	-1.5
44a	18	-1.5
44b	302	2.96
45a	476	5.69
45b	16	-1.5
46a	17	-1.5
46b	15	-1.5
48a	259	2.28
48b	267	2.41
49a	494	5.97
49b	491	5.92
51a	845	11.5
51b	943	13
52a	779	10.4
52b	745	14.1
53a	567	7.11
53b	861	11.7
54a	939	12.9
54b	913	15.7
55a	843	11.4
55b	814	14.3
56a	710	21.1
56b	789	17.1
58a	595	7.55
58b	851	11.6
59a	412	4.68
59b	741	9.84
61a	866	11.8
61b	876	12
62a	865	15.8
62b	670	21

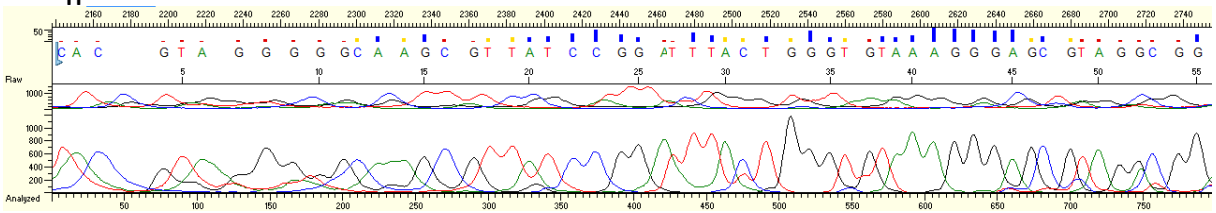
63a	630	8.1
63b	946	13.1
65a	820	16.2
65b	993	13.8
67a	459	5.42
67b	565	7.08
68a	989	13.7
68b	976	13.5
Negative control	23	-1.4



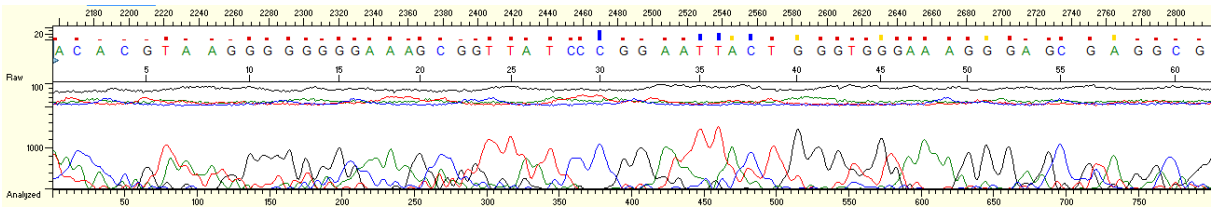
I



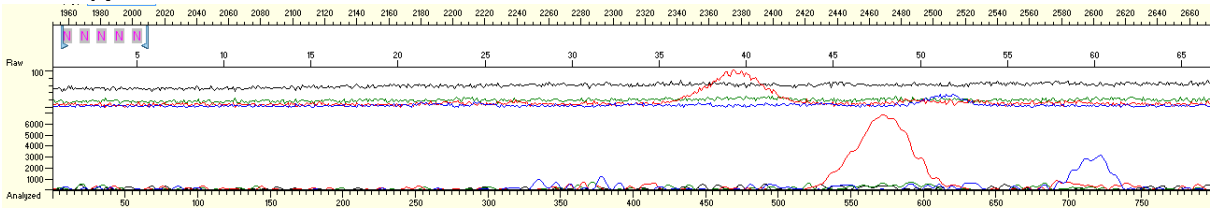
II



III



IV



Raw mixed sequence data as observed by ABI sequence scanner software v1.0.

"I" shows an example of high quality trace score.

"II" represents an example of medium quality trace score.

"III" represents an example of low quality trace score.

"IV" represents an example of no trace score (no readable nucleotide information)

Tests of Normality (none of the components are normally distributed)

	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Comp1	.352	115	.000	.368	115	.000
Comp2	.319	115	.000	.445	115	.000
Comp3	.369	115	.000	.331	115	.000
Comp4	.190	115	.000	.832	115	.000
Comp5	.254	115	.000	.675	115	.000
Comp6	.261	115	.000	.610	115	.000
Comp7	.221	115	.000	.738	115	.000
Comp8	.256	115	.000	.681	115	.000
Comp9	.288	115	.000	.581	115	.000
Comp10	.251	115	.000	.701	115	.000
Comp11	.193	115	.000	.826	115	.000
Comp12	.302	115	.000	.514	115	.000
Comp13	.211	115	.000	.791	115	.000
Comp14	.211	115	.000	.799	115	.000
Comp15	.239	115	.000	.741	115	.000
Comp16	.343	115	.000	.433	115	.000
Comp17	.297	115	.000	.595	115	.000
Comp18	.144	115	.000	.894	115	.000
Comp19	.302	115	.000	.547	115	.000
Comp20	.195	115	.000	.826	115	.000

Table 7b- Combination of forward and reverse primers for Illumina sequencing (samples number 31-56)

First samples	1	2	3	4	5	6	7
A	F1R1-34a-D	F1R2-40a-D	F1R3-45a-D	F1R4-51a-D	F1R5-55a-D	F1R6-61a-D	F1R7-67a-N
B	F2R1-34b-D	F2R2-40b-D	F2R3-45b-D	F2R4-51b-D	F2R5-55b-D	F2R6-61b-D	F2R7-67b-N
C	F3R1-35a-D	F3R2-41a-D	F3R3-46a-D	F3R4-52a-D	F3R5-56a-D	F3R6-62a-D	F3R7-68a-D
D	F4R1-35b-D	F4R2-41b-D	F4R3-46b-D	F4R4-52b-D	F4R5-56b-D	F4R6-62b-D	F4R7-68b-D
E	F5R1-36a-N	F5R2-43a-D	F5R3-48a-N	F5R4-53a-D	F5R5-58a-D	F5R6-63a-N	
F	F6R1-36b-N	F6R2-43b-D	F6R3-48b-N	F6R4-53b-D	F6R5-58b-D	F6R6-63b-N	
G	F7R1-37a-N	F7R2-44a-D	F7R3-49a-N	F7R4-54a-D	F7R5-59a-D	F7R6-65a-D	
H	F8R1-37b-N	F8R2-44b-D	F8R3-49b-N	F8R4-54b-D	F8R5-59b-D	F8R6-65b-D	

Table 7c- Combination of forward and reverse primers for Illumina sequencing (samples number 1-30)

Second samples	1	2	3	4	5	6	7	8
A	F9R9-1a -D	F9R10 -5a-N	F9R11 -9a-D	F9R12 -13a-N	F9R13 -17a-D	F9R14 -21a-D	F9R15 -25a-D	F9R16 -29a-D
B	F10R9-1b -D	F10R10 -5b-N	F10R11- 9b-D	F10R12 -13b-N	F10R13 -17b-D	F10R14 -21b-D	F10R15 -25b-D	F10R16 -29b-D
C	F11R9-2a -N	F11R10 -6a-D	F11R11 -10a-D	F11R12 -14a-D	F11R13 -18a-N	F11R14 -22a-D	F11R15 -26a-D	F11R16 -30a-D
D	F12R9-2b -N	F12R10 -6b-D	F12R11 -10b-D	F12R12 -14b-D	F12R13 -18b-N	F12R14 -22b-D	F12R15 -26b-D	F12R16 -30b-D
E	F1R9-3a -N	F1R10 -7a-N	F1R11 -11a-N	F1R12 -15a-D	F1R13 -19a-N	F1R14 -23a-D	F1R15- 27a-D	F1R16 -40a-D
F	F2R9-3b -N	F2R10 -7b-N	F2R11 -11b-N	F2R12 -15b-D	F2R13 -19b-N	F2R14 -23b-D	F2R15- 27b-D	F2R16 -40b-D
G	F3R9-4a -D	F3R10 -8a-N	F3R11 -12a-N	F3R12 -16a-N	F3R13 -20a-D	F3R14 -24a-N	F3R15- 28a-D	F3R16- 44a-D
H	F4R9-4b -D	F4R10 -8b-N	F4R11 -12b-N	F4R12 -16b-N	F4R13 -20b-D	F4R14 -24b-N	F4R15- 28b-D	F4R16- 44b-D